1 Data collection and documentation

1.1. What data will you collect, observe, generate or reuse?

This project will generate psychophysiological and performance data of 250 3rd year medical students tasked with delivering bad news (BBN) to a simulated patient.

We list the specific types of data, equipment, software, data storage formats, and data sharing formats for all research data in this project in Table 1 below.

Table1: Types of data, equipment, software, and formats for new data

Types	Equipment	Software	data storage format	data archiving/sharing format
Salivary data				
Raw data: salivary cortisol, dehydroepiandrosterone, alpha-amylase	Salicaps, enzyme-linked immunosorbent assay Spectrostar nano	MARS Data Analysis Software	.ruc, .csv	
Analysed data:		Excel, STATA	.xlsx, .dta	.csv
Cardiovascular data				
Raw data: Electrocardiography	VU-AMS	DAMS	.5FS, .bch	
Analyzed data:		Excel, STATA	.xlsx, .dta	.csv
Raw data: Impedance Cardiography	VU-AMS	DAMS	.5FS, .bch	
Analysed data:		Excel, STATA	.xlsx, .dta	.csv
Raw data: Blood pressure	Finometer	Beatscope	.fpf	
Analysed data:		Excel, STATA	.xlsx, .dta	.csv
Questionnaire data				
Raw data:	Questionnaires	unipark	.csv	
Analysed data:		STATA	.dta	.csv

Performacne data recording (Video)				
Raw data :	Video	Videokameras Sony PXW-Z150	MPEG-4 (H264 Videocodec und PCM Audiocodec)	MPEG-4
Performance data rating				
Raw data :	iPad	Examic eOSCE assessment system	.csv	
Analyzed data		Excel, SPSS	.xlsx,	.csv

Data volume over all (except video) are expected to remain below 20 GB. Video data will require roughly 100 GB.

1.2 How will the data be collected, observed or generated?

Methodologies for data collection / generation and data saving / naming for conducting high quality research
- Carrying out instrument /machine calibration and control processes according to the operating manual for all used devices mentioned under 1.1.
- Sticking to standard protocols and operating manuals for all used devices mentioned under 1.1 for data collection ensuring reliability and consistency.
- Designing appropriate experimental design, and following appropriate data recording and data cleaning ensuring internal validity.
- Ensuring training on techniques & data management for project staff (especially for PhD student and research assistant but also continuing education for long-time collaborators).
- Ensuring training on role play (simulated patients).
- Ensuring training on use of rating scales for assessing student performance in BBN.
- Review data quality after collection
- All data are saved to a repository accessible to all project members (not to personal accounts).

Organization and naming of data files
- Data files hierarchically organized into folders and sub-folders (raw data, transformed data and analyzed data in distinct folders) to make navigation and location easier.

    Folder structure:
       1_Foldername
           1.1_Subfoldername 1
               1.1.1 subsubfoldername 1

1.2_Subfoldername 2
            2_Foldername
                    2.1_Subfoldername 1
                    2.2_Subfoldername 2
                            2.2.1_subsubfoldername 1
                    Etc.

- ➤ Data naming and labeling as follows: short unique study identifier_summary of content_unique subject identifier. This helps (a) attributing a file to the correct project if it has been erroneously saved in a wrong project folder, (b) facilitating finding files, and (c) facilitating finding a given trial/subject.
- ➤ Versioning of continuing data analysis subsets distinguished with date (YYYYMMDD) prefixed to the file name and by a file extension indicating the version number (v01, v02, etc.) for reports.
- ➤ Delimiters are always underscores (_) and no other special characters (such as / % + etc.) in order to ensure compatibility between different computer systems.

The conventions of the project are set up, documented and shared with all team members to be consistent. The conventions are subject to peer review process (regular supervision and staff meetings) to ensure that procedures have been carried out correctly and that all data are properly recorded and saved.

1.3 What documentation and metadata will you provide with the data?

The study will be registered in BORIS Research Data at the University of Bern.
The following information will be provided:
name of the study, name of the PI, dates of study start and end, contact address, location of data, number of data sets overall and in subgroups, access conditions, keywords: breaking bad news, psychophysiological stress responses, arousal reappraisal, aim of the study
All publications will be registered within BORIS at the University of Bern.

Data documentation
We will document our research on the project level, file or database level, and variable or item level. The documentation on the project level will be given in the project proposal (for research questions, hypothesis, methodology and instruments, etc.), laboratory notebooks & experimental protocols, questionnaires, information about equipment settings & instrument calibration, database schema, methodology and technical reports.
To document our research on the file/database and variable/item level, we will use codebooks and data dictionaries, Stata software syntax, output files, a global driving and readme file and DMP.

Metadata

Readme XML file

Each database will be accompanied by a readme XML file, ensuring the understanding of the labelling and coding of variables as well as their operationalization. A unique Stata driving .do file will document how the various databases are related to each other and show the creation of the final database from scratch. The .do file will be saved as Readme XML file in order to ensure readability independent of the available computer software programs. This file will account for all the files and folders in the project and help ensure that the data can be correctly interpreted and reanalyzed by other researchers and adds contextual value to the datasets for future publishing and data sharing. The Readme XML file ensures compatibility with international standards and is human as well as machine-readable.

It will be written in plain text and will contain each filename, a short description of what data the file includes, creator's name and affiliation with contact information. Whenever adequate and possible, we will include key words, date of data creation, link to electronic notebook and associated datasets saved elsewhere, details on the methodology used and data processing steps (especially if not described in the publication), definitions of variables, vocabularies and units of measurement. Optionally included might be information on the size, format, version, access and funding.

Metadata for publishing datasets on nonprofit unstructured data repositories such as Zenodo or Dryad

➢ We will use similar standard XML metadata (DataCite Metadata Schema or Dublin Core) to publish and share unstructured datasets. The XML metadata that ensures machine readability / interoperability is generated after filing the repository submission form requesting intrinsic metadata. Metadata comprise, in addition, a persistent identifier, a publication date and conditions of access to (type of license) the dataset.

➢ We will upload, alongside the dataset, Readme files in text and/or XML (see previous paragraph) formats with more detailed information.

2 Ethics, legal and security issues

2.1 How will ethical issues be addressed and handled?

Before the start of data collection, we will search ethics approval from our human ethics committee (Kantonale Ethikkommission Bern), which comply with the regulations of Declaration of Helsinki. Health risks associated with our research protocol are minimal.

Participants will be asked to give an informed consent to the study participation after receiving written and oral information on the study content and procedures. Questionnaire, performance- and physiological-data will be treated confidentially and will be made anonymous.

For research data sharing, socio-demographic data will be made publicly available as long as they do not allow identifying

the participants. They might be presented as categories rather than as continuous variables (e.g. for age) in order to guarantee anonymity.

2.2 How will data access and security be managed?

Data access rights/permission to ensure the security of the data:
The data access rights will be given to the applicant, project partners, and researchers directly involved in the project (PhD student, research assistant).

The raw video data to be evaluated during the project will be collected on a password protected workstation at Institute for Medical Education. Video data is stored on a password protected server that is not accessible from outside (server is inaccessible from the internet). Psychophysiological data, that is generated during the experiments will only be identifyable by the study's participants ID and will also be stored on password protected server at the IML that is accessible for the project partners in Lausanne and Vienna. Participants' performance data are collected with Examic EOSCE software on tablets. All data is encrypted and signed on the client (tablet) with a RSA-512 bit key before being sent to Amazon S3 servers over a secure connection. Encryption applied is identical to the one used for data processed in national licensing exams. Electronic transfer or transfer with portable media from these collecting systems will comprise strong file encryption and safe key transfer. Workstations (to run statistics with SPSS, Excel etc.) used for the project are linked to a RAID system and data backups will be made every night.

2.3 How will you handle copyright and Intellectual Property Rights issues?

Intellectual property for datasets
According to the University of Bern contract, any data created or modified in the course of our professional activity as a University of Bern collaborator belongs to the University of Bern (except for copyrights belonging to the creator). This means that for any use other than scientific and academic purposes, University of Bern approval is necessary (e.g., for patenting and commercialization). We do not foresee any other than scientific and academic purposes in this project.

Open licenses for data
Once our analyses will have been completed and the manuscripts published, we will promote sharing and unlimited use of the data that we produced using http://opendefinition.org/licenses/. For sharing our data, we will use a creative common CC0 license as recommended by the University of Bern or a CC By license that is suitable for data sharing. The CC0 license is a 'public domain dedication', i.e. a waiver of all our rights including those of attribution. The CC By license allows others to distribute, remix, tweak and build upon our work, even commercially, as long as they credit us for the original creation. URL link: http://opendefinition.org/licenses/

3 Data storage and preservation

3.1 How will your data be stored and backed-up during the research?

The Institute for Medical Education (IML) owns an infrastructure including a local cloud with 100+ servers that share the existing resources. The IML provides appropriate protection of hosted data on its technical infrastructure, including restricted physical access, confidentiality and integrity of the data.
The virtual machines use a storage cluster. All server data of productive systems have a daily backup.

Filesharing/ collaboration
- We use the OpenSource Nextcloud that provideds acces to a project specific file share and collaboration tools (see https://nextcloud.com/).
- File and application access requires authentication with username and password
- Clients encrypt all communication (https)
- Port 443 (https) is open to enable sharing / collaboration with external partners

Data backup and safeguarding:
- The versioning of any saved file is a built-in feature of Nextcloud. A user can restore a wanted older version of a file without help of an administrator
- Storage
  ◦ The disc image of the Nextcloud instance is mounted on an OpenSource disc storage Ceph (https://ceph.io/)
  ◦ the storage has a redundancy of 3 copies of each data block of a file to prevent data loss as a result of damage of a disk or server.
- Backup (this is the default for all productive virtual machines):
  ◦ Daily file backup: 1 Full backup per month; incremental backups on other days. Retention 3 months
  ◦ Files will be encrypted by 128 characters long password locally on the server with OpenSouce software GnuPG (https://gnupg.org/)
  ◦ Validity of backups will be checked in regular intervals
  ◦ Encrypted backup data is stored on a storage located in another building

3.2 What is your data preservation plan?

Long-term data archiving: we will archive collected data on the (University of Bern-NAS) research storage space at least 10

years after publication.

Appropriate data archiving is the responsibility of the Principal investigator.

Published data:
Deposition and open data via Zenodo or Dryad repositories will ensure longevity of the data in the long-term. In addition, University of Bern-NAS (see 3.1) retains primary and secondary research data supporting published articles for at least 10 years after publication.

Unpublished data:
We will save unpublished high-quality final data generated during this project for re-use in our future projects. We will make some data available for use by researchers in future collaborations if our lab no longer uses it. Bad quality data will be permanently discarded at the end of the project.

4. Data sharing and reuse

4.1 How and where will the data be shared?

Data sharing at the latest at the time of publication, via non-profit digital repositories

Repositories making data available.
We will make supplementary files and key datasets accompanying a publication to demonstrate reproducibility openly available in appropriate digital data repositories that conform to the Fair Data principles and maintained by a non-profit organisation.

We will share specific datasets via domain-specific public repositories.
Unstructured data will be shared via data repository Zenodo or Dryad. These data repositories fulfill biomedical journals' and SNSF's requirements (allowing publishing FAIR data, non-commercial).

How we will make data available?
We will use specific formats for data sharing to ensure a file's preservation and re-usability (see table DMP1.1). In some cases, we will upload the "Original" file along with the converted file.
Datasets will be given a Digital Object Identifier (DOI) and associated metadata. The DOI corresponding to the datasets in the repository will be included in the article's reference list, allowing identification and access of any dataset in a publication.

Visibility and valorization of datasets

We will also link DOIs to appropriate records in the University's publication repository Boris, to enhance the dataset's visibility. Metadata about datasets will be publicly searchable and discoverable and will indicate how and on what terms the datasets can be accessed.

How will reuse of our data be valued?

We will share data using CC0 or CC BY licenses that will become citable products of research.

4.2 Are there any necessary limitations to protect sensitive data?

We do not anticipate that this study will generate patentable data.

The principal investigator will decide when to publish and make Research Data accompanying the article publicly available including whether to supply research data to a new user.

Use of the data and restrictions to data sharing

- This project will produce sensitive human personal data and will necessitate specific precautions and limitations for data sharing.
- We will share sensitive human data very carefully specifically due to legal, ethical and confidentiality issues. Therefore, socio-demographic data will be made publicly available as long as they do not allow identifying the participants. They might be presented as categories rather than as continuous variables (e.g. for age) in order to guarantee anonymity.
- Video recordings are not shared, because the confidentiality would be violated.
- We will share data between project collaborators without limitation.
- We may make encoded data available before publication upon demand by potential new collaborators.
- We will make anonymized datasets openly available on appropriate digital data repositories (see 4.1) at the latest at the time of publication.
- We may make data available before publication upon demand by potential new collaborators.
- We will put in place restrictions on data sharing to ensure sensitive data protection if we are unable to assure perfect data anonymization. In this case the data will be stored on the NAS at our institution to insure perfect data protection.

4.3 I will choose digital repositories that are conform to the FAIR Data Principles. YES

4.4 I will choose digital repositories maintained by a non-profit organization. YES