Sharing Longitudinal Data in Traumatic Stress Research: What Changes When You Add Time

A quick guide to how sharing longitudinal data differs from sharing cross-sectional data, and what to consider before you share.

The Challenge

Cross-sectional datasets capture a person once. Longitudinal datasets follow the same people across time, and that repetition changes more than the amount of data you have. It changes the privacy profile of the data. Time creates patterns, and patterns can create sources of identification. Many de-identification habits that work well for a single cross-sectional dataset do not transfer cleanly to longitudinal data measures. Thus, it is important to think about what is different before you share.

Repeated Values as an Identifying Signature

The central shift is that repeated values become a **signature**. A participant's sequence of PCL-5 scores across waves, or the shape of their recovery trajectory, can be nearly unique. These temporal signatures can be combined with stable characteristics (sociodemographics, trauma type on the LEC-5, occupation) to create a **mosaic effect**, such that when taken together, the multiple waves of longitudinal datasets can reveal more than any single wave. In practice this means a record that looks safe at one wave can become unique across the full series, and demographic fields that carry acceptably low risk cross-sectionally may need further generalization once they are tracked over time. It also means longitudinal data is, by design, pseudonymous rather than truly anonymous: you need a stable link between a person's waves, and that link is itself something to protect.

Time as an Additional Identifier

Time also enters the data as **dates and intervals**: assessment dates, time-since-trauma, the timing of an index event. These are quasi-identifiers, and a cohort tied to a specific, datable event (a named disaster, a single deployment) can be narrowed by the calendar alone. Because some trauma cohorts are small and event-defined, cross-tabulating several waves shrinks subgroups quickly. **Attrition** compounds this. Dropout in trauma samples is rarely random, since factors like baseline trauma symptoms severity or comorbid substance use tend to predict it, so who remains across waves is itself informative, and the group of completers can become small and more potentially identifible.

Accumulation of Identifiers and Measurement Changes

Two further considerations are specific to following people over time. First, **identifying information accumulates**. Each wave can add disclosures such as a new trauma, suicidality, or substance use, so a re-identified longitudinal record is not only more likely to be unique, it is also more harmful if exposed. Second, **instruments often change** across a long study (for example CAPS-IV to CAPS-5, or PCL for DSM-IV to PCL-5), which creates a small harmonization task inside a single dataset and affects how you document it for reuse.

Questions to ask before sharing your longitudinal data

1. Could the *sequence* or *trajectory* of a participant's scores be distinctive, even when no single score is?
2. Have you assessed re-identification risk across the full series of waves, not just one wave at a time?
3. Which variables stay fixed over time (anchors) and which vary (signature), and what might they reveal in combination?
4. Do any dates or intervals (assessment dates, time-since-trauma, event timing) narrow down who a participant could be?
5. Is your cohort tied to a specific, datable, or nameable event or population that makes the pool of possible participants small?
6. Is dropout in your sample non-random in ways that could make the remaining participants identifiable?
7. How much sensitive information accumulates per person across waves, and what is the potential harm if a record is re-identified?

Part of the Toolkit for FAIR Traumatic Stress Data. See also the "Checklist for Reducing Re-Identification Risk in Traumatic Stress Research"