edanz | INNOVATIVE SCIENTIFIC SOLUTIONS

edanz learning lab

- Data handling
- Data types
- Transparency policies

From the Edanz "Publishing and Ethics" course

Research data

learning.edanz.com

by Dr. Trevor Lane

Table of contents

Data Handling Mistakes to Avoid

Below are some data handing and presentation mistakes to avoid in your research. If you have any questions or require language or publishing support, please email global@edanzgroup.com.

Feature	Explanation/Example
Ethical data collection, handling, and presentation	• Don't collect animal/human data without prior ethics approval • Don't collect human data without informed consent (or permission from an ethics board to waive this requirement); don't present human data (images, genetic data, any other possibly identifying data) without written permission • Declare conflicts of interest; make sure all authors can access the data • Keep all data records confidential, anonymous, and securely stored • Retain data for inspection by your institution and target journal • Adhere to institutional, funding, and journal requirements on data accessibility and sharing • Cite all your data sources
Plagiarism or self-plagiarism, with or without copyright infringement	• Don't reuse results or illustrations from someone else's publication or your own previous publication, without (1) justifying why to the target journal during manuscript submission, (2) a citation and reference to the previous work, and, if needed, (3) copyright permission for reuse and a copyright credit line
Fabrication	• Don't make up any data • Don't reuse all or part of a digital image from another test/study with different labeling as a new image
Falsification	• Don't change or delete data • Don't change only part of a digital image or crop out parts • Don't piece together parts of a digital image from different sources without clear labeling and showing clear break-lines
Using misleading illustrations	• Don't use the wrong type of illustration for your data type • Don't use misleading axes (eg, graphs with unlabeled scales, *x* and *y* axes with transposed variables, *y* axis not starting at the origin) • Don't use 3D effects, which distort scales and lead to misreading values
Omitting important information	• Don't forget to provide the total N number for proportions • Don't omit the alpha (cutoff P-level for statistical significance) in the Methods; give a sample-size calculation; give confidence intervals
Using wrong tests	• Don't use inappropriate types of statistical test or analysis for your data type/s; plan analyses before the study
Cherry picking	• Don't select a sample, data, or analysis results just to support your point
P-hacking	• Don't use many statistical tests, collect more or exclude data/variables, or transform data just to find a positive result to report • Don't make multiple statistical comparisons of a dataset without correcting the cutoff P-level used
Fishing trip/ expedition, data dredging/fishing	• In hypothesis-testing studies, don't analyze large volumes of data just to find interesting results (especially without knowing the provenance of the dataset/s, collection methods, and any data transformations applied)
HARK	• Don't "hypothesize after the results are known": don't change or create your study hypothesis after the analyses show positive results

Types of Data

Below is a summary of the main differences between quantitative and qualitative data. If you have any questions or require language or publishing support, please email global@edanzgroup.com.

	Quantitative data	Qualitative data
Definitions	Data that are measurable numerical values	Data that can be observed but not measured
	Data that can be mathematically transformed and statistically analyzed to show trends and associations	Data that describe situations, properties, and characteristics; data commonly provide insight and in-depth understanding of people's lived experiences
	Whether single variables (one type of observation) or multiple variables (more than one type) are measured is decided before the study	Can be analyzed as themes and subthemes (content analysis) or types of language used such as words/pauses/sounds/gestures, intentions, and interactions between participants (discourse analysis, conversation analysis); coding and scoring systems may be decided before the study
	Commonly obtained in hypothesis-testing papers	Commonly obtained in hypothesis-generating studies
Types	Discrete (whole numbers) or continuous (scale can be subdivided) data from surveys, interviews, observations, experiments	Spoken/written/visual/audio (pictorial, textual, sound) data from structured or semi-structured questionnaires, interviews, observations, group discussions, literature/diaries/ newspapers, artifacts/objects
	Can be grouped into dichotomous nominal categories (eg, present/absent) or ordinal categories (eg, ranked quantities/scores or groups)	Can be classified into nominal categories (properties of equal hierarchy, eg, colors, pets) or ordinal categories (can be ordered, eg, comfort level, satisfaction level, agree/neutral/disagree)
	Can be parametric, with parameters/constants that can characterize and identify the distribution (eg, Normal distribution, with mean and standard deviation), or nonparametric, meaning without a predictable distribution type (eg, non-Normal distribution, with median and range). These data are analyzed statistically by parametric and nonparametric tests, respectively	Can come from exploratory, small-scale focused studies; can aim to show commonalities, consensus, as well as variation. Two or more categorical variables can be statistically tested by the chi-square test
Examples	Discrete: counts of people, number of family members	Data can be frequency counts: Hair color, eye color, sex/gender, race or ethnicity, nationality, species, types of

	Continuous: Height, weight, temperature, humidity, pressure, distance, time, volume, concentration, area, angles (can be classified as ordinal categories [low, medium, high, or increasing ranges of values] or nominal categories [old/young, high/low, hot/cold]	punctuation used, car type, names (of sports, disease type, possible risk factors, location)
	Can be different types of the same unit (eg, for money, the variable could be personal income, family income, disposable income)	Data can be quotes: Reasons for doing or not doing something; attitudes, beliefs, motivations, perceptions, rationales; qualities and descriptions of something (texture, smell)
Summary statistics	**Quantitative data** ***Examples***	**Qualitative data** ***Examples***
Counts	Numbers of households in a town	How many have pets of different types
Percentages	Percentage of households of married people	Percentage of households with at least one dog, or percentage with any pet
Mean (average value)	Mean age or number of people in a household	Mean level of how much people in a household like pets (conversion to numerical data needed)
Median (central value after ordering data)	Median age or number of people in a household	Median level of how much people in a household like pets (conversion to ordered numerical data needed)
Mode (most frequent value)	Modal age or number of people in a household	Most common name for a pet; most common type of pet

edanz INNOVATIVE SCIENTIFIC SOLUTIONS

Journal Transparency Policies

Use the checklist below to record what transparency policy your target journal has for each key area. If you have any questions or require language or publishing support, please email global@edanzgroup.com.

Journal: _____

Feature	*Not Implemented*	*Level I*	*Level II*	*Level III*
1. **Citation Standards** (citing your sources such as data and materials used)	Journal encourages citation of data, code, and materials, or says nothing. *Policy* ☐ *Article complies* ☐	Journal describes citation of data in guidelines to authors with clear rules and examples. *Policy* ☐ *Article complies* ☐	Article provides appropriate citation for data and materials used consistent with journal's author guidelines. *Policy* ☐ *Article complies* ☐	Article is not published until providing appropriate citation for data and materials following journal's author guidelines. *Policy* ☐ *Article complies* ☐
2. Data Transparency (the availability of your data)	Journal encourages data sharing, or says nothing. *Policy* ☐ *Article complies* ☐	Article states whether data are available, and, if so, where to access them. *Policy* ☐ *Article complies* ☐	Data must be posted to a trusted repository. Exceptions must be identified at article submission. *Policy* ☐ *Article complies* ☐	Data must be posted to a trusted repository, and reported analyses will be reproduced independently prior to publication. *Policy* ☐ *Article complies* ☐
3. Analytic Methods (Code) Transparency (the availability of your analytical code)	Journal encourages code sharing, or says nothing. *Policy* ☐ *Article complies* ☐	Article states whether code is available, and, if so, where to access it. *Policy* ☐ *Article complies* ☐	Code must be posted to a trusted repository. Exceptions must be identified at article submission. *Policy* ☐ *Article complies* ☐	Code must be posted to a trusted repository, and reported analyses will be reproduced independently prior to publication. *Policy* ☐ *Article complies* ☐
4. Research Materials Transparency (the availability of your materials)	Journal encourages materials sharing, or says nothing. *Policy* ☐ *Article complies* ☐	Article states whether materials are available, and, if so, where to access them. *Policy* ☐ *Article complies* ☐	Materials must be posted to a trusted repository. Exceptions must be identified at article submission. *Policy* ☐ *Article complies* ☐	Materials must be posted to a trusted repository, and reported analyses will be reproduced independently prior to publication. *Policy* ☐ *Article complies* ☐

Feature	*Not Implemented*	*Level I*	*Level II*	*Level III*
5. Design and Analysis Transparency (the availability of your full study design)	Journal encourages design and analysis transparency, or says nothing. *Policy* ☐ *Article complies* ☐	Journal articulates design transparency standards. *Policy* ☐ *Article complies* ☐	Journal requires adherence to design transparency standards for review and publication. *Policy* ☐ *Article complies* ☐	Journal requires and enforces adherence to design transparency standards for review and publication. *Policy* ☐ *Article complies* ☐
6. Study Preregistration (the preregistration of your study in an online registry)	Journal says nothing. *Policy* ☐ *Article complies* ☐	Article states whether preregistration of study exists, and, if so, where to access it. *Policy* ☐ *Article complies* ☐	Article states whether preregistration of study exists, and, if so, allows journal access during peer review for verification. *Policy* ☐ *Article complies* ☐	Journal requires preregistration of studies and provides link and badge in article to meeting requirements. *Policy* ☐ *Article complies* ☐
7. Analysis Plan Preregistration (the preregistration of your analysis in an online registry)	Journal says nothing. *Policy* ☐ *Article complies* ☐	Article states whether preregistration of study exists, and, if so, where to access it. *Policy* ☐ *Article complies* ☐	Article states whether preregistration with analysis plan exists, and, if so, allows journal access during peer review for verification. *Policy* ☐ *Article complies* ☐	Journal requires preregistration of studies with analysis plans and provides link and badge in article to meeting requirements. *Policy* ☐ *Article complies* ☐
8. Replication (study replication and submission of Registered Reports before the results are known)	Journal discourages submission of replication studies, or says nothing. *Policy* ☐ *Article complies* ☐	Journal encourages submission of replication studies. *Policy* ☐ *Article complies* ☐	Journal encourages submission of replication studies and conducts results blind review. *Policy* ☐ *Article complies* ☒	Journal uses Registered Reports as a submission option for replication studies with peer review prior to observing the study outcomes. *Policy* ☐ *Article complies* ☐

Source: Center for Open Science, "The TOP Guidelines" (CC BY), https://cos.io/our-services/top-guidelines/; Nosek, Brian A., George Alter, George C. Banks, Denny Borsboom, Sara D. Bowman, Steven J. Breckler, Stuart Buck, et al. 2016. "Transparency and Openness Promotion (TOP) Guidelines."