



# Built-in peer-reviewed science – from Nightingale’s metabolomics data to publications

Nightingale’s metabolomics platform has become the global standard in large-scale health studies. It has been built upon an unrivalled scientific foundation with over 250 peer-reviewed publications in top biomedical journals. Here, we showcase how the data transforms into a scientific publication. In addition, we explain the unique advantages of Nightingale’s metabolomics platform. Due to the physiologically meaningful biomarkers in the platform and easily operated quantitative data, it has already been used in highly multifaceted epidemiological studies and in numerous different study fields – and it still has the potential to reach novel areas.

## Advantages of quantitative metabolite data

One of the most important benefits of Nightingale’s service is that the metabolites are quantified in absolute concentration units. Furthermore, our platform is free of any batch effects. This means that the data is immediately ready for various scientific analyses, without a need to allocate researcher’s time and resources into tedious domain-specific data pre-processing. We provide biomarker concentrations in spreadsheet files, both in XLSX and CSV formats.

## Getting started

Our quantitative biomarker data can be analyzed as any other standard clinical chemistry measure using standard analysis software, such as R, SAS, SPSS and STATA.

To ensure that you get the most out of your data, we recommend performing the following standard pre-processing steps before proceeding with the biomarker analysis to answer your research questions.

- Outlier removal is recommended to ensure that the results are not driven by only a few data points, as regression models are typically sensitive to significant outliers.
- Data standardization is required to be able to compare associations across different biomarker concentration units and ranges. For example, the concentrations of all metabolic markers can be log-transformed and scaled to SD units to enable a meaningful comparison of results.

For R users, we have created a free tutorial to guide you through the basic steps of importing blood metabolomics data into R, combining them with genetic, clinical or other data, and performing basic epidemiological analysis. Basic understanding of R is enough to get started. You can access the tutorial from here: [NMR Data Analysis Tutorial](#).

In addition to that, we have generated a free data visualization tool for R users. The tool is called **ggforestplot** and it provides a simple solution for turning association results into figures that are ready for use in scientific publications and presentations. We present this tool in detail in the next section.

## Using ggforestplot

### Data analysis

For performing the data analysis, the ggforestplot R package provides three commonly used regression models: linear regression, logistic regression and Cox proportional hazards regression. These models are also implemented in many other statistical software tools, including SAS, SPSS and STATA. The choice of the model depends on the data at hand and the research question. Linear regression can be used for discovering biomarker associations with a continuous phenotype, such as BMI. Binary logistic regression is a method of choice when exploring the relationship between a metabolic biomarker and, for example, a binary disease indicator. As an example, Cox proportional hazards model is useful for analyzing biomarker associations with future risk of developing a disease. Note that relevant covariates, such as gender and age, should also be included in the model, as they can affect the biomarker concentrations.

## Visualisation and interpreting association results

Forest plot is a convenient way of visualizing metabolite association results. It illustrates the magnitude of the effect and significance of several metabolites at one glance.

With the following case example, we demonstrate how forest plot can facilitate results interpretation. In addition, the example also showcases how Nightingale's lipoprotein subclasses and lipid composition measures can be used for discovering new associations that clinical chemistry hasn't been able to identify.

### Showcase example

**Sliz et al. (2018)** examined the effects of statin therapy and genetic inhibition of PCSK9 on 228 metabolic biomarkers measured from the blood samples in the PROSPER (Prospective Study of Pravastatin in the Elderly at Risk) trial.

Both statins and PCSK9 inhibitors lower low-density lipoprotein cholesterol (LDL-C) levels to reduce risk of cardiovascular events but PCSK9 inhibitors were shown to be slightly less efficacious than statins for equivalent LDL-C reductions. This study was aiming to find out the reason for this difference.

As illustrated below, comprehensive metabolic profile analysis demonstrated that the effects of both drugs are similar, apart from weaker lowering of very low-density lipoprotein lipid measures (VLDL) by PCSK9 inhibitors compared to statins (Figure 1). This could potentially translate into smaller reductions in cardiovascular disease risk.



**Figure 1.** Effects of statin treatment and genetic inhibition of PCSK9 on lipoprotein and lipid levels. From Sliz et al. *Circulation* 2018;138:2499–512.

### Data analysis service

Nightingale Health also offers data analysis. The service includes association analysis between Nightingale's metabolomics data and the chosen research outcome, including visualisation of the results.

If you're interested, please get in touch with us for more information at [research@nightingalehealth.com](mailto:research@nightingalehealth.com).