

**METHOD FOR DETERMINING WHETHER A SUBJECT HAS A DISEASE OR
CONDITION OR IS AT RISK OF DEVELOPING A DISEASE OR CONDITION****TECHNICAL FIELD**

5 The present disclosure relates generally to methods for determining whether a subject has a disease or condition or is at risk of developing a disease or condition.

BACKGROUND

10 Various biomarkers may be useful for predicting whether a subject has a particular disease or condition or is at risk of developing the disease or condition. Such biomarkers may be measured from various biological samples, such as from blood samples or other biological fluids. However, certain sample
15 types may be more complex than others when quantitative measurements of biomarkers are to be obtained. There may also be a need for biomarkers that are predictive for e.g. severe diseases causing hospitalization or even death, such as severe infectious diseases and their complications, for example sepsis, pneumonia
20 and other lower respiratory infections.

SUMMARY

 A method for determining whether a subject has a disease or condition or is at risk of developing a disease or
25 condition is disclosed. The method may comprise

 determining in a biological sample obtained from the subject a quantitative value of at least one biomarker relative to a quantitative value of albumin in the biological sample; and
 comparing the quantitative value(s) of the at least one
30 biomarker to a control sample or to a control value;

 wherein an increase or a decrease in the quantitative value(s) of the at least one biomarker, when compared to the control sample or to the control value, is/are indicative of the subject having the disease or condition or having an increased
35 risk of developing the disease or condition.

 A method for determining whether a subject is at risk of developing an infectious disease or a complication thereof is also disclosed. The method may comprise

2

determining in a biological sample obtained from the subject a quantitative value of at least one biomarker of the following:

- glycoprotein acetyls,
- 5 - albumin,
- omega-6 fatty acids,
- monounsaturated fatty acids,
- saturated fatty acids,
- omega-3 fatty acids,
- 10 - apolipoprotein A1 (ApoA1),
- docosahexaenoic acid (DHA),
- leucine,
- acetate,
- alanine,
- 15 - apolipoprotein B (ApoB),
- glutamine,
- isoleucine,
- linoleic acid,
- phenylalanine,
- 20 - tyrosine,
- degree of unsaturation of fatty acids,
- valine,
- histidine,
- cholesterol in HDL (HDL-C),
- 25 - glucose,
- acetoacetate,
- 3-hydroxybutyrate,
- cholesterol in LDL (LDL-C),
- lactate,
- 30 - triglycerides in LDL (LDL-TG),
- pyruvate, or
- cholesterol in VLDL (VLDL-C); and

comparing the quantitative value(s) of the at least one biomarker to a control sample or to a control value;

- 35 wherein an increase or a decrease in the quantitative value(s) of the at least one biomarker, when compared to the control sample or to the control value, is/are indicative of the subject having an increased risk of developing the infectious disease or the complication thereof.

BRIEF DESCRIPTION OF THE DRAWINGS

The accompanying drawings, which are included to provide a further understanding of the embodiments and constitute a part of this specification, illustrate various embodiments. In the drawings:

Figure 1a shows the relation of baseline biomarker levels to future development of type 2 diabetes, when the biomarkers are analysed in absolute concentrations and when scaled relative to the concentration of albumin.

Figure 1b shows the relation of baseline biomarker levels to future development of severe pneumonia, when the biomarkers are analysed in absolute concentrations and when scaled relative to the concentration of albumin.

Figure 2a shows the relation of baseline biomarker levels to future development of pneumonia (severe or non-severe), when the biomarker concentrations are analysed in absolute concentrations.

Figure 2b shows the relation of baseline biomarker levels to future development of Other Lower Respiratory Diseases (acute bronchitis, acute bronchiolitis, and Unspecified acute lower respiratory infection), when the biomarker concentrations are analysed in absolute concentrations.

Figure 2c shows the relation of baseline biomarker levels to future development of sepsis, when the biomarker concentrations are analysed in absolute concentrations.

Figure 3a shows the relation of baseline biomarker levels for 3 plurality-biomarker-scores to future development of severe pneumonia for the whole study population and in the following subgroups of individuals: study participants without chronic respiratory or cardiometabolic disease at baseline, in different age groups (divided in age-tertiles of the study population), and separately for men and women. The relation of the of baseline biomarker levels for 3 different plurality-biomarker-scores with severe pneumonia occurring within 2 years after the blood samples were taken is also shown.

DETAILED DESCRIPTION

In one aspect, a method for determining whether a subject has a disease or condition or is at risk of developing a disease or condition is disclosed.

5 The method may comprise determining in a biological sample obtained from the subject a quantitative value of at least one biomarker relative to a quantitative value of albumin in the biological sample; and

10 comparing the quantitative value(s) of the at least one biomarker to a control sample or to a control value;

 wherein an increase or a decrease in the quantitative value(s) of the at least one biomarker, when compared to the control sample or to the control value, is/are indicative of the subject having the disease or condition or having an increased
15 risk of developing the disease or condition.

In a second aspect, a method for determining whether a subject is at risk of developing an infectious disease or a complication thereof is disclosed. The method may comprise

20 determining in a biological sample obtained from the subject a quantitative value of at least one biomarker (or a plurality of the biomarkers, for example at least two, or at least three, or at least four, or at least five, or at least six, or at least seven, or at least eight, or at least nine, or at least 10, or all) of the following:

- 25 - glycoprotein acetyls,
 - albumin,
 - omega-6 fatty acids,
 - monounsaturated fatty acids,
 - saturated fatty acids,
30 - omega-3 fatty acids,
 - apolipoprotein A1 (ApoA1),
 - docosahexaenoic acid (DHA),
 - leucine,
 - acetate,
35 - alanine,
 - apolipoprotein B (ApoB),
 - glutamine,
 - isoleucine,
 - linoleic acid,

- phenylalanine,
 - tyrosine,
 - degree of unsaturation of fatty acids,
 - valine,
 - 5 - histidine,
 - cholesterol in HDL (HDL-C),
 - glucose,
 - acetoacetate,
 - 3-hydroxybutyrate,
 - 10 - cholesterol in LDL (LDL-C),
 - lactate,
 - triglycerides in LDL (LDL-TG),
 - pyruvate, or
 - cholesterol in VLDL (VLDL-C); and
 - 15 comparing the quantitative value(s) of the at least one biomarker to a control sample or to a control value;
- wherein an increase or a decrease in the quantitative value(s) of the at least one biomarker, when compared to the control sample or to the control value, is/are indicative of the
- 20 subject having an increased risk of developing the infectious disease or the complication thereof.
- Any embodiments described below may be understood as relating to either one (or both) of the above aspects.
- The biological sample may be a dry blood sample, a dry
- 25 sample obtainable from blood, or a biological sample obtainable from a dry blood sample. However, in the context of this specification, the biological sample does not necessarily have to be a dry blood sample or obtainable from a dry blood sample, as will be set out in more detail below.
- 30 The dry blood sample may be a dry whole blood sample. However, the term may be understood, at least in some embodiments, to cover also dry samples containing serum and/or plasma (but not necessarily whole blood). The term "dry blood sample" or "dry sample obtainable from blood" may, in some
- 35 embodiments, refer to a dry sample obtainable from separated blood.
- The dry blood sample or the dry sample obtainable from blood may be obtainable from a fingertip of the subject, sampling of capillary blood from the upper arm of the subject,

and/or by venepuncture of the subject. However, they may be additional or alternative locations and/or methods of obtaining the blood.

The method may comprise determining in the biological sample quantitative values of a plurality of biomarkers. Thus any references to "at least one biomarker" in this specification may also be understood as referring to a plurality of biomarkers. For example, the plurality of the biomarkers may comprise 2, 3, 4, 5, 6, 7, 8, 9, 10, or more (i.e. at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, or at least 10) biomarkers. The term "plurality of the biomarkers" may thus, within this specification, be understood as referring to any number (above one) of biomarkers. The term "plurality of the biomarkers" may thus be understood as referring to any number (above one) and/or combination or subset of the biomarkers described in this specification. Determining a plurality of biomarkers may increase the accuracy of the prediction of whether the subject has or is at risk of developing a disease or condition. In general, it may be that the larger the number of the biomarkers, the more accurate or predictive the method. The plurality of biomarkers may be measured from the same biological sample or from separate biological samples and using the same analytical method or different analytical methods. In an embodiment, the plurality of biomarkers may be a panel of a plurality of biomarkers.

At least in an embodiment, in the context of this specification, the wording "comparing the quantitative value(s) of the biomarker(s) to a control sample or to a control value(s)" may be understood, as a skilled person would, as referring to comparing the quantitative value or values of the biomarker or biomarkers, to a control sample or to a control value(s) either individually or as a plurality of biomarkers (e.g. when a risk score is calculated from the quantitative values of a plurality of biomarkers), depending e.g. on whether the quantitative value of a single (individual) biomarker or the quantitative values of a plurality of biomarkers are determined.

In an embodiment, the method comprises determining in a biological sample obtained from the subject a quantitative value

of at least one biomarker (or a plurality of the biomarkers, for example at least two, or at least three, or at least four, or at least five, or at least six, or at least seven, or at least eight, or at least nine, or at least 10, or all) of the following relative to the quantitative value of the albumin in the biological sample:

- glycoprotein acetyls,
- omega-6 fatty acids,
- monounsaturated fatty acids,
- 10 - saturated fatty acids,
- omega-3 fatty acids,
- apolipoprotein A1 (ApoA1),
- docosahexaenoic acid (DHA),
- leucine,
- 15 - acetate,
- alanine,
- apolipoprotein B (ApoB),
- glutamine,
- isoleucine,
- 20 - linoleic acid,
- phenylalanine,
- tyrosine,
- degree of unsaturation of fatty acids,
- valine,
- 25 - histidine,
- cholesterol in HDL (HDL-C),
- glucose,
- acetoacetate,
- 3-hydroxybutyrate,
- 30 - cholesterol in LDL (LDL-C),
- lactate,
- triglycerides in LDL (LDL-TG),
- pyruvate, or
- cholesterol in VLDL (VLDL-C).

35 In an embodiment, the method comprises determining a quantitative value of glycoprotein acetyls.

In an embodiment, the method comprises determining a quantitative value of albumin.

In an embodiment, the method comprises determining a quantitative value of omega-6 fatty acids.

In an embodiment, the method comprises determining a quantitative value of monounsaturated fatty acids.

5 In an embodiment, the method comprises determining a quantitative value of saturated fatty acids.

In an embodiment, the method comprises determining a quantitative value of omega-3 fatty acids.

10 In an embodiment, the method comprises determining a quantitative value of apolipoprotein A1 (ApoA1).

In an embodiment, the method comprises determining a quantitative value of docosahexaenoic acid (DHA).

In an embodiment, the method comprises determining a quantitative value of leucine.

15 In an embodiment, the method comprises determining a quantitative value of acetate.

In an embodiment, the method comprises determining a quantitative value of alanine.

20 In an embodiment, the method comprises determining a quantitative value of apolipoprotein B (ApoB).

In an embodiment, the method comprises determining a quantitative value of glutamine.

In an embodiment, the method comprises determining a quantitative value of isoleucine.

25 In an embodiment, the method comprises determining a quantitative value of linoleic acid.

In an embodiment, the method comprises determining a quantitative value of phenylalanine.

30 In an embodiment, the method comprises determining a quantitative value of tyrosine.

In an embodiment, the method comprises determining a quantitative value of degree of unsaturation of fatty acids.

In an embodiment, the method comprises determining a quantitative value of valine.

35 In an embodiment, the method comprises determining a quantitative value of histidine.

In an embodiment, the method comprises determining a quantitative value of cholesterol in HDL (HDL-C).

In an embodiment, the method comprises determining a quantitative value of glucose.

In an embodiment, the method comprises determining a quantitative value of acetoacetate.

5 In an embodiment, the method comprises determining a quantitative value of 3-hydroxybutyrate.

In an embodiment, the method comprises determining a quantitative value of cholesterol in LDL (LDL-C).

10 In an embodiment, the method comprises determining a quantitative value of lactate.

In an embodiment, the method comprises determining a quantitative value of triglycerides in LDL (LDL-TG).

In an embodiment, the method comprises determining a quantitative value of pyruvate.

15 In an embodiment, the method comprises determining a quantitative value of cholesterol in VLDL (VLDL-C).

In an embodiment, the method comprises determining in a biological sample obtained from the subject a quantitative value of at least one biomarker (or a plurality of the biomarkers, for example at least two, or at least three, or at least four, or all) of the following relative to the quantitative value of the albumin in the biological sample:

- omega-6 fatty acids,
- monounsaturated fatty acids,
- 25 - saturated fatty acids, or
- omega-3 fatty acids.

In an embodiment, the method comprises determining in a biological sample obtained from the subject a quantitative value of at least one biomarker (or a plurality of the biomarkers, for example at least two, or at least three, or at least four, or at least five, or at least six, or at least seven, or at least eight, or at least nine, or at least 10, or at least 11, or at least 12, or at least 13, or at least 14, or all) of the following relative to the quantitative value of the albumin in the biological sample:

- 35 - apolipoprotein A1 (ApoA1),
- docosahexaenoic acid (DHA),
- leucine,
- acetate,

10

- alanine,
- apolipoprotein B (ApoB),
- glutamine,
- isoleucine,
- 5 - linoleic acid,
- phenylalanine,
- tyrosine,
- degree of unsaturation of fatty acids,
- valine, or
- 10 - histidine.

In an embodiment, the method comprises determining in a biological sample obtained from the subject a quantitative value of at least one biomarker (or a plurality of the biomarkers, for example at least two, or at least three, or at least four, or at least five, or at least six, or at least seven, or at least eight, or at least nine, or all) of the following relative to the quantitative value of the albumin in the biological sample:

- cholesterol in HDL (HDL-C),
- glucose,
- 20 - acetoacetate,
- 3-hydroxybutyrate,
- cholesterol in LDL (LDL-C),
- lactate,
- triglycerides in LDL (LDL-TG),
- 25 - pyruvate, or
- cholesterol in VLDL (VLDL-C).

In an embodiment, in particular of the second aspect, the method comprises

determining in the biological sample obtained from the subject a quantitative value of glycoprotein acetyls; and/or determining in the biological sample obtained from the subject a quantitative value of at least one biomarker of the following:

- omega-6 fatty acids,
- 35 - monounsaturated fatty acids,
- saturated fatty acids, or
- omega-3 fatty acids; and/or

11

determining in the biological sample obtained from the subject a quantitative value of at least one biomarker of the following:

- apolipoprotein A1 (ApoA1),
- 5 - docosahexaenoic acid (DHA),
- leucine,
- acetate,
- alanine,
- apolipoprotein B (ApoB),
- 10 - glutamine,
- isoleucine,
- linoleic acid,
- phenylalanine,
- tyrosine,
- 15 - degree of unsaturation of fatty acids,
- valine, or
- histidine; and/or

determining in the biological sample obtained from the subject a quantitative value of at least one biomarker of the following:

- cholesterol in HDL (HDL-C),
- glucose,
- acetoacetate,
- 3-hydroxybutyrate,
- 25 - cholesterol in LDL (LDL-C),
- lactate,
- triglycerides in LDL (LDL-TG),
- pyruvate, or
- cholesterol in VLDL (VLDL-C).

30 The subject may be human. The subject may, additionally or alternatively, be an animal, such as a mammal, for example, a non-human primate, a dog, a cat, a horse, a sheep, a goat, a bovine, a rabbit, a pig and/or a rodent, such as a mouse or a rat, or any other species.

35 In the context of this specification, the term "biomarker" may refer to a biomarker, for example a chemical or molecular marker, that may be found associated with a disease or a condition or the risk of having or developing thereof. It does not necessarily refer to a biomarker that would be statistically

fully validated as having a specific effectiveness in a clinical setting. The biomarker may be a metabolite, a compound, a lipid, a protein, a moiety, a functional group, a composition, a combination of two or more metabolites and/or compounds, a (measurable or measured) quantity thereof, a ratio or other value derived thereof, or in principle any measurement reflecting a chemical and/or biological component that may be found associated with a disease or condition or the risk of having or developing thereof. The biomarkers and any combinations thereof, optionally in combination with further measures, may be used to indicate or measure a biological process specific to and/or indicative of a specific disease or condition or the risk of developing the same, such as e.g. an infectious disease or a complication thereof, sepsis, pneumonia, other lower respiratory infection, or diabetes. Any "ratio" of two biomarkers or of at least one biomarker to albumin may refer to the ratio of the quantitative values of the biomarkers; or the ratio of the quantitative value of the biomarker and the quantitative value of the albumin; to a ratio of a quantitative value of a biomarker to another measure (e.g. a quantitative value of a metabolite or compound, or total quantitative value of metabolites or compounds of the same class); or to any combination thereof.

When the quantitative value of the at least one biomarker is determined relative to the quantitative value of the albumin in the biological sample, it may be possible to obtain relatively accurate quantitative value for the at least one biomarker. In other words, the (initial) quantitative value of the at least one biomarker (or the (initial) quantitative values of the plurality of biomarkers) may be normalized against or scaled to the quantitative value of the albumin. This may be possible also in cases in which there may be dilution effects, for example when using a dry blood sample that is subsequently diluted prior to determining the quantitative value of the at least one biomarker, and/or when the dry blood sample or other biological sample is obtainable such that the source of the sample may cause dilution effects, for example when using a blood sample obtainable from a fingertip of the subject. Fingertip samples may contain additional fluids, such as tissue

fluid, in addition to the blood drawn from the fingertip, the relative volume of which may be challenging to control.

Therefore, it may be possible to accurately measure the at least one biomarker from dry samples, such as dry blood
5 samples. The ratio of such measures may be considered to be in good correspondence to the values obtained from venous blood.

However, in some embodiments, in particular certain embodiments according to the second aspect, it may be possible to determine a quantitative value of at least one biomarker
10 relative to a quantitative value of a scaling biomarker (other than albumin) in the biological sample. Alternatively or additionally, certain other markers may also be useful for normalizing or scaling the quantitative value of the at least one biomarker.

15 The quantitative value of the at least one biomarker relative to the quantitative value of the albumin in the biological sample may be determined, for example, by calculating a ratio of an initial quantitative value of the at least one biomarker and the quantitative value of the albumin in the
20 biological sample. It may also be possible to use an additional mathematical transformation, for example for presentation of the results, e.g. to the subject.

The method may further comprise determining whether the subject has the disease or condition or is at risk of developing
25 the disease or condition using a risk score, hazard ratio, and/or predicted absolute risk calculated on the basis of the quantitative value(s) of the at least one biomarker or of the plurality of the biomarkers relative to the quantitative value of the albumin.

30 The risk score and/or hazard ratio and/or predicted absolute risk may be calculated based on any plurality, combination or subset of biomarkers described in this specification.

The risk score and/or hazard ratio and/or predicted
35 absolute risk may be calculated e.g. as shown in the Examples below. For example, the plurality of biomarkers measured using a suitable method, for example with NMR spectroscopy, may be combined using regression algorithms and multivariate analyses and/or using machine learning analysis. Before regression

analysis or machine learning, any missing values in the biomarkers may be imputed with the mean value of each biomarker for the dataset. A number of biomarkers, for example ten, that may be considered most associated with the onset of the disease or condition may be selected for use in the prediction model. Other modelling approaches may be used to calculate a risk score and/or hazard ratio and/or predicted absolute risk based on a subset of individual biomarkers, i.e. a plurality of the biomarkers.

An increase or a decrease in the risk score, hazard ratio, odds ratio, and/or predicted absolute risk and/or relative risk may be indicative of the subject having an increased risk of developing the disease or condition.

In the context of this specification, the term "glycoprotein acetyls", "glycoprotein acetylation", or "GlycA" may refer to the abundance of circulating glycosylated proteins, and/or to a nuclear magnetic resonance spectroscopy (NMR) signal that represents the abundance of circulating glycosylated proteins, i.e. N-acetylated glycoproteins. Glycoprotein acetyls may include signals from a plurality of different glycoproteins, including e.g. alpha-1-acid glycoprotein, alpha-1 antitrypsin, haptoglobin, transferrin, and/or alpha-1 antichymotrypsin. Glycoprotein acetyls and a method for measuring them is described e.g. in Kettunen et al., 2018, *Circ Genom Precis Med*. 11:e002234 and Soininen et al., 2009, *Analyst* 134, 1781-1785.

In the context of this specification, the term "albumin" may be understood as referring to serum albumin (often referred to as blood albumin). It is an albumin found in vertebrate blood. For example, human albumin (human serum albumin) is encoded by the *ALB* gene. Serum albumin is the most abundant blood protein in mammals. Albumin is a globular, water-soluble, un-glycosylated serum protein of approximate molecular weight of 65,000 Daltons. The measurement of albumin using NMR is described e.g. in publications by Kettunen et al., 2012, *Nature Genetics* 44, 269-276; Soininen et al., 2015, *Circulation: Cardiovascular Genetics* 8, 192-206 (DOI: 10.1161/CIRCGENETICS.114.000216) and Würtz et al., 2017, *American Journal of Epidemiology* 186 (9), 1084-1096 (DOI: 10.1093/aje/kwx016). Albumin may also be measured by various

other methods, for example in a clinical setting. Examples of such methods may include e.g. dye-binding methods such as bromocresol green and bromocresol purple.

In the context of this specification, the term "HDL" refers to high-density lipoprotein.

In the context of this specification, the term "LDL" refers to low-density lipoprotein.

In the context of this specification, the term "VLDL" refers to very-low-density lipoprotein.

In the context of this specification, the term "omega-6 fatty acids" may refer to total omega-6 fatty acids. Omega-6 fatty acids are polyunsaturated fatty acids. In omega-6 fatty acids, the last double bond in the fatty acid chain is the sixth bond counting from the methyl end.

In the context of this specification, the term "monounsaturated fatty acids" (MUFAs) may refer to total monounsaturated fatty acids. Monounsaturated fatty acids have one double bond in their fatty acid chain. The MUFAs may (mainly) include omega-9 and omega-7 fatty acids. Oleic acid (18:1 ω -9), palmitoleic acid (16:1 ω -7) and cis-vaccenic acid (18:1 ω -7) are examples of common MUFAs in human serum.

In the context of this specification, the term "saturated fatty acids" (SFAs) may refer to total saturated fatty acids. Saturated fatty acids may be or comprise fatty acids which have no double bonds in their structure. Palmitic acid (16:0) and stearic acid (18:0) are examples of abundant SFAs in human serum.

In the context of this specification, the term "omega-3 fatty acids" may refer to total omega-3 fatty acids. Omega-3 fatty acids are polyunsaturated fatty acids. In omega-3 fatty acids, the last double bond in the fatty acid chain is the third bond counting from the methyl end.

For any or all fatty acid measures, including omega-3, DHA, LA, MUFA, SFA, the fatty acid measures, for example the total omega-6-fatty acids, include blood (or serum/plasma) free fatty acids, bound fatty acids and esterified fatty acids. Esterified fatty acids may, for example, be esterified to glycerol as in triglycerides, diglycerides, monoglycerides, or phosphoglycerides, or to cholesterol as in cholesterol esters.

In the context of this specification, the term "apolipoprotein(s)" may refer to apolipoprotein molecules which are amphipathic proteins and key structural components in the surface area of lipoprotein particles. They may include
5 apolipoprotein B (ApoB) and apolipoprotein A1 (ApoA1).

In the context of this specification, the term "degree of unsaturation of fatty acids" may be understood as referring to the number of double bonds in total fatty acids, for example the average number of double bonds in total fatty acids.

10 In the context of this specification, the phrase "cholesterol in HDL", or cholesterol in any other lipoprotein class, such as in LDL or VLDL, may be understood as referring to total cholesterol in said lipoprotein class or subfraction.

In the context of this specification, the term
15 "quantitative value" may refer to any quantitative value characterizing the amount and/or concentration of a biomarker. For example, it may be the amount or concentration of the biomarker in the biological sample, or it may be a signal derived from nuclear magnetic spectroscopy (NMR) or other method
20 suitable for detecting the biomarker in a quantitative manner. Such a signal may be indicative of or may correlate with the amount or concentration of the biomarker. It may also be a quantitative value calculated from one or more signals derived from NMR measurements or from other measurements. Quantitative
25 values may, additionally or alternatively, be measured using a variety of techniques. Such methods may include mass spectrometry (MS), gas chromatography combined with MS, high performance liquid chromatography alone or combined with MS, immunoturbidimetric measurements, ultracentrifugation, ion
30 mobility, enzymatic analyses, colorimetric or fluorometric analyses, immunoblot analysis, immunohistochemical methods (e.g. *in situ* methods based on antibody detection of metabolites), and immunoassays (e.g. ELISA). Examples of various methods are set out below. The method used to determine the quantitative
35 value(s) in the subject should be the same method that is used to determine the quantitative value(s) in a control subject/control subjects or in a control sample/control samples.

In the context of this specification, the term "quantitative value of albumin" may refer to any quantitative

value characterizing the amount and/or concentration of the albumin. For example, it may be the amount or concentration of the albumin in the biological sample, or it may be a signal derived from nuclear magnetic spectroscopy (NMR) or other method suitable for detecting the albumin in a quantitative manner. Such a signal may be indicative of or may correlate with the amount or concentration of the albumin. It may also be a quantitative value calculated from one or more signals derived from NMR measurements or from other measurements. Quantitative values may, additionally or alternatively, be measured using a variety of techniques. Such methods may include mass spectrometry (MS), gas chromatography combined with MS, high performance liquid chromatography alone or combined with MS, immunoturbidimetric measurements, ultracentrifugation, ion mobility, enzymatic analyses, colorimetric or fluorometric analyses, immunoblot analysis, immunohistochemical methods (e.g. *in situ* methods based on antibody detection), and immunoassays (e.g. ELISA). Examples of various methods are set out below. The method used to determine the quantitative value(s) in the subject should be the same method that is used to determine the quantitative value(s) in a control subject/control subjects or in a control sample/control samples.

In the context of this specification, the term "control subject" may refer to a subject known not to suffer from the disease or condition, and/or known not to be at risk of having or developing the disease or condition. The control subject may be a matched control subject.

The disease may be an infectious disease or a complication thereof.

In an embodiment, the disease is a severe infectious disease, a severe infection or a (severe) complication thereof.

In an embodiment, the infectious disease or the complication thereof is sepsis.

In an embodiment, the infectious disease or the complication thereof is pneumonia.

In an embodiment, the infectious disease or the complication thereof is other lower respiratory disease (i.e. a lower respiratory tract disease).

In the context of this specification, the term "pneumonia" may be understood as referring to inflammation of one or both lungs of the subject. It may be caused e.g. by a bacterial, fungal and/or viral infection. The infection may involve inflammation of the air sacs in one or both lungs of the subject. The signs and symptoms of pneumonia may vary from mild to severe or potentially life-threatening, depending on factors such as the type of microorganism causing the infection, age and/or overall health of the subject. Pneumonia is the leading cause of death from infectious diseases worldwide.

In an embodiment, the pneumonia is severe pneumonia. Severe pneumonia may include or result in e.g. hospitalization and/or death.

In an embodiment, the method comprises determining a quantitative value of glycoprotein acetyls.

Glycoprotein acetyls may be useful for determining whether the subject is at risk of developing an infectious disease or a complication thereof, for example sepsis, pneumonia, such as severe pneumonia, or other lower respiratory infection.

In an embodiment, the disease or complication thereof is diabetes, such as type 2 diabetes.

The method may further comprise determining whether the subject is at risk of developing the infectious disease or the complication thereof using a risk score, hazard ratio, and/or predicted absolute risk calculated on the basis of the quantitative value(s) of the at least one biomarker or of the plurality of the biomarkers.

The risk score, hazard ratio, and/or predicted absolute risk may be calculated based on any plurality, combination or subset of biomarkers described in this specification.

The risk score, hazard ratio, and/or predicted absolute risk may be calculated e.g. as shown in the Examples below. For example, the plurality of biomarkers measured using a suitable method, for example with NMR spectroscopy, may be combined using regression algorithms and multivariate analyses and/or using machine learning analysis. Before regression analysis or machine learning, any missing values in the biomarkers may be imputed with the mean value of each biomarker for the dataset. A number

of biomarkers, for example ten, that may be considered most associated with the onset of the disease or condition may be selected for use in the prediction model. Other modelling approaches may be used to calculate a risk score and/or hazard ratio and/or predicted absolute risk based on a subset of individual biomarkers, i.e. a plurality of the biomarkers.

In an embodiment, a method for determining whether a subject has pneumonia or is at risk of developing pneumonia comprises

determining in a biological sample obtained from the subject a quantitative value of at least one biomarker relative to a quantitative value of albumin in the biological sample; and comparing the quantitative value(s) of the at least one biomarker to a control sample or to a control value;

wherein an increase or a decrease in the quantitative value(s) of the at least one biomarker, when compared to the control sample or to the control value, is/are indicative of the subject having pneumonia or having an increased risk of developing pneumonia.

In an embodiment of the above embodiment, the at least one biomarker may comprise or be glycoprotein acetyls.

In an embodiment of the above aspect, the biological sample may be a dry blood sample or obtainable from a dry blood sample. However, the biological sample may, alternatively or additionally, be any other biological sample, for example any biological sample described in this specification.

The quantitative value, or the initial quantitative value, of the at least one biomarker, or the plurality of the biomarkers, and/or the quantitative value of the albumin may be measured using nuclear magnetic spectroscopy (NMR), for example ^1H -NMR. The at least one additional biomarker, or the plurality of the additional biomarkers, may also be measured using NMR. NMR may provide a particularly efficient and fast way to measure biomarkers, including a large number of biomarkers simultaneously, and can provide quantitative values for them. NMR also typically requires very little sample pre-treatment or preparation. The biomarkers measured with NMR can effectively be measured for large amounts of samples using an assay for blood (serum or plasma) NMR metabolomics previously published by

Soininen et al., 2015, Circulation: Cardiovascular Genetics 8, 192-206 (DOI: 10.1161/CIRCGENETICS.114.000216); Soininen et al., 2009, Analyst 134, 1781-1785; and Würtz et al., 2017, American Journal of Epidemiology 186 (9), 1084-1096 (DOI: 10.1093/aje/kwx016). This provides data on 228 biomarkers per sample as described in detail in the above scientific papers.

In an embodiment, the (initial) quantitative value of the at least one biomarker and/or the quantitative value of the albumin is/are measured using nuclear magnetic spectroscopy.

However, quantitative values for various biomarkers described in this specification may also be performed by techniques other than NMR. For example, mass spectrometry (MS), enzymatic methods, antibody-based detection methods, or other biochemical or chemical methods may be contemplated, depending on the biomarker.

For example, glycoprotein acetyls can be measured or approximated by immunoturbidimetric measurements of alpha-1-acid glycoprotein, haptoglobin, alpha-1-antitrypsin, and transferrin (e.g. as described in Ritchie et al., 2015, Cell Syst. 28;1(4):293-301).

E.g. monounsaturated fatty acids and polyunsaturated fatty acids can be quantified (i.e. their quantitative values may be determined) by serum total fatty acid composition using gas chromatography (for example, as described in Jula et al., 2005, Arterioscler Thromb Vasc Biol 25, 1952-1959).

Cholesterol in lipoprotein fractions can be quantified using high performance liquid chromatography.

Apolipoprotein B, apolipoprotein A1, and glucose can be quantified by enzymatic clinical chemistry analysers, such as ROCHE COBAS 6000.

In the context of this specification, the term "sample" or "biological sample" may refer to any biological sample obtained from a subject or a group or population of subjects. The sample may be fresh, frozen, or dry.

The biological sample (in particular in embodiments in which the biological sample is not a dry blood sample) may comprise or be, for example, a blood sample, a plasma sample, a serum sample, or a sample derived therefrom. The biological sample may be, for example, a fasting blood sample, a fasting

plasma sample, a fasting serum sample, or a fraction obtainable therefrom. However, the biological sample does not necessarily have to be a fasting sample. The blood sample may be a venous blood sample.

5 The dry blood sample may be a dried whole blood sample, a dried plasma sample, a dried serum sample, or a dried sample derived therefrom.

 Some of the biomarkers may be determined from body fluids other than blood or fractions obtainable therefrom. The biological sample may, additionally or alternatively, comprise
10 or be a sample or one or more other body fluids or biofluids, for example an amniotic fluid sample, a urine sample, a saliva sample, a bile sample, a tear sample and/or a spinal fluid sample.

15 The method may comprise obtaining the biological sample from the subject prior to determining the quantitative value of the at least one biomarker. Taking a blood sample or a tissue sample of a subject or patient is a part of normal clinical practice. The collected blood or tissue sample can be prepared
20 and serum or plasma can be separated using techniques well known to a skilled person. Methods for separating one or more fractions from biological samples, such as blood samples or tissue samples, are also available to a skilled person. The term "fraction" may, in the context of this specification, also refer
25 to a portion or a component of the biological sample separated according to one or more physical properties, for instance solubility, hydrophilicity or hydrophobicity, density, or molecular size.

 The dry blood sample may be, in principle, any dried
30 blood sample.

 The dry blood sample may be a dried blood spot. The dry blood sample, for example for NMR analysis, may be provided as dried blood spot on filter paper or other substrate or media that is capable of absorbing blood. Examples of other substrates
35 or media may comprise or be e.g. various fibers and/or resins. The steps related to providing the sample and applying blood, which can be for example obtained either with venepuncture or fingerprick, on a paper, e.g. a filter paper, or other substrate or media that can absorb blood; and drying the blood on the

paper or other substrate or media where the blood is applied. During the drying process, the blood may also be passively separated to its components, for example to blood cells and plasma or serum, on the paper or other media where the blood is applied. The piece of the paper or other media where the blood or its components have dried may be punched or otherwise cut out of the paper or media. Alternatively or additionally, the dry blood sample may be provided as a dry blood sample in a HemaSpot HF device or other suitable device. Various other means for collecting and/or storing dry blood samples may also be contemplated.

The biological sample may be obtainable by further sample preparation from the dry blood sample.

The biological sample may be obtainable from the dry blood sample or from the dry sample obtainable from blood by extracting at least a part of the blood or one or more components thereof from the dry blood sample or from the dry sample obtainable from blood into a solvent. The subsequent sample preparation may include e.g. extracting the blood or its components from the dry blood sample, for example from the cut or punched piece of paper or other media. The blood or its components may be extracted with e.g. an aqueous buffer, or other suitable solvent, thus providing a liquid sample for NMR analysis or other suitable analytical method. A deuterated solvent may be added for ^1H NMR.

In the context of this specification, the term "control sample" may refer to a sample obtained from a subject and known not to suffer from the disease or condition or not being at risk of having or developing the disease or condition. The control sample may be matched. In an embodiment, the control sample may be a biological sample from a healthy individual or a generalized population of healthy individuals. The term "control value" may be understood as a value obtainable from the control sample and/or a quantitative value derivable therefrom. For example, it may be possible to calculate a threshold value from control samples and/or control values, above or below which the risk of developing the disease or condition is elevated. In other words, a value higher or lower (depending on the biomarker, risk score, hazard ratio, and/or predicted absolute

risk) than the threshold value may be indicative of the subject having an increased risk of developing the disease or condition.

An increase or a decrease in the quantitative value(s) of the at least one biomarker, or the plurality of the biomarkers, when compared to the control sample or to the control value, may be indicative of the subject having an increased risk of having or developing the disease or condition. Whether an increase or a decrease is indicative of the subject having an increased risk of developing the disease or condition, may depend on the biomarker.

A 1.2-fold, 1.5-fold, or for example 2-fold, or 3-fold, increase or a decrease in the quantitative value(s) of the at least one biomarker (or in an individual biomarker of the plurality of the biomarkers) when compared to the control sample or to the control value, may be indicative of the subject having an increased risk of developing the disease or condition.

In an embodiment, an increase in the quantitative value of glycoprotein acetyls, when compared to the control sample or to the control value, may be indicative of the subject having an increased risk of developing the infectious disease or the complication thereof, such as pneumonia, other lower respiratory infection, and/or sepsis.

In an embodiment, an increase in the quantitative value of glycoprotein acetyls relative to the quantitative value of the albumin in the biological sample, when compared to the control sample or to the control value, may be indicative of the subject having an increased risk of developing the infectious disease or the complication thereof, such as pneumonia, other lower respiratory infection, and/or sepsis.

In an embodiment, a decrease in the quantitative value of omega-6 fatty acids, when compared to the control sample or to the control value, may be indicative of the subject having an increased risk of developing the infectious disease or the complication thereof, such as pneumonia, other lower respiratory infection, and/or sepsis.

In an embodiment, a decrease in the quantitative value of omega-6 fatty acids relative to the quantitative value of the albumin in the biological sample, when compared to the control

sample or to the control value, may be indicative of the subject having an increased risk of developing the infectious disease or the complication thereof, such as pneumonia, other lower respiratory infection, and/or sepsis.

5 In an embodiment, an increase in the quantitative value of monounsaturated fatty acids, when compared to the control sample or to the control value, may be indicative of the subject having an increased risk of developing the infectious disease or the complication thereof, such as pneumonia, other lower
10 respiratory infection, and/or sepsis.

 In an embodiment, an increase in the quantitative value of monounsaturated fatty acids relative to the quantitative value of the albumin in the biological sample, when compared to the control sample or to the control value, may be indicative of
15 the subject having an increased risk of developing the infectious disease or the complication thereof, such as pneumonia, other lower respiratory infection, and/or sepsis.

 In an embodiment, a decrease in the quantitative value of saturated fatty acids, when compared to the control sample or
20 to the control value, may be indicative of the subject having an increased risk of developing the infectious disease or the complication thereof, such as pneumonia, other lower respiratory infection, and/or sepsis.

 In an embodiment, an increase in the quantitative value of saturated fatty acids relative to the quantitative value of the albumin in the biological sample, when compared to the control sample or to the control value, may be indicative of the
25 subject having an increased risk of developing the infectious disease or the complication thereof, such as pneumonia, other
30 lower respiratory infection, and/or sepsis.

 In an embodiment, a decrease in the quantitative value of omega-3 fatty acids, when compared to the control sample or to the control value, may be indicative of the subject having an increased risk of developing the infectious disease or the
35 complication thereof, such as pneumonia, other lower respiratory infection, and/or sepsis.

 In an embodiment, a decrease in the quantitative value of omega-3 fatty acids relative to the quantitative value of the albumin in the biological sample, when compared to the control

sample or to the control value, may be indicative of the subject having an increased risk of developing the infectious disease or the complication thereof, such as pneumonia, other lower respiratory infection, and/or sepsis.

5 In an embodiment, a decrease in the quantitative value of apolipoprotein A1 (ApoA1), when compared to the control sample or to the control value, may be indicative of the subject having an increased risk of developing the infectious disease or the complication thereof, such as pneumonia, other lower
10 respiratory infection, and/or sepsis.

In an embodiment, a decrease in the quantitative value of apolipoprotein A1 (ApoA1) relative to the quantitative value of the albumin in the biological sample, when compared to the control sample or to the control value, may be indicative of the
15 subject having an increased risk of developing the infectious disease or the complication thereof, such as pneumonia, other lower respiratory infection, and/or sepsis.

In an embodiment, a decrease in the quantitative value of docosahexaenoic acid (DHA), when compared to the control sample or to the control value, may be indicative of the subject having an increased risk of developing the infectious disease or the complication thereof, such as pneumonia, other lower
20 respiratory infection, and/or sepsis.

In an embodiment, a decrease in the quantitative value of docosahexaenoic acid (DHA) relative to the quantitative value of the albumin in the biological sample, when compared to the control sample or to the control value, may be indicative of the subject having an increased risk of developing the infectious disease or the complication thereof, such as pneumonia, other
25 lower respiratory infection, and/or sepsis.

In an embodiment, a decrease in the quantitative value of leucine, when compared to the control sample or to the control value, may be indicative of the subject having an increased risk of developing the infectious disease or the complication thereof, such as pneumonia, other lower respiratory
30 infection, and/or sepsis.

In an embodiment, a decrease in the quantitative value of leucine relative to the quantitative value of the albumin in the biological sample, when compared to the control sample or to

the control value, may be indicative of the subject having an increased risk of developing the infectious disease or the complication thereof, such as pneumonia, other lower respiratory infection, and/or sepsis.

5 In an embodiment, a decrease in the quantitative value of acetate, when compared to the control sample or to the control value, may be indicative of the subject having an increased risk of developing the infectious disease or the complication thereof, such as pneumonia, other lower respiratory
10 infection, and/or sepsis.

In an embodiment, a decrease in the quantitative value of acetate relative to the quantitative value of the albumin in the biological sample, when compared to the control sample or to the control value, may be indicative of the subject having an
15 increased risk of developing the infectious disease or the complication thereof, such as pneumonia, other lower respiratory infection, and/or sepsis.

In an embodiment, a decrease in the quantitative value of alanine, when compared to the control sample or to the control value, may be indicative of the subject having an
20 increased risk of developing the infectious disease or the complication thereof, such as pneumonia, other lower respiratory infection, and/or sepsis.

In an embodiment, an increase in the quantitative value of alanine relative to the quantitative value of the albumin in the biological sample, when compared to the control sample or to the control value, may be indicative of the subject having an
25 increased risk of developing the infectious disease or the complication thereof, such as pneumonia, other lower respiratory infection, and/or sepsis.
30

In an embodiment, a decrease in the quantitative value of apolipoprotein B (ApoB), when compared to the control sample or to the control value, may be indicative of the subject having an increased risk of developing the infectious disease or the
35 complication thereof, such as pneumonia, other lower respiratory infection, and/or sepsis.

In an embodiment, a decrease in the quantitative value of apolipoprotein B (ApoB) relative to the quantitative value of the albumin in the biological sample, when compared to the

control sample or to the control value, may be indicative of the subject having an increased risk of developing the infectious disease or the complication thereof, such as pneumonia, other lower respiratory infection, and/or sepsis.

5 In an embodiment, a decrease in the quantitative value of glutamine, when compared to the control sample or to the control value, may be indicative of the subject having an increased risk of developing the infectious disease or the complication thereof, such as pneumonia, other lower respiratory
10 infection, and/or sepsis.

In an embodiment, an increase in the quantitative value of glutamine relative to the quantitative value of the albumin in the biological sample, when compared to the control sample or to the control value, may be indicative of the subject having an
15 increased risk of developing the infectious disease or the complication thereof, such as pneumonia, other lower respiratory infection, and/or sepsis.

In an embodiment, a decrease in the quantitative value of isoleucine, when compared to the control sample or to the control value, may be indicative of the subject having an
20 increased risk of developing the infectious disease or the complication thereof, such as pneumonia, other lower respiratory infection, and/or sepsis.

In an embodiment, an increase in the quantitative value of isoleucine relative to the quantitative value of the albumin in the biological sample, when compared to the control sample or to the control value, may be indicative of the subject having an
25 increased risk of developing the infectious disease or the complication thereof, such as pneumonia, other lower respiratory infection, and/or sepsis.
30

In an embodiment, a decrease in the quantitative value of linoleic acid (LA), when compared to the control sample or to the control value, may be indicative of the subject having an
35 increased risk of developing the infectious disease or the complication thereof, such as pneumonia, other lower respiratory infection, and/or sepsis.

In an embodiment, a decrease in the quantitative value of linoleic acid relative to the quantitative value of the albumin in the biological sample, when compared to the control

sample or to the control value, may be indicative of the subject having an increased risk of developing the infectious disease or the complication thereof, such as pneumonia, other lower respiratory infection, and/or sepsis.

5 In an embodiment, an increase in the quantitative value of phenylalanine, when compared to the control sample or to the control value, may be indicative of the subject having an increased risk of developing the infectious disease or the complication thereof, such as pneumonia, other lower respiratory
10 infection, and/or sepsis.

 In an embodiment, an increase in the quantitative value of phenylalanine relative to the quantitative value of the albumin in the biological sample, when compared to the control sample or to the control value, may be indicative of the subject
15 having an increased risk of developing the infectious disease or the complication thereof, such as pneumonia, other lower respiratory infection, and/or sepsis.

 In an embodiment, an increase in the quantitative value of tyrosine, when compared to the control sample or to the control value, may be indicative of the subject having an increased risk of developing the infectious disease or the complication thereof, such as pneumonia, other lower respiratory
20 infection, and/or sepsis.

 In an embodiment, an increase in the quantitative value of tyrosine relative to the quantitative value of the albumin in the biological sample, when compared to the control sample or to the control value, may be indicative of the subject having an increased risk of developing the infectious disease or the complication thereof, such as pneumonia, other lower respiratory
25 infection, and/or sepsis.
30

 In an embodiment, a decrease in the quantitative value of degree of unsaturation of fatty acids, when compared to the control sample or to the control value, may be indicative of the subject having an increased risk of developing the infectious
35 disease or the complication thereof, such as pneumonia, other lower respiratory infection, and/or sepsis.

 In an embodiment, an increase in the quantitative value of degree of unsaturation of fatty acids relative to the quantitative value of the albumin in the biological sample, when

compared to the control sample or to the control value, may be indicative of the subject having an increased risk of developing the infectious disease or the complication thereof, such as pneumonia, other lower respiratory infection, and/or sepsis.

5 In an embodiment, a decrease in the quantitative value of valine, when compared to the control sample or to the control value, may be indicative of the subject having an increased risk of developing the infectious disease or the complication thereof, such as pneumonia, other lower respiratory infection,
10 and/or sepsis.

 In an embodiment, an increase in the quantitative value of valine relative to the quantitative value of the albumin in the biological sample, when compared to the control sample or to the control value, may be indicative of the subject having an
15 increased risk of developing the infectious disease or the complication thereof, such as pneumonia, other lower respiratory infection, and/or sepsis.

 In an embodiment, a decrease in the quantitative value of histidine, when compared to the control sample or to the
20 control value, may be indicative of the subject having an increased risk of developing the infectious disease or the complication thereof, such as pneumonia, other lower respiratory infection, and/or sepsis.

 In an embodiment, a decrease in the quantitative value of histidine relative to the quantitative value of the albumin in the biological sample, when compared to the control sample or to the control value, may be indicative of the subject having an
25 increased risk of developing the infectious disease or the complication thereof, such as pneumonia, other lower respiratory infection, and/or sepsis.
30 infection, and/or sepsis.

 In an embodiment, a decrease in the quantitative value of cholesterol in HDL (HDL-C), when compared to the control sample or to the control value, may be indicative of the subject having an increased risk of developing the infectious disease or
35 the complication thereof, such as pneumonia, other lower respiratory infection, and/or sepsis.

 In an embodiment, a decrease in the quantitative value of cholesterol in HDL (HDL-C) relative to the quantitative value of the albumin in the biological sample, when compared to the

control sample or to the control value, may be indicative of the subject having an increased risk of developing the infectious disease or the complication thereof, such as pneumonia, other lower respiratory infection, and/or sepsis.

5 In an embodiment, an increase in the quantitative value of glucose, when compared to the control sample or to the control value, may be indicative of the subject having an increased risk of developing the infectious disease or the complication thereof, such as pneumonia, other lower respiratory
10 infection, and/or sepsis.

In an embodiment, an increase in the quantitative value of glucose relative to the quantitative value of the albumin in the biological sample, when compared to the control sample or to the control value, may be indicative of the subject having an
15 increased risk of developing the infectious disease or the complication thereof, such as pneumonia, other lower respiratory infection, and/or sepsis.

In an embodiment, an increase in the quantitative value of acetoacetate, when compared to the control sample or to the control value, may be indicative of the subject having an
20 increased risk of developing the infectious disease or the complication thereof, such as pneumonia, other lower respiratory infection, and/or sepsis.

In an embodiment, an increase in the quantitative value of acetoacetate relative to the quantitative value of the albumin in the biological sample, when compared to the control sample or to the control value, may be indicative of the subject having an increased risk of developing the infectious disease or the complication thereof, such as pneumonia, other lower
25 respiratory infection, and/or sepsis.
30

In an embodiment, an increase in the quantitative value of 3-hydroxybutyrate, when compared to the control sample or to the control value, may be indicative of the subject having an increased risk of developing the infectious disease or the
35 complication thereof, such as pneumonia, other lower respiratory infection, and/or sepsis.

In an embodiment, an increase in the quantitative value of 3-hydroxybutyrate relative to the quantitative value of the albumin in the biological sample, when compared to the control

sample or to the control value, may be indicative of the subject having an increased risk of developing the infectious disease or the complication thereof, such as pneumonia, other lower respiratory infection, and/or sepsis.

5 In an embodiment, a decrease in the quantitative value of cholesterol in LDL (LDL-C), when compared to the control sample or to the control value, may be indicative of the subject having an increased risk of developing the infectious disease or the complication thereof, such as pneumonia, other lower
10 respiratory infection, and/or sepsis.

In an embodiment, a decrease in the quantitative value of cholesterol in LDL (LDL-C) relative to the quantitative value of the albumin in the biological sample, when compared to the control sample or to the control value, may be indicative of the
15 subject having an increased risk of developing the infectious disease or the complication thereof, such as pneumonia, other lower respiratory infection, and/or sepsis.

In an embodiment, an increase in the quantitative value of lactate, when compared to the control sample or to the
20 control value, may be indicative of the subject having an increased risk of developing the infectious disease or the complication thereof, such as pneumonia, other lower respiratory infection, and/or sepsis.

In an embodiment, an increase in the quantitative value of lactate relative to the quantitative value of the albumin in the biological sample, when compared to the control sample or to the control value, may be indicative of the subject having an
25 increased risk of developing the infectious disease or the complication thereof, such as pneumonia, other lower respiratory infection, and/or sepsis.
30

In an embodiment, an increase in the quantitative value of triglycerides in LDL (LDL-TG), when compared to the control sample or to the control value, may be indicative of the subject having an increased risk of developing the infectious disease or
35 the complication thereof, such as pneumonia, other lower respiratory infection, and/or sepsis.

In an embodiment, an increase in the quantitative value of triglycerides in LDL (LDL-TG) relative to the quantitative value of the albumin in the biological sample, when compared to

the control sample or to the control value, may be indicative of the subject having an increased risk of developing the infectious disease or the complication thereof, such as pneumonia, other lower respiratory infection, and/or sepsis.

5 In an embodiment, an increase in the quantitative value of pyruvate, when compared to the control sample or to the control value, may be indicative of the subject having an increased risk of developing the infectious disease or the complication thereof, such as pneumonia, other lower respiratory
10 infection, and/or sepsis.

In an embodiment, an increase in the quantitative value of pyruvate relative to the quantitative value of the albumin in the biological sample, when compared to the control sample or to the control value, may be indicative of the subject having an
15 increased risk of developing the infectious disease or the complication thereof, such as pneumonia, other lower respiratory infection, and/or sepsis.

In an embodiment, a decrease in the quantitative value of cholesterol in VLDL (VLDL-C), when compared to the control
20 sample or to the control value, may be indicative of the subject having an increased risk of developing the infectious disease or the complication thereof, such as pneumonia, other lower respiratory infection, and/or sepsis.

In an embodiment, a decrease in the quantitative value
25 of cholesterol in VLDL (VLDL-C) relative to the quantitative value of the albumin in the biological sample, when compared to the control sample or to the control value, may be indicative of the subject having an increased risk of developing the infectious disease or the complication thereof, such as
30 pneumonia, other lower respiratory infection, and/or sepsis.

EXAMPLES

Reference will now be made in detail to various
embodiments, an example of which is illustrated in the
35 accompanying drawings.

The description below discloses some embodiments in such a detail that a person skilled in the art is able to utilize the embodiments based on the disclosure. Not all steps or features of the embodiments are discussed in detail, as many

of the steps or features will be obvious for the person skilled in the art based on this specification.

EXAMPLE 1

5 Biomarker measures quantified by nuclear magnetic resonance (NMR) were investigated as to whether they could be predictive of the following infectious diseases and their complications even many years after the blood sampling: Pneumonia and severe
10 Pneumonia, Other lower respiratory diseases, and sepsis. In addition, biomarker associations with type 2 diabetes were examined to explore the effect of scaling the individual biomarker concentrations relative to albumin. All analyses were conducted based on the UK Biobank, with over 100,000 study participants with blood biomarker data from NMR available.

15

Study population

Details of the design of the UK Biobank have been reported by Sudlow et al 2015, PLoS Med. 2015;12(3):e1001779. Briefly, UK Biobank recruited 502 639 participants aged 37-73
20 years in 22 assessment centres across the UK. All participants provided written informed consent and ethical approval was obtained from the North West Multi-Center Research Ethics Committee. Blood samples were drawn at baseline between 2007 and 2010. No selection criteria were applied to the sampling.

25

NMR metabolomics

From the entire UK Biobank population, a random subset of baseline plasma samples from 118 466 individuals were measured using the Nightingale high-throughput NMR metabolomics platform. This provides simultaneous quantification of routine
30 lipids, lipoprotein subclass profiling with lipid concentrations within 14 subclasses, fatty acid composition, and various low-molecular weight metabolites including amino acids, ketone bodies and gluconeogenesis-related metabolites in molar concentration units. Technical details and epidemiological
35 applications have been reviewed (Soininen et al 2015, Circ Cardiovasc Genet; 2015;8:192-206; Würtz et al 2017, Am J Epidemiol 2017;186:1084-1096).

The quantified metabolomic biomarker data from UK Biobank samples were curated and approved mid-May 2020. Values

outside four interquartile ranges from median were considered as outliers and excluded.

Analyses of biomarker relations with disease risk

5 The blood biomarker associations with disease risk were conducted based on UK Biobank data. Analyses focused the relation of biomarker concentration on disease occurrences after the blood samples were collected, to determine how each individual biomarker could predict future disease risk. Examples of how multi-biomarker scores, in the form a weighted plurality of biomarkers, were also explored whether they could be predictive of future disease risk, even stronger than each individual biomarker. Information on disease events occurring after the blood samplings for all study participants were recorded from UK Hospital Episode Statistics data, death registries as well as primary care records and self-reports. All analyses are based on first occurrence of diagnosis, so individuals with recorded diagnosis of the given disease prior to blood sampling were omitted from the statistical analyses. The following diagnoses were used to define the disease endpoints: For type 2 diabetes diagnosis, ICD-10 E11 and operative procedures were used, as described in Rao et al. Circ Genom Precis Med. 2018;11(7):e002162. Self-reported diabetes was not included in the endpoint. Pneumonia-related diagnoses were based on any occurrence of ICD-10 diagnoses J12-J18. Sepsis was defined as a diagnosis of either A40 (Streptococcal sepsis) or other sepsis (A41), with diagnoses recorded in primary care, hospital or death registries; Other Lower Respiratory Infections were defined as acute bronchitis, acute bronchiolitis, or unspecified acute lower respiratory infection, with respective ICD-10 diagnosis codes J20, J21 or J22. For analyses on severe pneumonia, individuals with pneumonia diagnoses recorded in primary care settings and as self-reports were excluded from the analyses; in other words, severe pneumonia was defined as having a diagnosis J12-J18 in hospital or death registries. There were approximately 110 000 individuals for available for the analyses of biomarkers with these diseases. The registry-based follow-up was from blood sampling in 2007-2010 through to 2016-2017, depending on UK Biobank assessment centre (approximately 872 000 person-years). For analyses on severe

pneumonia, further analyses were conducted to demonstrate that the results were consistent for men and women, across different ages, and for people with and without chronic disease conditions at the time of the blood sampling. Specifically, the analyses on biomarker scores vs severe pneumonia were repeated when individuals with chronic respiratory and cardiometabolic diseases (cardiovascular disease, diabetes, lung cancer, chronic obstructive pulmonary disease, liver diseases, and renal failure) were omitted from the analyses. Analyses of biomarker scores were also conducted vs short-term risk of severe pneumonia, defined as pneumonia diagnoses recorded from hospital or death registries during the first 2 years after the blood samples were taken.

For biomarker association testing, Cox proportional-hazard regression models adjusted for age, sex, and assessment centre were used. The biomarkers were studied for the magnitude of association without and with scaling relative to albumin concentration. Results were plotted in magnitudes per standard deviation of each biomarker measure to allow direct comparison of association magnitudes. The relations of 3 specific plurality-biomarker-scores (scores calculated on the basis of a plurality of biomarkers) vs risk for severe pneumonia were examined in further detail, in form of analyses of different subsets of the study population (with and without individuals who had chronic respiratory or cardiometabolic disease at time of blood sampling, different age groups of the study population, men and women separately, and short-term compared to long-term risk of severe pneumonia). In addition, the same 3 plurality-biomarker-scores were plotted in form of gradient percentile plots, showing the proportion of individuals who developed severe pneumonia during follow-up when binning individuals into the percentiles of the biomarker levels. We also examined these 3 plurality-biomarker-scores using Kaplan-Meier plots of the cumulative risk for pneumonia during follow-up, for quintiles of the plurality-biomarker-scores and further stratified for extreme quantiles of the two biomarkers.

Summary of results

Baseline characteristics of the study population for biomarker analyses vs future disease risk are shown in Table 1.

The number of disease cases for severe pneumonia is exemplified. Among the 108 449 study participants with complete data available to calculate the biomarker scores, there were 2531 pneumonia events recorded in hospital or death registries after the baseline blood sampling (median follow-up time 8.1 years). For other diseases studied here, the number of individuals who developed the given disease after the blood samples were taken are listed in the Figures that illustrate the results, and the clinical characteristics are similar to the data in Table 1. For the plurality-biomarker scores shown in Figure 3a and Figure 3b, only half of the study population was used in the plots, since the other half was used to derive the weights for the combinations of biomarkers in the plurality-biomarker scores.

15

Table 1: Clinical characteristics of study participants.

	Severe Pneumonia (hospitalisation or death)	
	Incident cases	Controls
Individuals with NMR biomarker measures	2531	105 918
Age (median)	62.0	57.0
Females (%)	44 %	54 %
Cardiovascular disease (%)	18 %	6 %
Diabetes (%)	10 %	4 %
Lung cancer (%)	1 %	0.2 %
Chronic obstructive pulmonary disease (%)	6 %	1 %
Liver diseases (%)	2 %	1 %
Renal failure (%)	4 %	1 %

20 Biomarker associations for future diabetes and pneumonia when scaling concentrations relative to albumin concentration

Fig. 1a shows the hazard ratios for 29 blood biomarkers for future onset of type 2 diabetes. The results are based on

statistical analyses of over 110,000 individuals from the UK Biobank, out of whom over 4,000 developed type 2 diabetes during a median of 8.1 year follow-up. The results are shown for absolute concentration of the biomarkers (grey circles), and when the biomarker concentrations are scaled relative to the concentration of albumin measured from the same blood sample (black triangles). The analyses were adjusted for age, sex, and UK Biobank assessment centre in Cox proportional-hazard regression models. The results for both unscaled and albumin-scaled biomarkers are shown per 1-standard deviation increase in the given biomarker measure.

All the 29 biomarkers were statistically significantly associated with future onset of type 2 diabetes, both when analysed as absolute concentrations, and when the concentrations were scaled relative to albumin. All associations were in the same direction in relation to type 2 diabetes risk with and without scaling concentrations relative to albumin.

Figure 1b shows the hazard ratios for 29 blood biomarkers for future onset of severe pneumonia. The results are based on statistical analyses of over 100,000 individuals from the UK Biobank, out of whom over 2,400 developed severe pneumonia (defined as diagnosis J12-J18 in the national hospital registries or death records) during a median of 8.1 year follow-up. The results are shown for absolute concentration of the biomarkers (grey circles), and when the biomarker concentrations are scaled relative to the concentration of albumin measured from the same blood sample (black triangles). The analyses were adjusted for age, sex, and UK Biobank assessment centre in Cox proportional-hazard regression models. The results for both unscaled and albumin-scaled biomarkers are shown per 1-standard deviation increase in the given biomarker measure.

All 29 biomarkers were associated with future onset of severe pneumonia when analysed either as absolute concentrations, or when the concentrations were scaled relative to albumin. 23 out of the 29 biomarkers were statistically significantly associated with future onset of severe pneumonia when analysed as absolute concentrations. There were also 23 out of biomarkers that were statistically significantly associated with

future onset of severe pneumonia when the concentrations were scaled relative to albumin.

Overall, the analyses illustrate that a broad span of blood biomarkers measured by NMR spectroscopy are indicative of the risk for future type 2 diabetes and severe pneumonia in general population settings. The biomarker associations for type 2 diabetes are similar if the concentrations are scaled relative to albumin concentration. The biomarker associations for severe pneumonia are altered to some extent when concentrations are scaled relative to albumin, but many biomarkers are still strongly associated with risk for severe pneumonia after the albumin-scaling, or become even stronger associated with severe pneumonia after the albumin scaling.

Abbreviations used in the Figures:

VLDL: Very-Low-Density Lipoprotein
LDL: Low-Density Lipoprotein
HDL: High-Density Lipoprotein
C: cholesterol
TG: Triglycerides
LA: Linoleic acid
DHA: Docosahexaenoic acid
SFA: Saturated fatty acids
MUFA: Monounsaturated fatty acids
SD: standard deviation

EXAMPLE 2

The study population and statistical methods for Example 2 are as for Example 1.

Summary of Results on Biomarkers for Infectious diseases and their complications

Figure 2a shows the hazard ratios for 29 blood biomarkers for future onset of pneumonia (either severe or non-severe), when the biomarkers are analysed in absolute concentrations. The association for a plurality-biomarker-score termed "NGH infection biomarker score", comprising a weighted sum of the individual biomarkers, is also shown. The results are based

on statistical analyses of over 100,000 individuals from the UK Biobank, out of whom over 2,650 developed pneumonia (defined as diagnosis J12-J18 in the primary care records, or in the national hospital registries, or in the death records, or self-reported) during a median of 8.1 year follow-up. Filled circles denote that the P-value for association was $P < 0.001$ (corresponding to multiple testing correction), and open circles that the P-value for association was $P > 0.001$. The analyses were adjusted for age, sex, and UK Biobank assessment centre in Cox proportional-hazard regression models. The results are shown per 1-standard deviation increase in the given biomarker measure. 22 out of the 29 biomarkers were statistically significantly associated with future onset of pneumonia, based on $P < 0.001$. The strongest association with risk for future development of pneumonia was observed for the plurality-biomarker-score termed "NGH infection biomarker score".

Figure 2b shows the relation of baseline biomarker levels to future development of Other Lower Respiratory Diseases (acute bronchitis, acute bronchiolitis, and Unspecified acute lower respiratory infection), when the biomarker concentrations are analysed in absolute concentrations. shows the hazard ratios for 29 blood biomarkers for future onset of sepsis, when the biomarkers are analysed in absolute concentrations. The association for a plurality-biomarker-score termed "NGH infection biomarker score", comprising a weighted sum of the individual biomarkers, is also shown. The results are based on statistical analyses of over 100,000 individuals from the UK Biobank, out of whom over 5,900 developed sepsis (defined as diagnosis J20 (acute bronchitis) or J21 (acute bronchiolitis) or J22 (Other acute lower respiratory infections) in the primary care records, self-reported, or in the national hospital registries, or in the death records) during a median of 8.1 year follow-up. Filled circles denote that the P-value for association was $P < 0.001$, and open circles that the P-value for association was $P > 0.001$. The analyses were adjusted for age, sex, and UK Biobank assessment centre in Cox proportional-hazard regression models. The results are shown per 1-standard deviation increase in the given biomarker measure. 21 out of the 29 biomarkers were statistically significantly associated with future onset of sepsis, based on

P<0.001. The strongest association with risk for future development of sepsis was observed for the plurality-biomarker-score termed "NGH infection biomarker score".

5 **Figure 2c** shows the hazard ratios for 29 blood biomarkers for future onset of sepsis, when the biomarkers are analysed in absolute concentrations. The association for a plurality-biomarker-score termed "NGH infection biomarker score", comprising a weighted sum of the individual biomarkers, is also
10 shown. The results are based on statistical analyses of over 100,000 individuals from the UK Biobank, out of whom over 1,250 developed sepsis (defined as diagnosis A40 (Streptococcal sepsis) or A41 (Other sepsis) in the primary care records, self-reported, or in the national hospital registries, or in the death
15 records) during a median of 8.1 year follow-up. Filled circles denote that the P-value for association was P<0.001, and open circles that the P-value for association was P>0.001. The analyses were adjusted for age, sex, and UK Biobank assessment centre in Cox proportional-hazard regression models. The results
20 are shown per 1-standard deviation increase in the given biomarker measure. 15 out of the 29 biomarkers were statistically significantly associated with future onset of sepsis, based P<0.001. The strongest association with risk for future development of sepsis was observed for the plurality-biomarker-score
25 termed "NGH infection biomarker score".

Figure 3a shows the relation of baseline biomarker levels for 3 plurality-biomarker-scores to future development of severe pneumonia. Specifically, the plurality biomarker the following: 1) a weighted combination of up to 29 biomarkers in absolute concentrations (denoted NGH infection biomarker score; no
30 biomarker scaling), 2) a weighted combination of up to 29 biomarkers in concentrations scaled relative to albumin (denoted NGH infection biomarker score; albumin scaled biomarkers), and 3) Glycoprotein acetyls scaled to albumin concentration (denoted
35 GlycA/Alb). The weighted combinations of biomarkers used in the plurality-biomarker-scores were derived using statistical regression analyses and machine learning, based on 50% of the study population (derivation set; approximately 53,000 individuals). The predictive performance of the 3 plurality-biomarker-

scores was then evaluated in the other 50% of the study population (validation set; 53477 individuals) to avoid overfitting of the results. The associations of the 3 plurality-biomarker-scores are shown for all individuals in the validation part of the study population, and in the subset of individuals without chronic respiratory or cardiometabolic disease at baseline. The results are also shown in different age groups (divided in age-tertiles of the study participants when they gave the blood samples), and separately for men and women. The relation of the baseline biomarker levels for 3 different plurality-biomarker-scores with severe pneumonia occurring within 2 years after the blood samples were taken is also shown.

These results show the following:

- that the 3 plurality-biomarker-scores are stronger associated with future development than any single biomarker (see Figure 1b for results on single biomarkers).
- that Glycoproteins relative to albumin, as well as a plurality-biomarker-score based on biomarkers scaled relative to albumin concentration, are strongly predictive of future development of severe pneumonia
- that the results are similar irrespective of age of study participants.
- that the results are similar for men and women.
- that the results are even stronger for prediction of short-term risk for severe pneumonia (disease occurring first 2 years after blood sampling), compared to long-term risk for severe pneumonia (disease occurring on average up to 8 years after the blood sampling)

Figure 3b shows the risk gradient for future severe pneumonia according to percentiles of the 3 plurality-biomarker-scores (upper half of the figure). Each dot corresponds to approximately 500 individuals. are also shown. We also calculated the cumulative risk for severe pneumonia after the blood samples were taken in various quantiles of the 3 plurality-biomarker-scores using Kaplan-Meier plots (lower half of the figure). The plots are shown for the validation set part of the study population,

i.e. 50% which was not included for derivation of the plurality-biomarker-scores (n=53477 individuals).

These results show the following:

- 5 - The relation of the 3 plurality-biomarker-scores with risk for future onset of severe pneumonia is particularly elevated for individuals with the very highest levels of the plurality-biomarker-scores, i.e. non-linear effects.
- 10 - The increase in risk for individuals with the very highest levels of the plurality-biomarker-scores is present within the first few years after the time of blood sampling.

It is obvious to a person skilled in the art that with the advancement of technology, the basic idea may be implemented
15 in various ways. The embodiments are thus not limited to the examples described above; instead they may vary within the scope of the claims.

The embodiments described hereinbefore may be used in any combination with each other. Several of the embodiments may
20 be combined together to form a further embodiment. A method, a product, or a use, disclosed herein, may comprise at least one of the embodiments described hereinbefore. It will be understood that the benefits and advantages described above may relate to one embodiment or may relate to several embodiments. The
25 embodiments are not limited to those that solve any or all of the stated problems or those that have any or all of the stated benefits and advantages. It will further be understood that reference to 'an' item refers to one or more of those items. The term "comprising" is used in this specification to mean
30 including the feature(s) or act(s) followed thereafter, without excluding the presence of one or more additional features or acts.