

Data Sharing Agreement for Participation in a Statistical Analysis Commons as part of the CHARGE (Cohorts for Heart and Aging Research in Genomic Epidemiology) Consortium

enter MEMBER study name or USER institution name here

Note: This agreement was developed from the template provided by the NHLBI, version 8/23/2013.

Introduction

The name of the data sharing activity is the Statistical Analysis Commons Consortium. It is an activity associated with the CHARGE consortium and is serving as a pilot study of the TOPMed Program for the analysis of whole-genome sequence and other -omics data. Members of the Statistical Analysis Commons Consortium may or may not be a member of the CHARGE consortium, but they must agree to follow the principles set out by the CHARGE consortium. Additional collaborations are likely to occur in the future. The CHARGE consortium has developed agreements for collaborations, working groups and data sharing. The key documents, which are included by reference, are:

- CHARGE Principles: Principles for collaboration in the consortium, approved Jan 2008 by founding cohorts.
- CHARGE Working Group Principles: Principles for collaboration in CHARGE working groups, new participants in working groups or CHARGE projects are asked to review and agree to these principles before joining.
- Aggregate Results-Sharing Principles: Guidelines for sharing results of GWAS, for participation in working groups and collaborative meta-analyses.

These principles were developed before the concept of the Statistical Analysis Commons was formulated. As needed, they may be updated with the implementation of the Statistical Analysis Commons. These documents can be found on the CHARGE public website: <http://www.chargeconsortium.com/main/Consortium-Documents>

ONGOING UPDATES WEBSITE:

To avoid repeated revisions of this Data Sharing Agreement, updated lists of Statistical Analysis Commons members, components, participants, users, authorized accounts, member's data transfer groups, phenotypes, consent groups, secure data location site(s), and data coordinating center(s) are maintained at the **ONGOING UPDATES WEBSITE:**

<http://www.chargeconsortium.com/commons>

A brief description of the Consortium's purpose.

The purpose of the consortium is to foster collaboration among well-phenotyped studies that have both extensive genotype and phenotype data and to improve the discovery of genetic variants associated with a variety of phenotypes. The primary aim is to conduct high-quality analyses that produce timely, reliable and valid findings across multiple cardiovascular and aging-related phenotypes. CHARGE and other consortia have traditionally used meta-analysis to combine study-specific results across participating studies. The genotype data include not only whole exome sequence and whole-genome sequence data, but also genome-wide single-nucleotide polymorphism (SNP) and exome chip SNP data. For some analyses of these large-scale data, especially permutation testing, the use of pooled data from collaborating studies will be simpler and may provide more efficiency and ease of use than the traditional method of meta-analysis. This agreement will facilitate the creation of pooled data for the Statistical Analysis Commons. This Statistical Analysis Commons also serves as a pilot study of the TOPMed Program for the analysis of whole-genome sequence data.

The studies contributing data to the Statistical Analysis Commons are defined as “MEMBERS.” The MEMBERS or parties to this agreement include:

1. The Atherosclerosis Risk in Communities (ARIC) Study
2. The Cardiovascular Health Study (CHS)
3. The Framingham Heart Study (FHS)
4. The Old Order Amish Study (OOA)
5. The Jackson Heart Study (JHS)
6. The Multi-Ethnic Study of Atherosclerosis (MESA)
7. The Rotterdam Study (RS)

Over time, other studies are expected to join the Statistical Analysis Commons as MEMBERS or as USERS (centers that develop methods or analyze data) who agree to participate and abide by this agreement. A current list of MEMBERS and USERS is available at the ONGOING UPDATES website.

A description of the structure of the Consortium.

The Statistical Analysis Commons was created to develop methods and a setting for pooled analysis of large-scale genotype-phenotype data. It is guided by a “Commons Steering Committee,” which includes one voting member for each MEMBER study contributing to the Statistical Analysis Commons.

New MEMBERS and USERS join the Consortium after approval by the Commons Steering Committee.

The Statistical Analysis Commons also includes a data coordinating center (DCC) and serves as a pilot study for the TOPMed Program. Transparency, disclosure, and communications about all collaborations, additional follow-up experiments, or efforts to obtain additional funding are essential to developing, ensuring, and maintaining trust within the consortium. The University of Texas Health Science Center at Houston (UT) has agreed to serve initially as the DCC for the Statistical Analysis Commons Consortium. (See ONGOING UPDATES WEBSITE for current DCC.) The DCC will serve under the guidance of the Commons Steering Committee.

Each MEMBER must submit a copy of the signed Consortium data-sharing agreement to the DCC. The Statistical Analysis Commons Consortium will consist of those studies that have submitted their copy of the executed data-sharing agreement to the DCC.

The Statistical Analysis Commons ONGOING UPDATES WEBSITE maintains a complete list of MEMBER studies and USERS that will have access to the pooled data. The MEMBER Studies will maintain a list of individuals approved to access the Statistical Analysis Commons, and they agree to provide updated lists to the Statistical Analysis Commons website.

Data and plans for data distribution in the Statistical Analysis Commons.

Study MEMBERS are responsible for harmonizing the phenotype data. The phenotype data available in the Statistical Analysis Commons, including any participant-consent related restrictions on data use (if applicable), are listed on the ONGOING UPDATES WEBSITE.

The genetic data to be included in the Statistical Analysis Commons include available genetic data, including array assays (GWAS, exome chip), exome-sequence data, and whole- genome sequence data as well as other array or -omics data such as methylation, RNA, expression, or metabolomics. For ease of use, we nonetheless refer to “genetic” data throughout.

The designated data-transfer group from each study, typically but not necessarily the cohort’s coordinating center, will be responsible for transferring both the genotype and the harmonized phenotype data to the Statistical Analysis Commons or to DCC at the UT. The list of MEMBERS’ data-transfer groups are available at the website for ONGOING UPDATES (provided on page 1 of the agreement).

In addition to the harmonized data, data elements will need to include any participant-consent related restrictions on the use of the data. As needed, the Statistical Analysis Commons may create a Commons Study Number for each participant, and the MEMBER study’s designated data transfer group (or Coordinating Center) will maintain the link between the Commons Study Number and the MEMBER’s study participant number. The DCC will be responsible for creating the pooled datasets in consultation with the Commons Steering Committee and domain experts in the working groups.

It is not the purpose of the Statistical Analysis Commons to serve as a data distribution hub, similar to dbGaP. Rather, it is to create a shared infrastructure to bring data, analysis tools, and investigators together to expedite coordinated analyses and discovery. In special circumstances, an investigator may request to download the data for local computing for a limited time, for example for methods development and validation. Such downloading would require IRB approval, permission of the Commons Steering Committee, and may require individual cohort approval.

Secure data location.

Data security and participant confidentiality are paramount to each of the cohorts separately and to the Statistical Analysis Commons Consortium. Importantly, none of the 18 HIPAA identifiers will be part of the Statistical Analysis Commons. The pooled datasets may be placed on secure data servers maintained by an agreed-upon cloud service provider (e.g. DNAnexus), which will provide a secure password-protected site for data storage and large-scale data analysis, and whose employees may provide administrative support. An updated list of secure data locations is included on the ONGOING UPDATES website.

Individual-level access to the pooled data hosted on the cloud service provider is governed by an Authorized Access Agreement (ONGOING UPDATES WEBSITE).

Definitions

For purposes of this agreement:

1. **MEMBERS** refers to the parties to this agreement, i.e., the entities that compose the Consortium. Each MEMBER is a legal entity and is an organization in which a performance site (such as a data coordinating center) for the MEMBER Study or the Statistical Analysis Commons Consortium is located.
2. **MEMBER STUDY** may be a single or multi-center study or a consortium of studies.
3. **MEMBER REPRESENTATIVE** refers to the investigator who is representing the MEMBER Study in the Statistical Analysis Commons Consortium.
4. **COMPONENTS** refers to field centers or other Study affiliated centers that are represented by the official consortium MEMBERS. All COMPONENTS must have a relationship to their respective MEMBERS. COMPONENTS are not required to sign this data sharing agreement but are bound to its terms through internal agreements specific to the MEMBER Study.
5. **USERS** refers to centers that develop methods and/or conduct analyses of Statistical Analysis Commons data and are approved to participate by all MEMBERS of the Commons Steering Committee. USERS sign this data sharing agreement and are bound to its applicable Terms and Conditions. Each USER must provide a copy of its IRB approval and signed Consortium Data Sharing Agreement to the Commons DCC.
6. **USER REPRESENTATIVE** refers to the investigator who is representing the USER institution in the Statistical Analysis Commons Consortium.

AGREEMENT

This agreement is for: ___USER
 ___MEMBER

USER INSTITUTION or MEMBER study name: xxxx

USER or MEMBER representative: xxxx

FOR MEMBERS, THE DESIGNATED DATA-TRANSFER GROUP: xxxx

For MEMBERS, other groups, either Field Centers or study-affiliated centers (“COMPONENTs” of the MEMBER), that have agreed to participate and abide by this agreement include: xxxx

All investigators agree to abide by the CHARGE principles incorporated on the first page.

Scope of Work

The primary work involves an evaluation of the association of large-scale genomic data with various heart, lung, blood, sleep and aging-related phenotypes or the evaluation of potential gene-environment interactions. Working Groups are responsible for the development of pre-specified analysis plans, and they are encouraged to use guidelines provided by the Analysis and Bioinformatics Committee or to seek additional consultation from that committee. Each Working Group is expected to focus on its phenotypes of interest and to respect any and all participant-consent-related limitations on the use of the data. If a Working Group has interest in other phenotypes for the evaluation of pleiotropy or gene-environment interactions, the relevant Working Groups are expected to develop a collaborative plan. Phenotype-only or genotype-only analyses across all participating MEMBER studies are possible, but they are not the primary purpose of the Statistical Analysis Commons and so would require review and approval by the Commons Steering Committee and perhaps the parent cohorts in the form of an ancillary study. With approval by each participating MEMBER study, other harmonized data can be added to the pooled dataset listed on the ONGOING UPDATES website. Each MEMBER study can decide whether to add or not to add the new data.

Terms and Conditions

1. Sharing among Consortium MEMBERS

Each MEMBER agrees to share data with other Statistical Analysis Commons MEMBERS, the MEMBERS' COMPONENTS, and USERS according to CONSORTIUM policies. The data to be shared are listed on the ONGOING UPDATES WEBSITE. These data may be freely shared among MEMBERS of the Statistical Analysis Commons, their respective COMPONENTS, and USERS. Sharing for purposes outside the scope of the Consortium identified above must receive appropriate permissions from all MEMBERS that generated the data, as long as the Statistical Analysis Commons is active. The Statistical Analysis Commons shall be deemed "active" until it is terminated, as described in Term 12.

2. Sharing among non-Consortium MEMBERS

As long as the Statistical Analysis Commons is active, MEMBERS agree not to share data generated by other MEMBERS with any party that is not a MEMBER, COMPONENT, or USER, unless the party joins the Statistical Analysis Commons, agrees to its terms and conditions, and signs this agreement. Principal Investigators at each MEMBER, COMPONENT, and USER may work with individuals at their respective institutions as they would on typical research projects. MEMBERS and COMPONENTS may continue to share data derived at their respective institutions according to usual policies, practices, and procedures.

3. Conveying obligations to COMPONENTS

MEMBERS consisting of multiple COMPONENTS agree to share the terms of this agreement with their respective COMPONENTS, so researchers at each COMPONENT use and handle data according to the provisions of this agreement. MEMBERS with COMPONENTS certify that their COMPONENTS have agreed to be bound by the terms of this agreement and that COMPONENTS that have not agreed will not share or access Consortium data under this agreement.

4. Complying with Consents

Each MEMBER and USER agrees to comply with all limitations accompanying the MEMBER- or COMPONENT-contributed data and to convey any such limitations to other Statistical Analysis Commons MEMBERS and USERS that access these contributed data through the Statistical Analysis Commons. Limitations should appropriately reflect the informed consent of Study participants from whom the data shared under this agreement were collected and derived. Each MEMBER will list any limitations (e.g., cardiovascular disease research only) on use of the data it is sharing with the Consortium in Appendix 1, and, when in doubt about the appropriateness of sharing, will consult with an ethics board or IRB.

5. Respecting Participant Privacy and Ensuring Data Confidentiality

In order to respect the privacy of the participants from whom data were collected or derived, MEMBERS and USERS agree not to contact or make any effort to identify individuals, families, communities, tribes, or populations that are or may be the sources of the data. Participant identities may be known to researchers at the MEMBER Study that collected the data, but under no circumstances should HIPAA identifiers be shared with other MEMBERS or USERS. Data are confidential, and breaches to their security may subject MEMBERS or USERS to legal action.

6. Publishing and Presenting

The CHARGE Consortium's policies will guide the development and submission of the Statistical Analysis Commons Consortium publications and presentations. MEMBERS and USERS will provide and receive standard acknowledgement language to include in all publications and presentations resulting from membership in the Consortium.

7. Withdrawing Membership in the Consortium

Any MEMBER or USER may withdraw from the Statistical Analysis Commons at any time by providing all other MEMBERS with its intent to withdraw. If a MEMBER withdraws, it agrees to work with the Consortium to handle the disposition of its data (from MEMBER and its COMPONENTS) and the future conduct of studies using its data.

Options for disposition of MEMBER- or COMPONENT-contributed data after withdrawal of a MEMBER include:

- Removing the withdrawing MEMBER'S data from the Consortium dataset immediately
- Completing current analyses and not using withdrawing MEMBER'S data for future analyses
- Continuing to use the withdrawing MEMBER'S data as long as the Consortium is active

In addition, the withdrawing MEMBER will abide by Statistical Analysis Commons policy on disposition of Statistical Analysis Commons data.

8. Removing a MEMBER

If any MEMBER or USER believes that any other MEMBER or USER is in violation of one or more terms of this agreement, the MEMBER or USER who suspects the violation will either consult with other MEMBERS of the Commons Steering Committee or notify the offending MEMBER or USER of the suspected violation. If the violation is not remedied within 30 days of written notice to the offending MEMBER or USER, or if a breach of data security or confidentiality is discovered and not remedied within 30 days of notice to the offending MEMBER or USER, the majority of MEMBERS of the Commons Steering Committee may terminate the offending MEMBER'S or USER'S participation in the Consortium. Once a MEMBER'S participation is terminated, data disposition will be handled according to Term 7.

9. No Warranties

All data shared are provided without warranties, but MEMBERS certify that data were collected in accordance with academic or industry standards, as applicable, and, for research with human subjects, with 45 CFR 46 (or with applicable legal and ethical standards, for MEMBERS based outside the United States).

10. Liability

No indemnification for any loss, claim, damage, or liability is intended or provided by any MEMBER. Each MEMBER or USER shall be liable for any loss, claim, damage, or liability that it incurs as a result of its activities under this agreement.

11. Data Security

Each MEMBER and USER agrees to safeguard data it accesses as it would its own data. An updated list of secure data locations is included on the ONGOING UPDATES website. Individual-level access to the pooled data hosted on the cloud service provider is governed by an Authorized Access Agreement (ONGOING UPDATES WEBSITE).

12. Terminating the Consortium

This agreement shall go into effect as of the date of the final signature and will remain in effect until a majority of the MEMBERS agree to dissolve the Statistical Analysis Commons. MEMBERS of the Commons Steering Committee will review its progress at 5 years to determine whether the activity should continue and whether a new or modified agreement should be signed.

- When any of the above conditions for termination of the Consortium occurs, the MEMBERS remaining at the time of termination will work together on a plan to:
 - Complete any remaining analyses
 - Destroy data or keep it to share with former MEMBERS or non- MEMBERS via Study-approved new collaborations or Consortia
- If Consortium data may be used after termination of the Statistical Analysis Commons, MEMBERS will set conditions for use and will consider issues such as:
 - Whether non-MEMBERS may access Consortium data, and under what circumstances, such as
 - If the non-MEMBER collaborates with a former MEMBER
 - If other former MEMBERS are required to agree to the non-MEMBER's use, and if so, how many must agree
 - How best to ensure that future analyses conform with the original consents
 - Whether data may be adapted for use in a repository for long-term access by any qualified researcher without the permission of former MEMBERS
 - How long into the future former MEMBERS' input should be required for continued use of Consortium data

Signatures

The USER or MEMBER REPRESENTATIVE has received authorization for the USER or MEMBER REPRESENTATIVE's Institution to sign on their behalf. The MEMBER REPRESENTATIVE has received authorization from its Study to join the Statistical Analysis Commons. Each MEMBER certifies that it has the authority to distribute the data to other Consortium MEMBERS.

Electronic signatures may count as originals.

Name of MEMBER (study name); USERS -- leave blank

Name of USER or MEMBER REPRESENTATIVE'S institution

Institution's Authorized Official's name and title (printed)

Institution's Authorized Official's signature

Date

Read and understood by:

USER or MEMBER REPRESENTATIVE name and title

USER or MEMBER REPRESENTATIVE signature

Date