Nut-22.2. Online material: provide data collection tools and data as online material or explain how they can be accessed or why they cannot be provided.

Example 1. “Dietary information was collected by using a 121-item, self-administered FFQ. Details of which foods were included in each food group are listed in the online Appendix”

Explanation.

Traditionally, efforts to share research have focused on manuscripts that provide a narrative summary of the conducted study and the results. However, other research outputs, such as protocols, data collection instruments, software, and algorithms, are essential for the interpretation of findings and the reproduction of the research project. An increasing number of journals allow researchers to upload supplementary online material or to link objects to online sources or repositories. This is an opportunity to maximize the build-up of scholarly knowledge and is increasingly recognized as an integral part of good research practice and academic culture. There are ongoing discussions about ethical aspects of data-sharing: for instance, within European data infrastructure initiatives [e.g., Biobanking and BioMolecular Resources Research Infrastructure–European Research Infrastructure Consortium (www.bbmri-eric.eu) (152) and the European Clinical Research Infrastructure Network (www.ecrin.org);see also Nut-22.1]. Data-sharing could vary from sharing information with regard to the study design, the mode of data collection and outcomes, the number of participants, and a full list of available measurements, up to the software used and individual, anonymized data. A tangible benefit of sharing protocols and hypotheses for epidemiologic studies in public repositories is that duplication of efforts potentially would be avoided. Legitimate data exploration and discovery could be cost-effective and maximize the impact of available epidemiologic data and should not be limited by preregistration of protocols. Important caveats apply, however, and whether epidemiologic studies should be preregistered or not is debated.

Researchers are encouraged to provide access to the data needed to reproduce the results. Various research-funding agencies, universities, and scientific journals have adopted policies that allow data to be accessible for the reproduction of the study findings; and different data repositories are being developed for such purposes. To ensure an effective use of data for future research and scientific discovery, data-sharing needs to be organized so that interaction with computational agents is facilitated. In other words, data shoul be “findable, accessible, interoperable, and reusable” (FAIR).

Research information with regard to humans should bemanaged with the highest and most appropriate ethical standards (see Nut-22.1). Efforts to ensure that research data are available include collection and storage of high-quality information with long-term validity. In order to do this, data must be well documented, so that other researchers can access, understand, and use these data, and add value to the original data independently of the original investigators. Furthermore, there is a need for high-quality stewardship of scientific data and adequate procedures, including long-term care, quality control, and adequate commitment as well as resources to handle the data. Funding bodies such as the United Kingdom Medical Research Council, the NIH, and the European Commission now explicitly require statements around data management. All applicants submitting funding proposals to these (and many other) funding agencies are required to include a Data Management Plan as an integral part of their application.