CHECKLIST:
Reducing Re-Identification Risk in Traumatic Stress Research Data

From the Global Collaboration on Traumatic Stress FAIR Data Workgroup

A. Address content of dataset

☐ Remove direct identifiers

☐ If required by your national / regional regulations, replace subject identification numbers with a new set of identification numbers in a different random sequence

☐ Remove free text containing personally identifiable information

    ☐ If needed for analytic value (data re-use), recode information into aggregated categories

☐ List quasi-identifiers

    ☐ Considering the context of your study, list those of greatest potential concern (alone or in combination) for re-identification risk

B. Assess potential for harm if participants were re-identified

☐ Were participants promised confidentiality? (true for nearly all trauma research)
- Potential for harm to trust in research, researchers, institutions.

☐ Would disclosure of study participation or of specific attributes cause harm or stigma?

    ☐ List items of greatest concern for harm / stigma if connected to identified individuals

    ☐ Characterize level of harm if disclosed

- Low risk – Would not cause more than minimal harm / would not be of use to attackers.
- High risk – Potential to embarrass or otherwise cause harm to respondents.

C. Assess relationship between sample (cases in dataset) and population (reidentification frame)

☐ Potential to construct a reidentification frame (Could someone construct a list of people who might be in the study / dataset?)

☐ Estimated size and nature of re-identification frame?

☐ Relationship of sample size to re-identification frame

In relation to the re-identification frame, cases in dataset represent:	Risk related to sample size / proportion
Complete or nearly complete sample of the re-identification frame.	Very high
More than 10% sample of re-identification frame (10:1 ratio).	High
Between 10% and 1% sample of re-identification frame.	Medium
1% or less sample of re-identification frame (100:1 ratio).	Low
0.1% or less of re-identification frame (1000:1 ratio or higher)	Very low

D. Conduct systematic analysis to assess and mitigate risk within the dataset

☐ Initial inspection for k-anonymity

☐ Penetration testing – check univariate and bivariate combinations of select quasi-identifiers for outliers, unusual cases, and small groups

☐ Use the above to make recommendations for changes to improve anonymity

    ☐ Data reduction – generalize, suppress, delete specific variables?

    ☐ Balance data reduction with analytical value for data re-use
(consider more restrictive sharing / access if data reduction harms data usefulness)

☐ Reassess after changes.

E. Consider how data will be shared / made accessible for re-use.

Select method that is compatible with re-identification risk assessment as well as funder and legal / regulatory requirements

☐ High potential for harm and high relative risk based on sample relationship to reidentification frame? → Implement more restrictive conditions for data sharing and re-use.

☐ For all data sharing methods: Inform data users of their responsibility to respect participant privacy and to inform the original research team or data repository in the case of any inadvertent re-identification.

Range of possibilities includes:

- Openly available data, with stated terms of use for public download.

- Data in repository that is accessible upon request / application. Might require user registration with written / digital agreement to terms of use.

- Restricted data with greater access controls. Might require formal inter-institutional agreements that include terms of use for data.

RESULTS & RECOMMENDATIONS

Study / Dataset name:

Theoretical study population	[Describe target population]
Potential harm	**Relative risk of potential harm IF re-identified = high / medium / low**
Potential to construct a re-identification frame	[Is it possible to construct a list of individuals who could have been invited to participate in study? Describe.]
Relationship of survey sample to reidentification frame	[Sample N]| [estimated size of re-identification frame] **Relative risk based on sample proportion = high / medium / low**
Direct identifiers	[List any that remain in dataset – will they be deleted?]
Free text with personally identifiable information	[List any that remain in dataset –will they be deleted? recoded?]
Select demographic quasi-identifiers of concern	Variables of potential concern: • [List here]
Initial inspection for k-anonymity	[Is it possible / appropriate to assess formally for k-anonymity? -- based on sample size and number of quasi-identifiers of concern] [Summarize results of k-anonymity testing]
Penetration testing – univariate and bi-variate checks of select quasi-identifiers for outliers and small groups	[Describe results of penetration testing] **Relative risk: [Were unusual cases, outliers, high-risk combinations found?]**
Recommendation	[List specific action recommendations here] **Assessment: [Can risk be reduced if recommendations implemented?}** **Assessment after changes are made:**
Appropriate means of data sharing / access	[Describe options for method(s) of making data accessible that fit *re-identification risk assessment*, as well as *funder & legal / regulatory requirements*] • Publicly available // Repository with clear access options // Other [Describe how users will be informed of and agree to terms of use that protect against re-identification]

Version 1.0 – October 2024