# Nightingale Health Metabolic Biomarkers

Companion document

## Introduction

Nightingale Health Plc. is performing metabolic biomarker profiling of EDTA plasma samples from all UK Biobank participants. This document provides an outline of the methodology and the metabolic biomarker data in the UK Biobank resource. It describes the earlier phase 1 release, the current phase 2 release, and the small updates between the releases.

In the phase 1 release, we shared metabolic biomarker data from approximately 118,000 participants at baseline recruitment and 5,000 at repeat assessment (around 1,500 of these had both a baseline and repeat assessment). These samples were measured between June 2019 and April 2020 in Nightingale Health's laboratories in Finland. The phase 2 release covers metabolic biomarker data from an additional 157,000 participants at baseline recruitment and 12,000 at repeat assessment. These samples were measured between April 2020 and June 2022. This release brings the current number of samples with available metabolic biomarker data to 292,000, of which 275,000 are from baseline recruitment and 17,000 from repeat assessment (around 15,500 of these now have both a baseline and repeat assessment).

Nightingale Health's metabolic biomarker platform is based on high-throughput nuclear magnetic resonance (NMR) spectroscopy. A total of 249 metabolic measures, of which 168 absolute levels and 81 ratio measures, are quantified per EDTA plasma sample, covering both routine biomarkers and emerging biomarkers with medical relevance. The biomarkers include detailed measures of cholesterol metabolism, fatty acid compositions, and various low-molecular weight metabolites, such as amino acids, ketones, and glycolysis metabolites. For 14 lipoprotein subclasses, the lipid concentrations and composition are measured in terms of triglycerides, phospholipids, total cholesterol, cholesterol esters, and free cholesterol, and total lipid concentration within each subclass. Most of the biomarkers are measured in absolute concentration units (mmol/L). With this phase 2 release, two new derived variables were shared in addition to the standard set of 249 biomarkers (see Section Addition of derived variables below for more information).

## Biomarker measurements by NMR

EDTA plasma samples from aliquot 3 were measured using Nightingale Health's NMR-based metabolic biomarker profiling platform. If comparisons are done to the clinical biochemistry markers available in the UK Biobank, please note that those were measured from serum samples, primarily from aliquot 1. The phase 1 and 2 samples were

both a random subset of the full cohort.

The samples were prepared directly in 96 well-plates by the UK Biobank. At least 85 µL of plasma was aliquoted in each well using TECAN freedom EVO 150 robotic liquid handlers, which have coefficients of variation in pipetting volume at <0.75% across 8 tips. Plasma samples were shipped to Nightingale Health's laboratories in the 96-well plates on dry ice in sample batches of ~5,000-20,000.

Details of the metabolic biomarker profiling platform and experimentation have been described previously (1–4). In brief, EDTA plasma samples were stored in a freezer at -80ºC. Before preparation, frozen samples were slowly thawed at +4ºC overnight, and then mixed gently and centrifuged (3 min, 3400´g, +4ºC) to remove possible precipitate. Aliquots of each sample were transferred into 3-mm outer-diameter NMR tubes and mixed in 1:1 ratio with a phosphate buffer (75mM $Na_2HPO_4$ in 80%/20% $H_2O/D_2O$, pH 7.4, also including 0.08% sodium 3-(trimethylsilyl) propionate-2,2,3,3-$d_4$ and 0.04% sodium azide) automatically with an automated liquid handler (PerkinElmer Janus Automated Workstation).

The prepared samples were loaded onto a cooled sample changer, which maintains the temperature of samples waiting to be measured at +6ºC. Two NMR spectra were recorded for each plasma sample using a 500 MHz NMR spectrometer (Bruker AVANCE IIIHD). The first spectrum is a presaturated proton NMR spectrum, which features resonances arising mainly from proteins and lipids within various lipoprotein particles. The other spectrum is a T2-relaxation-filtered spectrum where most of the broad macromolecule and lipoprotein lipid signals are suppressed, leading to enhanced detection of low-molecular-weight metabolites. Automated quality control of the spectral data was performed. The metabolic biomarkers were quantified using Nightingale Health's proprietary software (Nightingale Health biomarker quantification library 2020). Appendix 1 shows biomarker distributions for the fourteen consecutively measured batches of ~10,000-25,000 samples.

As an indication of the appropriate laboratory quality management, Nightingale Health's metabolic biomarker profiling platform has qualified for CE-marking according to the required laboratory standards for clinical use (IVD 98/79/EC; Directive on In Vitro Diagnostic Medical Devices). As a result, our laboratory information management system is accredited according to ISO 13485. Comparison of Nightingale Health's biomarker measurements with clinical biochemistry in cohort studies also indicate high consistency, with correlation coefficients commonly >0.9 (3, 5). The consistency in the UK Biobank is slightly reduced due to sample dilution (see Section Sample dilution issue below) as well as difference in the aliquot and sample type (serum vs. EDTA plasma) between the clinical biochemistry panel and Nightingale Health assays.

## Sample dilution issue

UK Biobank plasma samples from aliquot 3 analysed by Nightingale Health are expected to be 5-10% diluted. The dilution is believed to come from mixing of participant sample with water due to seals that failed to hold a system vacuum in the automated

liquid handling systems. While this issue may have an impact on some of the biomarker concentration values, it is expected to have limited impact on the epidemiological analyses. However, we recommend that this aspect is considered when conducting analyses that rely on absolute concentrations, such as stratification based on specific biomarker concentrations. The dilution may also cause challenges in comparing the biomarker concentration distributions with those observed in other cohort studies. We therefore caution against using the concentrations observed in UK Biobank as reference levels for translational applications. This issue is illustrated in Figure 1, showing the consistency between lipids, apolipoproteins, creatinine, albumin, and glucose measured by routine clinical chemistry and Nightingale Health NMR for UK Biobank (phase 1) and FinHealth 2017 cohort (figure adapted from (3)).

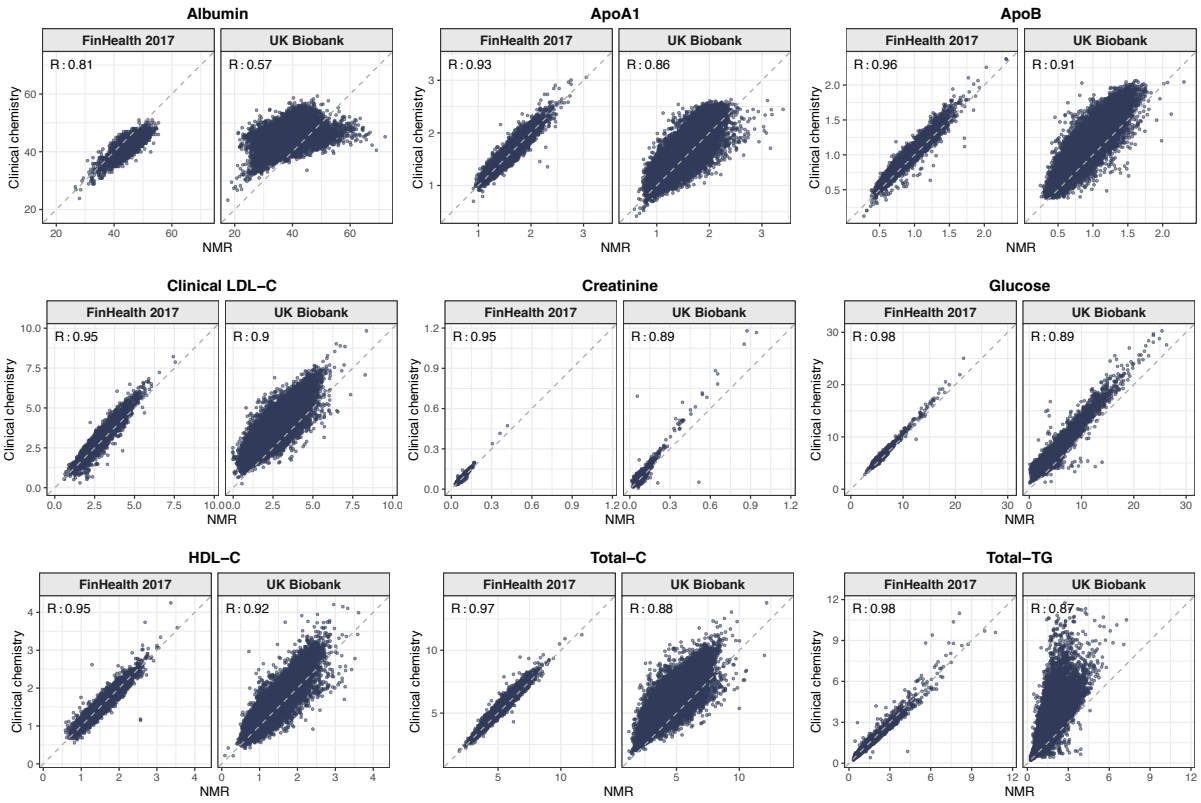


**Figure 1. Comparison of the NMR biomarker measurements to clinical chemistry.** Scatterplots are shown for lipids and other routine biomarkers for which both NMR and clinical chemistry measurements are available in the UK Biobank (phase 1, N = 118,000) and the FinHealth 2017 cohort (N = 6,000). The more pronounced deviations in correlations and absolute concentrations in UK Biobank compared to FinHealth 2017 samples are primarily due a known dilution issue in the UK biobank samples. Figure adapted from (3).

# Quality control

Metabolic biomarker profiling by Nightingale Health's NMR platform provides consistent results over time and across spectrometers. Furthermore, the sample preparation is minimal in the Nightingale Health's metabolic biomarker platform, circumventing all extraction steps. These aspects result in highly repeatable biomarker measurements. Pre-specified quality metrics were agreed between UK Biobank and Nightingale Health to ensure consistent results across the samples, and pilot measurements were conducted.

Nightingale Health performed real-time monitoring of the measurement consistency within and between spectrometers throughout measurements of the UK Biobank samples. Two control samples provided by Nightingale Health were included in each 96-well plate to track the consistency across multiple spectrometers. Furthermore, two blind duplicate samples provided by the UK Biobank were included in each well plate, with the position information unlocked only after results delivery.

Coefficient of variation (CV) targets across the metabolic biomarker profile were pre-specified for both Nightingale Health's internal control samples and UK Biobank's blind duplicates. The targets were met for each consecutively measured batch. CVs across spectrometers based on Nightingale Health's internal control samples and blind duplicates that were measured as part of the phase 1 and 2 data release are summarised in Figure 2. For many of the metabolic biomarkers the CVs are below 5%. Consistency of the blind duplicate measures for individual biomarkers is shown in Appendix 2. In addition to technical repeatability, biological stability of the biomarker measures over time between the samples from baseline and repeat assessment is presented in Appendix 3.

## Quality control flags

Nightingale Health's integrated quality procedures verify the sample quality by reporting signs of degradation and contamination issues. These are reported as flags along with the measurement results. Issues affecting the whole sample are reported as sample-level flags, and issues affecting only certain biomarkers are reported as biomarker-level flags, which are provided as a separate data field for each biomarker. In general, if a biomarker has a flag but the biomarker value is still provided, it indicates that the presence of the interfering substance is low and deemed not to interfere with the quantification of the biomarker (i.e., the value can be trusted). Appendix 4 provides detailed information on the available quality flags. There is no need to a priori remove any biomarker values based on the flags, but researchers may consider performing sensitivity analyses as described below.

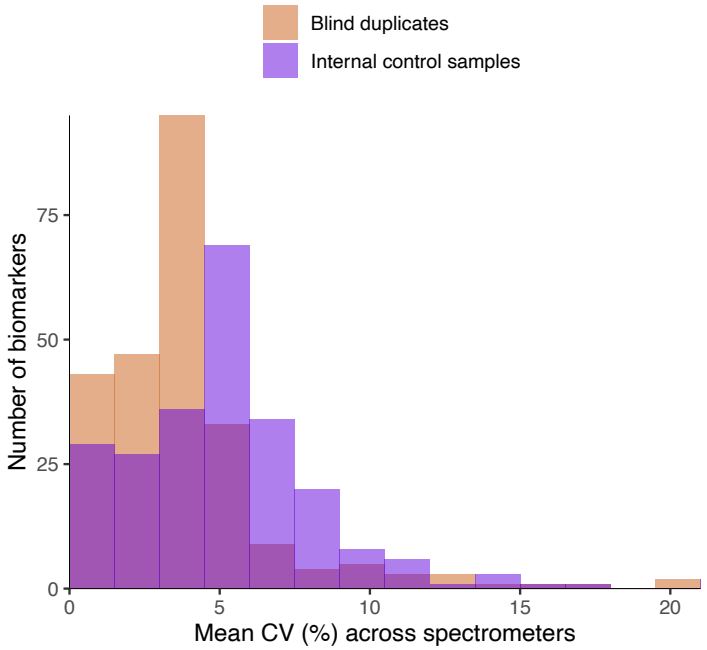


**Figure 2. Distributions of the coefficients of variation (CV) for the 249 metabolic measures.** The distributions are shown for UK Biobank's blind duplicate samples (orange) and Nightingale Health's internal control samples (purple) in the combined phase 1 and phase 2 dataset. The CVs are shown across the eight NMR spectrometers used for the measurements. CVs of individual biomarkers are available in Appendix 2.

# Recommended approaches for data preprocessing and analysis

## Data pre-processing and considerations for epidemiological analyses

Nightingale Health's metabolic biomarker profiling platform is known for high repeatability over time and absence of batch effects. This can be seen in Appendix 1, showing the biomarker distributions from fourteen consecutively measured sample batches. The metabolic biomarker data can generally be used for epidemiological analyses without any pre-processing and can be analysed in the same manner as the clinical biochemistry data available in the UK Biobank.

Minor differences in biomarker concentrations may arise from the different spectrometers used for the biomarker measurements. As shown in Appendix 5, the differences in the biomarker distributions between spectrometers are generally minor. However, researchers may consider correcting or adjusting for the spectrometer in epidemiological analyses. With the release of the phase 2 data, we have added a new spectrometer-corrected variable for alanine, which is the single most sensitive biomarker to spectrometer differences (see Section Addition of derived variables below). A detailed report on quality control and removal of technical variation has been published earlier (6).

### Quality control flags

Biomarker values substantially affected by interfering substances have been removed during the quality control procedures. However, the researcher may consider performing sensitivity analyses by excluding samples flagged with “Low protein”, which may indicate more severe sample dilution. Biomarker values flagged with “Below limit of quantification” may also be omitted in sensitivity analyses, since this flag indicates that the concentration of the given biomarker is smaller than the range where the Nightingale Health’s metabolic biomarker profiling platform considers quantification highly accurate.

## Changes between the phase 1 and phase 2 data release

### Recalibration of one spectrometer

Spectrometer no. 3 went through a recalibration procedure during the measurements of phase 2 samples, to provide a better match with the other spectrometers. Subsequently, the data from phase 1 was updated to align with the same calibration factors. This may result in minor absolute biomarker value differences between the original phase 1 release and the current release for samples that were measured using this specific spectrometer. The correlations between the phase 1 biomarker values before and after the recalibration are all exactly 1, as shown in appendix 6.

### Correction of sample data from six swapped well-plates

In the complete set of data for phase 1 and 2, the metabolic biomarker data for three pairs of 96-well plates (564 samples) were identified to have been inadvertently swapped and were consequently corrected. Two of these plate pairs were part of the phase 1 release. These corrections have also been noted in the data showcase, although they are expected to have little to no effect on epidemiological analyses.

### Addition of derived variables

Two new derived variables have been shared along with the phase 2 data release in addition to the standard 249 biomarkers to facilitate research: a composite glucose-lactate variable and spectrometer-corrected alanine. The composite glucose-lactate variable is derived from the combination of lactate and glucose concentrations in the sample, accounting for anaerobic glycolysis that takes place during the period from sample collection to sample processing. Therefore, it provides a measurement that may be better correlated with the glucose in the sample at the at the time of collection.

The spectrometer-corrected alanine reflects alanine concentration with the effect from the spectrometer removed. Alanine is the single most sensitive biomarker to spectrometer differences (6). The correction was done by matching the means and standard deviations of alanine concentration distributions across all spectrometers to those of a chosen master spectrometer.

## GWAS summary statistics

GWAS summary statistics of metabolic biomarkers measured by Nightingale Health have been made available through MR-Base database, developed by the MRC Integrative Epidemiology Unit at the University of Bristol.

## References


1. P. Soininen, A. J. Kangas, P. Würtz, T. Suna, M. Ala-Korpela, Quantitative Serum Nuclear Magnetic Resonance Metabolomics in Cardiovascular Epidemiology and Genetics. *Circulation: Cardiovascular Genetics*. **8**, 192–206 (2015).

2. A. V. Ahola-Olli, L. Mustelin, M. Kalimeri, J. Kettunen, J. Jokelainen, J. Auvinen, K. Puukka, A. S. Havulinna, T. Lehtimäki, M. Kähönen, M. Juonala, S. Keinänen-Kiukaanniemi, V. Salomaa, M. Perola, M.-R. Järvelin, M. Ala-Korpela, O. Raitakari, P. Würtz, Circulating metabolites and the risk of type 2 diabetes: a prospective study of 11,896 young adults from four Finnish cohorts. *Diabetologia*. **62**, 2298–2309 (2019).

3. H. Julkunen, A. Cichońska, M. Tiainen, H. Koskela, K. Nybo, V. Mäkelä, J. Nokso-Koivisto, K. Kristiansson, M. Perola, V. Salomaa, P. Jousilahti, A. Lundqvist, A. J. Kangas, P. Soininen, J. C. Barrett, P. Würtz, Atlas of plasma NMR biomarkers for health and disease in 118,461 individuals from the UK Biobank. *Nat Commun*. **14**, 604 (2023).

4. Nightingale Health Biobank Collaborative Group, Metabolomic and genomic prediction of common diseases in 477,706 participants in three national biobanks. *Medrxiv* doi:10.1101/2023.06.09.23291213 (2023).

5. E. Tikkanen, V. Jägerroos, M. V. Holmes, N. Sattar, M. Ala-Korpela, P. Jousilahti, A. Lundqvist, M. Perola, V. Salomaa, P. Würtz, Metabolic Biomarker Discovery for Risk of Peripheral Artery Disease Compared With Coronary Artery Disease: Lipoprotein and Metabolite Profiling of 31 657 Individuals From 5 Prospective Cohorts. *J Am Heart Assoc*. **10**, e021995 (2021).

6. S. C. Ritchie, P. Surendran, S. Karthikeyan, S. A. Lambert, T. Bolton, L. Pennells, J. Danesh, E. Di Angelantonio, A. S. Butterworth, M. Inouye, Quality control and removal of technical variation of NMR metabolic biomarker data in ~120,000 UK Biobank participants. *Sci Data*. **10**, 64 (2023).