Clinical Therapeutics xxx (xxxx) xxx

Contents lists available at ScienceDirect

Clinical Therapeutics

journal homepage: www.elsevier.com/locate/clinthera

Original Research

Unsupervised cluster analysis reveals distinct subtypes of ME/CFS patients based on peak oxygen consumption and SF-36 scores

Marcos Lacasa[1,*], Patricia Launois[5], Ferran Prados[1,2,3,4], José Alegre[5], Jordi Casas-Roma[1]

[1] *e-Health Center, Universitat Oberta de Catalunya, Barcelona, Spain*
[2] *Center for Medical Image Computing, University College London, London, United Kingdom*
[3] *National Institute for Health Research Biomedical Research Centre at UCL and UCLH, London, United Kingdom*
[4] *Queen Square MS Center, Department of Neuroinflammation, UCL Institute of Neurology, Faculty of Brain Sciences, University College London, London, United Kingdom*
[5] *Myalgic Encephalomyelitis / Chronic Fatigue Syndrome Unit, Division of Rheumatology, Vall d'Hebron Hospital Research Institute Universitat Autònoma de Barcelona, Barcelona, Spain*

ARTICLE INFO

Keywords:
Biomarker
Chronic fatigue syndrome
Cardiopulmonary exercise test
Clustering
K-means

ABSTRACT

Purpose: Myalgic encephalomyelitis, commonly referred to as chronic fatigue syndrome (ME/CFS), is a severe, disabling chronic disease and an objective assessment of prognosis is crucial to evaluate the efficacy of future drugs. Attempts are ongoing to find a biomarker to objectively assess the health status of (ME/CFS), patients. This study therefore aims to demonstrate that oxygen consumption is a biomarker of ME/CFS provides a method to classify patients diagnosed with ME/CFS based on their responses to the Short Form-36 (SF-36) questionnaire, which can predict oxygen consumption using cardiopulmonary exercise testing (CPET).

Methods: Two datasets were used in the study. The first contained SF-36 responses from 2,347 validated records of ME/CFS diagnosed participants, and an unsupervised machine learning model was developed to cluster the data. The second dataset was used as a validation set and included the cardiopulmonary exercise test (CPET) results of 239 participants diagnosed with ME/CFS. Participants from this dataset were grouped by peak oxygen consumption according to Weber's classification. The SF-36 questionnaire was correctly completed by only 92 patients, who were clustered using the machine learning model. Two categorical variables were then entered into a contingency table: the cluster with values {0,1} and Weber classification {A, B, C, D} were assigned. Finally, the Chi-square test of independence was used to assess the statistical significance of the relationship between the two parameters.

Findings: The results indicate that the Weber classification is directly linked to the score on the SF-36 questionnaire. Furthermore, the 36-response matrix in the machine learning model was shown to give more reliable results than the subscale matrix ($p - value < 0.05$) for classifying patients with ME/CFS.

Implications: Low oxygen consumption on CPET can be considered a biomarker in patients with ME/CFS. Our analysis showed a close relationship between the cluster based on their SF-36 questionnaire score and the Weber classification, which was based on peak oxygen consumption during CPET. The dataset for the training model comprised raw responses from the SF-36 questionnaire, which is proven to better preserve the original information, thus improving the quality of the model.

Introduction

Myalgic encephalomyelitis (ME), commonly referred to as chronic fatigue syndrome (CFS), is a serious, complex, chronic, multisystem illness of unknown etiology. It is also known as post-viral fatigue syndrome as it is often triggered by a persistent viral infection. Research on the prevalence of the disease has focused predominately on meta-analysis due to the variety of this type of data and the complexity of calculations. ME/CFS is characterized by unexplained and persistent post-exertional fatigue that is not relieved by rest, is exacerbated by physical and mental exertion, and shows other core symptoms such as cognitive, immunometabolic, autonomic, and neuroendocrine dysfunc-

Abbreviations: ME/CFS, Myalgic encephalomyelitis chronic fatigue syndrome; CPET, cardiopulmonary exercise test; VO2 peak, peak oxygen consumption; VO2 VT1, oxygen consumption at the anaerobic threshold; RPE, rate of perceived exertion; RER, respiratory exchange ratio.

* Corresponding author: ADaS Lab - e-Health 5Center, Universitat Oberta de Catalunya, Rambla del Poblenou, 156, 08018 Barcelona. (SPAIN)
E-mail address: mlacasaca@uoc.edu (M. Lacasa).

https://doi.org/10.1016/j.clinthera.2023.09.007
Accepted 9 September 2023
Available online xxx

Please cite this article as: M. Lacasa, P. Launois, F. Prados et al., Unsupervised cluster analysis reveals distinct subtypes of ME/CFS patients based on peak oxygen consumption and SF-36 scores, Clinical Therapeutics, https://doi.org/10.1016/j.clinthera.2023.09.007

tion.[1] Published data indicates a prevalence of between 0.89 and 1.14% among the population of the United States.[2] Additional data includes an analysis of adolescents in the United Kingdom with an estimated prevalence range of between 1.47% and 2.99%.[3] In contrast to these studies, recent research carried out in China estimates the prevalence among the Chinese population to be 12.54%.[4] The prevalence in Australia was estimated at 0.79%.[5] ME/CFS is a major cause of disability, with many patients feeling unable to fulfil their family responsibilities and forced to limit their social activities. The condition significantly impacts their work (difficulty performing their job) and daily life (for example, climbing stairs, lifting or carrying groceries, moving a table, or pushing a vacuum cleaner).[6] Patients also have characteristic inflammatory and muscular symptoms, sleep disturbances, and altered cognitive function.[7] Muscular symptoms include pain, generalized muscle weakness, post-exertional fatigue, and neurological (mental fatigue, impaired cognition, psychomotor slowing, disturbed sleep, hypersensitivity to noise, light, and odours, headache, pain, paresthesia and severe dysautonomia).[8] Neurocognitive symptoms have also been reported (slow reaction time, indicating the likelihood of an ADHD-like pattern of functioning).[9]

The most widely used international definition for ME/CFS is the Fukuda criteria, developed in 1994 by the Center for Disease Control and Prevention (CDC) in Atlanta, Georgia, USA[10]. As some symptoms in the Fukuda criteria overlap with depression, it has been criticized for its lack of specificity.[11] In contrast, the Canadian Consensus Criteria distinguish patients with ME from those who are depressed, identifying patients who are more physically debilitated and have greater physical and cognitive dysfunction.[12] These criteria were updated in 2011 to include post-exertional fatigue.[13]

Cardiopulmonary exercise testing (CPET) provides information that helps identify abnormal values in cardiac and/or ventilatory responses that rule out ME/CFS, thus differentiating these patients from those likely to be affected by other etiologies causing exercise intolerance. We used a gas analyser to measure peak oxygen consumption (VO2 peak) and anaerobic threshold oxygen consumption (VO2 VT1) during exercise. These measurements determine the degree of deterioration of the patient's functional capacity (simultaneously the cardiovascular and ventilatory response to a known metabolic stress are examined) and early transition to anaerobic metabolism during exercise, respectively. Both can be measured during a single exercise test which provides information on the efficiency of the metabolism in response to physiological exercise,[14] with VO2 peak showing a strong predictive and prognostic value. Patients with chronic heart failure were evaluated with a single CPET using Weber and Janicki's[15] established classification system.

Attempts are still ongoing to identify a reliable, objective biomarker for the health status of ME/CFS patients. Currently, the most widely used technique to measure ME/CFS patient status is internationally validated self-administered questionnaires such as the SF-36 questionnaire. Several studies have attempted to identify patterns of association between these questionnaires and the pathophysiology of specific systems and links to contemporary multidisciplinary molecular pathology, including comparative MRI,[16] exploring symptom co-occurrence using network analysis,[17] and investigating the relationship between oxygen consumption and the physical subscale of the SF-36 questionnaire in retrospective studies.[18] Physicians currently base ME/CFS diagnosis and prognosis on exclusion and subjective clinical interpretation, and to date, a reliable, objective method using accessible inexpensive tests that can identify ME/CFS patients with worse prognosis due to the wide-ranging symptoms accompanying the disease and subjective interpretation of results has not yet been found. Therefore, validating an objective biomarker such as oxygen consumption in CPET would be helpful as a primary outcome of trials developing new treatments and for a deeper understanding of the evolution of the disease.

This aim of this study was twofold: first, to prove that oxygen consumption is a potential biomarker of ME/CFS; and second, to develop a new machine learning-based method to identify patients diagnosed with ME/CFS and at high risk of physical impairment. Our proposed model predicts CPET scores from self-administered SF-36 questionnaires in the primary healthcare setting and assists in the early identification of patients in need of referral to a dedicated hospital unit.

Patients and Methods

Study Population

This study involved two clinical trials, both approved by the ethics committee of the Vall d'Hebron University Hospital, Barcelona, Spain: 1) "Population-based Registry of Chronic Fatigue Syndrome Patients", approved 10/18/2006; and 2) "Study of exercise intolerance in adult patients with chronic fatigue syndrome using a cardiopulmonary exercise stress test (CPET)", approved 09/22/2020.

The 36-Item Short Form Health Survey

This prospective cross-sectional study included 2,522 patients diagnosed with ME/CFS from the Vall d'Hebron University Hospital (Barcelona, Spain), made up of 90.5% women (mean age 48.11±10.31 years) and 9.5% men (mean age 44.41±11.35 years). Data from the Spanish version of the SF-36 questionnaire[19] between 2008 and 2022 was obtained and recorded. Both the Spanish and English versions of the SF-36 can be found in Supplementary Material.

The SF-36 questionnaire consisted of 36 questions requiring the patient to choose one option. The options ranged from 1 to 6, as illustrated in Table 1. Eight subscales were defined and calculated based on the weighted sum of a small number of responses. The score for each question was coded so that higher scores reflected better health.

Net rank *equals the difference between the maximum and minimum theoretical values. The minimum is equal to the number of questions formed by each subscale, as the minimum value of each question is always 1.*

The formula to calculate each subscale value:

$$Subscale = \frac{(Real\ punctuation - minimum)}{Net\ rank} \times 100$$

If the minimum value of each question were zero, the procedure could be proposed as a weighted sum expressed as a percentage. This was essential point as it meant the matrices could be converted to a linear application. Refer to Supplementary Material for further details.

Decoded SF-36 questionnaire

The items and scales on the SF-36 are scored so that a higher score represents better health. For example, the function scales are scored so that a higher score indicates better function, and the pain scale is scored so that a high score suggests the patient is pain-free. After data entry, the items and scales were scored in three steps:

1. Re-encode the 10 items listed in the SF-36 manual.
2. Compute the scale score by summing the items in the same scale (raw scale score).
3. Convert the raw scale scores to a 0-100 scale (converted scale scores).

Modulus

The 36 responses to the SF-36 questionnaire were analysed, with the eight subscales investigated separately. The 36 responses for each participant were collected in columns, and each row represented one participant's anonymized data. In both cases, a 36-value vector (in the case of the 36-response matrix) or an 8-dimension vector (in the case of the subscales) was computed for each participant. For further details, see Supplementary Material.

Table 1
Subscales calculation scheme.

Subscale	Number of questions	Rank punctuations (min-max)	Net rank
Physic Function	10	10-30	20
Physic Rol	4	4-8	4
Body Pain	2	2-12	10
General Health	5	5-25	20
Vitality	4	4-24	20
Social Function	2	2-10	8
Emotional Rol	3	3-6	3
Mental Health	5	5-30	25

For each vector, the modulus was calculated as follows:

$\text{modulus}_i = \sqrt{\Sigma_{j=1}^{d} x_j^2}$

$modulus_i$: *the* i^{th} *register.*

d : *number of dimensions.* 36 *for our model.*

x_j : *the* j^{th} *element of vector.*

Single CPET test dataset

The "Study of exercise intolerance in adult patients with chronic fatigue syndrome using a cardiopulmonary exercise test (CPET)" is a prospective, cross-sectional, enrolled cohort study initiated in 2020. Details of the study are set out below:

- **Objectives**: To evaluate exercise intolerance in adult patients diagnosed with ME/CFS using CPET, taking into consideration cardiovascular, ventilatory, muscular, and metabolic variables to determine functional capacity and metabolic efficiency of exercise.
- **Inclusion criteria**: Participants aged between 18 and 60 and diagnosed (in accordance with the Fukuda[10] and Carruthers[12] criteria) with ME/CFS by the physician of the Central Sensitization Syndromes Unit, Vall d'Hebron Hospital, Barcelona. Participants must have clinical manifestations of the symptomatic muscle group of exercise intolerance, be able to perform the exercise test, and agree to take part in the study by signing the informed consent form.
- **Exclusion criteria**: Patients with contraindications for CPET or who do not provide informed consent to participate in the study.

Two hundred and thirty-nine participants were enrolled and referred to the Cardiopulmonary Rehabilitation Exercise Laboratory to perform CPETs. The study included 85.7% women (mean age 50.15±8.61 years) and 14.3% men (mean age 46.94±9.82 years). Patients were advised to take their regular medication and to avoid strenuous exercise on the test day. They were also asked to fast for two hours before the test and to drink only water. The participants were connected to an electrocardiogram, a pulse oximeter, and a blood pressure cuff to continuously monitor their vital signs during all study phases. They underwent a symptom-limited maximal exercise test using an electronically braked bicycle ergometer (Ergoline GmbH ER 800 S, Bitz, Germany). Exhaled air was collected via a two-way breathing valve attached to a mask that covered the participant's nose and mouth. Respiratory gasses breath-by-breath were analyzed using a Vyntus CPX gas analyzer (Ergoespirometer, Vyaire, Hoechberg, Germany) and SentrySuite 3.0 software. The spirometer was calibrated under ambient conditions before testing. Version V-781239 V 06.02 of the technical manual is available on the SentrySuite website. We used cardiopulmonary exercise testing based on work by Dr. Wasserman and colleagues at the University of California, Los Angeles.[20]

CPET Protocol Phases

The CPET tests were performed in the Cardiopulmonary Rehabilitation Exercise Laboratory under the supervision of a physician and a nurse and following the standardized protocols of the Central Sensitization Syndromes Unit of the Vall d'Hebron Hospital, Barcelona. These tests were performed in a hospital setting at no additional cost to the patient. The tests involved the following 3 phases:

Phase 1. Determine baseline cardiovascular and respiratory values as follows:

- Simple spirometry: three consecutive manoeuvres were performed to obtain reproducible data.
- Slow spirometry: one assessment was performed to determine maximum voluntary ventilation (MVV).

Phase 2. Maximum incremental cycle ergometer test: starting with three minutes of rest, followed by two minutes of unloaded pedalling, then adding an incremental load at a rate of 10 W/min (Watts / minutes) while maintaining a pedalling cadence of between 40-50 and 50-60 W/min (depending on the training status and previous fatigue level of the participant) until exhaustion due to muscular fatigue and/or dyspnoea, or according to the operator's medical criteria as electrocardiographic changes or symptoms contraindicating continuation of the test.

Phase 3. Recovery. On cycling completion, recovery from exercise was monitored for 3 minutes and the reasons for stopping CPET recorded. Continuous electrocardiographic monitoring of heart rate (12-lead) and blood pressure was performed every 2 minutes throughout the test. A Borg test was performed at baseline and at maximal exercise.

Maximal criteria test

Gas exchange data were recorded during CPET and recovery phases. The criteria used to determine whether participants had reached maximal physiological effort were as follows: plateau in oxygen consumption with increasing workload; modified rate of perceived exertion (RPE) > 8 (scale of 0-10); respiratory exchange ratio (RER) > 1.1, reaching at least 85% of age-predicted maximal heart rate, or a peak blood lactate > 8 mmol. When two of the three criteria were met, it was considered that the patient had exerted maximal effort.

The key measure sought in this study was the cardiorespiratory fitness of patients with ME/CFS, as determined by peak oxygen uptake (VO2 peak: measured during incremental exercise) and representing the maximum aerobic power during cumulative effort. The VO2 peak was measured in millilitres of oxygen per kilogram of body weight per minute ($mL^{*}Kg^{-1}{}^{*}min^{-1}$). This parameter is described as the maximum energy capacity achieved by aerobic metabolism per unit of time (aerobic capacity) during an incremental CPET. Any pathophysiological situation that impairs oxygen transport from the air to the mitochondria and its utilization during exercise will reduce the predicted values of peak oxygen uptake according to age and sex.[21]

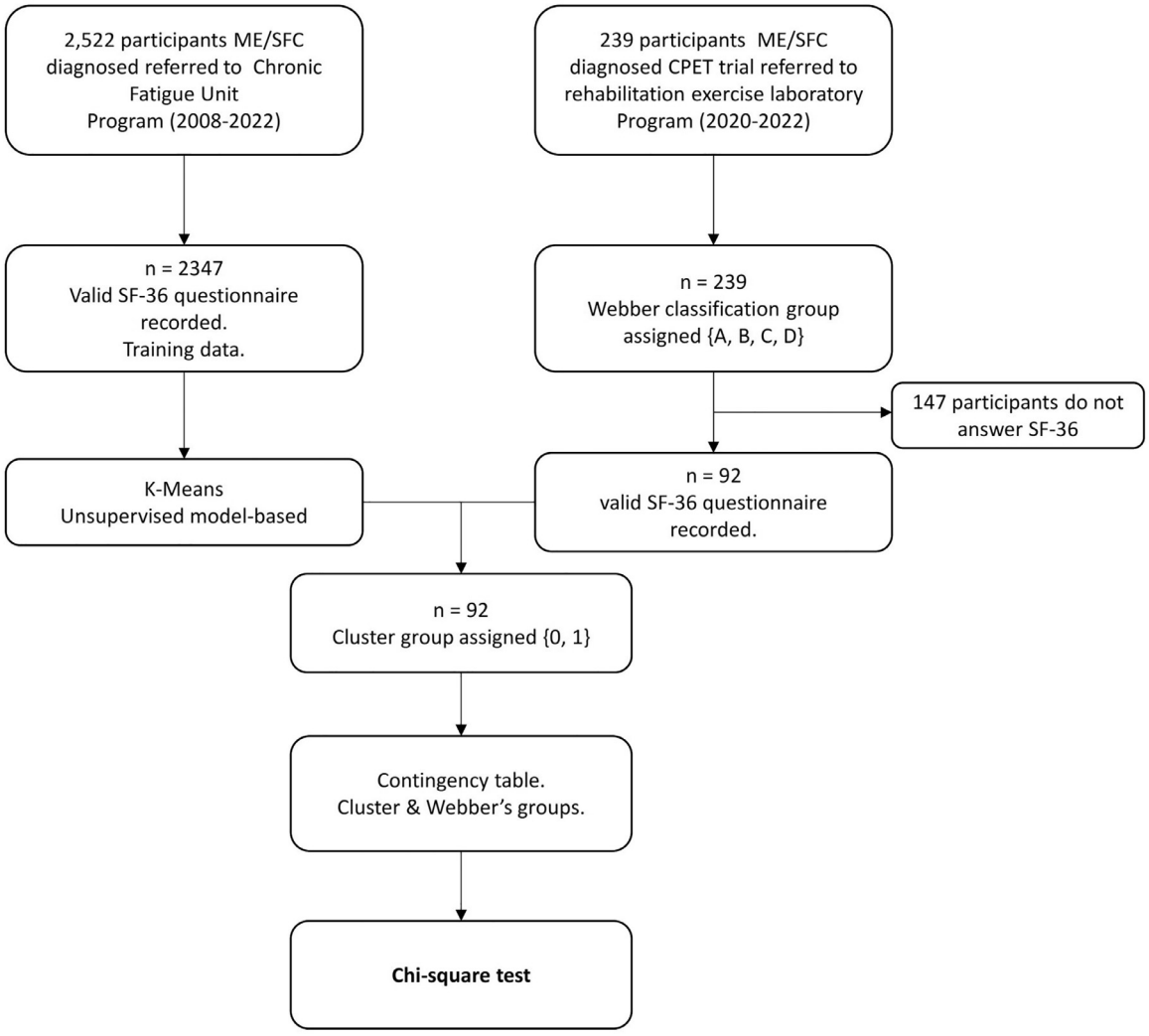

Figure 1. CONSORT diagram.

Patient enrolment procedure

This research presents data from two studies (see Figure 1): the first, "Population-based Registry of Chronic Fatigue Syndrome Patients", was approved 10/18/2006 and is used for automatic patient clusters using machine learning; the second, "Study of exercise intolerance in adult patients with chronic fatigue syndrome using a cardiopulmonary exercise stress test (CPET)", was used as a validation set.

The first study, "Population-based Registry of Chronic Fatigue Syndrome Patients" included SF-36 questionnaires collected from 2522 patients recruited at the Chronic Fatigue Syndrome Unit of the Vall d'Hebron Hospital, Barcelona, between 2008 and 2022, with their informed consent to participate in research. From this initial group, only 2347 of the questionnaires were valid. These 2347 questionnaire responses were used to train the proposed machine learning model to identify two different ME/CFS clusters.

The second study: "Study of exercise intolerance in adult patients with chronic fatigue syndrome using a cardiopulmonary exercise stress test (CPET)" included 239 CPET tests collected between 2020 and 2022 from ME/CFS patients referred to the Cardiopulmonary Rehabilitation Exercise Laboratory from the Central Sensitization Syndromes Unit of the Vall d'Hebron Hospital, Barcelona. The CPET results were collected according to Weber's classification. From the 239 patients, 92 completed both studies (CPET and SF-36 questionnaire). A contingency table of the 92 selected datasets was created and the previously trained clustering model applied, thus classifying each of the 92 patients as belonging to cluster 0 or cluster 1. Finally, as a validation step, the Chi-square test of independence was performed to determine whether there was a relationship between the two categorical variables (Weber's classification and clusters).

Statistical Analysis

Contingency table difference analysis

The creation of the contingency table is based on two categorical variables. The first variable is computed by the previously trained machine learning model, which assigns a cluster value {0, 1} based on all the answers from each input SF-36 questionnaire. The second variable is denoted by a letter {A, B, C, D} contingent on the peak oxygen consumption measured in the CPET.

- Mild to none (A) when Peak VO2 value is > 20 $mL^{*}Kg^{-1}{}^{*}min^{-1}$
- Mild to moderate (B) when Peak VO2 value range is 16-20 $mL^{*}Kg^{-1}{}^{*}min^{-1}$
- Moderate to severe (C) when Peak VO2 value range 10-16 $mL^{*}Kg^{-1}{}^{*}min^{-1}$
- Severe (D) when Peak VO2 value is < 10 $mL^{*}Kg^{-1}{}^{*}min^{-1}$

This is evaluated by groups {A, B, C, D} of patients who performed the exercise test, n=239. The self-administered SF-36 questionnaire was

completed by 92 patients only. All the SF-36 responses from these 92 patients were assigned to clusters {0, 1} in the model. These data were then used to create a contingency table for the 92 participants. In summary, each participant is assigned to a cluster (0 or 1) and given a letter (corresponding to the peak oxygen consumption: A, B, C or D) that defines the two categorical variables in our analysis (Figure 1).

The Chi-square test of independence is a statistical hypothesis test used to determine whether two categorical or nominal variables are likely to be related and can be used when counting the values of two categorical variables. The parameters were analysed for a more in-depth evaluation of the results. The likelihood ratio chi-square tests provide a range of parameterizations that support accurate selections for various distributions and sample sizes. Pearson's and Cressie-Read's chi-squared tests often tend to select overly complex bivariate parameterizations (up to 4 options in Weber's classification). This is because chi-squared statistics are overly inflated when models are fitted to sparse bivariate distributions and minor expected frequencies are used as divisors.[22] We decided to compare different strategies using different values of lambda. Python package Pingouin has used (version 0.5.2) to see if they differed.[23]

Parameters analysed in the contingency table:

- **Lambda**. A measure of association that reflects the proportional reduction in error when the values of the independent variable are used as predictors of the values of a dependent variable. A value equal to 1 indicates that the independent variable is a perfect predictor of the dependent variable. A value equal to 0 indicates that the independent variable has no contribution to the prediction of the dependent variable: Pearson (lambda=1), Cressie-Read (lambda=0.67) and Log-likelihood (lambda=0).
- **dof** is the Chi-square's degrees of freedom, and is calculated using the equation $dof = (r - 1) \times (c - 1)$, where r is the number of rows and c is the number of columns.
- The **p-value** is the probability of obtaining a chi-square equal to or greater than that obtained in the current experiment in the current experiment, given that the null hypothesis is true. Generally, a p-value of 0.05 or greater is considered critical. Anything less indicates significant variances, and the hypothesis must be rejected.
- **Cramer's V** is an effect size measurement for the Chi-square test of independence and measures how strongly two categorical fields are associated. The degree of freedom (dof) is 3. A value of between 0.06 and 0.17 is considered small-medium; up to 0.30, medium-large; and greater than 0.30, large.[28] Cramer's V is based on Pearson's Chi-squared statistic and was published by Harald Cramer.[29]
- The **power of the goodness of fit** (δ) or **Chi-square independence test**. High power means there is a low probability of concluding that no effect exists when there is one. Statistical power depends upon effect size and the sample size.

$$\delta = d \times \sqrt{n}$$

Where d is Cohen's coefficient and n is the number of registers.[24].

Clustering analysis

The clustering analysis was implemented in Python (version 3.7.14) using and comparing the decoded SF-36 answers and subscale matrices. The dimensions were 2347 × 36 and 2347 × 8, respectively. To select the optimal number of clusters, some models were fitted with values in the range.[2,6] for k (Birch and Spectral Clustering) using the elbow method.[25] and the Calinski and Harabasz metric (see Supplementary Materials for further details). Three validation metrics were proposed using the scikit-learn package (version 1.0.2) to evaluate the performance of each tested model when the true labels are unknown:

- Silhouette Coefficient
- Calinski–Harabasz Index[26]
- Davies–Bouldin Index[27]

Results

Matrix Analysis

The two databases for analysis were compared and contrasted to determine which provided the best information for classifying patients with ME/CFS; first, the 36 raw responses, followed by the eight subscales (for all results, see Supplementary Material). No correlation was found between the variables in the two matrices. However, when the modulus of each matrix was compared, this yielded different results, with the modulus of the raw responses showing a different health status to the modulus of the matrix of the subscales. This finding is to argue that using one matrix cannot be used as an equivalent to the other, as the model would give different results for each. A PCA analysis shows that in both matrices, emotional role (ER) and mental health (MH) are correlated and symmetric with those parts of the questionnaire dealing with physical function and vitality. Thus, the differences between the matrices were given by weighting the subscales, which significantly modified a single value of the SF-36 questionnaire. It shows that the subscales cannot be interpreted globally and must be analysed explicitly. This study demonstrates that the matrix of the 36 raw responses used as training data improves the results of the unsupervised machine learning model.

Cluster Analysis

The best-performing model for clustering MS/CFS patients was the K-means model, with the optimal number of clusters being two. The model was trained using the 36 raw answers of the 2,347 validated questionnaires to classify each subject into one of the two labelled groups with values {0, 1}. See Supplementary Material for further details.

Validation group Analysis

A total of 239 validated records were included. The Weber Classification results of the study population that completed the CPET are presented below. Participants' physical data is shown, and the peak oxygen consumed during the test is analysed and labelled as stated above:

- Mild to none (A) when Peak VO2 value is > 20 $mL^{*}Kg^{-1}{}^{*}min^{-1}$
- Mild to moderate (B) when Peak VO2 value range is 16-20 $mL^{*}Kg^{-1}{}^{*}min^{-1}$
- Moderate to severe (C) when Peak VO2 value range is 10-16 $mL^{*}Kg^{-1}{}^{*}min^{-1}$
- Severe (D) when Peak VO2 value is < 10 $mL^{*}Kg^{-1}{}^{*}min^{-1}$

A total of 239 records were collected and validated. The results are shown in Table 2.

Contingency table for single CPET: Weber's classification and clustering analysis

Only 92 records were selected. These corresponded to patients with validated SF-36 questionnaire responses and CPET. A similar analysis to the previous group analysis was performed on each of the 92 patients individually. Table 3 shows the records grouped according to Weber's classification. Note that there are no class D records for the male group as all collected values for the VO2 Peak variable are greater than 10.

Following this, the 92 selected patients were classified into two clusters using the previously trained machine learning model. Table 4 shows the results of the data analysed according to the cluster assigned by the machine-learning-based model.

Table 2
Weber Classification Results (n=239). Following Weber's classification, class was defined according to the VO2 value of a single CPET. Weber's classification stratifies patients based on peak VO2 and anaerobic threshold to define functional exercise capacity[15]: Mild to none (A) when Peak VO2 value is > 20; Mild to moderate (B) when Peak VO2 value range is 16-20; Moderate to severe (C) when Peak VO2 value range is 10-16; and Severe (D) when Peak VO2 value is < 10. **n**: Number of participants in each subgroup. Arithmetic means and standard deviation (in brackets) are shown for other parameters. Age in years, weigh in kg, height in cm and VO2 peak in $mL^{*}Kg^{-1}{*}min^{-1}$.

Gender	Class	n	Age	Weight	Height	VO2 peak
Female	A	35	47.37 (10.07)	56.85 (8.46)	160.48 (6.80)	24.56 (3.89)
	B	50	49.48 (9.23)	65.36 (13.92)	161.01 (6.02)	17.32 (1.01)
	C	95	51.47 (7.78)	73.05 (14.39)	160.75 (6.26)	13.13 (1.60)
	D	25	50.40 (7.51)	74.68 (17.06)	163.24 (5.26)	8.54 (1.09)
Male	A	15	47.0 (12.10)	78.02 (11.30)	174.26 (8.38)	25.19 (3.65)
	B	8	42.75 (9.13)	82.25 (11.75)	174.50 (13.58)	17.58 (1.29)
	C	9	50.22 (6.18)	88.88 (17.60)	177.11 (6.62)	12.76 (1.78)
	D	2	48.50 (2.12)	77.45 (16.19)	177.5 (0.70)	8.80 (0.84)

Table 3
Weber Classification Contingency Table Analysis (n=92). Weber's classification defined the class according to the VO2 value of a single CPET. Weber's Classification stratifies patients based on peak VO2 and anaerobic threshold to define functional exercise capacity[15]: Mild to none (A) when Peak VO2 value is > 20; Mild to moderate (B) if Peak VO2 value range is 16-20; Moderate to severe (C) when Peak VO2 value range is 10-16; and Severe (D) when Peak VO2 value is < 10. **n**: Number of participants in each subgroup. Arithmetic means and standard deviation (in brackets) are shown for other parameters. Age in years, weigh in kg, height in cm and VO2 peak in $mL^{*}Kg^{-1}{*}min^{-1}$. **Modulus**: subgroup means modulus value.

Gender	Class	n	Age	Weight	Height	VO2 peak	Modulus
Female	A	7	46.85 (10.25)	60.31 (11.15)	159.85 (9.92)	23.15 (0.8)	12.61 (1.89)
	B	22	51.54 (9.59)	61.85 (12.35)	161.18 (4.44)	17.51 (0.8)	13.98 (2.35)
	C	44	51.59 (7.15)	72.79 (14.92)	160.87 (5.88)	12.87 (1.54)	12.20 (2.46)
	D	8	51.87 (8.44)	75.0 (22.4)	160.87 (4.64)	8.12 (1.44)	11.73 (1.99)
Male	A	2	51.0 (7.07)	85.0 (14.12)	169.0 (12.73)	24.45 (2.19)	17.17 (1.61)
	B	5	44.6 (5.31)	86.24 (13.51)	177.60 (12.76)	17.18 (1.07)	13.65 (2.3)
	C	2	47.0 (7.07)	71.5 (12.02)	181.0 (1.41)	14.1 (0.84)	13.47 (1.23)

Table 4
Clustering Classification Contingency Table Analysis (n = 92). Cluster: number defined by the Euclidean K-means result with matrix A. Cluster 0 corresponds to better health status because of the higher modulus. Similarly, group one corresponds to patients with worse health status. n: Number of participants in each subgroup. Arithmetic means and standard deviation (in brackets) are shown for other parameters. Age in years, weight in kg, height in cm and VO2 peak in $mL^{*}Kg^{-1}{*}min^{-1}$. **Modulus**: mean of modulus value of the subgroup.

Gender	Cluster	n	Age	Weight	Height	VO2 peak	Modulus
Female	0	32	53.12 (7.87)	64.96 (16.63)	159.75 (6.45)	16.50 (3.52)	15.06 (1.38)
	1	49	49.94 (8.28)	71.57 (14.57)	161.77 (5.17)	13.28 (3.94)	11.12 (1.59)
Male	0	8	47.50 (5.45)	83.37 (13.97)	178.00 (5.17)	18.56 (4.63)	14.79 (2.23)
	1	1	39.00 (-)	77.20 (-)	164.00 (-)	16.50 (-)	11.26 (-)

Table 5
Chi-square test results. All parameters defined in the method section.

Test	Lambda	Chi2	dof	*p*-value	Cramer	Power
Pearson	1.00	19.06	3	0.000266	0.46	0.96
Cressie-read	0.67	19.57	3	0.000208	0.46	0.97
Log-likelihood	0.00	22.29	3	0.000005	0.49	0.98

Finally, Chi-square tests were used to determine whether there is a statistically significant difference between the observed and expected values. Three tests were performed to analyse the independence of the variables, as shown in Table 5. It should be emphasised that the three tests (Pearson, Cressie-read and Log-likelihood) are positively correlated, as a p-value<0.01 indicates a clear relationship between the categories analysed (clusters and Weber's Classification). This implies that the model correctly predicts the CPET scores with a precision equal to or higher than 99%. As a general guideline for consistency, the observed and expected contingency tables should not have cells with frequencies less than 5.

Discussion

The purpose of this study was to determine whether there is a relationship between the results of self-administered SF-36 questionnaires and CPET oxygen consumption values. CPET results were analysed, focusing on peak VO2 as a determinant of functional capacity and stratified according to Weber's classification.[15] The one-day test has been examined in other pathologies such as cardiac and pulmonary,[30] and more recently in long COVID-19 syndrome.[31] Previous studies have observed decreased cardiovascular response and increased global and maximal heart rate in ME/CFS patients using the one-day exercise test, which may be pathology-specific, resulting in early-onset fatigue, dysfunctional exercise capacity, inconsistent response, or lack of motivation.[32] In contrast, other studies[33] suggest that reduced exercise capacity may be related to autonomic dysfunction, as ME/CFS patients have difficulty reaching their age-predicted maximal heart rates. This may be one of the reasons why their physical performance is impaired.[33] Data analysis shows that ME/CFS patients have lower cardiorespiratory fitness levels than healthy control subjects,[34] and another recent study revealed that results from an analysis of various factors obtained after a single CPET could be used as biomarkers for diagnosing ME/CFS.[35] Although the exact mechanisms associated with low exercise capacity in patients with ME/CFS have not been determined, a single CPET can be a reliable and accessible test for patients that provides objective physiological data on their response to exercise.

It has been suggested that ME/CFS could have an autoimmune etiology, as antibodies against beta2-adrenergic receptors (β2AdR) and muscarinic acetylcholine receptors (M3 AChR and M4 AChR) have been

identified in symptoms such as cognitive deficits, autonomic dysregulation, and immune activation.[36] To assess this, patients must undergo laboratory testing, which is considerably more costly to the healthcare system than filling in a SF-36 questionnaire. Furthermore, laboratory testing is not considered an optimal method for diagnosing a cognitive or physical impairment. The procedure that we propose, however, is based on analysing answers to a SF-36 questionnaire, which is easily accessible in primary care and has the advantage that only those patients with specific results (classified as cluster 1 in the proposed model) would be evaluated for referral to a dedicated unit.

An earlier study demonstrated that low VO2 max values are directly correlated with a subscale of the SF-36 questionnaire.[18] The differences in peak VO2 values between the male and female populations observed in our study are in line with those found in previous studies,[37,38] in that women presented lower peak VO2 values than men on the first CPET and a worse classification on the Weber scale. This is confirmed in the clustering analysis shown in Table 4.

Results of the study were based on a chi-square test of independence (see Table 5) using three tests with different lambda values and indicate clearly that worse physical condition on exertion corresponds to a worse response to the SF-36 questionnaire (p-value < 0.001, for the three tests). These results suggest that a CPET can be used as a biomarker to measure oxygen consumption and objectively assess the health status of patients diagnosed with ME/CFS. The relationship between the SF-36 and CPET results of this study may have important implications for clinical practice. The SF-36 questionnaire is readily available for all ME/CFS patient assessment services (both primary care and dedicated units). Using the questionnaire would help predict oxygen consumption and could initiate a referral from primary care to a dedicated unit to begin early multidisciplinary evaluation and measure patient progress after treatments. These results are consistent with other studies discussed in this paper and highlight the relationship between responses to quality of life questions the SF-36 questionnaire and CPET results from a machine learning perspective.

Limitations

The main limitation of our study is that it is performed in only one hospital centre with an unbalanced gender population. It would be appropriate to contrast our results with a larger number of participants from multiple centres.

Conclusions

Low oxygen consumption on a CPET could be considered a diagnostic biomarker in patients with ME/CFS. Findings of our study reveal a relationship between the SF-36 questionnaire and Weber's classification. Clustering datasets from health questionnaires such as the SF-36 should be performed on raw response data, which preserves the original information and improves the quality of the model. Adopting this procedure in primary care by analysing responses to a SF-36 questionnaire may help early referral of potential patients with ME/CFS to a dedicated unit (accredited by the Department of Health of the Generalitat de Catalunya requires a group of professionals, internists, nurses, physiatrists, psychologists and physiotherapists, who carry out multidisciplinary work, with the application of pharmacological and non-pharmacological treatments, such as programmed physical exercise and cognitive-behavioral therapy). Further research could focus on confirming the results presented in a multi-centre scenario.

Ethics Disclosure

Protection of human and animal subjects. We confirm that the procedures this study were in accordance with the regulations established by the Clinical Research and Ethics Committee and the World Medical Association's Declaration of Helsinki. The research protocols were approved by the Research Ethics Committee of the Vall d'Hebron University Hospital for both studies: 1) "Population-based Registry of Patients with Chronic Fatigue Syndrome", approved 18/10/2006; and 2) PR-AG9 01/2020 "Study of exercise intolerance (IEF) in adult patients with chronic fatigue syndrome (CFS) using a cardiopulmonary exercise stress test (CPET)", approved 22/09/2020.

Confidentiality of data. We confirm we have followed the protocols of their work centre on publication of patient data, right to privacy and informed consent. The authors obtained the informed consent of all patients and/or subjects mentioned in the article.

Authorship. We confirm that all named authors of the manuscript have participated in the conception and design of the study; collection, analysis, and interpretation of the data; and the submitted manuscript has been written, revised and approved by all of us.

Declaration of Conflicting Interests. We wish to confirm there are no potential conflicts of interest concerning this article's research, authorship, and/or publication.

Funding

This research received no specific grant from public, commercial, or non-profit funding agencies.

Acknowledgments

We wish to thank Amelia Marquino for her dedication and excellent work managing and maintaining the database. We are also grateful to the Fundación Mútua Madrileña for financing the creation of the database through a grant awarded in 2006 for the "Population-based Registry of Patients with Chronic Fatigue Syndrome" project.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.clinthera.2023.09.007.

References

1. Morris G, Puri BK, Walker AJ, Maes M, Carvalho AF, Walder K, et al. Myalgic encephalomyelitis/chronic fatigue syndrome: From pathophysiological insights to novel therapeutic opportunities. *Pharmacological Research*. 2019:104450. doi:10.1016/j.phrs.2019.104450.
2. Lim E-J, Ahn Y-C, Jang E-S, Lee S-W, Lee S-H, Son C-G. Systematic review and meta-analysis of the prevalence of chronic fatigue syndrome/myalgic encephalomyelitis (CFS/ME). *Journal of Translational Medicine*. 2020. doi:10.1186/s12967-020-02269-0.
3. Norris T, Collin SM, Tilling K, Nuevo R, Stansfeld SA, Sterne JA, et al. Natural course of chronic fatigue syndrome/myalgic encephalomyelitis in adolescents. *Arch Dis Child*. 2017;102:522–528. doi:10.1136/archdischild-2016-311198.
4. Wu Q, Gao J, Bai D, Zhong Y, Yang Z, Jiang X. Prevalence of chronic fatigue syndrome in China: a meta-analysis. Youjiang Med J.
5. Lloyd AR, Hickie I, Boughton CR, Spencer O, Wakefield D. Prevalence of chronic fatigue syndrome in an Australian population. *Med J Aust*. 1990;153:522–528. doi:10.5694/j.1326-5377.1990.tb126191.x.
6. Committee on the Diagnostic Criteria for Myalgic Encephalomyelitis/Chronic Fatigue Syndrome, Board on the Health of Select Populations, Institute of Medicine. Background. National Academies Press (US); 2015. Available: https://www.ncbi.nlm.nih.gov/books/NBK284897/.
7. Maes M, Twisk FN. Why myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) may kill you: disorders in the inflammatory and oxidative and nitrosative stress (IO&NS) pathways may explain cardiovascular disorders in ME/CFS. *Neuro Endocrinol Lett*. 2009;30:677–693. Available. https://www.ncbi.nlm.nih.gov/pubmed/20038921.
8. Wirth KJ, Scheibenbogen C, Paul F. An attempt to explain the neurological symptoms of Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. *J Transl Med*. 2021;19:471. doi:10.1186/s12967-021-03143-3.
9. Fernández-Quirós J, Lacasa-Cazcarra M, Alegre-Martín J, Sanmartín-Sentañes R, Almirall M, Launois-Obregón P, et al. The Conners Continuous Performance Test CPT3TM: Is it a reliable marker to predict neurocognitive dysfunction in Myalgic encephalomyelitis/chronic fatigue syndrome? *Front Psychol*. 2023;14:1127193. doi:10.3389/fpsyg.2023.1127193.
10. Fukuda K, Straus SE, Hickie I, Sharpe MC, Dobbins JG, Komaroff A. The chronic fatigue syndrome: a comprehensive approach to its definition and study. International Chronic Fatigue Syndrome Study Group. *Ann Intern Med*. 1994;121:953–959. doi:10.7326/0003-4819-121-12-199412150-00009.

11. Christley Y, Duffy T, Martin CR. A review of the definitional criteria for chronic fatigue syndrome. *J Eval Clin Pract*. 2012;18:25–31. doi:10.1111/j.1365-2753.2010.01512.x.
12. Carruthers BM, Jain AK, De Meirleir KL, Peterson DL, Klimas NG, Martin Lerner A, et al. Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. *Journal of Chronic Fatigue Syndrome*. 2003:7–115. doi:10.1300/j092v11n01_02.
13. Carruthers BM, van de Sande MI, De Meirleir KL, Klimas NG, Broderick G, Mitchell T, et al. Myalgic encephalomyelitis: International Consensus Criteria. *J Intern Med*. 2011;270:327–338. doi:10.1111/j.1365-2796.2011.02428.x.
14. Stevens S, Snell C, Stevens J, Keller B. VanNess JM. Cardiopulmonary Exercise Test Methodology for Assessing Exertion Intolerance in Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. *Front Pediatr*. 2018;6:242. doi:10.3389/fped.2018.00242.
15. Weber KT, Janicki JS. Cardiopulmonary exercise testing for evaluation of chronic cardiac failure. *Am J Cardiol*. 1985;55:22A–31A. doi:10.1016/0002-9149(85)90792-1.
16. Asprusten TT, Sletner L, Wyller VBB. Are there subgroups of chronic fatigue syndrome? An exploratory cluster analysis of biological markers. *J Transl Med*. 2021;19:48. doi:10.1186/s12967-021-02713-9.
17. Kujawski S, Słomko J, Newton JL, Eaton-Fitch N, Staines DR, Marshall-Gradisnik S, et al. Network Analysis of Symptoms Co-Occurrence in Chronic Fatigue Syndrome. *Int J Environ Res Public Health*. 2021;18. doi:10.3390/ijerph182010736.
18. van Campen CLMC, Rowe PC, Visser FC. Validation of the Severity of Myalgic Encephalomyelitis/Chronic Fatigue Syndrome by Other Measures than History: Activity Bracelet, Cardiopulmonary Exercise Testing and a Validated Activity Questionnaire: SF-36. *Healthcare (Basel)*. 2020;8. doi:10.3390/healthcare8030273.
19. Alonso J, Prieto L, Antó JM. The Spanish version of the SF-36 Health Survey (the SF-36 health questionnaire): an instrument for measuring clinical results. *Med Clin*. 1995;104:771–776. Available. https://www.ncbi.nlm.nih.gov/pubmed/7783470 .
20. Wasserman K, Hansen JE, Sue DY, Stringer WW, Whipp BJ. Principles of exercise testing and interpretation: including pathophysiology and clinical applications. *Med Sci Sports Exercise*. 2005;37:1249.
21. Mezzani A. Cardiopulmonary Exercise Testing: Basics of Methodology and Measurements. *Ann Am Thorac Soc*. 2017;14:S3–S11. doi:10.1513/AnnalsATS.201612-997FR.
22. Moses T, Holland PW. A comparison of statistical selection strategies for univariate and bivariate log-linear models. *Br J Math Stat Psychol*. 2010;63:557–574. doi:10.1348/000711009X478580.
23. Vallat R. Pingouin: statistics in Python. *J Open Source Softw*. 2018;3:1026. doi:10.21105/joss.01026.
24. Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2nd Edition. Routledge; 1988. doi:10.4324/9780203771587.
25. Bengfort B, Gray L, Bilbro R, Roman P, Deziel P, McIntyre K, et al. Yellowbrick v1.5. 2022. doi:10.5281/zenodo.7013541
26. Kozak M. A Dendrite Method for Cluster Analysis" by Caliński and Harabasz: A Classical Work that is Far Too Often Incorrectly Cited. *Communications in Statistics - Theory and Methods*. 2012;41:2279–2280. doi:10.1080/03610926.2011.560741.
27. Halkidi M, Batistakis Y, Vazirgiannis M. On Clustering Validation Techniques. *J Intell Inf Syst*. 2001;17:107–145. doi:10.1023/A:1012801612483.
28. Rousseeuw PJ. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J Comput Appl Math*. 1987;20:53–65. doi:10.1016/0377-0427(87)90125-7.
29. Novikov A. PyClustering: Data Mining Library. *J Open Source Softw*. 2019;4:1230. doi:10.21105/joss.01230.
30. Goulart C da L, Dos Santos PB, Caruso FR, Arêas GPT, Marinho RS, Camargo P de F, et al. The Value of Cardiopulmonary Exercise Testing in Determining Severity in Patients with both Systolic Heart Failure and COPD. *Sci Rep*. 2020;10:4309. doi:10.1038/s41598-020-61199-5.
31. Ferreira EVM, Oliveira RKF. Mechanisms of exercise intolerance after COVID-19: new perspectives beyond physical deconditioning. *J Bras Pneumol*. 2021;47:e20210406. doi:10.36416/1806-3756/e20210406.
32. Inbar O, Dlin R, Rotstein A, Whipp BJ. Physiological responses to incremental exercise in patients with chronic fatigue syndrome. *Med Sci Sports Exerc*. 2001;33:1463–1470. doi:10.1097/00005768-200109000-00007.
33. De Becker P, Roeykens J, Reynders M, McGregor N, De Meirleir K. Exercise capacity in chronic fatigue syndrome. *Arch Intern Med*. 2000;160:3270–3277. doi:10.1001/archinte.160.21.3270.
34. Zambolin F, Duro-Ocana P, Faisal A, Bagley L, Gregory WJ, Jones AW, et al. Fibromyalgia and Chronic Fatigue Syndromes: A systematic review and meta-analysis of cardiorespiratory fitness and neuromuscular function compared with healthy individuals. *PLoS One*. 2022;17:e0276009. doi:10.1371/journal.pone.0276009.
35. Pifarré F, Rosselló L, Hileno R, Palmi J, Bañeres L, Planas A, et al. The use of oxygen as a possible screening biomarker for the diagnosis of chronic fatigue. *Apunts Sports Medicine*. 2022;57:100379. doi:10.1016/j.apunsm.2022.100379.
36. Gravelsina S, Vilmane A, Svirskis S, Rasa-Dzelzkaleja S, Nora-Krukle Z, Vecvagare K, et al. Biomarkers in the diagnostic algorithm of myalgic encephalomyelitis/chronic fatigue syndrome. *Front Immunol*. 2022;13:928945. doi:10.3389/fimmu.2022.928945.
37. van Campen C (linda) MC, Rowe PC, Visser FC. Two-Day Cardiopulmonary Exercise Testing in Females with a Severe Grade of Myalgic Encephalomyelitis/Chronic Fatigue Syndrome: Comparison with Patients with Mild and Moderate Disease. *Healthc Pap*. 2020;8:192. doi:10.3390/healthcare8030192.
38. van Campen CLMC, Visser FC. Comparing Idiopathic Chronic Fatigue and Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS) in Males: Response to Two-Day Cardiopulmonary Exercise Testing Protocol. *Healthcare (Basel)*. 2021:9. doi:10.3390/healthcare9060683.