GBG Guidelines for Biomarker Projects

Information for cooperation partners

GBG Forschungs GmbH | Dornhofstraße 10 | 63263 Neu-Isenburg
Tel. +49 6102 7480-0 | Fax +49 6102 7480-440 | info@GBG.de | www.GBG.de

Introduction

The analysis of clinically relevant properties of biomarkers is a research focus of GBG. Typically, corresponding projects are executed together with external cooperation partners processing the biomaterial. Although using biomaterial from GCP-regulated GBG studies, most such projects are retrospective and independent, i.e. not covered by the study protocols. Therefore, different regulations apply.

GBG is an academic research organization with excellent international reputation and publishes frequently in high-impact scientific journals. To defend and strengthen this position of widely recognized high-quality research GBG established several internal working instruction documents although not required by regulations. GBGs cooperation partners are encouraged to learn from this document how the cooperation should be structured and what GBGs most important expectations are.

This document describes typical biomarker projects not covered by any study protocol. Deviations can be made, but require an appropriate rationale, discussion, and documentation. In most projects the cooperation partner is the project leader and also the corresponding author for publications.

Project Schedule

Biomarker projects consist of the following steps:

1. To apply for a project, fill in the “Cooperation Proposal Form”[1] and send it to GBG (You probably did this already).
2. If the project is approved by GBG, transfer of biomaterial to the laboratory of the cooperation partner is organized by GBGs TraFo department. Samples are usually double-pseudonymized before shipment according to GBGs bio-banking policy.
3. With new cooperation partners, a contract in form of a Material Transfer Agreement (MTA) has to be set up (GBG can provide a template). This document covers the terms and conditions under which the cooperation will take place.
4. A statistical analysis plan (SAP) is created; GBG can provide a template. The project leader is responsible for the first SAP draft, to guide the discussion, and to implement all changes into

[1] https://gbg.de/en/research/trafo.php

the final document. The final SAP document must be signed by all responsible persons before data transfer (end of step 5).

5. The cooperation partner performs measurements as specified in the SAP and transfers the biomarker data to GBG.
6. GBG combines the biomarker data with the clinical data[2] and performs statistical analyses as specified in the SAP. The statistical report is shared with the cooperation partner.
7. The main aim is to publish results in a full paper; optionally an abstract may be submitted to a symposium. Similar to the SAP, the corresponding author is responsible for the first draft, to guide the discussion with the coauthors, and to implement all changes into the final document. GBG creates a list of all authors according to GBGs guidelines[3].

Value of Biomaterial

GBG's biomaterial has a high value because it was collected within a well-controlled clinical study; detailed clinical data are available. The cooperation partner is expected to be aware of this.

GBG expects each partner to have all required assays established in its laboratory before biomaterial from GBG is processed. GBG's biomaterial should not be used for assay development or assay validation.

Machine Learning

GBG also participates in projects using machine learning techniques to develop clinically relevant biomarkers. Each project should aim for a concrete improvement for the patients; GBG is not interested in projects demonstrating the feasibility of machine learning approaches only.

GBG prefers projects aiming to validate a (machine learned) biomarker but may also provide data for training; see section "Data" below for requirements on the data transfer. Training data should consist of 500 or more patients per study and 1000 patients or more in total. If data from different studies (e.g. GBG and non-GBG) are combined appropriate measures should be defined to ensure that the biomarkers describes the biology but not the cohort; please consider the assays, technical devices such as scanners, study inclusion criteria, treatment schemes, and others. For the GBG the scientific value of the machine learning project needs to be high compared to the strategic value (novelty) of the data transferred.

If GBG data is used to validate a biomarker trained on non-GBG data, the statistical evaluation of the results is preferred to be performed at GBG to minimize the transfer of valuable data. In addition, this approach ensures that the biomarker can be clearly classified as pre-defined in a publication. Even if GBG data is used for validation only, GBG would like to know details about the training of the biomarker to estimate the chances of success of the project: mathematical training algorithms,

[2] clinical data includes baseline data, treatment data, longitudinal data, and patient outcome data
[3] https://gbg.de/en/research/publications.php, https://gbg.de/wAssets/docs/forschung/Guideline-for-Authorship.pdf

cohorts used including patient numbers and clinical characteristics, performance on training set, and other potentially relevant information.

SAP

The SAP is a crucial document for project execution in several aspects. Its content describes the pre-defined statistical results as opposed to post-hoc results. To ensure scientifically appropriate statistics the SAP must be finalized before transfer of the biomarker data. The SAP also (implicitly) defines project steps, assigns responsibilities, and documents procedures such as sample selection and processing.

The SAP must comply with the "Cooperation Proposal Form" and GBGs project approval decision. It must be signed by at least one representative of GBG and each cooperation partner. After signing the content of the SAP is binding.

The project leader creates the first draft of the SAP; GBG provides a template; the first draft should at the very least describe the objectives. The project leader coordinates the development of the SAP document. All team members except the project leader use MS Word's track change mode and/or comments to edit the SAP document. The project leader accepts, rejects, or modifies these changes and/or removes the comments after processing.

The statistical methods and details of reporting[4] depend on the objectives, variables and patient sets to be analyzed. Therefore, they should not be described in the first draft; they are specified by the statistician after discussion of the objectives and other specifications.

Statistical Analysis

Statistical analysis is performed by GBG according to GBGs internal working instructions which are part of GBGs quality management system. GBG shares clinical data with a cooperation partner only if required by the project needs; lack of resources or trust is not an argument for data sharing.

GBG generates a statistical report based on the SAP. Additional statistical analyses may be requested but should be avoided for two reasons: First, additional analyses consume much more resources compared to having them included in the SAP right from the beginning. The project will be delayed. Second, fishing for better results is not an acceptable scientific procedure, and the statistician may reject additional analyses for this reason. Pre-defined statistical analyses are favored over exploratory and post-hoc analyses.

Data

Patient-derived data must only be used for the documented purposes of the project. This applies to clinical data and to biomarker data, even if pseudonymized or anonymized. Access to patient-derived

[4] sections "Statistical Methods for Analysis" and "Tables and Figures" in GBGs template

data must be restricted as far as possible at each partner's site: preferably to project team members only.

Project activities are subject to the General Data Protection Regulation (GDPR), cooperation partners and GBG have shared responsibility. GBGs processes are based on double-pseudonymization of biomaterial samples: Patient numbers from the study are replaced by pseudonyms. Biomarker data not specified in the study protocol must be identified by the pseudonyms only. The process of double-pseudonymization is planned as part of the project activities.

In case you receive clinical or biomarker data that is associated to the patient number or any other information that could be used to identify the patient: Do not distribute this data, reduce the number of people having access to the data as far as possible, and notify GBG (but do not send the data).

Do not perform any actions that aim to de-pseudonymize or de-anonymize patient data.

In addition, data privacy pseudonyms also ensure that biomarker data is generated blindly, i.e., measurements in a laboratory cannot be tuned to yield favorable results.

Publications

GBG is interested to publish the results of a project regardless of whether the results are good or bad. Negative results are also important and should be communicated to other researchers as well.

Full paper manuscripts are the preferred type of publication, in addition abstracts for posters or oral presentations may be submitted to certain symposia. The list of authors is determined according to GBGs guidelines[5].

The corresponding author performs the following steps: create the first draft, discuss the document with the core[6] authors, and then discuss the document with all coauthors (deadline: two weeks). Similar to the SAP the corresponding author is responsible to guide the discussion. Other authors contribute to the document using the track change mode and/or comments; the corresponding author accepts, rejects, or modifies these changes and/or removes the comments after processing.

For a manuscript the corresponding author is responsible for submissions including further processing after journal decision (revision or adjustments of the manuscript for submission to different journals).

For abstracts, posters and oral presentations certain submission deadlines are defined by the symposium. The corresponding author sends a first draft to GBG five weeks prior to the respective submission deadline to allow time for internal and co-author review. GBG can provide templates.

[5] https://gbg.de/en/research/publications.php, https://gbg.de/wAssets/docs/forschung/Guideline-for-Authorship.pdf
[6] the 2-5 authors contributing most, including the statistician

GBGs medical writing team should be included in each email related to publications: abstracts@gbg.de. The corresponding author should also specify this email address in the submission or forward each email from the journal or symposium to abstracts@gbg.de.