Effective Biomolecular Reading and Writing

Valerie Matarese

2017, 2022

Matarese Valerie Ann

Effective Biomolecular Reading & Writing, 2nd edition

ISBN (print) 979-12-210-0995-8

First edition published in 2017.

Published by Matarese Valerie Ann, Via Roma 10, I-31020 Vidor (TV), Italy.

Printed and bound in Italy.

Table of Contents

Tables

Figures

Boxes

Preface to the new edition

Five years have passed since the first edition of this handbook was compiled from a mass of lecture notes for Effective Biomedical Writing, an intensive practical course designed for doctoral students and early career researchers. During this period, several topics of research writing and publishing underwent substantial changes. Moreover, through successive editions of the course, it became clear that certain chapters needed revision to provide more effective instruction.

The second edition of *Effective Biomolecular Reading and Writing* presents new and revised content. In particular, Chapter I.5, "Reporting quantities and statistics", has been entirely rewritten to incorporate important changes in the International System of Units, provide greater guidance on sentence structure for statistics, and explain and illustrate an emerging paradigm of statistical reporting now required by some medical journals. Chapter II.2, "Preparing the displays: tables and figures", has been revised to provide more advice and instruction on choosing types of charts, making tables and figures, and writing table titles and figures legends. Short bibliographies of suggested reading material have been added to conclude some chapters, and new tables, figures, and boxes address niche topics that were proving difficult to course participants.

This volume is now accompanied by a supplement with additional illustrations, examples, and practice questions. Course participants receive the supplement as PDF files in installments, before most lessons. Participants are requested to read the relevant chapter (indicated in the course program) both *before* each lesson, to familiarize themselves with the terminology, and *afterward*, to reinforce the concepts. When the course is over, the handbook will serve as a handy reference to participants as they advance in their research careers.

I wish to express my gratitude to prior course participants for their questions, comments, and writing examples, which have informed me of the real needs of early career researchers learning to write research articles in good scientific English. I hope the revised edition provides clearer explanations and better examples from which to learn this enormous, powerful skill of scientific communication.

Valerie Matarese
Vidor, April 2022

Preface

This handbook is intended as a support text for “Effective Biomedical Writing,” an intensive practical course designed for doctoral students and early career researchers. It is written for young scholars who use English as an additional language, especially those unfamiliar with the speech of a native English speaker. Accordingly, it presents the content of the lectures but not all the examples and exercises used in class. The language used is American English but the text addresses issues of writing in British English as well, with explanations of differences in terminology and usage between the two languages.

Course participants are recommended to read the relevant lecture notes both *before* each lesson (to familiarize themselves with the vocabulary) and *afterward* (to reinforce the concepts). They are also asked to bring this handbook to every lesson. The handbook will also serve as a reference for course participants after the course is over. For this reason, it is organized in two parts, the first dealing with the issues and ethics of research publication and the second describing methods for writing scientific articles. This is not the same organization as that of the class work, as each day of the course addresses topics from both parts of the handbook. The accompanying program indicates which topics are dealt with on each day.

This volume is a work in progress. It has developed from several editions of the course given in different formats over 10 years. I gratefully acknowledge the comments and feedback of course participants, both those of prior editions and of the present, for sharing their views on the effectiveness of the explanations and for providing teaching examples from their writings. I am also indebted to my international colleagues who have shared their own approaches to teaching research writing through their publications, their presentations at professional meetings, and their contributions to *Supporting Research Writing*, a book I edited and published in 2013. This handbook nonetheless still needs improvement. As researchers do when drafting research articles, I will revise, and revise and revise!

Valerie Matarese
Vidor, February 2017

Part I. Publication issues and ethics

I.1 Aims and challenges of publishing research

The research article, or research paper, is today considered a researcher's main output, that is, the product of a researcher's activities that counts the most towards status and career. Researchers produce other things, of course, such as review papers, commentary, peer review reports, patents, software, and data sets, and there are efforts to give these types of output greater attention. But for now, in this course, our focus is on the research article.

Why do researchers write and publish research articles? Many of us say that, by publishing, we **advance knowledge** by adding new findings to the scientific knowledge base. We leave a **permanent record** of our work, so that we can get **due credit** for our discoveries: we lengthen our CVs and we get recognized, by being invited to present our work at meetings, being promoted to higher academic positions, and so on. By publishing, we also get **feedback**, as the article is critiqued in letters to the editor and, increasingly, in comments posted online. And, hopefully, the study gets **independent verification** by being repeated by others, who confirm the findings. Only once a study is replicated and validated in this way can it really advance knowledge – this is a tenet of the scientific method.

By publishing, we communicate with a broad group of readers. In the narrowest sense, we communicate with researchers working in the same field as us, namely our colleagues and competitors. They form a minority of all our readers, however, and can be thought to represent the innermost of a set of concentric circles of readers (Figure I.1). Broadening out to the next circle, we communicate with researchers in other fields, some closely related but many more working in distant fields wishing to understand the broader picture of biological phenomena. Our readers are not only scientists but also persons whose work intersects with research, for example grant funders and university administrators who make decisions that affect our careers. The literature is also read by science writers, journalists, health policy-makers, patients, activists, and even the average person (perhaps to understand a disease). The outer circle of readers has different

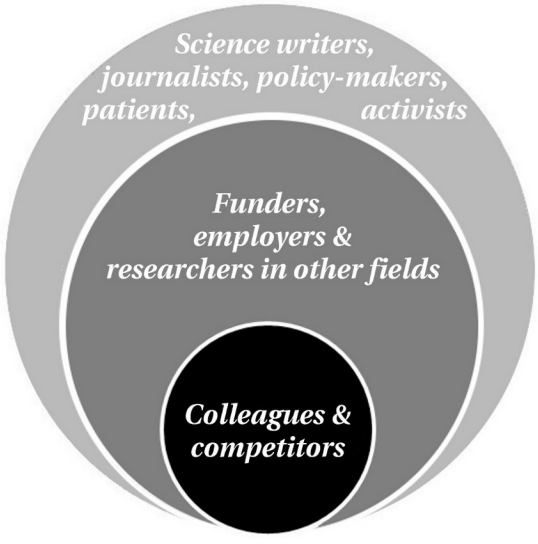

Figure I.1: Concentric circles of readers

information needs and reads research articles differently from how the inner circle reads. If we address all these readers when writing, our work will have greater impact, and the visibility of our research efforts will be optimized.

So what makes a good research article? To start, good research: Our study addressed a timely, important scientific topic. We formulated a clear study question and designed a series of powerful experiments able to provide a reliable answer. Our research materials were of high quality, the individual experiments were skillfully executed, and the data were accurately collected, analyzed, and interpreted, leading to meaningful conclusions. But there's more.

Equally important is good reporting: The Introduction section of our research article clearly states the context of the research and justifies why the study was done, helping readers understand the study's **relevance**. The materials and methods are accurately described in enough detail for the research to be repeated and, ideally, confirmed by multiple teams of other researchers; this supports the study's **reproducibility**. The results are fully documented (in figures and tables) and clearly explained, providing the **evidence**. The conclusions are based on the reported evidence: this is our **scientific advance**. And, overall, the writing is clear (easily and quickly understood), precise (not vague or ambiguous), original (not copied or recycled), and efficient (not unnecessarily verbose or sophisticated). The text has a forward flow of information, the arguments are sound, and the prose is reader-friendly: this is our **scientific voice** emerging from the page and communicating with our readers.

Many published research articles lack these features. Even though journals choose articles to publish on the basis of a positive peer review, not all journals set high-enough standards for acceptance. We may therefore find ourselves reading articles with ambiguous wording or insufficient methodological details. We may judge a study to be underpowered or have an imperfect statistical analysis. We may even repeat the reported experiments in our own lab but not get the same results, concluding that part or all of the study is irreproducible.

Irreproducible research articles are often the product of poor scientific practices. Studies may be poorly designed: the methods or biological models may be inappropriate for the study question, or the controls may be inadequate and the statistical tests unsuited for the type of data. The chemical reagents and biological specimens may be impure, contaminated, degraded, or mislabeled, or they may be uncharacterized or not validated for the specific use in the study. The methods may not be executed to perfection, with uncontrolled environmental variables and instrumental error interfering with the collection of reliable data, and the data analysis may be corrupted by sample mix-up and spreadsheet errors. Finally, the investigators' objectivity in analyzing the data may be clouded due to a fervent belief in the tested hypothesis, leading to a biased interpretation.

Irreproducibility can also be due to poor reporting. In the published report, the methods may be described in an unclear, incomplete or inaccurate manner, and the results may be only a selection of the "best" data that fit a hypothesis but that do not

tell the whole story. Sometimes, irreproducibility is due to **research misconduct –** behavior that is improper and not conforming to standards and expectations of a scientist. Misconduct may involve negligence (i.e. lack of care or attention), dishonesty, or fraud (i.e. deliberate deception), for example. Two particular types of misconduct can lead to irreproducibility:

- **Falsification** is the manipulation of an experiment or its data to show a desired effect or to prove a hypothesis. It may involve altering the research materials or instruments to generate false data or, after an experiment, manually changing numerical data or using graphics software to alter images of gel blots, for example.
- **Fabrication** is the making up of data and reporting them as if they were real. It includes both inventing numerical data and claiming that images (e.g. of gels or cells) are different than they really are.

So, research articles may be irreproducible because of poor research practices, poor reporting, or misconduct. One study estimated that over 50 % of published preclinical research articles have problems with reproducibility.[1] Most often, the problems are not clearly identified and made public. Sometimes, corrections to problematic articles are published. Only in rare cases are articles of irreproducible research removed from the literature by a formal process called retraction. **Retraction** is not an easy process. It requires an investigation by the journal, often involving the research institute too, and difficult editorial decisions. A retraction notice must be published, and the authors may dispute the decision. Only about 0.01 % of all articles indexed in Medline have been retracted, for various reasons (not limited to irreproducibility). Retraction is certainly not a solution for irreproducibility.

The very real problem of irreproducibility corresponds to an enormous amount of **research waste**, defined as the ineffective, unproductive spending of research funds and researchers' time and effort. Irreproducible articles cause other researchers to waste time and money doing experiments that cannot succeed; research advances more slowly than it should, and our desired goals – a new therapy, for example, take longer to achieve. The topic of irreproducibility and research waste is getting much attention in both the scientific community and general press, and journals, funders, and universities are working to help researchers identify the causes and implement solutions.

Initiatives being implemented today for reproducible research include: removal of word limits on the Methods section of research articles to allow full reporting; development of checklists to help authors include all the necessary information for a study's replication; encouragement of the sharing of data, materials, and software; and the establishment of best practices for reporting biological reagents such as antibodies and cell lines. These changes aim to develop a more rigorous publication culture among researchers, and they require researchers to be ever more familiar with

1 See Freedman et al. (PLoS Biol. 2015;13(6):e1002165) and references therein

the requirements of “publishability” in terms of both the **language skills** needed for reporting and the **skills in communicating scientific-technical content** according to current best practices. Therefore, this course dedicates equal time to issues of writing for publication and to methods for presenting content in a research article according to international standards. For this reason, this course manual is divided into two parts, and the lessons alternate between both parts of the manual as the course progresses.

I.2 Research journals in the biomedical sciences

This course is about writing biomedical research articles for publication, so we start by discussing the medium in which they are published: research journals in the biomedical sciences. First, some definitions.

A *journal* is a periodical publication that publishes, at regular intervals, a collection of articles written by different persons. A *research journal* is a journal that publishes primarily articles of original research, although it may also publish other content, e.g. editorials and reviews. A *peer-reviewed research journal* is a research journal that welcomes spontaneous submissions from any and all researchers (there's no need to be invited) and that uses external review to select manuscripts to publish. The scientific community considers peer-reviewed research journals the most appropriate medium in which to formally present new discoveries and put forward new hypotheses. These journals, and the research articles they publish, comprise the *primary research literature:* they are primary sources of research information.

In the biomedical sciences, there are many thousands of journals. We'll limit our discussion to journals listed in Medline and PubMed Central, two digital resources for searching the biomedical literature developed and maintained by the US National Library of Medicine (Box I.1). The journals included in these two overlapping resources have been selected based on a rigorous scientific–editorial assessment.[2] First of all, the journals have **biomedical scope**, i.e. they report on research in health, medicine, or life sciences. They have **clear editorial policies**, especially on research ethics, peer review, conflicts of interest. and sponsorship, which safeguard the quality of the content, and they demonstrate adherence to best practices in scholarly publishing as set out in several research and reporting guidelines. Their published articles demonstrate **scientific rigor**, i.e. the "strict application of the scientific method to ensure unbiased and well-controlled experimental design, methodology, analysis, interpretation, and reporting of results."[3] Finally, the journals exhibit professional quality in production (i.e. printing or digital publishing) and administrative practices. Journals must maintain these quality criteria to remain included in these two digital resources.

Most of the journals listed in Medline and PubMed Central are research journals, but these resources also include some review journals, i.e. journals that publish predominantly review articles and no original research articles. Review journals and their articles comprise the *secondary research literature* because they are based on the primary literature. Also included are some general science journals (mixing news, commentary and research; examples are *Science* and *Nature*), general medical journals (also with a variety of article types; most notably *JAMA*, *Lancet*, and *BMJ*), and clinical practice journals. Our focus, however, remains on research journals.

2 See https://www.ncbi.nlm.nih.gov/books/NBK518737/
3 https://grants.nih.gov/policy/reproducibility/index.htm