IFSO®

ORIGINAL CONTRIBUTIONS

Check for updates

Metabolic Health Index (MHI): Assessment of Comorbidity in Bariatric Patients Based on Biomarkers

Saskia L. M. van Loon [1,2,3] [iD] · Ruben Deneer [1,3,4] · Simon W. Nienhuijs [5] · Anna Wilbik [2] · Uzay Kaymak [2] · Natal van Riel [4] · Volkher Scharnhorst [1,3,4] · Arjen-Kars Boer [1,3,4]

Abstract

Purpose The focus of bariatric surgery is reduction of weight, reflected in body mass index (BMI). However, the resolution of comorbidity is a second important outcome indicator. The degree of comorbidity is hard to quantify objectively as comorbidities develop gradually and are interdependent. Multiple scoring systems quantifying comorbidity exist but they lack continuity and objectivity. In analogy with BMI as index for weight, the Metabolic Health Index (MHI) is developed as objective quantification of metabolic health status. Laboratory data were used as comorbidities affect biomarkers. Conversely, laboratory data can be used as objectively obtained variables to describe comorbidity.

Methods Laboratory data were collected and crosschecked by national quality registry entries. Machine learning was applied to develop an ordinal logistic regression model, using 4 clinical and 32 laboratory input variables. The output was mathematically transformed into a continuous score for intuitive interpretation, ranging from 1 to 6 (MHI).

Results In total, 4778 data records of 1595 patients were used. The degree of comorbidity is best described by age at phlebotomy, estimated Glomerular Filtration Rate (eGFR), and concentrations of glycated hemoglobin (HbA1c), triglycerides, and potassium. The model is independent of day of sampling and type of surgery. Mean MHI was significantly different between patient subgroups with increasing number of comorbidities.

Conclusion The MHI reflects severity of comorbidity, enabling objective assessment of a bariatric patient's metabolic health state, regardless day of sampling and surgery type. Next to weight-focused outcome measures like %TWL, the MHI can serve as outcome measure for metabolic health.

Keywords Bariatric surgery · METABOLIC SURGERY · Hypertension · Diabetes · Dyslipidemia · Metabolic syndrome · Machine learning · Biomarkers · Value based health care · Outcome measure

✉ Saskia L. M. van Loon
  saskia.s.v.loon@catharinaziekenhuis.nl

1 Department of Clinical Chemistry, Catharina Hospital,
  P.O. Box 1350, 5602, ZA Eindhoven, The Netherlands

2 Department of Industrial Engineering and Innovation Sciences,
  Eindhoven University of Technology, Eindhoven, The Netherlands

3 Expert Center Clinical Chemistry, Eindhoven, The Netherlands

4 Department of Biomedical Engineering, Eindhoven University of
  Technology, Eindhoven, The Netherlands

5 Department of Surgery, Catharina Hospital,
  Eindhoven, The Netherlands

Introduction

More than 10% of the worldwide population is considered obese, having a body mass index (BMI) ≥ 30 kg/m² [1]. Obesity is associated with comorbidities like type 2 diabetes mellitus (T2D), hypertension, and dyslipidemia, which are responsible for > 2.5 million deaths per year worldwide [2, 3]. The only treatment modality with proven long-term effect is bariatric surgery [2]. However, historically, only patients with morbid obesity (BMI > 40 kg/m²) are offered bariatric surgery. Lately, also patients with a BMI between 35 and 40 kg/m² with clinically relevant comorbidities like T2D, hypertension, or dyslipidemia have become eligible for bariatric surgery. This new indication for surgery is most often referred to as metabolic surgery, which focuses on improving obesity-related comorbidities next to weight loss [4–8]. European

guidelines already suggest metabolic surgery for patients with dysregulated T2D and BMI between 30 and 35 kg/m^2 to support T2D improvement or remission [8].

Nowadays, the concept of value-based health care is becoming increasingly important and the outcome of provided care should be weighed against the costs. Weight loss is a clear outcome measure for bariatric surgery. Weight loss can be objectively quantified by multiple measures like BMI, percentage of total weight loss (%TWL), or percentage of excess weight loss (%EWL) [4]. Conversely, improving obesity-related comorbidity is difficult to measure objectively. Comorbidities like T2D, hypertension, and dyslipidemia develop gradually and are interdependent. Together they reflect the continuum of the metabolic syndrome (MetS) or syndrome X. MetS is a widely accepted concept identifying the centrally obese patient with increased risk for cardiovascular disease and diabetes [6, 9–15]. Therefore, it is important to assess the comorbidities T2D, hypertension, and dyslipidemia. There are criteria to diagnose each comorbidity. However, objectively expressing improvement of comorbidity is more difficult [8]. In clinical practice, this is often left to the physician's assessment. Current criteria for improvement of comorbidity are dichotomous and do not cover minor steps towards remission. For example, a diabetic patient that went from insulin injections before surgery to oral medication only after surgery is not considered as "partial remission" according to the criteria in the European guidelines [8]. However, this change in pharmacotherapy can still be seen as an improvement. Ideally, an outcome measure should absorb these minor improvements of comorbidity. A measure on a *continuous* scale that is also derived from *objective* input would be preferred here.

For single comorbidities, quantitative measures are already available [11, 13–15]. However, half of the comorbid bariatric patients in our population had not one but multiple comorbidities (multimorbidity). In addition, comorbidities interact, and therefore, their cumulative impact needs to be quantified together to assess their overall impact. Already, multiple scoring systems for the assessment of MetS severity have been developed, like the Edmonton Obesity Staging System (EOSS) [16, 17]. The EOSS relies on definitions of risk or of comorbid conditions and on the interpretation of the terms like "mild," "moderate," or "severe." The nature of MetS as a continuous spectrum of disease is by the EOSS divided in five separate classes [18, 19]. Examples of continuous MetS scores are the continuous metabolic syndrome risk score (cMSy), the simple metabolic syndrome (siMS) score, and the metabolic syndrome severity score (MetSSS), all acknowledging the MetS continuum [20–22]. The MetS scores use input variables similar to the diagnosis criteria for MetS, i.e., triglyceride level, high-density lipoprotein (HDL) cholesterol level, glucose

level, waist circumference, and both systolic and diastolic blood pressure [10, 14, 15]. So, next to objective laboratory test results, these MetS severity scores also rely on input data that are susceptible to interobserver variation due to, for example, the manual way in which they are normally obtained.

In search for the ideal performance indicator, an outcome measure is needed that is both continuous and objective. These current MetS scores do not meet up to both criteria. A new measure must describe comorbidity-related metabolic health status by integrating the impact of multimorbidity and being observer-independent. Therefore, the aim of this study is to develop such an outcome measure, called the Metabolic Health Index (MHI). As comorbidities affect biomarkers, we hypothesized that biomarkers would be ideal to describe the severity of comorbidities like T2D, hypertension, and dyslipidemia. A data-driven approach was chosen to mine extensive laboratory data from our high-volume bariatric center. The laboratory data consist of biomarkers reflecting hematological status, kidney function, liver function, thyroid function, glucose metabolism, lipid profile, and inflammation parameters. All variables were measured both before and several times after surgery. Machine learning was applied, offering the opportunity to reveal otherwise unrecognized patterns in the data describing the overall impact of the multiple comorbidities. This approach seeks to develop an easily interpretable model based on objective input. The output, called MHI, should be numeric and continuous. Similar to the BMI describing a patient's weight status, the MHI could help in the objective assessment of comorbidity in bariatric patients summarizing the metabolic health burden due to comorbidities T2D, hypertension, and dyslipidemia.

Methods

All patients provided informed consent on the use of their medical data for scientific purposes. Statistic software R version 3.5.0 was used for data analysis and model building [23, 24].

Data

Comprehensive laboratory data were collected from patients who underwent primary laparoscopic sleeve gastrectomy (SG) or Roux-and-Y gastric bypass (LRYGB) surgery between January 2012 and January 2015 at the Catharina Hospital obesity center (Eindhoven, The Netherlands). Dutch guidelines were followed for mandatory screening and follow-up program. All patients periodically visited the outpatient clinic and had blood drawn for laboratory workup (at any time of the day). The variables included in the laboratory workup are hemoglobin, hematocrit, erythrocytes, mean

corpuscular hemoglobin, mean corpuscular volume, thrombocytes, leukocytes (all XE-5000, Sysmex); bilirubin, aspartate aminotransferase (AST), alanine aminotransferase (ALT), lactate dehydrogenase, alkaline phosphatase, gamma-glutamyl transferase, urea, creatinine, potassium, sodium, phosphate, albumin, C-reactive protein, cholesterol (total), high-density lipoprotein cholesterol, triglycerides, glucose, parathormone, ferritin, folic acid (all Cobas, Roche DX); prothrombin time (international normalized ratio (INR); Stago, STA-R MAX); and hemoglobin A1c (HbA1C; Tosoh G8, Sysmex). Additional calculated variables were AST/ALT ratio, cholesterol/HDL-ratio, estimated glomerular filtration rate (eGFR, using Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) formula, replacing creatinine as input variable), and albumin-corrected calcium (corrected calcium (mmol/L) = total calcium (mmol/L) + 0.02 × (40 − albumin (g/L)). This resulted in 32 biomarkers being included for modeling. In addition, data on patient's gender, age at phlebotomy, BMI, applied surgical technique, and date of surgery were collected from the hospital's Electronic Medical Records (EMR). The dates were categorized into periods with respect to surgery, i.e., 3 months before surgery or 6, 12, 24 months after surgery. Laboratory workups performed at 3 months before surgery and at 6, 12 and 24 months after surgery were included. This resulted in 34 variables for modeling, i.e., all 32 biomarkers, gender, and age. The remaining variables "BMI," "period," and "type of surgery" were only used for sub-analyses.

The laboratory data were merged with bariatric national quality registry entries extracted from the Dutch Audit for Treatment of Obesity (DATO) [25]. The DATO entries included data on the presence or absence of comorbidities T2D, hypertension, and dyslipidemia, both before and after surgery (see Supplemental Materials, describing used criteria). These data were collected during periodic clinical visits, where the physician or nurse practitioner assessed comorbidity and considered as current "gold standard" method. An ordinal output variable was defined by adding the three binary DATO labels for each comorbidity and stating comorbidity presence as "none," "one," or "multiple." It is assumed here that patients having multiple comorbidities are less healthy compared with those with fewer or no comorbidities.

Upon merging the data, the maximal time window allowed between a laboratory workup and a DATO entry was set at 90 days. For each patient and for each period, the pair of laboratory and DATO data (further referred to as "data record") having the smallest number of days between them was chosen as record for modeling. Laboratory workups and DATO entries that did not suffice the merging criterion of 90 days were excluded. Incomplete data records (e.g., missing biomarker data or missing comorbidity labels) were excluded. In addition, patients without pre-

surgical data were excluded, as effect of surgery could then not be assessed.

Model Building

The preferred modeling technique was ordinal logistic regression because it allows regression of an ordinal outcome variable on both continuous and categorical input variables. Of the available logistic regression models, the penalized extended continuation ratio model was used. This model is based on conditional probabilities and fits ordinal responses when subjects have to "pass through" one category to get to the next [23, 26, 27].

Internal model validity was evaluated by applying 10-fold cross-validation. Cross-validation is a common method to assess the generalization power of a model. The dataset was randomly split in 10 equally sized, representative disjoint subsets. In each subset, the distribution of comorbidities was similar to the total dataset. A model was built on a merged set of 9 of the 10 subsets. The remaining subset was used to evaluate the model's performance. By building 10 models, one gets a better estimation of the quality of the model, allowing for calculations of the error on the model's output.

In the cross-validation, each model was initially built using all 34 input variables ("full" model). A selection of variables was made to reduce the number of input variables in the final model and to eliminate variables that were of limited informative value to the outcome of the model. Here, Harrell's "model approximation" procedure was used [23]. Input variables were removed from the "full" equation in a repetitive, step-wise manner until the simplified linear predictor approximated the linear predictor of the "full" model to a certain goodness of fit (R^2). The R^2 was determined after the removal of each single input variable. The simplification was stopped when an R^2 of 0.95 was reached. This threshold for approximation was set at 0.95 to limit the number of variables being included in the final model while maintaining the model's discriminative power (feature selection). The variables were thus selected in a data-driven manner, not guided by medical domain knowledge.

The outcome of the final model was a predicted probability of having comorbidity. The assumption was that the predicted probability corresponds to the patient's metabolic health burden. Because the model's logistic function (S-curve) is not linear, it is less intuitive to interpret. Therefore, the predicted probability was mathematically transformed onto a linear scale and recalculated into an intuitive, continuous score of increasing metabolic health burden, the MHI. The final model was also visualized in a nomogram for interpretation of the effect size of the individual variables. In this nomogram, each effect in the model was converted to a 0 to 100 scale. Points

can be assigned to each variable depending on its obtained value.

Model Evaluation

Model performance was assessed for discriminative power using the area under the receiver operating characteristic (ROC) curve (AUC). The prediction of comorbidity by the model was compared with the categorical comorbidity labels "none," "one," or "multiple." Mean AUC was estimated by averaging the AUC of each model developed in the 10-fold cross-validation. Using the mean AUC and corresponding variability around the mean, an estimate was provided of the stability of the applied modeling technique. The distinctive character of the MHI was also assessed by comparing the MHI between different patient groups with respect to the number and type of comorbidity. After performing an ANOVA test to assess overall difference in mean MHI, Tukey's honest significant difference test was used to compare mean MHI between groups.

The robustness of the model was assessed by comparing the MHI as calculated by the final model with the MHI calculated by a model fitted on different subsets of the data, i.e., pre-surgery and post-surgery data (time dependency) or SG and LRYGB data (procedure dependency).

Clinical Potential

%TWL is a common "weight-centric" outcome measure in the assessment of the procedure's performance. The MHI can be used as "comorbidity-centric" outcome measure, allowing for a second dimension in this same assessment. Therefore, the study population was categorized into subgroups based on %TWL and MHI reduction at 12 months after surgery. A clear definition of "success" has not been uniformly defined [4]. Therefore, the 25th and 75th percentiles of both %TWL and MHI reduction 12 months post-surgery were derived from the study population and used to identify "poor" and "good" performers, rather than defining "success" or "failure" based on certain values for either %TWL or MHI. Thus, the study population was divided in subgroups "poor" performers (below 25th percentile), "average" performers (between 25th and 75th percentile), and "good" performers (above 75th percentile) for both %TWL and MHI reduction after 12 months.

The MHI can also serve as an institutional performance indicator to quantify the improvement in metabolic health status by the care provided. This possible application of the MHI on institutional level is shown by comparing mean MHI before and after surgery. In addition, two subgroups were identified having different pre-surgical conditions, determining their eligibility for surgery. The discrimination was made

based on the current two eligibility criteria, resulting in "bariatric" patients (BMI $\geq$ 40 kg/m^2) and "metabolic" patients (BMI between 35 and 40 kg/m^2 with at least one obesity-related comorbidity like T2D, hypertension, or dyslipidemia) [2, 4, 5].

Results

Exploratory Data Analysis

In the period between January 2012 and January 2015, 1739 patients underwent bariatric surgery at the Catharina hospital. A minority of 144 patients (8%) were excluded because of missing data, mostly due to missing data prior to surgery. This was attributed to including patients based on surgery date. In January 2012, a standard laboratory workup panel was introduced at our bariatric center. Of patients having surgery in the beginning of the inclusion period, no standard laboratory workup data were available prior to surgery. Of the remaining 1595 patients, 4778 complete sets of laboratory workup data and DATO entries were obtained (data records). Patient group sizes differed between the defined time points as data were retrospectively collected over a cross-section of the bariatric patient population (Table 1). Compared with the initial cohort, 694 patients (44%) were lost in follow-up after 24 months. These dropped-out patients did not differ significantly in BMI or MHI at baseline from patients still available in the 24-month follow-up, respectively, 44,1 kg/m^2 versus 43,8

Table 1 Patient characteristics at different time points

	Pre (n = 1595)	6M (n = 1263)	12M (n = 1019)	24M (n = 901)
Female	79%	80%	81%	81%
Age in years ± SD	41 ± 11.3	42 ± 11.2	43 ± 10.9	44 ± 10.8
Weight in kg ± SD	126 ± 20.8	92 ± 16.6	84 ± 15.8	86 ± 16.6
BMI in kg/m^2 ± SD	44 ± 5.7	32 ± 4.9	29 ± 4.7	30 ± 5.1
GBP	48%	46%	48%	46%
SG	52%	54%	52%	54%
T2D	18%	13%	9%	7%
Hypertension	34%	26%	21%	15%
Dyslipidemia	17%	13%	12%	9%
No comorbidities	57%	66%	72%	79%
One comorbidity	24%	19%	18%	13%
Multiple comorbidities	20%	14%	10%	8%

Pre, pre-surgery; *6M*, 6 months after surgery; *12M*, 12 months after surgery; *24M*, 24 months after surgery; *n*, number of patients; *BMI*, body mass index; *LRYGB*, laparoscopic Roux-en-Y gastric bypass; *SG*, sleeve gastrectomy; *T2D*, type 2 diabetes mellitus

kg/m² and MHI 3,1 versus 3,1. Also, the ratio SG/LRYGB did not differ significantly between both groups (51/49 vs 54/46). Therefore, the drop-out was considered random.

The majority of the patients (80%) was female (Table 1). At surgery, overall mean (± standard deviation (SD)) age was 41 (± 11) years and overall mean BMI was 44 (± 5.7) kg/m². Also, 43% (n = 692) of the patients had comorbidities, of which 55% (n = 379) had one comorbidity, 28% (n = 197) had two comorbidities, and 17% (n = 116) had all three comorbidities. With respect to the comorbidities of interest, 79% (n = 549) had hypertension, 42% (n = 294) had T2D, and 40% (n = 278) had dyslipidemia. Correlations between variables were analyzed. Expected high correlations were observed, e.g., between HbA1c and glucose (r = 0.78, Pearson), introducing multicollinearity in the logistic regression model. In general, multicollinearity does not reduce the reliability of the outcome of the model but affects the interpretation regarding individual coefficients.

Model

The "full" model using all 34 input variables was simplified by omitting variables one-by-one in a repetitive, step-wise manner. This step-wise omission of variables was stopped when the approximation of the "full" model by the simplified model dropped to an R^2 of 0.95. At this point in the data-driven simplification of the "full" model, five variables remained in the model: HbA1c, age, triglycerides, eGFR, and potassium. Excluding one of these variables from the model resulted in $R^2 < 0.95$. These variables were thus considered of most informative value to the outcome of the model, i.e., describing metabolic health status. The linear predictor

$(X\hat{\beta})$ of the final, simple logistic regression model was given by the following equation:

$$X\hat{\beta} = -3.64 + \left(1.00 \times 10^{-1} \times HbA1c\,[mmol/mol]\right)$$
$$+\left(5.98 \times 10^{-2} \times age[years]\right) + \left(4.49 \cdot 10^{-1} \cdot triglycerides[mmol/L]\right)$$
$$-\left(8.23 \cdot 10^{-1} \cdot potassium[mmol/L]\right) - \left(1.06 \cdot 10^{-2} \cdot eGFR\left[ml/\min/1.73m^2\right]\right)$$

$$(1)$$

Equation (1) was visualized in a nomogram for ease of interpretation (Supplemental Figure 1).

In order to compare different patient groups of increasing metabolic health burden, the MHI was categorized in classes I to VI (Table 2). Classes were defined using quintiles of predicted probability, with an additional category for the 95th percentile. The mean level (± standard deviation, SD) of HbA1c increased from 34 ± 3 mmol/mol in MHI class I to 74 ± 14 mmol/mol in MHI class VI (reference interval: 20–42 mmol/mol). Mean age went from 33.5 ± 8.1 years in MHI class I to 54.1 ± 8.2 years in MHI class VI. The mean level of triglycerides increased from 1.1 ± 0.5 mmol/L in MHI class I to 3.4 ± 1.9 mmol/L in MHI class VI (target value: < 2.0 mmol/L). Mean potassium levels were similar in all MHI classes, i.e., 4.0 ± 0.3 mmol/L (reference interval: 3.5–5.0 mmol/L). The mean eGFR decreased from 110 ± 13 mL/min/1.73m² to 88 ± 25 mL/min/1.73m² (reference value: > 90 mL/min/1.73m²).

The distribution of patients with either none, one, two, or three comorbidities was evaluated over the continuous range of the MHI (Fig. 1). Among the patients with MHI of 1, 98.5% had no comorbidity. This proportion decreases to 6.8% for patients with MHI of 6. Vice versa, no patients with MHI of 1 had three comorbidities, while a majority of the patients with

Table 2 MHI classes and model variables

Probability	< 0.20	0.20–0.40	0.40–0.60	0.60–0.80	0.80–0.95	> 0.95
Points	< 100	100–118	119–134	135–152	153–182	> 182
MHI	1–2	2–3	3–4	4–5	5–6	> 6
MHI class	I	II	III	IV	V	VI
n	1946	1427	681	351	226	147
HbA1c (mmol/mol)	34 ± 3	37 ± 3	40 ± 4	44 ± 6	55 ± 8	74 ± 14
Age at phlebotomy (years)	33.5 ± 8.1	45.9 ± 6.9	50.8 ± 7.8	53.8 ± 8.1	53.5 ± 7.9	54.1 ± 8.2
Triglycerides (mmol/L)	1.1 ± 0.5	1.3 ± 0.6	1.7 ± 0.8	2.1 ± 1.0	2.5 ± 1.3	3.4 ± 1.9
Potassium (mmol/L)	4.0 ± 0.3	4.0 ± 0.3	4.0 ± 0.3	4.0 ± 0.3	4.0 ± 0.4	4.0 ± 0.4
eGFR (CKD-EPI) (mL/min/1.73m²)	110 ± 13	98 ± 13	92 ± 15	89 ± 18	89 ± 21	88 ± 25

Classification of MHI according to percentile ranges, reflecting severity of metabolic health burden due to comorbidity (mean ± SD). Shown are predicted probability and corresponding point range according to the nomogram (see Supplemental Figure 1). Number of measurements (n), hemoglobin A1c (HbA1c), estimated glomerular filtration rate (eGFR, using Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) formula)

Fig. 1 Frequency distribution of number of comorbidities per MHI point. An introduction to statistical learning - with applications in R. The distribution of patients with either none, one, two, or three comorbidities is plotted for each MHI point in the range between 1 and 6. The number of comorbidities per patient was based on registry data from the Dutch Audit of Treatment of Obesity (DATO)
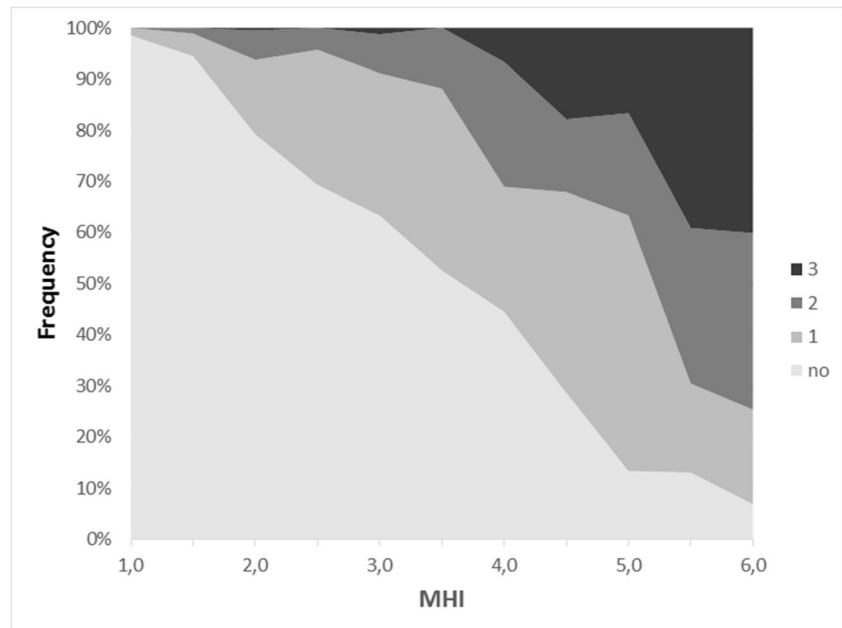
MHI of 6 had three comorbidities (40%). These observations support our assumption that the model's predicted probability, which transformed into the MHI, mimics the metabolic health burden.

Model Performance

The predictions of the 10 models in the cross-validation were used to construct a ROC curve (Fig. 2). The mean (± 2SE) AUC of the 10 simplified models of the cross-validation was 0.82 (± 0.02). The added value of combining multiple variables to quantify metabolic health burden was shown by constructing ROC curves of the model's individual input variables (Fig. 2). The AUC of the final model was significantly larger than the AUCs of the individual variable models.
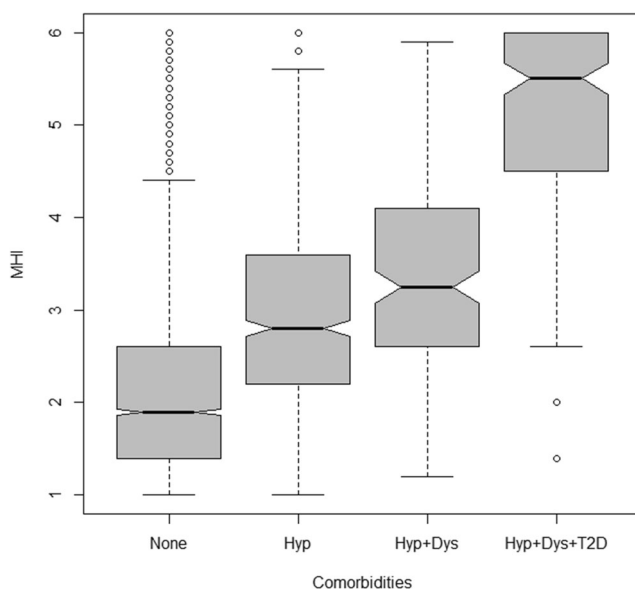
The distinctive character of the MHI was further endorsed by comparing the MHI between patient groups with differences in the number and type of comorbidity, as defined according the gold standard (i.e., DATO). The mean increase in MHI, including 95% confidence interval (CI), between patients without comorbidity and patients with only dyslipidemia or only T2D was respectively 0.7 (0.4–0.9) and 2.0

Fig. 2 ROC curve of MHI. Receiver operating characteristic (ROC) curve of the MHI model is plotted together with the ROC curves of each single variable in the MHI model. The mean ROC curve out of 10-fold cross-validation is shown with error bars reflecting standard error (SE). Mean (± 2SE) area under the ROC curve (AUC) for the MHI model is 0.82 (± 0.02) compared with mean (± SE) AUCs of 0.74 (± 0.02), 0.73 (± 0.02), 0.69 (± 0.02), 0.64 (± 0.02), and 0.53 (± 0.02) of respectively HbA1c, age, triglycerides, eGFR (CKD-EPI), and potassium. The added value of combining variables into the MHI is highlighted
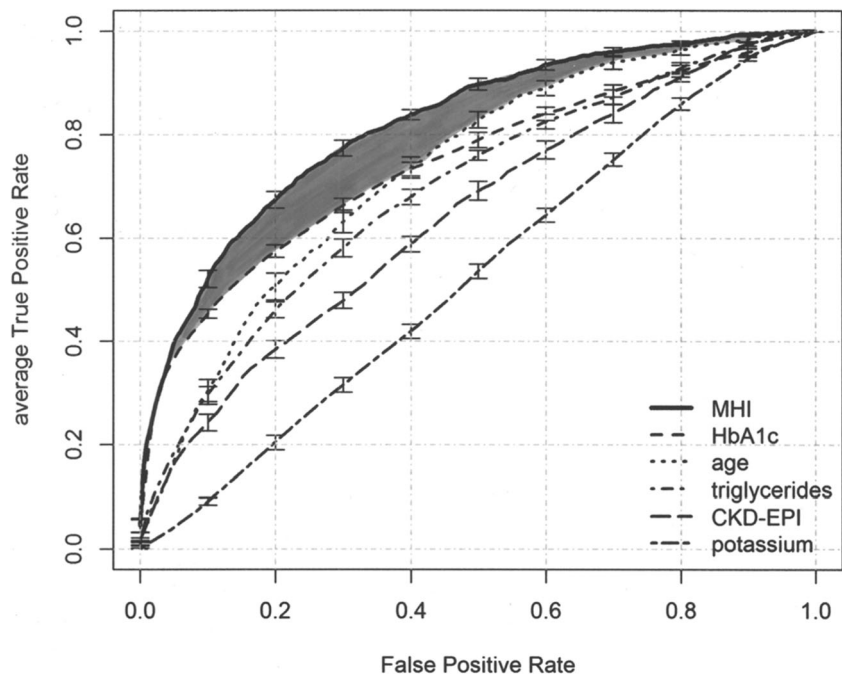
(1.7–2.3). As triglyceride and HbA1c levels are used to diagnose respectively T2D and dyslipidemia, this discrimination is obvious. However, for patients with only hypertension, the mean increase in MHI compared with non-comorbid patients was 0.9 (0.7–1.0) (Fig. 3). Here, the variables included in the MHI model are not used to diagnose hypertension. The observed differences in mean MHI were all significant ($p <$ 0.001). The ability to discriminate between patient groups with and without hypertension was also observed in cases of multimorbidity. In patients with either dyslipidemia or T2D, the MHI increased on average 0.6 (0.2–1.0; $p < 0.001$) when also hypertension was diagnosed (Fig. 3). On top of that, the mean increase in MHI in the case of T2D and dyslipidemia without and with hypertension was 0.7 (0.3–1.1; $p < 0.001$). The model can thus predict hypertension, even without including blood pressure measurements or specifying any additional treatment.

The robustness of the model was assessed with respect to dependency on time (pre- and post-surgery) and surgical procedure (SG and LRYGB). The median difference in MHI, including interquartile range (IQR), between the final model and the model fitted only on pre-surgical data ($n = 1595$) was − 0.1 (− 0.2–0.0). For post-surgical data ($n = 3183$), the median difference in MHI with respect to the final model was 0.0 (0.0–

0.1). This procedure was repeated with respect to surgical procedure. The median MHI difference for SG ($n = 2534$) was − 0.1 (− 0.1–0.0) and for LRYGB ($n = 2244$) 0.1 (0.0–0.2). In all four cases, the median difference in MHI was relatively small. In addition, all IQR included zero and, thus, the observed differences were considered not significant. Therefore, the final MHI model was considered robust and independent with respect to time and type of surgery.

Applications of the MHI

The 25th and 75th percentile-based patient categorization with respect to %TWL and MHI reduction 12 months post-surgery resulted in the identification of "poor" performers (below 25th percentile), "average" performers (between 25th and 75th percentile), and "good" performers (above 75th percentile) in both dimensions (Table 3). According to %TWL alone, 271 patients out of 1019 would be considered as "poor" performers. However, using the MHI as the second outcome measure, a subgroup of 57 patients could be identified underperformed with respect to weight loss but overperformed with respect to improving their metabolic health status. When considering only %TWL, this subgroup would be considered as "poor" performing. However, by adding the second dimension in the assessment using the MHI, one might change this statement. Especially in the patient group with BMI 35–40 kg/m^2, where the surgery, next to weight loss, is mainly intended to improve metabolic health status, this second dimension in performance assessment is of added value.

On institutional level, the improvement in metabolic health of treated bariatric patients was quantified with the MHI and used to compare outcome in different subgroups (Fig. 4). Within our dataset, the mean (± SD) baseline MHI was 3.1 (± 1.5) for the overall population. Six months after surgery, the mean MHI decreased to 2.3 (± 1.2) and remained constant afterwards, i.e., 2.3 (± 1.1) and 2.3 (± 1.1) at respectively 12 and 24 months postoperative. There is clearly an improvement in metabolic health status over time, which is mainly achieved within the first 6 months after surgery. Closer evaluation of the total population reveals two subgroups, i.e., "bariatric" patients and "metabolic" patients. Evaluation of the bariatric patients ($n = 1227$) revealed a mean baseline MHI of 3.0 (± 1.5) and an overall decrease of 0.7 MHI points within 6 months post-surgery. The metabolic patients ($n = 195$) had a mean baseline MHI of 4.3 (± 1.5) with an overall decrease of 1.1 MHI points after 6 months (Fig. 4). The metabolic patients indeed have a different (higher) starting point with respect to MHI compared with the bariatric patients. Their improvement in MHI also is larger over the same time period, emphasizing the impact of bariatric surgery on improving comorbidity in this patient group.

Fig. 3 MHI and hypertension. Boxplot showing the increase in MHI for categories of comorbidity including hypertension. Boxes represent the area between the first (Q1) and third quartile (Q3). Medians are plotted within the boxes. Notches represent confidence interval around the median. The whiskers display 1.5 times interquartile range (IQR, Q3–Q1) or minimum/maximum when inside IQR. The open dots represent outliers (> 1.5 IQR). None = no comorbidity ($n = 3280$), Hyp = hypertension ($n = 630$), Hyp+Dys = hypertension and dyslipidemia ($n = 180$), T2D+Hyp+Dys = T2D and hypertension and dyslipidemia ($n = 199$)

Table 3 MHI as outcome measure next to %TWL

		% total weight loss (12M)			
		Q1 (≤ 28%TWL)	Q2–Q3 (28–38%TWL)	Q4 (> 38%TWL)	Total
MHI reduction (12 M)	Q1 (ΔMHI ≤ 0.2)	67	99	56	222
	Q2–Q3 (ΔMHI 0.2–1.3)	148	273	121	542
	Q4 (ΔMHI > 1.3)	57	131	67	255
	Total	272	502	244	1019

Improvement in MHI as outcome measure in addition to percentage of total weight loss (%TWL) at 12 months after surgery (12 M). The population is categorized into three groups based on percentile ranges, where Q1 is the 25th percentile ("poor" performers) and Q4 is the 75th percentile ("good" performers). Quartiles 2 and 3 are combined and reflect the median 50% of the population ("average" performers)
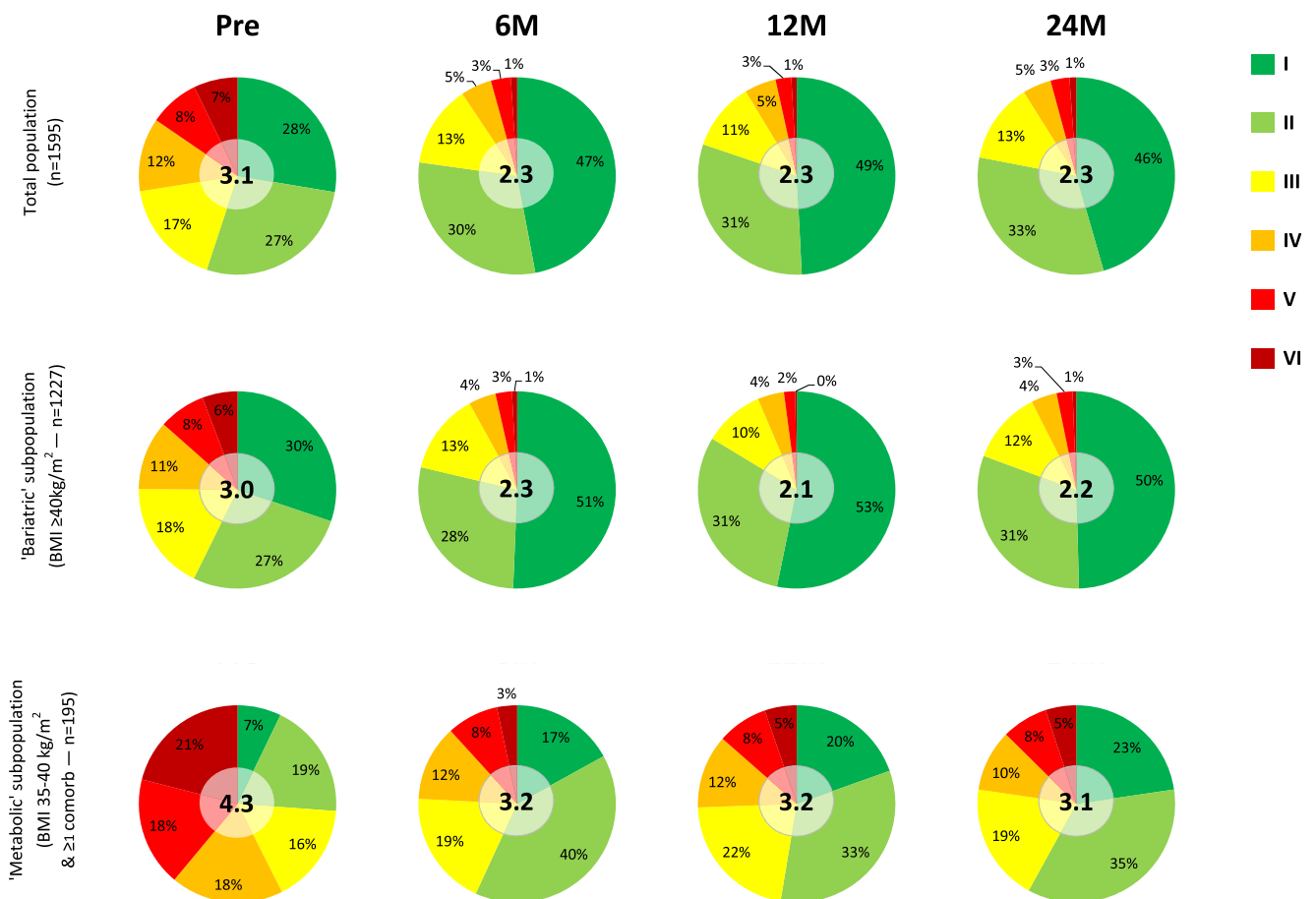
Fig. 4 MHI diagrams. Diagrams (pie charts) of the distribution of MHI classes per period in time for the total population (top row), the "bariatric" subpopulation with BMI ≥ 40 kg/m² (middle row), and the "metabolic" subpopulation with BMI between 35 and 40 kg/m² with at least one obesity-related comorbidity like T2D, hypertension, or dyslipidemia (bottom row). Mean overall MHI is given in the center of each chart. Number of patients (n)

Discussion

Quantification of patient status through data-driven indices is a recent trend in medicine [28]. The development of the MHI is in line with this trend. The MHI enables the transition from a binary label, indicating the presence or absence of comorbidity, into a continuous index quantifying severity of metabolic health burden due to comorbidity. Instead of three separate outcome labels for the comorbidities T2D, hypertension, and dyslipidemia, one single objective measure is presented. The MHI includes the crosstalk between the three comorbidities, reflecting the continuum of the metabolic syndrome, and is independent with respect to day of sampling and surgery type. The MHI is capable of discriminating patients with increasing number of comorbidities, including hypertension which is not diagnosed based on biomarkers. Having one single, easy to interpret measure simplifies comparisons with other outcome measures like BMI or %TWL, compared with having to analyze outcome per comorbidity separately.

The MHI can be used to refine the readily available information that a patient has comorbidity, in the sense that the severity of comorbidity can now be quantified using a continuous score. Also improvement in comorbidity and resolution can be quantified using the MHI. The MHI correlates with the number of drugs a hypertensive patient uses, i.e., mean MHI increased with increasing number of antihypertensive drugs (data not shown). The same phenomenon was observed in diabetic patients using oral medication only compared with insulin. These findings support the statement that the MHI allows for objective quantification of the severity of comorbidity and the identification of minor improvements in comorbidity, not captured by the current criteria.

Compared with other MetS scores like EOSS, cMSy, siMS, and MetSSS, the MHI has the advantage that it can be used as outcome measure because it is both continuous and based on variables that can be objectively obtained. The MHI is, besides age at phlebotomy, solely based on objectively obtained laboratory measurements. This objective character is a main asset of the MHI compared with all other MetS scores. In addition, the MHI is relatively easy to implement as additional laboratory variable as all five variables are already available in the laboratory information system. The MHI could be automatically reported as a calculated variable next to routine analytical variables. This makes the MHI beneficial over other MetS scoring algorithms that use separate platforms and, therefore, are not easily available to automatically calculate the outcome.

The MHI is largely intuitive, despite being built through a data-driven approach. The combination of HbA1c, age, triglycerides, eGFR, and potassium results in a higher discriminative power with respect to identifying comorbidity compared with each variable separately. Where HbA1c or triglyceride levels alone each resulted in an AUC of 0.74 and 0.69

respectively, combining them with age, eGFR, and potassium into the MHI increased the AUC to 0.82. The data-driven approach selected these five well-recognized variables as optimal input. The selection of biomarkers HbA1c and triglycerides is intuitive as they both are used in the diagnosis of T2D and dyslipidemia respectively. It is also well known that age is related to prevalence of comorbidity in general. The involvement of renal function, expressed by eGFR according to the CKD-EPI formula, can be explained by the end-organ damage induced by hypertension and/or T2D. The inclusion of potassium as a variable in the MHI model is less obvious. In the approximation of the full MHI model, the exclusion of potassium as the fifth variable decreased the R^2 only from 0.95 to 0.94, so one may argue its role in the final MHI model. Nonetheless, sub-analysis showed that the subgroup of patients with hypertension had significantly lower potassium levels compared with subgroups without or with other (combinations of) comorbidity. Potassium levels can be decreased upon use of diuretics in the treatment of hypertension. It can even result in hypopotassemia depending on the type and number of diuretics used. The potassium level is more likely to decrease when multiple antihypertensive drugs are used. We therefore believe that in the MHI model, potassium is reflecting the use of (multiple) antihypertensive drugs and is related with the severity (or persistency) of hypertension.

The approach of including only objective laboratory data has several limitations. Only comorbidities T2D, hypertension, and dyslipidemia were used in the development of the MHI as these represent the MetS. Additional data on these comorbidities such as physical fitness or perceived quality of life could improve the quantification of disease burden [29–31]. However, these types of data contain a degree of subjectiveness which makes them harder to compare. Also, no information on medical treatment was taken into account. One could argue that higher HbA1c levels without insulin are not even worse than a lower value while still using antidiabetics. On the other hand, the glycemic status is the goal irrespective to its treatment (e.g., lifestyle advice an/or bypass and/or metformin) and surpasses the difficulties encountered in scoring and rating the medications.

The MHI can serve as a clinical performance indicator next to weight-focused markers like %TWL, as shown in Table 3. Although the definition of "poor," "average," and "good" performer was data-driven, a group of patients was identified by using the MHI, which was considered "poor" with respect to %TWL but "good" when considering reduction in MHI. The MHI can also be used to compare subgroups, e.g., based on performed type of surgery like SG or LRYGB, within a single institution as well as between different institutions. However, more research is needed with respect to the generalizability of the MHI and its application in different populations. Laboratory tests are not standardized and clinical decision-making regarding diagnosis and treatment may depend on

local and not universal practice. Prospective studies for examining the additional value for the individual patient have not been performed, but one might argue that there is a potential role for the MHI as a tool to personalize individual counseling in postoperative monitoring. Currently, there has not been a trial where treatment regimens with or without MHI were compared. As age is one of the variables to calculate the MHI, one might consider an age-corrected MHI to rule-out the effect of aging during monitoring. Furthermore, the MHI might be used to manage patient expectations and may improve compliance as complying to one measure summarizing your metabolic health status is intuitively beneficial over three separate outcomes for each comorbidity.

There are currently no clear criteria with respect to the assessment of improvement of metabolic health status. In daily practice, this assessment now relies on the physician's observation, which may be subject to observer bias. The influence of changing medication regimes also troubles the assessment of improvement in metabolic health. Despite its limitations, the physician's observation is currently the gold standard in the follow-up of metabolic health status after bariatric surgery. As the physician's assessment was used to develop our model, the MHI never will replace the physician's judgment, but definitely will support the physician in a more objective assessment of comorbidity.

Conclusions

In conclusion, the MHI enables objective assessment of metabolic health status in bariatric patients concerning comorbidities T2D, hypertension, and dyslipidemia. Hereby, additional value has been created through merging and mining available clinical and laboratory data. Although built by a data-driven approach, our model is largely intuitive. The MHI does not only have the potential to support clinical decision-making, but could also serve as a useful tool for research and policy makers to better define the impact and improvement of obesity-related comorbidities on metabolic health status.

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.

Informed Consent Informed consent was obtained from all individual participants included in the study.

Ethical Approval For this type of study formal consent is not required.

References

1. Ng M, Fleming T, Robinson M, et al. Global, regional, and national prevalence of overweight and obesity in children and adults during 1980–2013: a systematic analysis for the Global Burden of Disease Study 2013. The Lancet. 2014;384(9945):766–81.
2. Buchwald H, Avidor Y, Braunwald E, et al. Bariatric surgery: a systematic review and meta-analysis. JAMA. 2004;292(14):1724–37.
3. The GBD 2015 Obesity Collaborators, Health effects of overweight and obesity in 195 countries over 25 years. N Engl J Med. 2017;377(1):13–27.
4. Frühbeck G. Bariatric and metabolic surgery: a shift in eligibility and success criteria. Nat Rev Endocrinol. 2015;11(8):465–77.
5. Sjöholm K, Anveden Å, Peltonen M, et al. Evaluation of current eligibility criteria for bariatric surgery: diabetes prevention and risk factor changes in the Swedish Obese Subjects (SOS) study. Diabetes Care. 2013;36(5):1335–40.
6. Kaur J. A comprehensive review on metabolic syndrome. Cardiol Res Pract. 2014;2014:1–21.
7. Adams TD, Davidson LE, Litwin SE, et al. Weight and metabolic outcomes 12 years after gastric bypass. N Engl J Med. 2017;377(12):1143–55.
8. Fried M, Yumuk V, Oppert JM, et al. Interdisciplinary European guidelines on metabolic and bariatric surgery. Obes Surg. 2014;24(1):42–55.
9. Monteiro R, and Azevedo I. Chronic inflammation in obesity and the metabolic syndrome. Mediators Inflamm. vol. 2010, Article ID 289645, 10 pages, 2010. https://doi.org/10.1155/2010/289645
10. Han TS, and Lean ME. A clinical perspective of obesity, metabolic syndrome and cardiovascular disease. JRSM Cardiovasc Dis. 2016;5 https://doi.org/10.1177/2048004016633371
11. O'Neill S, Bohl M, Gregersen S, et al. Blood-based biomarkers for metabolic syndrome. Trends Endocrinol Metab. 2016;27(6):363–74.
12. Eckel RH, Grundy SM, Zimmet PZ. The metabolic syndrome. The Lancet. 2005;365(9468):1415–28.
13. Sikaris KA. The clinical biochemistry of obesity. Clin Biochem Rev. 2004;25(3):165–81.
14. Alberti KGM, Zimmet P, Shaw J. The metabolic syndrome—a new worldwide definition. The Lancet. 2005;366(9491):1059–62.
15. Eckel RH, Alberti K, Grundy SM, et al. The metabolic syndrome. The Lancet. 2010;375(9710):181–3.
16. Sharma AM, Kushner RF. A proposed clinical staging system for obesity. Int J Obes. 2009;33(3):289–95.
17. Gill RS, Karmali S, Sharma AM. The potential role of the Edmonton obesity staging system in determining indications for bariatric surgery. Obes Surg. 2011;21(12):1947–9.
18. DeBoer MD, Gurka MJ. Clinical utility of metabolic syndrome severity scores: considerations for practitioners. Diabetes Metab Syndr Obes Targets Ther. 2017;10:65–72.
19. Cubeddu LX, Hoffmann IS. Metabolic syndrome: an all or none or a continuum load of risk? Metab Syndr Relat Disord. 2011;10(1):14–9.
20. Wijndaele K, Beunen G, Duvigneaud N, et al. A continuous metabolic syndrome risk score. Diabetes Care. 2006;29(10):2329.
21. Soldatovic I, Vukovic R, Culafic D, et al. siMS Score: simple method for quantifying metabolic syndrome. PLOS ONE. 2016;11(1): e0146143.
22. Wiley JF, Carrington MJ. A metabolic syndrome severity score: a tool to quantify cardio-metabolic risk factors. Prev Med. 2016;88: 189–95.
23. Harrell FE. Regression modeling strategies. 2nd ed. Switzerland: Springer International Publishing; 2015.

24. R Core. Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical. Computing. 2019

25. Poelemeijer YQ, Liem RS, Nienhuijs SW. A Dutch nationwide bariatric quality registry: DATO. Obes Surg. 2018;28(6):1602-10.

26. James G, Witten D, Hastie T, Tibshirani R. An introduction to statistical learning - with applications in R. 1st ed. New York; Springer-Verlag; 2013

27. Steyerberg E. Clinical prediction models - a practical approach to development. 1st ed. New York: Springer-Verlag; 2009.

28. Rothman MJ, Rothman SI, Beals J. Development and validation of a continuous measure of patient condition using the electronic medical record. J Biomed Inform. 2013;46(5):837–48.

29. Cohen DD, Gómez-Arbeláez D, Camacho PA, et al. Low muscle strength is associated with metabolic risk factors in Colombian children: the ACFIES study. PLoS ONE. 2014;9(4):e93150

30. Kim J, Tanabe K, Yokoyama N, et al. Association between physical activity and metabolic syndrome in middle-aged Japanese: a cross-sectional study. BMC Public Health. 2011;11:624.

31. Ekblom Ö, Ekblom-Bak E, Rosengren A, et al. Cardiorespiratory fitness, sedentary behaviour and physical activity are independently associated with the metabolic syndrome, results from the SCAPIS pilot study. PLOS ONE. 2015;10(6):e0131586.