Impact of exercise training on sarcopenia associated with non-alcoholic fatty liver disease (NAFLD) in humans: A systematic review and meta-analysis

Andrea Gonzalez1,2,3, Mayalen Valero-Breton1,2,3, Camila Huerta-Sanhueza1,2,3, Oscar Achiardi6, Felipe Simon2,4,5, and Claudio Cabello-Verrugio1,2,3*

1 Laboratory of Muscle Pathology, Fragility and Aging, Department of Biological Sciences, Faculty of Life Sciences. Universidad Andres Bello, Santiago, 8370146, Chile.
2 Millennium Institute on Immunology and Immunotherapy, Santiago, 8370146, Chile.
3 Center for the Development of Nanoscience and Nanotechnology (CEDENNA), Universidad de Santiago de Chile, Santiago, 8350709, Chile.
4 Millennium Nucleus of Ion Channels-Associated Diseases (MiNICAD), Universidad de Chile, Santiago 8370146, Chile
5 Laboratory of Integrative Physiopathology, Department of Biological Sciences, Faculty of Life Sciences. Universidad Andres Bello, Santiago, 8370146, Chile
6 Escuela de Kinesiología, Facultad de Ciencias, Pontificia Universidad Católica de Valparaíso. Avenida Brasil 2950, Valparaíso, Chile.

* Corresponding author: Claudio Cabello-Verrugio. Laboratory of Muscle Pathology, Fragility, and Aging. Department of Biological Sciences. Faculty of Life Sciences. Universidad Andres Bello, Postal

ABSTRACT

Objective: To determine the efficacy of physical exercise on strength, muscle mass and physical function in adult patients with non-alcoholic fatty liver disease (NAFLD).

Design: We conducted a systematic review and meta-analysis of seven studies to investigate exercise interventions on training in muscle strength, muscle mass and physical performance.

Data sources: We identified relevant randomised controlled trials (RCTs) in electronic databases (PubMed, CINAHL and Scopus).

Eligibility criteria: We analysed data from 66 RCTs investigating the effects of exercise on at least one variable of sarcopenia (muscle strength, muscle mass, physical function) in patients with NAFLD.

Results: Physical performance criteria improved in the exercise groups (mean differences [MD] 8.26 ml/kg*min [95% CI 5.27 to 11.24 ml/kg*min], p < 0.0001) versus the control; muscle mass, determined as lean body mass (LBM), showed no evidence of the beneficial effects of exercise versus the control (MD 1.01 kg [95% CI -1.78 to 3.80 kg], p = 0.48); we did not include muscle strength, as none of the selected studies evaluated it.

Summary/conclusion: Exercise training is a useful intervention tool to treat sarcopenia in patients with NAFLD; it increases their physical performance in the form of aerobic capacity but does not affect LBM. Future research should include muscle strength assessments and resistance training to evaluate the effects of exercise on sarcopenia in NAFLD patients.

Keywords: sarcopenia, non-alcoholic fatty liver disease (NAFLD), exercise training, muscle mass, muscle strength, physical performance
INTRODUCTION
Non-alcoholic fatty liver disease (NAFLD) is the most common type of chronic liver disease (CLD). It includes a broad spectrum of disorders ranging from accumulating lipids in the liver (steatosis) to the progressive inflammation denoted as non-alcoholic steatohepatitis (NASH) to advanced stages of damage such as fibrosis and cirrhosis\(^1\)\(^2\).

Sarcopenia is one of the most common complications associated with NAFLD, with a prevalence of 30—70\(^3\). According to the European Working Group on Sarcopenia in Older People (EWGSOP), a sarcopenia diagnosis is based on three criteria: i) low muscle strength (the primary indicator of sarcopenia), ii) small muscle quantity or quality and iii) low physical performance\(^4\). Sarcopenia negatively affects NAFLD progression\(^5\); its severity is more pronounced in the advanced stages of NAFLD, considered as an independent predictor of pre- and post-liver transplant complications and mortality\(^6\)\(^7\). Therefore, therapeutic interventions to prevent, revert or improve sarcopenia associated with NAFLD are essential.

Since sarcopenia involves a decline in muscle mass, strength and physical function, physical exercise is a promising tool for treating sarcopenia\(^8\). Due to low-quality evidence, the results of studies that evaluate the effect of exercise on patients with cirrhosis-associated sarcopenia are highly variable and inconclusive\(^9\). Despite evidence that training has a positive effect on liver dysfunction parameters in patients with NAFLD\(^10\)\(^11\), no studies evaluate the impact of training on sarcopenia parameters in NAFLD patients. In this context, a recent narrative review concluded that there is no consensus on the best type of exercise to improve muscle strength and physical function in NAFLD patients. Thus, more evidence is required to determine exercise specifications (type, intensity, frequency, supervised vs domiciliary)\(^2\).

Considering all the antecedents, we conclude that the effects of exercise on sarcopenia associated with NAFLD are poorly determined, and that there is no consensus on the best type of exercise for these patients. Therefore, our objective is to assess the efficacy of physical exercise on strength, muscle mass and physical function in adult patients with NAFLD.

METHODOLOGY
Protocol and Registration
We conducted our systematic review and meta-analysis following the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) statement. We registered the review protocol in the International Prospective Register of Systematic Reviews (PROSPERO reference number CRD42020191471), available from 3 September 2020.

Search strategy
We performed a systematic search in PubMed, CINAHL and Scopus from March—May 2020. We included articles published from January 2000—March 2020 containing the critical concepts of sarcopenia, fatty liver disease, exercise and related words. The essential concepts of the search were different for each database. For example, for PubMed, we used the keywords (‘atrophic muscular disorders’ OR ‘muscle atrophy’ OR ‘muscle degeneration’ OR ‘muscle fiber atrophy’ OR ‘muscle fiber degeneration’ OR ‘muscle wasting’ OR ‘muscular wasting’ OR ‘muscular atrophy’ OR ‘muscular atrophies’ OR ‘muscular degeneration’ OR ‘sarcopenia’) AND (‘non-alcoholic steatosis’ OR ‘non-alcoholic fatty liver disease’ OR ‘fatty liver’ OR ‘hepatic fat’ OR ‘liver fibrosis’ OR ‘liver disease’ OR ‘fatty liver disease’ OR ‘obesity’) AND (‘exercise’ OR ‘physical activity’ OR ‘exercise intervention’ OR ‘training’) (see Supplementary Material 2). We filtered the results for clinical randomised controlled trials (RCTs). Then, AG reviewed all titles and abstracts from the articles selected and only downloaded those mating the inclusion criteria.

Inclusion/exclusion criteria
The inclusion criteria are based on study population, intervention, comparison question (control) and outcomes (PICO)\(^12\).
• Investigate adults (18 or older) with liver disease of non-alcoholic origin (specifically, NAFLD or NASH) as confirmed by at least one of these parameters: hepatic biopsy, ultrasonography, computed tomography (CT), nuclear magnetic resonance spectroscopy (MRS), serum bile acids, gamma-glutamyl transpeptidase (GGT), aminotransferases, the ratio of aspartate transaminase (AST)/alanine transaminase (ALT) or bilirubin, alkaline phosphatase, dyslipidaemia\(^{13-17}\) (Population).

• Apply resistance training, endurance training or both. Combined exercise is defined as interventions that simultaneously used resistance and endurance training (Intervention).

• Apply exercise alone or in combination with other interventions (educational, nutritional, etc.); must have a control group and a group with exercise only (to determine its independent effect) (Control). Control groups cannot have any intervention that could influence study outcome measures (e.g. nutritional or physical intervention). Moreover, participants of the exercise group must perform three or more sessions of training per week for at least four weeks to determine the chronic effect of exercise on the variables. Studies must conduct pre- and post-intervention assessments to identify changes in the variables.

• Assay at least one variable included in the consensus definition of sarcopenia\(^2\) \(^18\); thus, the studies must evaluate any of three criteria: muscle strength, muscle mass, physical function (Table 1) (Outcome).

• Be a peer-reviewed study

• Be written in English

Exclusion criteria:

• Bibliographic reviews, not controlled or randomised clinical trials, animal models

• The participants had comorbidities such as type II diabetes mellitus, viral hepatitis, cardiovascular disease, lung disease or neurological disease.

• Interventions that would interfere with identifying any exercise-mediated effects

Two researchers (CH-S and MVB) independently performed title and abstract screening and further reviewed the abstracts using the full-text version of the reports. They resolved discrepancies/disagreements by consensus – when this was not possible, two independent researchers acted as referees (AGR and CCV).

**Table 1. Assessment criteria for the diagnosis of sarcopenia in NAFLD**

<table>
<thead>
<tr>
<th>Body composition (muscular mass)</th>
<th>Muscular Strength</th>
<th>Physical Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extremity circumferences (Thigh, Arm)</td>
<td>Handgrip strength</td>
<td>6-MWT (6-minute walk test)</td>
</tr>
<tr>
<td>Thigh US (ultrasound)</td>
<td>Knee flexion/extension</td>
<td>2-MST (2-minute step test)</td>
</tr>
<tr>
<td>BIA (bioelectrical impedance analysis)</td>
<td>Dynamometer</td>
<td>CPET (cardiopulmonary exercise testing)</td>
</tr>
<tr>
<td>CSA (the cross-sectional area from magnetic resonance imaging)</td>
<td>1 maximum repetition (1RM)</td>
<td>SPPB (Short physical performance Battery)</td>
</tr>
<tr>
<td>Measurement of sarcopenia criteria in NAFLD patients: muscle strength, muscle mass and physical performance in clinical practice and research.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------------------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Table 1</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>DXA (dual-energy x-ray absorptiometry)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>10 maximum repetition (10RM)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Usual gait speed</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Anthropometry</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Isokinetic evaluation</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Chair stands</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>MAMA (middle-arm muscle area)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Peak expiratory flow (specific to respiratory)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Timed get-up-and-go test</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Stair climb power test</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Data extraction**

We collected the data from seven final papers (n = 7), which we tabulated and ordered into a Microsoft Excel 2016 database. One researcher extracted the data, and two different researchers reviewed it to ensure data processing accuracy (Supplementary Material 1). The data included the following parameters: first author’s surname; publication year; number of participants by group; participants’ sex and age; method used to diagnose NAFLD and its severity; type, intensity, frequency and duration of the exercise intervention; sarcopenia criteria assessment methods. When studies presented intention-to-treat (ITT) values, we only extracted ITT data. Data extraction and synthesis for the outcome of interest are shown in the data synthesis section.

**TESTEX rating scale application**

We assessed study quality using the TESTEX rating scale (Table 2)\(^9\). It includes criteria to assess methodological quality as well as ranking the whole article. The scale comprises 12 assessment criteria for a maximum score of 15 points. Higher ratings reflect better study quality and reporting\(^9\).

TESTEX does not yet provide a validated cutoff score\(^9\), so we categorised the studies depending on the median score\(^10,20\). We classified studies above the median score as ‘high-quality’ and those below as ‘low-quality’. AG and OA performed TESTEX scoring separately. When different evaluations occurred, they discussed discrepancies until reaching a consensus score.

**Table 2. TESTEX study quality assessment**

<table>
<thead>
<tr>
<th>Study</th>
<th>Score</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>2012 Sullivan</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td>2013 Pugh</td>
<td></td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td>2014 Pugh</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>2015 Hallsworth</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2016 Shojae-Moradie</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>2017 Cheng</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>13</td>
</tr>
<tr>
<td>2017 Hughton</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><strong>Median</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>11</td>
</tr>
</tbody>
</table>

**Table 2. Study quality.**

*Notes.* 1) eligibility criteria specified (1 point); 2) randomisation defined (1 point); 3) allocation concealment (1 point); 4) groups similar at baseline (1 point); 5) assessor blinding in study reporting (1 point); 6) outcome measures assessed in 85% of patients (3 points); 7) intention-to-treat analysis (1 point); 8) between-group statistical comparisons reported (2 points); 9) point measures and measures of variability for all reported outcome measures (1 point); 10) activity monitoring in the control group (1 point); 11) relative exercise intensity remained constant (1 point); 12) exercise volume and energy expenditure (1 point).
Data synthesis
When at least three included studies reported the same outcome, we pooled data to perform a meta-analysis using Review Manager 5.4 (RevMan 5.4) as per the Cochrane Manual for Systematic Reviews of Interventions.\(^1\) We used a random-effects model for heterogeneity to address study participants’ different backgrounds and the variety of exercise interventions (which create differences in training variables such as exercise intensity and duration, session frequency, etc.). Outcome measures also differ among studies depending on the available technology to quantify muscle mass and physical performance. Finally, we expected methodological heterogeneity, as exercise training research uses different designs.

We used inverse variance because we analysed continuous data to compare exercise vs control or conventional care groups using mean and standard deviations from post-intervention values for each outcome. When other forms of variability measures such as standard error or confidence intervals (CIs) appeared, we calculated the standard deviations using RevMan 5.4.\(^1\)

We did not plan to investigate subgroups of patients, as we expected the participants to be homogeneous (see inclusion criteria); however, considering the heterogeneous nature of exercise training protocols, we expected to find different types of exercise training, which are classified as either endurance or resistance. In the case of this variety, we planned to perform a subgroup analysis comparing endurance training vs resistance training, investigating heterogeneity across subgroup results and analysing the variability in effect estimates between types of exercise using an I\(^2\) statistic.

For each outcome analysed, we present the data as mean differences (MD) with a 95% CI. We used standardised mean differences (SMD) to represent effect sizes and facilitate interpretations, with a value of 0.2 set as small, 0.5 as moderate and 0.8 or higher as a large effect size.\(^2\) Positive values indicated that it was favourable to the exercise intervention, and vice versa. Zero value suggested that there was no effect on the analysed variable.

Heterogeneity
We quantified heterogeneity using the \(\chi^2\) test, where a significant \(p\)-value was indicative of more considerable heterogeneity. We also applied a complementary inconsistency test \(I^2\) and presented this with its corresponding 95% CI to assess the degree of heterogeneity. Higher \(I^2\) values indicated greater heterogeneity.\(^3\)

Bias risk
We assessed all the included RCTs using the Cochrane collaboration risk of bias tool to search for elements that could over- or under-estimate the intervention’s effect. We evaluated the following for each study: selection bias (random sequence and generations, allocation concealment), performance bias (blinding of participants and research staff), detection bias (blinding of outcome evaluation), attrition bias (incomplete outcome data), reporting bias (selective reporting), other sources of bias.\(^1\) Two independent reviewers (AG and OA) performed the assessments; they discussed discrepancies until reaching a consensus.

RESULTS
Review studies
We initially identified 66 articles in PubMed, CINAHL and Scopus. Of these, we excluded one by duplication and 29 after analysing the title and abstract. Among the remaining 36 articles, we excluded 29 due to the combination of exercise with another intervention (\(n = 15\)), participants under 18 years (\(n = 2\)), article written in Spanish (\(n = 1\)), reports not RCTs (\(n = 5\)), no determination or evaluation of sarcopenia criteria (\(n = 2\)) and patients were not diagnosed with NAFLD or had comorbidities (\(n = 4\)). Supplementary Material 3 lists the excluded studies and causes for the exclusion.\(^5\) We included seven articles in the final analysis\(^5\) (Figure 1).
Description of included studies
A total of 182 participants were included in the analysis. Among them, 98 participated in exercise interventions and 84 in control groups (see Table 2). Five studies used a moderate-intensity protocol of endurance training53-56,59. Three studies used a combined exercise protocol (endurance and resistance)56-58; all seven studies the exercise interventions were supervised. All the studies had a control group composed of patients with NAFLD following standard or conventional care, which did not include exercise intervention or any other component that could have influenced the outcomes of interest53-59.

Exercise intervention duration ranged from 12–16 weeks, with training frequency of 2–5 times per week lasting 20–60 min per session. Endurance training intensity varied from 45–75% of oxygen consumption peak (VO2peak), 30–60% of heart rate reserve (HRR) or 16–18/20 (Borg Scale) rating of perceived exertion (RPE). Endurance training varied: cycle ergometer, treadmill, Nordic brisk walking, walking and self-selected gymnasium exercise routines.

Two articles described high-intensity interval training (HIIT) as a form of endurance training. One of these articles used 5 intervals of 2 min, adding 10 s to each range per week with 3 min of recovery and an RPE (Borg Scale) intensity of 16–17 (very hard)57. The other article used cycling intervals with an RPE (Borg scale) intensity of 16–1858. The remaining five studies carried out continuous training53-56,59.

Among the three articles that carried out combined training, one performed a moderate-intensity endurance protocol (40%–60% HRR) combined with resistance training; however, the details of the latter training were not described56. Another study used combined training protocol by performing endurance interval training with an intensity based on RPE (Borg scale) of 16–18 (very hard) together with resistance training with an RPE intensity (Borg scale) of 14–16 (hard). The exercises included hip and knee extensions, horizontal rows, chest presses, vertical rows and knee extensions58. One article required participants to perform an interval on a cycle ergometer, followed by a light-band-resisted upper body exercise (60 s) in the following order: face-pull, horizontal push, horizontal pull, 30° push. However, it was unclear whether this protocol was a combined workout57.

Quality assessment analysis
The TESTEX scale analysis yielded a median score of 11 out of 15 possible points for the studies10, with values closer to 15 of higher quality than values closer to 1. We evaluated assessment quality by organising the studies as low-quality if their score was less than the median and high-quality if the score was greater than or equal to the median10. Accordingly, six studies were high quality and one article was low-quality (Table 2). The main weaknesses TESTEX revealed were the supervision of physical activity in the control groups (0 of 7 articles), intention-to-treat analysis (1 of 7 studies) and blinding of the assessor for at least one key outcome (1 of 7 articles).

Risk of bias assessment
The risk of bias was low in the RCTs included in this systematic review and meta-analyses for every outcome. However, we had some concerns about their analysis of physical performance regarding blinding the assessors. Only one of the studies performed this key methodological feature, indicating a potential risk of detection bias for articles that used a cardiopulmonary exercise test (CPET) to assess physical performance, as assessors might influence patients’ performance in this type of evaluation. The details about the risk of bias assessment can be found in Supplementary Materials 4 and 5.

Change in muscle strength
We excluded the muscle strength criterion for sarcopenia because none of the selected studies evaluated it.
Change in physical performance

For the physical performance criterion, we performed a meta-analysis with the results obtained post-exercise by comparing the control group vs the exercise group. Data on physical performance were present in four studies\(^{53-56}\), all of which determined the direct maximum oxygen consumption (VO\(_{2\text{max}}\)) or oxygen consumption peak (VO\(_{2\text{peak}}\)) based on CPET and expressed in mL/Kg*min. In this regard, we excluded one article because it performed the 2 km walking test, which does not directly evaluate physical performance\(^{59}\).

Three studies showed significant increases in VO\(_{2\text{max}}\) or VO\(_{2\text{peak}}\). One study showed no change; here, the authors recognised that the results could be explained by a type II statistical error because of the small number of participants\(^{53}\).

The pooled analysis of the studies indicated a change in physical performance in favour of exercise (MD 8.26 mL/Kg*min [95% CI 5.27–11.24 mL/Kg*min], \(p < 0.0001\), Figure 2) and a large effect size (SMD 1.10 [0.60–1.60], \(p < 0.0001\), I\(^2\) = 0%). Studies that analysed physical performance had low heterogeneity (Chi\(^2\) 2.43, \(p = 0.49\); I\(^2\) 0%; Tau\(^2\) = 0.00).

Change in muscle mass

Four studies determined muscle mass as LBM\(^{53-57,59}\). They evaluated LBM by different methods, including dual-energy X-ray absorptiometry (DXA) and air displacement plethysmography. Despite this, we could compare the values they reported because they expressed the results in kilograms (Kg). None of these studies reported changes in LBM.

The studies we used to analyse this outcome had low heterogeneity (Chi\(^2\) 1.74, \(p = 0.63\); I\(^2\) 0%; Tau\(^2\) = 0.00); our meta-analysis found no evidence of a difference in the effect between exercise and control groups (MD 1.01 Kg [95% CI -1.78 to 3.80 Kg], \(p = 0.48\), Figure 3). There was a small effect size (SMD 0.09 [-0.27 to 0.44], \(p = 0.64\), I\(^2\) = 0%).

DISCUSSION

Following our results, we established the effect of exercise training in two of the three criteria for sarcopenia in NAFLD patients. Exercise training has a positive impact on physical capacity and no effect on the LBM of NAFLD patients. It was not possible to determine the impact of exercise on NAFLD patients’ muscle strength.

Effect of exercise on muscle strength

Some antecedents show the relationship between NAFLD status and muscle strength\(^{60}\). A recent study suggests that a higher hepatic steatosis index (HIS), a clinical feature of NAFLD patients, is associated with lower muscle strength in both sexes\(^{60}\). This antecedent suggests the relevance of muscle strength on NAFLD. Therefore, it is crucial to establish and analyse the influence of factors that could improve muscle strength in patients with NAFLD. There is consistent evidence that resistance training increases muscle strength\(^{61-63}\). Despite clear antecedents that link increased muscle strength with resistance exercise, none of the RCTs included here featured resistance training as a sole intervention. Three studies used a combined exercise protocol (endurance and resistance)\(^{56-58}\). Despite the combined training, we expected them to evaluate physical qualities such as aerobic capacity and muscular strength; however, all the articles only assessed aerobic capacity, and none of them evaluated muscle strength. For this reason, we could not analyse muscle strength as a criterion of sarcopenia.

Beside the conceptual importance of measuring muscle strength in studies that include resistance training, there are also methodological reasons. In general, muscle strength assessments are more straightforward and require less equipment than direct VO\(_{2\text{max}}\) or VO\(_{2\text{peak}}\) determination. Concerning the methodology to measure muscle strength, the handgrip strength (HGS) test is one of the most used...
evidence-based methods for evaluating muscle strength. HGS is a better predictor of adverse clinical outcomes than muscle mass. If HGS cannot be used, several alternative methodologies are available: knee flexion/extension, dynamometer, one maximum repetition (1RM), ten maximum repetitions (10RM), isokinetic evaluation and peak expiratory flow (specific to respiration).

One recent study evaluated the relationship between NAFLD status and HGS using a cohort declared as representative of the general Korean population. The results indicate that a higher HIS is associated with lower HGS in both sexes. Therefore, HGS, which is an assessment of general strength in sarcopenia and NAFLD, is quick and easy to perform and should be used as a base procedure for evaluating muscle strength in patients with liver pathologies.

Future research should include muscle strength assessments to investigate the effects of exercise, especially resistance training, in the sarcopenia of NAFLD patients.

**Exercise on physical performance**

Physical performance, one of the criteria for a sarcopenia diagnosis, is widely evaluated using the CPET, which provides a global assessment of integrative physiological responses. Results from the meta-analysis in our study indicate that most of the articles showed increased physical performance after the training. Only one study found no changes in VO$_{2peak}$ after exercise, which may be explained by the small sample size.

The overall analysis revealed low heterogeneity and an improvement in physical performance under exercise regardless of its features (type, intensity, duration). Three studies performed endurance training without another intervention, while one article performed combined resistance and endurance training. The endurance training parameters applied in these studies were consistent with the clinical practice guidelines for managing NAFLD from European associations for the study of the liver, diabetes and obesity (EASL, EASD, EASO); they describe a comprehensive lifestyle approach for treating these patients.

All the studies performed CPET and reported the VO$_{2max}$ or VO$_{2peak}$ as indicators of maximal aerobic capacity, which reflects the respiratory and circulatory systems’ ability to supply oxygen to skeletal muscles during exercise. They predicted an increase in oxygen consumption after endurance training, which induces adaptations in the cardiovascular system and increases mitochondrial biogenesis and capillary density in the skeletal muscles, improving the transport and use of oxygen to generate energy. Zenith et al. in 2014 observed increases in VO$_{2peak}$ in cirrhotic patients after eight weeks of endurance training. This antecedent is consistent with the fact that the primary adaptation to endurance training is cardiorespiratory fitness improvement. Therefore, our meta-analysis showed that planned and supervised physical exercise improved physical performance in patients with sarcopenia associated with NAFLD. It is essential to mention that these patients should perform regular exercise to preserve its beneficial effects. This recommendation is based on the substantial benefits exercise has on hepatic metabolism and the fast loss induced by endurance training. Thus, training is a therapeutic strategy for improving cardiorespiratory fitness in fatty liver disease. Furthermore, regular physical activity before liver transplantation is essential to combatting patients’ immediate stress post-transplant; it is also a critical determinant in long-term health after hepatic transplant.

CPET is the gold standard for assessing cardiorespiratory fitness and functional capacity. Unlike tests that estimate oxygen consumption, CPET measures respiratory, cardiovascular and neuromuscular system function. We excluded one study incorporated in our meta-analysis (in the muscle mass outcome) from the physical performance results, as oxygen consumption was determined from a 2 km walking test. In addition to the CPET assessment, the studies on patients with sarcopenia associated with NAFLD could include other validated muscle function tests for primary sarcopenia. For example, the short physical performance battery test (SPPB) evaluates gait, strength and balance through performance scoring system – it is an easy and quick test to implement. SPPB can assess the severity of sarcopenia
as well as predicts mortality on the liver transplant waitlist\textsuperscript{78,79}. Future studies should use the SPPB to gain additional information on muscle function in NALFD patients, as the European Association for the Study of the Liver (EASL) suggests\textsuperscript{80}.

**Effect of exercise on muscle mass**

As shown in Figure 3, exercise does not affect NAFLD patients’ LBM – a conclusion based on our analysis of four studies\textsuperscript{53,57-59}. These studies determined LBM using plethysmography and DXA. Both methods are widely used to assess body composition in research. The studies showed a large variability, but the variability across the studies was very consistent. The lack of significance in the overall effect may be because DXA determinations have low reproducibility, dependent upon the equipment used. The gold standard methods suggested for measuring body composition, especially for muscular mass, are CT and magnetic resonance imaging (MRI)\textsuperscript{18}.

Considering that all the studies in the meta-analysis mainly comprised endurance training protocols as the central exercise intervention, it is possible that endurance training does not affect NAFLD patients’ LBM. Our results are consistent with the findings of a systematic review in cirrhotic patients trained with endurance protocols, which reported no changes in LBM\textsuperscript{81}. Contrary to these results, a study in cirrhotic patients featuring resistance training showed an increase in muscle size\textsuperscript{82}. This suggests that the type of exercise training determines the effects on muscle mass, and that resistance training could have a positive impact on muscle mass in NAFLD patients.

**Exercise conditions to regulate sarcopenia**

Therapies to revert muscular mass loss in NAFLD patients include pharmacological and nutritional approaches. However, exercise-based treatments should not be excluded because they can improve the patients’ generalised physical condition and life quality\textsuperscript{8}. It is widely established that resistance training is the most promising method to increase muscle strength and mass as well as balance\textsuperscript{83}; further, it can induce muscular hypertrophy, translating to increased muscle mass\textsuperscript{83}. Based on these antecedents, resistance training should be indicated as a therapy, alone or in combination, for sarcopenia in NAFLD patients. It is necessary to consider the conditions of resistance training to use in NAFLD patients. The standard conditions suggest that a high load induces a higher hypertrophic response in skeletal muscle\textsuperscript{84}; however, loads below 30% of 1RM seem sufficient to trigger a hypertrophic response. The current recommendation for hypertrophic training is an intensity of 40–80% of the individual 1RM, with loads >60% to increase maximal force and muscular mass\textsuperscript{85}. Thus, the best conditions for resistance training in NAFLD patients require further study.

**Limitations**

As mentioned, the main limitations of the RCTs included here were difficulty in blinding participants, therapists and assessors and participants’ non-adherence to the assigned intervention regimes. Blinding participants and researchers is a significant challenge in exercise RCTs, as lifestyle interventions require patients and therapists to be fully aware of the type of intervention to be delivered. Nonetheless, reducing the risk of bias is crucial to blinding the outcome assessors, which must be clearly stated in the RCT report. As for non-adherence, researchers should include an intention to treat analyses in a pre-specified fashion, considering that attrition is a common feature in exercise RCTs.

Another limitation was the lack of supervision for not performing the physical activity in the control group. This monitoring is essential to ensuring a low level of physical activity that does not influence the results. If this variable is not controlled, then the sedentary controls could indeed be individuals who belong to a group that performs an unsupervised and undeclared exercise. Thus, limitation should be considered in future RCTs.
The lack of categorisation of NAFLD severity in the RCT participants is another major limitation of the studies included here. This categorisation is fundamental to determining if the effect of exercise on sarcopenia depends on the stage of the disease. Future RCTs should include this classification when planning and executing exercise protocols.

The exercise training programs were significantly different among the studies regarding exercise intensity, total duration, session frequency and modality (endurance or combined training). Among these, training modality had special consideration, since resistance training is known to be one of the most potent interventions for increasing muscle size and quality in healthy and ill subjects. This is caused by increased synthesis of myofibrillar and mitochondrial proteins in untrained individuals and to an increase in strength, muscle mass and effort tolerance. Nevertheless, none of the studies used resistance training to treat NAFLD patients. Therefore, we strongly recommend including resistance training as the main component of exercise protocols in future research on sarcopenia in NAFLD patients.

CONCLUSION
Exercise training is a useful tool to treat sarcopenia in NAFLD patients; it increases physical performance in the form of aerobic capacity. Nevertheless, it does not affect LBM. Future research should include muscle strength assessments and resistance training to evaluate the effects of exercise on sarcopenia in NAFLD patients.

WHAT IS ALREADY KNOWN?
- Sarcopenia negatively affects NAFLD progression.
- Previous RCTs and meta-analyses have reported beneficial effects from exercise on liver dysfunction in NAFLD patients.

WHAT ARE THE NEW FINDINGS?
- There are statistically significant effects of exercise interventions on sarcopenia in NAFLD patients.
- Exercise training has a positive effect on physical performance (aerobic capacity) but does not affect NAFLD patients’ LBM. We did not include muscle strength as an outcome.

Contributors
All authors comply the ICMJE Recommendations of authorship. A.G.R., C.C.V., F.S. and O.A.T. had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis, O.A.T. contributed with the study design, to statistical analysis and interpretation of the results, and C.H.S., F.S. and M.V.B contributed with the study design and writing of the manuscript. All authors approved the final version to be published.

Competing interests
None declared.

Funding
The manuscript was supported by research grants from the National Fund for Science and Technological Development (FONDECYT 1200944 [CCV], 1201039 [FS]), the Millennium Institute on Immunology and Immunotherapy (P09-016-F [CCV, FS]) and Basal Grant CEDENNA (AFB180001 [CCV]). The Millennium Nucleus of Ion Channels-Associated Diseases (MiNICAD) is a Millennium Nucleus supported by the Iniciativa Científica Milenio of the Ministry of Economy, Development and Tourism (Chile).

Acknowledgment
We acknowledge Orlando Flores Guerrero PhD., Faculty of Health and Behavioural Sciences, The University of Queensland, for his input while drafting the manuscript.
FIGURE LEGENDS

Figure 1. Flow chart of search results and study selection according to the PRISMA statement.

Figure 2. A meta-analysis of pooled effect size and confidence intervals (CIs) (95%) of the interventions with exercise versus control group on physical performance evaluated by the cardiopulmonary exercise test (CPET) (mL/Kg * min).

Figure 3. A meta-analysis of pooled effect size (ES) and confidence intervals (CIs) (95%) of the intervention with exercise versus control group on lean mass evaluated by dual-energy X-ray absorptiometry (DXA) and air displacement plethysmograph (Kg).

Supplementary Material 1. Details of the studies included in the systematic review and meta-analysis.

Supplementary Material 2. Search strategy in PubMed, CINAHL and Scopus databases.

Supplementary Material 3. Causes for exclusion for each excluded randomised controlled trial.

Supplementary Material 4. Methodological quality of the included studies in physical performance analysis. Methodological quality of the randomised controlled trials (n = 4) was assessed using the Cochrane risk of bias tool (six evaluation-critical methodological components).

Supplementary Material 5. Methodological quality of the included studies in lean body mass (LBM) analysis. Methodological quality of the randomised controlled trials (n = 4) was assessed using the Cochrane risk of bias tool (six evaluation-critical methodological components).
REFERENCES


71. Nolan PA


76. Dunn MA-Q, Rogal SA


Records identified through database searching (n=66)

- Pubmed (n=44)
- CINAHL (n=21)
- Scopus (n=1)

Duplicate articles (n=1)

Records screened (n=65)

Records excluded based on review of title and/or abstract (n=29)

Full text articles assessed for eligibility (n=36)

Criteria to exclude full text articles (n=29):
- Exercise combined with another intervention (n=15)
- Subjects under 18 years of age (n=2)
- Article in Spanish (n=1)
- Not RCT (n=5)
- Not evaluate at least 1 sarcopenia criteria (n=2)
- Lack clearly diagnosed NAFLD or that had comorbidities (n=4)

Studies included in final analysis (n=7)

Figure 1
<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>Exercise</th>
<th>Control</th>
<th>Mean Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Total</td>
</tr>
<tr>
<td>Sullivan 2012</td>
<td>24.8</td>
<td>5.2</td>
<td>12</td>
</tr>
<tr>
<td>Pugh 2013</td>
<td>36.5</td>
<td>15.2</td>
<td>6</td>
</tr>
<tr>
<td>Pugh 2014</td>
<td>33.4</td>
<td>12.2</td>
<td>13</td>
</tr>
<tr>
<td>Shojaee-Moradie 2016</td>
<td>33</td>
<td>5.8</td>
<td>15</td>
</tr>
<tr>
<td>Total (95% CI)</td>
<td>46</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Heterogeneity: Tau² = 0.00; Chi² = 2.45, df = 3 (P = 0.49); I² = 0%
Test for overall effect: Z = 5.42 (P < 0.00001)
<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>Exercise</th>
<th>Control</th>
<th>Mean Difference</th>
<th>Mean Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sullivan 2012</td>
<td>62 12.8</td>
<td>65 12</td>
<td>5.4%</td>
<td>-3.00 [-15.03, 9.03]</td>
</tr>
<tr>
<td>Hallsworth 2015</td>
<td>55.6 11.1</td>
<td>58.4 7.3</td>
<td>12.9%</td>
<td>-2.80 [-10.55, 4.95]</td>
</tr>
<tr>
<td>Houghton 2017</td>
<td>58 10</td>
<td>57 7</td>
<td>16.3%</td>
<td>1.00 [-5.91, 7.91]</td>
</tr>
<tr>
<td>Cheng 2017</td>
<td>42.5 6.6</td>
<td>40.4 6.8</td>
<td>65.4%</td>
<td>2.10 [-1.35, 5.55]</td>
</tr>
<tr>
<td>Total (95% CI)</td>
<td>64</td>
<td>59</td>
<td>100.0%</td>
<td>1.01 [-1.78, 3.80]</td>
</tr>
</tbody>
</table>

Heterogeneity: $\tau^2 = 0.00; \text{Chi}^2 = 1.74, \text{df} = 3 (P = 0.63); I^2 = 0\%$

Test for overall effect: $Z = 0.71 (P = 0.48)$

**Figure 3**