

Notes from discussions at the data harmony KEW - 16th January 2013

Feedback from breakout sessions

Group 1

When carrying out harmonisation, need to answer a specific question rather than harmonise for the sake of it.

Training and capacity building should use real examples to demonstrate a range of principles for harmonisation.

Published papers should clearly explain the harmonisation and analysis process.

An additional benefit of harmonisation is that commonly used variables can supplement gaps in cohort data and provides an improved quantitative statistical aspect.

Lessons from current work across studies should be fed into future surveys.

Group 2

The level of harmonisation may depend on the question being addressed and researchers should be helped to understand why this is the case.

As harmonised datasets need to be produced, there may be a need for multiple versions with different levels of harmonisation. A guide (possibly using a flowchart) would be required to help researchers decide which datasets to use for which questions.

Harmonisation should be carried out so that it is possible for other cohorts to carry out the same procedure. This requires production of detailed documentation and syntax in multiple packages.

There is a need to consider the role of CLOSER in calibration studies e.g. of different measuring devices. We could consider applying for use of the innovation panel of Understanding Society to help with, for example, calibration studies.

Group 3

There is a need to advertise what is available in the cohorts, e.g. biosamples resource. This could be done through, for example, giving papers at conferences and workshops linked to conferences.

Training and capacity building could take the form of intensive data analysis workshops modelled on Friday Harbor (<http://alzheimer.ucdavis.edu/fhpsych/>). This is a helpful model because there are clear outputs (i.e. written papers) as well as the TCB function. Also it can help to foster a community of researchers working on similar topics and/or datasets. The Friday Harbor workshops are for advanced researchers and if a model was to be used with less experienced researchers there would need to be lots of preparation of datasets in advance. One model is to have a single research question and investigate it using multiple data resources or to focus on a single resource and explore different research questions.

It would be very helpful to document the best measures within a cohort and provide the syntax for the derivation of these variables. In particular this is useful for variables that are frequently used by those from a different discipline, e.g. epidemiologists would find it helpful to know which is the best measure of parental social class to use in some of the older cohorts. When something is measured on several occasions during childhood but only a simple summary measure is needed it would be helpful for users to have guidance on the best summary measure to use.

It is important to demonstrate why harmonisation matters i.e. demonstrate via sensitivity analyses. Users need to understand what the practical implications will be of harmonising or not harmonising variables across studies or across sweeps – i.e. what difference would it make to substantive results.

Ideally we will also get the research community to share syntax they have created for harmonising variables – however for this to be effective it needs to be possible to understand other people's syntax/do files and so some standards for documentation of variables would be a helpful output from CLOSER. This could also include a checklist for creating a new variable.

Group 4

Distinguish between harmonisation as abstract versus research question. Need to produce harmonised variables (e.g. SEP) as control variables – then can harmonise without a scientific question.

There needs to be some quality control when harmonising to the lowest common denominator. What is the lowest quality threshold? The lowest common denominator is likely to be different for different purposes and we may decide to not try to harmonise across some data sets if the quality threshold becomes too low.

Consider an international perspective when harmonising data. Consider how relevant variables can be harmonised internationally.

Need for experts in relevant disciplines. Consider how to involve these groups and consider need to give them specific tasks. Specifically user (researcher) involvement will be critical and it will need to be inter-disciplinary.

We thought working to a common template at an early stage would also ensure that when harmonising we were considering the same issues as other groups harmonising different types of variables.

We agreed that the SES harmonisation would include Understanding Society and would in no way just be restricted to the cohort data sets.

Discussion

The Friday Harbor type workshop was discussed and supported. This would be a residential workshop, focus on a particular topic where the aim is to produce a set of papers. It would bring together methodologists, cohort researchers and experts in the substantive topic. This could be tried at tier 2 first before opening to a tier 3 event. It was noted that this type of event requires good preparation to be successful. The idea of an intensive residential workshop could also be applied at tier 1 level. It was agreed that his idea would be investigated and pursued further.

It was agreed that the “tagging” of preferred variables within studies was a good idea to be pursued further.

The idea of promoting use of the CLOSER cohorts could be done through CLOSER symposia at conferences.

Other issues raised during the day

Level of harmonisation:

What do we gain by additional harmonisation when we can carry out harmonised data analysis across multiple cohorts?

Consider balance of harmonisation to the lowest common denominator versus the richness within the individual studies.

Need to be careful not to harmonise out any interesting cohort effects.

Need to consider the composition and sampling in studies and possible harmonisation of the population samples.

Calibration:

Do we need calibration studies? Consider the use of external datasets for validation.

Acknowledge that methods of analysis for biological structure and function and biosamples change. Analysis required to assess comparability.

Should we model “corrections” or add/subtract a constant from data? This is a balance between the expertise of the user and ease of use.

Wider CLOSER issues:

Consider the co-ordination of the use of expert groups across the WPs.

Contextual information to help inform about, for example changing SEP context, will be part of the CLOSER USP.

Ethics and legal issues are not part of a CLOSER WP, but could be considered for discussion at the leadership team meetings and as the subject of a future CLOSER symposium.