**Risk analysis for the export/publication of semi-aggregated data (n < 10)**

**Aim:**

This document aims to instruct workspace/HPC users how to assess the risk of semi-aggregated data to (inadvertently) reveal the identity or expose results of individual Lifelines participants.

**Background:**

Lifelines normally does not export individual data (i.e. results organized by pseudonimized participant) outside of our own secure infrastructure. This has several reasons:

1. Lifelines has agreed with participants (via the Informed Consent) that collected data and samples are only used for the purpose of research in the field of healthy ageing (“purpose limitation” in the sense of the GDPR).
2. Lifelines makes careful ethical decisions regarding the return of personal results to participants, especially if they a) may cause worry and stress and b) are not clinically valid (which is often the case in research, for example the risk score to develop a disease).

To minimize the risk that personal information about participants (both contact information and health-related data) is used for the wrong purposes (i.e. commercial or political), or that participants learn about (unwanted) personal results via a publication, Lifelines functions as a gatekeeper and only releases personal data to users in a secure environment and under a signed contract (DMTA).

Lifelines allows the export of data outside of the LL infrastructure in the following cases:

1. Individual data may be exported for a specific project IF there is a valid scientific reason for the request (i.e. the regular infrastructure is not compatible with the required analyses) and IF the data is placed in a secure environment with similar safety precautions as the normal infrastructure (i.e. researchers take over the role of gatekeeper). In these situations an Individual Data Export Request must be filled in and a Code of Conduct must be signed on top of the DMTA;
2. Aggregated data (in tables with groups of n > 10 participants) may be exported and publically shared without any limitation, for example in peer-reviewed papers;
3. Individual participants may request to receive their own results (i.e. data they provided to Lifelines).

**Export of semi-aggregated data**

It is possible that you need to export (& publish) semi-aggregated data, i.e. data organized in very small participant groups (n = 1-9). If you have such a situation and you have a scientifically valid argument to share or publish these data, Lifelines will perform a more in-depth content analysis to assess the risks described above.

You can help us by doing this yourself in advance, following these instructions:

1. Identify all results for participant groups of n < 10
2. For each of these participant groups, note down the following characteristics:
	1. All selection criteria, i.e. characteristics that participants need to have in order to be part of the group. This starts with “all persons in this group are Lifelines participants” and could include things like “all persons in this group are female”, “all persons in this group filled in questionnaire 1B”, “all persons in this group have blood levels of factor x above 3.4” or “all persons in this group have mutation y”.
* The result itself, i.e. the characteristic that you discovered in your research that is true for this group of participants. In case of a simple frequency distribution, there is no result.
1. Now look at the list of selection criteria and decide, per criterium, in which of the following three categories they fall:
	1. It is only possible for you, the researcher, to know which (pseudonimized) participants adhered to this criterium (for example: blood levels that you analysed, gene mutations you discovered, complex sum scores you developed, or part of a linkage that only you have access to).
	2. It is also possible for participants themselves to know if they adhere to this criterium (for example: having a certain disease or lifestyle habit, or participation in a specific Lifelines assessment).
	3. It is also easy for “outsiders” to know who adheres to this criterium (for example: rare ethnicities or professions, extreme length/weight or old age, sparsely inhabited postal codes – especially when these are combined). Keep in mind that participants often share things on social media.
2. Then look at the result characteristic, if any, and determine if this is already known to participants (i.e. it is a characteristic that people generally know about themselves, or Lifelines has the policy to return the result to participants), or not. When in doubt, ask us. Keep in mind that participants may be highly educated in a medical field and able to interpret scientific results!
3. With all characteristics categorized, use the following flowchart to assess the risk of exposure or identification:

****

**Outcome:**

If you decide, based on the flowchart, that your export is low risk, then please present the table and your analysis to Lifelines so that we can double check.