
Seminar 2

Field specific research ethics.

Fredrik Gustafsson (fredrik.gustafsson@it.uu.se)
Eduardo José Gómez Hernández (eduardojose.gomez@um.es)
Djurdjija (Durdija) Dzodan (djurdjija.dzodan@kemi.uu.se)
Tuan Anh Dao (tuananh.dao@it.uu.se)
Benjamin Weber (benjamin.weber@it.uu.se)

Background

- We have access to a very large medical dataset of brain MRI scans.
- The dataset can be "anonymized" by removing all names, national identification numbers etc. and leaving only the brain MRI scans themselves.
- If publicly released, this "anonymized" dataset is believed to hugely benefit the scientific community and lead to, e.g., much improved cancer screening methods.
- However, we do not have permission from the patients to publicly release this data in any form.

Dilemma

Should we still release the "anonymized" dataset?

Does the benefit to the scientific community and the potential to help save lives outweigh the privacy of the patients?

1. Release the entire "anonymized" dataset

Dataset Size

✓ All available data is included, dataset size is maximized.

✗ The dataset can be treated as compromised data, and thus never be used in practise by other researchers. We might be told by authorities to take the dataset down.

Scientific Community

✓ Immediate increase of the medical sample databases, implying possible new knowledge, opens new research possibilities.

✗

Privacy

✓ The dataset is basically anonymized, it should be very difficult to identify an individual from his/her MRI scan.

✗ The current data could let identificative traces behind of the patients. Even if they can't be identified, patients might NOT want their MRI scan to be included in a public dataset.

Future Patients

✓ The released data can encourage a lot of research and help introduce medical improvements earlier (and possibly save lives).

✗

1. Release the entire "anonymized" dataset

Our Reputation

- ✓ Can increase our reputation as we are contributing to public databases with new samples. Can get a high-impact publication.
- ✗ Can be treated as a negligence action, preventing future research to be taken seriously. Might lose our jobs as researchers.

Financial Cost

- ✓
- ✗ Might be sued by patients. Might be fined. Might lose our jobs.

Required Time

- ✓ Can release the dataset more or less immediately.
- ✗

2. Request permission from all patients, release data only for patients who grant permission

Dataset Size

✓ A large number of patients (e.g. 90-95%) might say yes, if we have good arguments for why the dataset will be useful to researchers.

✗ Only a small number of patients (e.g. 5%) might say yes or even respond, thus resulting in a much smaller dataset.

Scientific Community

✓ Immediate increase of the medical sample databases, implying possible new knowledge, opens new research possibilities.

✗ A small dataset would be much less useful.

Privacy

✓ A patient's MRI scan is included in the dataset only if he/she has given us permission to do so.

✗

Future Patients

✓ The released data can encourage a lot of research and help introduce medical improvements earlier (and possibly save lives).

✗ A small dataset would not encourage as much research, might not get any improved medical procedures.

2. Request permission from all patients, release data only for patients who grant permission

Our Reputation

- ✓ We would be considered serious researchers who follow good research practice.
- ✗ A small dataset would be less useful to other researchers, and our work would thus be less impactful.

Financial Cost

- ✓
- ✗ Might be very expensive, need to contact 10,000s of people.

Required Time

- ✓
- ✗ Might be very time consuming.

3. Request permission from a small random subset of patients (100-1000 patients), release the entire dataset if and only if a significant majority (e.g. 75%) of patients grant permission

Dataset Size

- ✓ We might be able to release all available data, maximizing the dataset size.
- ✗ Might not be able to release any data whatsoever.

Scientific Community

- ✓ Immediate increase of the medical sample databases, implying possible new knowledge, opens new research possibilities.
- ✗ No data whatsoever might be released.

Privacy

- ✓ The dataset is basically anonymized, it should be very difficult to identify an individual from his/her MRI scan. It seems like a majority of the patients doesn't have any privacy concerns.
- ✗ The current data could let identificative traces behind of the patients. Even if they can't be identified, patients might NOT want their MRI scan to be included in a public dataset.

Future Patients

- ✓ The released data can encourage a lot of research and help introduce medical improvements earlier (and possibly save lives).
- ✗ No data whatsoever might be released, thus no future benefits.

3. Request permission from a small random subset of patients (100-1000 patients), release the entire dataset if and only if a significant majority (e.g. 75%) of patients grant permission

Our Reputation

- ✓ Can increase our reputation as we are contributing to public databases with new samples. Can get a high-impact publication.
- ✗ Can be treated as a negligence action, preventing future research to be taken seriously. Might lose our jobs as researchers.

Financial Cost

- ✓
- ✗ A small cost to contact the subset of patients. Might be sued by patients. Might be fined. Might lose our jobs.

Required Time

- ✓
- ✗ Somewhat time consuming to contact the subset of patients.

4. Do not release any of the data, instead start requesting permission from all future patients, and gradually release data only for those who grant permission

Dataset Size

- ✓ The dataset can eventually grow very large, even larger than the currently available data.
- ✗ The currently available data cannot be released, the initial versions of the dataset will probably be very small in comparison.

Scientific Community

- ✓ Small, but constant, increase of medical samples database
- ✗ The small initial datasets would not be particularly useful.

Privacy

- ✓ A patient's MRI scan is included in the dataset only if he/she has given us permission to do so.
- ✗

Future Patients

- ✓ If the dataset eventually becomes large enough, it will encourage a lot of research and help improve medical improvements.
- ✗ This new dataset of properly gathered data can be very small, and thus fail to encourage a lot of research.

4. Do not release any of the data, instead start requesting permission from all future patients, and gradually release data only for those who grant permission

Our Reputation

- ✓ We would be considered serious researchers who follow good research practice.
- ✗ The dataset could be too small to be relevant, having less impact.

Financial Cost

- ✓ Would probably not add much cost to the previously used data collection procedure.
- ✗

Required Time

- ✓
- ✗ Might take a very long time to build a dataset of decent size.

5. Request permission from a subset of patients + start requesting permission from all future patients, and gradually release data only for those who grant permission

Dataset Size

- ✓ Can get a quite large dataset that eventually can grow very large, even larger than the currently available data.
- ✗ Only a subset of the currently available data can be released, the initial versions of the dataset will probably be quite small.

Scientific Community

- ✓ Immediate quite large increase + small, but constant, increase of medical samples database
- ✗ If too small, the initial datasets would not be particularly useful.

Privacy

- ✓ A patient's MRI scan is included in the dataset only if he/she has given us permission to do so.
- ✗

Future Patients

- ✓ If the dataset eventually becomes large enough, it will encourage a lot of research and help improve medical improvements.
- ✗ This new dataset of properly gathered data can be very small, and thus fail to encourage a lot of research.

5. Request permission from a subset of patients + start requesting permission from all future patients, and gradually release data only for those who grant permission

Our Reputation

- ✓ We would be considered serious researchers who follow good research practice.
- ✗ The dataset could be too small to be relevant, having less impact.

Financial Cost

- ✓ Would probably not add much cost to the previously used data collection procedure (after the initial stage).
- ✗ Small-medium cost to contact the subset of patients.

Required Time

- ✓
- ✗ Might take a quite long time to build a dataset of decent size.

6. Try to prove that the data should not be considered sensitive, release the entire dataset if and only if this is successfully proven

Dataset Size

- ✓ All available data is included, dataset size is maximized.
- ✗ Might not be able to release any data whatsoever (if we are unable to prove it).

Scientific Community

- ✓ Can make other researchers to release their own private datasets, as they are not sensitive anymore.
- ✗ No data whatsoever might be released.

Privacy

- ✓
- ✗ Even if they can't be identified, some patients might NOT want their MRI scan to be included in a public dataset.

Future Patients

- ✓ As much data can be released, if we are able to proof this, more research is encouraged, helping to introduce medial improvements.
- ✗ No data whatsoever might be released, thus no future benefits.

6. Try to prove that the data should not be considered sensitive, release the entire dataset if and only if this is successfully proven

Our Reputation

- ✓ Proving this, can increase our reputation as we removed the hustle of requesting permission.
- ✗ Can be viewed as we don't care about privacy issues, especially if we fail to prove it.

Financial Cost

- ✓
- ✗ Needs more research time, and including other research fields, increasing the project's cost.

Required Time

- ✓ Can save time in future data gathering.
- ✗ Medically prove this, can be very time consuming, is not a trivial task.

Best Possible Solution?

5. Request permission from a subset of patients + start requesting permission from all future patients, and gradually release data only for those who grant permission.

Discussion

Should we publish the "anonymized" dataset publicly even when it was obtained un-properly? Given that huge benefits to the scientific community are expected.

- Yes
- No

Assuming an extreme situation, for example that the world is facing a global pandemic and this dataset is believed to play an important role in the safe development of an efficient vaccine.

If it is not released soon, the world could end up in a much worse situation. Should we still prioritize the patients' privacy?

- Yes
- No