Supplementary Online Content

Durstenfeld MS, Sun K, Tahir P, et al. Use of cardiopulmonary exercise testing to evaluate long COVID-19 symptoms in adults: a systematic review and meta-analysis. *JAMA Netw Open*. 2022;5(10):e2236057. doi:10.1001/jamanetworkopen.2022.36057

This supplementary material has been provided by the authors to give readers additional information about their work.

eTable 1. Search Strategies for PubMed, Web of Science, and EMBASE

DATABASE	SEARCH STRATEGY
PubMed	("COVID-19"[Mesh] OR COVID OR "SARS-CoV-2"[Mesh] OR SARS-CoV-2) AND (“cardiopulmonary exercise test*” OR CPET OR CPX OR CPEX OR exercise capacity OR VO2 OR "Anaerobic Threshold"[Mesh] OR anaerobic threshold)
Web of Science	(COVID OR SARS-CoV-2) AND (“cardiopulmonary exercise test*” OR CPET OR CPX OR CPEX OR exercise capacity OR VO2 OR anaerobic threshold)
Embase	('coronavirus disease 2019'/exp OR 'coronavirus disease 2019') AND ('cardiopulmonary exercise test'/exp OR 'cardiopulmonary exercise test' OR 'cardiopulmonary exercise testing'/exp OR 'cardiopulmonary exercise testing' OR cpet OR cpx OR cpex OR 'exercise capacity'/exp OR 'exercise capacity' OR vo2 OR 'anaerobic threshold'/exp OR 'anaerobic threshold')

eTable 2. Quality Assessment and Potential Threats to Validity Among Studies Included in Comparison of Peak $\dot{V}O_2$ Among Those With and Without Symptoms >3 Months After SARS-CoV-2 Infection

First Author, Year	Study Participation	Study Attrition	CPET Protocol, Execution & Interpretation	Assessment of LC Symptoms	Confounding	Statistical Analysis & Reporting	Key Threats to Validity
Aparisi, et al,[51] 2021	**Moderate** Mostly hospitalized	**Moderate** 53/522 (10%) of hospitalized and few non-hospitalized	**Moderate** Treadmill ramp Low average RER	**Moderate** Used standardized non-COVID tools for dyspnea	**High** Not addressed	**High** No adjusted models	• Selection Bias • Low average RER suggests submaximal CPET • Confounding • Lack of interpretation of individual studies
Barbagelata et al,[41], 2021	**High** Retrospective EHR-based study without explanation for why individuals without LC underwent CPET	**Moderate** No information provided	**High** Treadmill with individualized Bruce/ modified Bruce High proportion low RER studies	**High** Defined as dyspnea or fatigue >45 days after symptom onset but ascertained through chart review	**Moderate** Adjust for gender, cardiovascular history, and use of beta blockers	**Moderate** Data-driven variable selection	• Retrospective EHR-based study without clarity regarding comparison group of people without LC—why CPETS were performed on 88 individuals "without LC" at exactly the same time after COVID diagnosis is not explained • High proportion of non-maximal studies
Brown, et al,[47] 2022	**Moderate** Only hospitalized	**Moderate** No information provided	**Low** Novel CPET-CMR protocol	**Moderate** Use of self-reported exercise capacity may not reflect LC	**Low** Use of restriction/ exclusion	**Moderate** No adjusted models, but well-matched	• Only included hospitalized individuals • Matched on key confounders, but no adjusted models
Durstenfeld, et al, [28] 2022	**Moderate** Mostly non-hospitalized convenience sample	**Moderate** Only 39/120 (33%) completed CPET although differences appear minimal	**Low** Cycle ergometer targeting 10 minute test, few stopped early, interpretation well-described	**Low** Defined as new symptoms consistent with WHO; sensitivity analyses performed	**Low** Did not assess pre-COVID fitness	**Low** Adjusted models with likely confounders	• Selection Bias • Attrition • Confounding by pre-COVID fitness

First Author, Year	Study Participation	Study Attrition	CPET Protocol, Execution & Interpretation	Assessment of LC Symptoms	Confounding	Statistical Analysis & Reporting	Key Threats to Validity
Ladlow et al,[34] 2022	**Low** Includes active duty military with appropriate controls	**Low** 113/150 (75)	**Low** Cycle ergometer targeting 10 minute test	**Moderate** Presence of one or more symptoms may be overly sensitive and not specific	**High** Stratification by severity of illness; did not account for BMI differences	**High** No adjusted models	• Even though all participants had prior exercise testing, these results are not reported or used to adjust for pre-COVID fitness • No adjustment for confounders (ie BMI)
Margalit et al,[40] 2022	**Moderate** Mostly non-hospitalized individuals attending COVID recovery clinic	**Moderate** Included 113/462 (24%) of those randomly sampled	**Moderate** Treadmill Low average RER No individual interpretation of studies	**Low** Well described assessment of LC fatigue, included sensitivity analyses	**Moderate** Extensive measurement of possible confounders, but unclear if incorporate into models	**Moderate** No description of variables included in models	• Selection bias: Most of the randomly sampled individuals from within the LC clinic were ineligible or did not agree to participate • Low average RER suggests submaximal CPET • Lack of description of statistical models
Schaeffer et al,[48,69] 2021	**Moderate** Only hospitalized	**Low** 49/91 (54%) completed CPET	**Low** 15 W/min cycle ergometer	**Low** Binary fatigue variable does not account for pre-COVID fatigue	**Moderate** Excluded comorbidities, but higher BMI in fatigue group	**High** No adjusted models, but sensitivity analysis with % predicted	• Selection bias (only hospitalized) • Did not account for confounders in analysis, but reported both absolute and percent predicted
Skjørten et al,[36] 2021	**Moderate** Only hospitalized	**Low** 156/236 (66%) completed "adequate" CPET and not excluded	**Moderate** Treadmill, modified Bruce Low average RER Wasserman algorithm	**Moderate** Use mMRC dyspnea scale 0 vs 1-4	**High** Excluded comorbidities, but higher BMI in dyspnea group	**High** Only adjust for age & sex	• Selection bias (only hospitalized) • Low average RER suggests submaximal CPET • Adjusted models only adjust for age and sex
Szekely et al,[32] 2021	**Moderate** Emergency department during acute infection and attended LC Clinic	**Low** 71/165 (43%); flowchart, but differences between those	**Low** Semi-supine cycle ergometer targeting 10 minute test	**Moderate** No description of how dyspnea & fatigue were assessed	**Moderate** Forced age & sex into models, but did not include BMI, severity, and	**High** Stepwise multivariable analysis left out confounders and	• Selection bias from only including those who sought care acutely and followed up in LC Clinic • Data-driven analysis left out important confounders (BMI, for example) and adjusted for likely

		assessed and not			other confounders	adjusted for mediators	mediators (stroke volume, TAPSE, HR, A-Vo_2 difference)

eTable 3. Quality Assessment and Potential Threats to Validity Among Studies Included in Assessment of Limitations of Exercise Capacity

First Author, Year	Reduced Def.	Reduced Peak VO_2, n (%)	Study Participation	Study Attrition	CPET Protocol, Execution & Interpretation	LC Symptoms	Confound-ing	Statistical Analysis & Reporting	Key Threats to Validity Pertinent to Classification
Abdallah/ Schaeffer et al,[48,69] 2021	<85%	41/63 (65)	**Moderate** Prospective cohort of only hospitalized	**Low** 49/91 (54) completed CPET	**Low** Cycle ergometer fixed protocol 15 W/min step	**Low** Binary fatigue variable does not account for pre-COVID fatigue	**Moderate** Excluded comorbidities, but higher BMI in fatigue group	**High** No adjusted models, but sensitivity analysis with % predicted	• Selection bias • Confounding (pre-existing medical comorbidities, beta blockers) • CPET interpretation not described
Alba et al,[45] 2021	<80%	6/18 (33)	**High** Retrospective cohort referred for CPET from LC Clinic	**High** Not reported	**Low** Upright cycle ergometer, excluded low RER	**Moderate** mMRC dyspnea scale	**High** Not addressed	**High** No adjusted models	• Small samples size • Selection bias • High proportion with preexisting cardiopulmonary disease
Ambrosino et al,[59] 2022	<20 ml/kg/min	28/36 (78)	**High** Pulmonary rehab after severe COVID-19, mostly on long-term oxygen	**Moderate** 36/112 (32)	**Low** Cycle ergometer, no low RER (or excluded)	**N/A**	**Low** Adjusted	**Low** Adjusted models include most confounders	• Selection bias: all severe COVID mostly still on oxygen • Unclear time after infection • Lack of interpretation of individual studies
Aparisi, et al,[51] 2021	NR		**Moderate** Prospective cohort mostly hospitalized	**Moderate** 53/522 (10) of hospitalized and few non-hospitalized	**Moderate** Treadmill ramp Low average RER	**Moderate** Used standardized non-COVID tools for dyspnea	**High** Not addressed	**High** No adjusted models	• Selection Bias • Low average RER suggests submaximal CPET • Confounding • Lack of interpretation of individual studies
Barbagelata et	<85%	39/112 (35)	**High** Retrospective EHR-	**Moderate** No	**High** Treadmill with individualized	**High** Dyspnea or fatigue >45	**Moderate** Adjust for gender,	**Moderate**	• High proportion of non-maximal studies (RER<1.1 for 47% of

al,[41], 2021			based study without explanation for why individuals without LC underwent CPET	information provided	Bruce/ modified Bruce High proportion low RER studies	days after symptom onset but ascertained through chart review	cardiovascular history, and use of beta blockers	Data-driven variable selection	studies and 49% did not reach anaerobic threshold) • High prevalence of cardiovascular disease and risk factors
First Author, Year	Reduced Def.	Reduced Peak VO_2, n (%)	Study Participation	Study Attrition	CPET Protocol, Execution & Interpretation	LC Symptoms	Confound-ing	Statistical Analysis & Reporting	• Key Threats to Validity Pertinent to Classification
Borrego Rodriguez et al,[70] 2021	<100 %	32/57 (56)	**Low** Non-hospitalized health care workers	**Moderate** Not reported	**Moderate** Details not reported	**Moderate** Dyspnea on exertion >3 months after infection	**High** Excluded structural heart disease	**High** No adjusted models	• Confounding • Use of unconventional <100% cutoff • Interpretation not described (abstract only)
Brown, et al,[47] 2022	Self-reported reduced exercise capacity	20/40 (50)	**Moderate** Prospective cohort of hospitalized without ICU stay, myocardial injury, or comorbidities	**Moderate** Not reported	**Low** Novel CPET-CMR protocol using supine cycle ergometer	**Moderate** Use of self-reported exercise capacity may not reflect LC	**Low** Use of restriction/ exclusion	**Moderate** No adjusted models, but well-matched	• Only included hospitalized individuals
Cassar et al,[29,71] 2021	<80%	6/31 (19)	**Moderate** Prospective cohort after COVID hospitalization	**Low** 46/58 (79) retained	**Moderate** Cycle ergometer 10W/min ramp, 26% submaximal tests	**Low** Use validated scales and longitudinal symptom assessment	**Low** Group matched controls	**Moderate** Details of adjusted analyses are not provided	• Only included hospitalized individuals • Confounding
Clavario et al,[27] 2021	<85%	99/200 (50)	**Moderate** Prospective cohort after COVID	**Low** 200/225 (89)	**Low** Cycle ergometer targeting 10 minute test	**N/A**	**High** Included patients with HF,	**Moderate** Data-driven variable selection	• Only included hospitalized individuals • Confounding

			hospitalization		Independent interpretation by 2 reviewers		COPD, MI		
First Author, Year	Reduced Def.	Reduced Peak VO_2, n (%)	Study Participation	Study Attrition	CPET Protocol, Execution & Interpretation	LC Symptoms	Confound-ing	Statistical Analysis & Reporting	Key Threats to Validity Pertinent to Classification
de Boer et al,[43] 2021	<84%	16/50 (32)	**High** Retrospective case series of clinically referred CPETs for PASC	**High** Not reported	**Low** Cycle ergometer ramp	**N/A**	**Moderate** Address through stratification	**High** No adjusted models	• Focus on compromised mitochondrial function estimated from stoichiometric equations • Selection bias
Debeaumont et al,[33]2021	<85%	12/23 (52)	**High** Retrospective case series of hospitalized COVID patients referred for CPET	**High** Not reported	**Low** Cycle ergometer customized to target	**Low** Use mMRC scale for dyspnea	**High** Not addressed	**High** No adjusted models	• Only included hospitalized individuals subsequently referred for CPET
Dorelli et al, [52,53]2021	NR	NR	**High** Prospective cohort post-hospitalization <65 years old without comorbidities	**Moderate** 28/130 (22)	**Low** Cycle ergometer targeting 8-12 minute test	**N/A**	**Moderate** Restricted patients with comorbidities including obesity	**High** Unclear why authors want to use models to predict ventilatory inefficiency and no justification for variables considered	• Primary comparison is between those with and without exercise ventilatory inefficiency • Lack of interpretation of individual studies
Durstenfeld, et al, [28] 2022	<85%	15/39 (38)	**Moderate** Prospective cohort mostly non-hospitalized	**Moderate** Only 39/120 (33%) completed	**Low** Cycle ergometer targeting 10 minute test,	**Low** Defined as new symptoms consistent	**Low** Did not assess pre-	**Low** Adjusted models with likely confounders	• Selection Bias • Confounding by pre-COVID fitness

			convenience sample without cardiovascular disease	CPET although differences appear minimal	few studies stopped early, interpretation well-described	with WHO; sensitivity analyses performed	COVID fitness		
First Author, Year	Reduced Def.	Reduced Peak VO_2, n (%)	Study Participation	Study Attrition	CPET Protocol, Execution & Interpretation	LC Symptoms	Confound-ing	Statistical Analysis & Reporting	Key Threats to Validity Pertinent to Classification
Evers et al,[58] 2022,	<100 % predicted work	11/30 (37)	**High** Retrospective case series of patients referred for post-COVID exercise limitation or dyspnea	NR 16/30 (53) underwent repeat CPET	**Low** Cycle ergometer targeting <12 minute test	**Low** mMRC dyspnea scale	**High** Not addressed	**High** No adjusted models	• Selection bias
Frésard, et al,[54] 2022	>84% NR		**High** Retrospective cohort of clinical CPETs referred for LC and persistent dyspnea	**High** Not reported	**Low** Cycle ergometer target 10 minute test	**Moderate** Use validated scales from non-COVID settings	**High** Not addressed	**High** No adjusted models	• Primary comparison is dysfunctional breathing (mostly mild-moderate COVID) compared to ventilatory limitation (mostly severe COVID)
Godinho et al,[72] 2021	NR	5/10 (50)	**High** Case series of non-hospitalized patients with persistent exercise limitations referred for clinical CPET	**High** Not reported	**High** No information provided	**N/A**	**High** Not addressed	**High** No adjusted models	• Very small case series with lack of adequate details to assess quality from abstract and no preprint or manuscript available

Jahn et al,[73] 2021	<83%	19/35 (54%)	**Moderate** Case series of patients with severe COVID pneumonitis attending post-hospitalization pulmonary rehab	**Low** 35/44 (80)	**Low** Semi-recumbent cycle ergometer, interpretation described	**Moderate** Use validated scales from non-COVID settings, but 60% missing	**High** Not addressed, did not exclude prior disease	**High** No adjusted models	• Selection bias (severe COVID only) • Confounding
First Author, Year	Reduced Def.	Reduced Peak VO_2, n (%)	Study Participation	Study Attrition	CPET Protocol, Execution & Interpretation	LC Symptoms	Confound-ing	Statistical Analysis & Reporting	Key Threats to Validity Pertinent to Classification
Johnsen et al, [74] 2021	<84%	16/31 (52)	**High** Case-series of post COVID clinic referrals for CPET to evaluate symptoms	**High** 34/117 (29) + 23 outpatients, but unclear which 31 were included for CPET	**High** Minimal information provided	**Moderate** Detailed clinical phenotyping but not described for those who underwent CPET	**High** Not addressed for CPET	**High** Adjusted models for symptom variables for age and sex, but not for CPET	• Focus of paper is clinically phenotyping LC; does not provide adequate detail about CPET
Kersten et al,[57] 2021	NR	17/35 (55)	**High** Case-series of post COVID clinic referrals for CPET if initial testing abnormal or not revealing	**High** 36/231 (16) targeted for symptomatic	**High** Treadmill ramp, interpretation strategy not described and only summary CPET findings reported	**Moderate** Minimal information provided	**High** Not addressed	**High** Descriptive only	• Selection bias • High attrition • Those who underwent CPET are not well described • CPET data are not reported, only categorization of reason for limitation
Ladlow et al,[34] 2022	<85%	4/61 (7)	**Low** Prospective cohort of active-duty military	**Low** 113/150 (75)	**Low** Cycle ergometer targeting 10 minute test	**Moderate** Presence of one or more symptoms	**High** Stratification by severity of illness; did	**High** No adjusted models	• Selection bias: active military personnel • No description of interpretation

			personnel with appropriate controls			may be overly sensitive and not specific	not account for sex, age, BMI		
Liu et al,[56] 2021	NR	NR	**Moderate** Prospective post-hospitalization cohort	**High** Not reported	**Moderate** Treadmill, interpretation not described or reported	**N/A**	**High** Not addressed	**High** Adjusted models to predict pulmonary fibrosis at 7 months, but model development strategy not described	• Focus of paper is pulmonary fibrosis at 7 months; does not provide adequate detail about CPET findings or interpretation or classify participants by symptoms
First Author, Year	Reduced Def.	Reduced Peak VO_2, n (%)	Study Participation	Study Attrition	CPET Protocol, Execution & Interpretation	LC Symptoms	Confound-ing	Statistical Analysis & Reporting	Key Threats to Validity Pertinent to Classification
Mancini et al,[46] 2021	<80%	24/41 (59)	**High** Case-series of LC clinic referrals for CPET for dyspnea with normal cardiopulmonary testing	**High** Not reported	**Low** Cycle ergometer 25 W/3 minute step, subset with invasive ("hemodynamic") CPET; classification well described	**Low** Interview for ME/CFS symptoms	**Moderate** Used % predicted; excluded known cardiopulmonary disease; high proportion on beta blockers, not held; other confounders not addressed	**High** No adjusted models	• Selection bias (LC Clinic referrals) • Confounding (ie beta blocker use)
Margalit et	NR		**Moderate** Nested case-control	**Moderate** 113/462 (24)	**Moderate** Treadmill with low average	**Low** Well described	**Low** Extensive measurem	**Moderate** No description	• Selection bias: All sampled individuals were from LC Clinic; most of

al,[40] 2022			study within COVID recovery cohort of mostly non-hospitalized individuals attending COVID recovery clinic	randomly sampled	RER; no individual interpretation	assessment of LC fatigue, included sensitivity analyses	ent of possible confounders, and well-balanced groups	of variables included in models	the randomly sampled individuals from within the LC clinic were ineligible or did not agree to participate • Low average RER suggests submaximal CPET
First Author, Year	*Reduced Def.*	*Reduced Peak VO_2, n (%)*	*Study Participation*	*Study Attrition*	*CPET Protocol, Execution & Interpretation*	*LC Symptoms*	*Confound-ing*	*Statistical Analysis & Reporting*	*Key Threats to Validity Pertinent to Classification*
Mohr et al,[44] 2021	<85%	8/10 (80)	**High** Retrospective case-series of post COVID clinic referrals for CPET for dyspnea	**High** 10/42 (24)	**High** CPET methods and interpretation not described	**High** Not described	**High** Not addressed	**High** Descriptive only; no adjusted models	• Small sample size • Selection bias • Inadequate description of CPET methods and interpretation • Heterogeneity within sample without addressing likely confounders
Motiejunaite et al,[50]2021	<85%	86/114 (75)	**High** Prospective cohort but target population and recruitment not well-described	**High** Not reported	**Moderate** Cycle ergometer, interpretation well described	**High** Symptom assessment not described	**High** Not addressed	**High** Compared reduced to preserved diffusing capacity; no adjusted models	• Analytic focus is comparing those with reduced vs. preserved diffusing capacity
Moulson et al,[61] 2022	<80%	3/21 (14)	**High** Case-series of young athletes referred for cardiopulmonary	**Moderate** 13/21 (62) retained	**Moderate** Treadmill or cycle ergometer, protocols & interpretation well described	**Low** Interview for symptoms	**Low** Only included young athletes without comorbidities and	**High** Descriptive only; no adjusted models for symptoms	• Selection bias: only included symptomatic athletes • Attrition for longitudinal CPET

			symptoms after COVID				compared to similar reference group of athletes		
Parkes et al,[75]2021	<85%	10/12 (83)	**High** Retrospective cohort of clinical CPETs	**High** 12/600 (2)	**High** Not described; sub-max tests are hinted at	**High** Not described	**High** Not addressed	**High** No adjusted models	• Small sample size • Selection bias • Inadequate description of CPET methods and interpretation
Plegue zuelos, et al,[60] 2021	NR		**High** Case series of survivors of ARDS from COVID pneumonia requiring mechanical ventilation & tracheostomy	**High** Not reported	**Low** Cycle ergometer targeting 6-12 minute test	**High** Not described	**High** Not addressed, but compared to multiple reference groups	**High** No adjusted models	• Focus is comparing mechanical efficiency among those with severe COVID to those with COPD, ischemic heart disease, and healthy controls • Selection bias: only included patients requiring prolonged ICU care • Confounding
First Author, Year	Reduced Def.	Reduced Peak VO_2, n (%)	Study Participation	Study Attrition	CPET Protocol, Execution & Interpretation	LC Symptoms	Confound-ing	Statistical Analysis & Reporting	Key Threats to Validity Pertinent to Classification
Ribeiro Baptista, et al,[39] 2022	<80%	37/105 (35)	**Moderate** Prospective cohort of severe COVID requiring hospitalization >7 days and oxygen (43% ICU)	**Moderate** 105/220 (48)	**Moderate** Cycle ergometer 10-20 W/min; interpretation not described	**Moderate** mMRC dyspnea scale	**High** Not addressed	**Moderate** Stepwise backward selection for models to assess associations with reduced VO_2	• Selection bias: only included patients with severe COVID • Confounding
Rinaldo, et al[38,76] 2021	<85%	41/75 (55)	**Moderate** Prospective cohort post-	**High** Not reported	**Moderate** Cycle ergometer with	**Moderate** mMRC dyspnea scale	**High** Not addressed	**High** No adjusted models	• Selection bias • Overly simplistic interpretation of abnormal studies

			hospitalization		individualized protocol, but classification by breathing reserve and heart rate reserve is overly simplistic				• Confounding not addressed
Singh et al,[42] 2021	<80%	NR	**High** Prospective cohort of patients referred for CPET from LC Clinic for unexplained exercise intolerance with negative initial workup	**High** Not reported	**Moderate** Invasive CPET including pulmonary artery and radial artery lines with cycle ergometer with individualized protocol, but tests terminated at RER>1.1 or HR>85% predicted	**N/A**	**High** Matching by age and sex but not other potential confounders (ie BMI higher in COVID than controls)	**High** No adjusted models	• Selection bias • Termination of exercise based on submaximal criteria
First Author, Year	Reduced Def.	Reduced Peak VO_2, n (%)	Study Participation	Study Attrition	CPET Protocol, Execution & Interpretation	LC Symptoms	Confound-ing	Statistical Analysis & Reporting	Key Threats to Validity Pertinent to Classification
Skjørten et al,[36] 2021,	<80%	49/156 (31)	**Moderate** Multicenter prospective cohort only hospitalized	**Low** 156/236 (66%) completed “adequate” CPET and were not excluded for	**Moderate** Treadmill, modified Bruce Low average RER Wasserman algorithm	**Moderate** Use mMRC dyspnea scale 0 vs 1-4	**High** Excluded comorbidities, but higher BMI in dyspnea group	**High** Only adjust for age & sex	• Low average RER suggests submaximal CPET

				comorbidity					
Szekely et al,[32] 2021	<85%	49/71 (69)	**Moderate** Prospective cohort of individuals who went to emergency department for acute COVID-19 and attended LC Clinic	**Low** 71/165 (43%) with clear flowchart, but with some differences between those assessed and those not assessed	**Low** Semi-supine cycle ergometer targeting 10 minute test	**Moderate** No description of how dyspnea & fatigue were assessed	**Moderate** Forced age & sex into models, but did not include BMI, severity, and other confounders	**High** Stepwise multivariable analysis left out confounders and adjusted for mediators	• Selection bias • Interpretation of individual studies not described
First Author, Year	Reduced Def.	Reduced Peak VO_2, n (%)	Study Participation	Study Attrition	CPET Protocol, Execution & Interpretation	LC Symptoms	Confound-ing	Statistical Analysis & Reporting	Key Threats to Validity Pertinent to Classification
Vannini et a;,[77] 2021	<80%	19/41 (46)	**Moderate** Prospective cohort post-hospitalization	**High** Not reported	**Moderate** Cycle ergometer 10W/min ramp; interpretation not described	**N/A**	**High** Stratification by severity of acute COVID; other confounders not addressed	**High** No adjusted models	• Selection Bias • Confounding not adequately addressed • CPET interpretation not described
von Gruenewaldt et al,[55] 2022	<80%	2/20 (10)	**High** Retrospective cohort of clinical CPETs	**High** Not reported	**Moderate** Cycle ergometer 10 or 20W/min ramp targeting 10 minute test; interpretation focused on dysfunctional breathing	**High** Symptoms assessed through records; participants without PCR verified diagnosis	**High** Not addressed	**High** No adjusted models	• Small sample size • A priori focus is abnormal breathing pattern • Unclear if other interpretations were considered

Vonbank et al,[37] 2021	NR		**Low** Prospective cohort including full spectrum of acute SARS-CoV-2 infection	**High** Not reported	**Moderate** Cycle ergometer targeting 8-12 minute test; no interpretation of individual studies	**N/A**	**Moderate** Addressed through adjusted model, but not all included	**High** Stepwise multivariable analysis left out confounders and adjusted for mediators	• Focus is comparing exercise capacity by severity of acute illness to healthy controls • Interpretation not described

eMethods. Study Protocol

The full, pre-registered Protocol is available at https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42021299842.
This review was conducted according to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis)[25] guidelines and was registered prospectively on PROSPERO prior to beginning the search.

Condition being studied: post-acute sequelae of COVID-19, also known as Long COVID, which according to the WHO definition is >3 months after acute infection with SARS-CoV-2.

Inclusion criteria: all studies of adults with confirmed COVID-19 at least 3 months after onset that include cardiopulmonary exercise testing with measurement of peak VO_2 published since 2020 will be included. Baseline cardiopulmonary exercise testing from interventional or randomized controlled trials will also be included if they meet the other inclusion criteria.

Exclusion criteria: studies of children, studies of conditions other than COVID-19/SARS-CoV-2, studies in the acute or early post-acute phase (<3 months after infection), review articles, case reports.

Intervention/exposure: Cardiopulmonary exercise testing, which includes measurement of metabolic gases with either treadmill or cycle ergometer exercise.

Participants/population: We are interested in all adults with COVID without respect to hospitalization status or severity of acute illness.

Inclusion/Exclusion criteria: adults with confirmed COVID-19 at least 3 months after onset that include cardiopulmonary exercise testing with measurement of peak VO_2 will be included. We excluded studies of children, studies of conditions other than COVID-19/SARS-CoV-2, studies in the acute phase (<3 months after infection).

Comparators/control: We will include case series without controls, as well as studies with healthy controls, control participants with unexplained dyspnea, or that compare those who have fully recovered from COVID compared to those reporting ongoing symptoms.

Types of studies to be included: We will include observational studies including case series, cross-sectional studies, case-control studies, and cohort studies. We will also include randomized trials of interventions, in which case we will use baseline CPET data. We will exclude case reports and review articles.

Context: We will include studies that include the full spectrum of COVID-19; specifically, we will not restrict to only studies of those requiring ICU or hospitalization during acute infection.

Main Outcomes: The primary outcome will be peak VO_2 (in ml/kg/min and % predicted). If meta-analysis is possible, studies that do not include this measure will be excluded from meta-analysis. We will report the difference in peak VO_2 between those with and without COVID and among those with COVID between those with and without post-acute sequelae.

Additional outcomes: Additional outcomes will include the proportion with exercise limitation <80 or 85% of predicted (different studies use different cutoffs), difference in exercise capacity between those with and without cardiopulmonary symptoms (absolute and relative difference with 95% confidence intervals and p value), common features among those with limitations (i.e., reduced oxygen pulse pressure, reduced chronotropic response). We will likely report these effect measures in odds-ratios as we expect that many of the studies may be case-control studies.

Search Strategy & Information Sources: A comprehensive, electronic search strategy will be used to identify studies published since 2020 and indexed in PubMed, EMBASE, and Web of Science by a research librarian (PT) with extensive experience in systematic reviews. Unpublished abstracts from conference proceedings and indexed preprints will be included as part of our gray literature search. We will also review references from studies selected for data extraction. The search strategy will include terms and synonyms for the following: (COVID or SARS-CoV-

2) AND ("cardiopulmonary exercise test*" OR (CPET or CPX or CPEX) OR exercise capacity OR VO2 OR anaerobic threshold). Searches will be tailored to each database depending on indexing terminology. Searches were conducted on December 20, 2021, and rerun prior to the final analysis on May 24, 2022; pre-prints were searched through June 9, 2022. Abstracts were reviewed for inclusion by two independent reviewers (MSD & KS); if there is disagreement after consensus discussion, a third reviewer will be consulted. All data extraction was done independently, in duplicate, using REDCap for data entry.

Gray literature plan: see search strategy for details; we will review conference abstracts, pre-prints, and references from studies that meet the inclusion criteria.

Data Extraction (Selection & Coding): Data including authors, title, date of study, location of study, sample size (including total with COVID, total with Cardiopulmonary Long COVID, and COVID-negative controls, if included), median time since acute infection and interquartile range, inclusion criteria (with particular attention to inclusion of hospitalized/ICU/ambulatory during acute illness and those with specific comorbidities or populations of interest), comparator group, exercise modality (treadmill or cycle ergometer), peak VO2 (in ml/kg/min and % predicted), proportion with exercise limitation <85% of predicted, difference in exercise capacity between those with and without cardiopulmonary symptoms (absolute and relative difference with 95% confidence intervals and p value), common features among those with limitations (i.e., reduced oxygen pulse pressure, reduced chronotropic response). If available, other cardiopulmonary parameters will be recorded including echocardiographic, pulmonary function tests, chest computed tomography, and cardiac magnetic resonance imaging.

Data Management: Studies identified through the searches will be managed using Covidence. Data extracted will be recorded using REDCap.

Quality Assessment: We will use Cochrane's Quality in Prognostic Studies (QUIPS) tool to assess for bias of included studies. We will assess study populations (especially choice of control groups), study attrition for non-cross-sectional studies, peak VO_2 assessment quality, outcome measurement, study confounding, and statistical analysis and reporting. We will use Cochrane's Quality in Prognostic Studies (QUIPS) tool to assess for bias of included studies.

Data synthesis: Overall findings of each study will be summarized in a table. If possible, a meta-analysis will be performed to compare the peak VO2 among those with and without COVID. An odds ratio of having reduced exercise capacity may also be estimated if possible. Heterogeneity will be assessed using I^2. The primary subgroup we plan to investigate is to compare peak VO2 (and the other explanatory variables for reduced exercise capacity) among those with and without PASC/Long COVID. If possible, we may also compare those with severe acute infection requiring hospitalization and/or ICU care with those who were asymptomatic or had mild acute infection. Lastly, we may compare the early post-acute period (3-6 months), medium term (6-12 months), and long term (>12 months). Analyses will be performed using STATA version 17.

Analysis of subgroups: The primary subgroup we plan to investigate is to compare peak VO2 (and the other explanatory variables for reduced exercise capacity) among those with and without PASC/Long COVID. If possible, we may also compare those with severe acute infection requiring hospitalization and/or ICU care with those who were asymptomatic or had mild acute infection. Lastly, we may compare the early post-acute period (3-6 months), medium term (6-12 months), and long term (>12 months).

Risk of Bias/Quality Assessment: Risk of bias will be assessed at both the study and the outcome level for each included study. Publication bias will be assessed using a Funnel Plot. The strength of the body of evidence will be assessed using the GRADE (Grading of Recommendations, Assessment, Development, and Evaluations) framework.

eAppendix. Study Findings and Quality Form

Record ID	_____
Extractor	○ Matt ○ Kevin
First author's last name	_____
all authors	_____
Title	_____
year	○ 2020 ○ 2021 ○ 2022
Journal	_____
Type of Publication	○ Full Manuscript ○ Research Letter ○ Abstract/Conference Proceedings ○ Non-English Full Manuscript ○ Review ○ Other (comments)
Duplicate with (enter other 1st author/Journal)	_____
study type	○ prospective cohort (research CPETs) ○ retrospective cohort ○ case-control ○ case-series (ie patients referred for clinically ordered CPETs) ○ other
study type comments	_____

Study Location (City/State/Country/Region)	_____
Start date of study	_____
End date of study	_____
total sample size who underwent CPET	_____
Median time since acute infection (days) -if reported in months multiply by 30 -if reported in weeks multiply by 7	_____
Interquartile Range time since acute infection (days) -if reported in months multiply by 30 -if reported in weeks multiply by 7	_____
If median/IQR time since infection not reported, then put mean and standard deviation here	_____
Inclusion Criteria	
Inclusion Criteriano testing required How was COVID diagnosed?PCR confirmed acute infection	◯ ◯ ◯ ◯ antibody testing other
Inclusion Criteria: age	_____
Mean or median age	_____
age standard deviation	_____
number female	_____

female %	_____
Inclusion hospitalization	○ Included patients irrespective of hospitalization ○ Included only patients hospitalized for acute disease ○ INcluded only ○ Other (note in comments)
Inclusion ICU	○ Included patients irrespective of ICU admission ○ Included only patients admitted to ICU for acute disease ○ Other (note in comments)
Inclusion Criteria Are only athletes	○ No being an athlete is not required ○ Yes only athletes
Inclusion Comorbidities/Special	○ No specific comorbidities required for entry ○ Specific comorbidities required (ie heart failure)
Inclusion Comorbidities required for inclusion	
Primary comparison	_____
Sample Size of control group WITHOUT COVID	_____
Peak VO2 (ml/kg/min) among controls WITHOUT COVID	_____
Peak VO2 (% predicted) among controls WITHOUT COVID	_____
Among those without COVID, proportion with exercise limitation (0 to 1.00)	_____
Sample Size who had COVID	_____

Peak VO2 (ml/kg/min) among all WITH COVID	_____
Peak VO2 (% predicted) among all WITH COVID	_____
Among those WITH COVID, proportion with exercise limitation (0 to 1.00)	_____
Sample Size with COVID but without PASC/Long COVID	_____
Peak VO2 (ml/kg/min) WITH COVID but without PASC	_____
Peak VO2 (% predicted) WITH COVID but WITHOUT PASC	_____
Sample Size with PASC/Long COVID	_____
Peak VO2 (ml/kg/min) among those WITH PASC/Long COVID	_____
Peak VO2 (% predicted) among those WITH PASC/Long COVID	_____
Number with reduced exercise capacity	_____
Among those WITH PASC/LONG COVID, proportion with exercise limitation (0 to 1.00)	_____
Definition of Exercise Limitation	

Exericse modality cycle ergometer	_____ ◯ treadmill ◯ other (list in comments)
Difference in peak VO2 (ml/kg/min) Cardiopulmonary PASC vs no PASC	_____
Confidence interval of Difference in peak VO2 (ml/kg/min) Cardiopulmonary PASC vs no PASC	_____
Difference in peak VO2 (% predicted) Cardiopulmonary PASC vs no PASC	_____
Confidence interval of Difference in peak VO2 (% predicted) Cardiopulmonary PASC vs no PASC	_____
Relative exercise capacity (RR) among those with PASC vs no PASC	_____
Confidence interval of relative exercise capacity Cardiopulmonary PASC vs no PASC	_____
Primary etiology of reduced exercise capacity in PASCNo	◯ primary etiology Deconditioning ◯ Ventilatory Limitation ◯ Cardiac Limitation ◯ Chronotropic ◯ Multifactorial ◯ Other ◯ Peripheral
Proportion with PASC with deconditioning	_____
Proportion with PASC with ventilatory limitation	_____
Proportion with PASC with cardiac limitation	_____

Proportion with PASC with peripheral limitation (oxygen extraction/utilization)	_____
Proportion with PASC with chronotropic incompetence	_____
Other reason for limitation reported	_____
Proportion with PASC with Other Limitation	_____
Other objective data available	☐ None ☐ Rest echo ☐ Stress echo ☐ Chest CT ☐ CMR ☐ Inflammatory markers ☐ cardiac biomarkers ☐ right heart cath ☐ PFTs ☐ 1st pass ventriculography ☐ lactate/arterial blood gas
Primary analytic comparison reported	_____

Study Quality Assessment
1. Study Participation
Goal: To judge the risk of selection bias (likelihood that relationship between PF and outcome is different for participants and eligible non-participants).

Source of target population The source population or population of interest is adequately described for key characteristics (age, sex, hosp/ICU, time since COVID, special populations comorbidities, precovid fitness)	○ No ○ Partial ○ Yes ○ Unsure
Method used to identify population: The sampling frame and recruitment are adequately described, including methods to identify the sample sufficient to limit potential bias (number and type used, e.g., referral patterns in health care)	○ No ○ Partial ○ Yes ○ Unsure
Period of recruitment is adequately described	○ No ○ Partial ○ Yes ○ Unsure
Place of recruitment (setting and geographic location) are adequately described	○ No ○ Partial ○ Yes ○ Unsure
Inclusion and exclusion criteria are adequately described (e.g., including explicit diagnostic criteria or "zero time" description).	○ No ○ Partial ○ Yes ○ Unsure
Adequate study participation	○ No ○ Partial ○ Yes ○ Unsure
The baseline study sample (i.e., individuals entering the study) is adequately described for key characteristics: (age, sex, hosp/ICU, time since COVID, special populations comorbidities, precovid fitness).	○ No ○ Partial ○ Yes ○ Unsure

Overall comments on study populations _____

Overall comments on control groups? _____

Summary Study Participation The study sample represents the population of interest on key characteristics, sufficient to limit potential bias of the observed relationship between PF and outcome	○ Low risk ○ Moderate risk ○ High risk

2. Study Attrition

Goal: To judge the risk of attrition bias (likelihood that relationship between PF and outcome are different for completing and non-completing participants).

Proportion of baseline sample available for analysis Response rate (i.e., proportion of study sample completing the study and providing outcome data) is adequate.	○ No ○ Partial ○ Yes ○ Unsure
number eligible	_____
number included	_____
proportion retained	_____
Attempts to collect information on participants who dropped out of the study are described.	○ No ○ Partial ○ Yes ○ Unsure
Reasons for loss to follow-up are provided.	○ No ○ Partial ○ Yes ○ Unsure
Participants lost to follow-up are adequately described for key characteristics (hosp, symptoms, etc)	○ No ○ Partial ○ Yes ○ Unsure
There are no important differences between key characteristics (age, sex, time since COVID, severity of acute illness, comorbidities, persistent symptoms) and outcomes in participants who completed the study and those who did not.	○ No ○ Partial ○ Yes ○ Unsure
Study Attrition Summary Loss to follow-up (from baseline sample to study population analyzed) is not associated with key characteristics (i.e., the study data adequately represent the sample) sufficient to limit potential bias to the observed relationship between PF and outcome.	○ Low Risk ○ Moderate Risk ○ High Risk
Overall comments on Study Attrition	_____

3. Prognostic Factor Measurement **Goal: To judge the risk of measurement bias related to how PF was measured (differential measurement of PF related to the level of outcome).**	
Definition of the PF (CPET) A clear definition or description of CPET is provided (e.g., including exercise modality & protocol, stopping criteria, assessment of submaximal tests (ie RER, Borg, HR, double product), and clear specification of the method of measurement and classification of limitations)	◯ No ◯ Partial ◯ Yes ◯ Unsure
Valid and Reliable Measurement of PF (CPET) Method of PF measurement is adequately valid and reliable to limit misclassification bias (e.g., may include relevant outside sources of information on measurement properties, also characteristics, such as blind measurement and limited reliance on recall) Especially how tests are interpreted, how anaerobic threshold is identified	◯ No ◯ Partial ◯ Yes ◯ Unsure
Continuous variables are reported or appropriate cut-points (i.e., not data-dependent) are used.	◯ No ◯ Partial ◯ Yes ◯ Unsure
The method and setting of measurement of PF is the same for all study participants	◯ No ◯ Partial ◯ Yes ◯ Unsure
Adequate proportion of the study sample has complete data for PF variable.	◯ No ◯ Partial ◯ Yes ◯ Unsure
Appropriate methods of imputation are used for missing 'PF' data.	◯ No ◯ Partial ◯ Yes ◯ Unsure
PF (CPET) Measurement Summary PF is adequately measured in study participants to sufficiently limit potential bias. Overall comments on CPET quality	◯ Low risk ◯ Moderate risk ◯ High risk _____

4. Outcome Measurement

Goal: To judge the risk of bias related to the measurement of outcome (differential measurement of outcome related to the baseline level of PF).

A clear definition of outcome (PASC/Long COVID/Symptoms) is provided, including duration of follow-up and level and extent of the outcome construct	○ No ○ Partial ○ Yes ○ Unsure
Valid and Reliable Measurement of Outcome The method of outcome measurement used is adequately valid and reliable to limit misclassification bias (e.g., may include relevant outside sources of information on measurement properties, also characteristics, such as blind measurement and confirmation of outcome with valid and reliable test).	○ No ○ Partial ○ Yes ○ Unsure
Method and Setting of Outcome Measurement The method and setting of outcome measurement is the same for all study participants.	○ No ○ Partial ○ Yes ○ Unsure
Outcome (Symptoms) Measurement Summary Outcome of interest is adequately measured in study participants to sufficiently limit potential bias.	○ Low risk ○ Moderate risk ○ High risk
Overall comments on assessment of PASC/Long COVID/symptoms	_____

5. Study Confounding

Goal: To judge the risk of bias due to confounding (i.e. the effect of PF is distorted by another factor that is related to PF and outcome).

Important Confounders Measured All important confounders, including treatments (key variables in conceptual model: LIST), are measured.	○ No ○ Partial ○ Yes ○ Unsure
Definition of the confounding factor Clear definitions of the important confounders measured are provided (e.g., including dose, level, and duration of exposures).	○ No ○ Partial ○ Yes ○ Unsure
Valid and Reliable Measurement of Confounders Measurement of all important confounders is adequately valid and reliable (e.g., may include relevant outside sources of information on measurement properties, also characteristics, such as blind measurement and limited reliance on recall).	○ No ○ Partial ○ Yes ○ Unsure

Method and Setting of Confounding Measurement The method and setting of confounding measurement are the same for all study participants.	○ No ○ Partial ○ Yes ○ Unsure
Method used for missing data Appropriate methods are used if imputation is used for missing confounder data.	○ No ○ Partial ○ Yes ○ Unsure
Appropriate Accounting for Confounding Important potential confounders are accounted for in the study design (e.g., matching for key variables, stratification, or initial assembly of comparable groups).	○ No ○ Partial ○ Yes ○ Unsure
Important potential confounders are accounted for in the analysis (i.e., appropriate adjustment).	○ No ○ Partial ○ Yes ○ Unsure
Study Confounding Summary Important potential confounders are appropriately accounted for, limiting potential bias with respect to the relationship between PF and outcome .	○ Low risk ○ Moderate risk ○ High risk

Overall comments for counfounding

_____

6. Statistical Analysis and Reporting Goal: To judge the risk of bias related to the statisticalanalysis and presentation of results	
Presentation of analytical strategy There is sufficient presentation of data to assess the adequacy of the analysis	○ No ○ Partial ○ Yes ○ Unsure
Model development strategy The strategy for model building (i.e., inclusion of variables in the statistical model) is appropriate and is based on a conceptual framework or model.	○ No ○ Partial ○ Yes ○ Unsure
The selected statistical model is adequate for the design of the study.	○ No ○ Partial ○ Yes ○ Unsure
Reporting of results There is no selective reporting of results.	○ No ○ Partial ○ Yes ○ Unsure

Statistical Analysis and Presentation Summary

- ◯ Low Risk
- ◯ Moderate Risk
- ◯ High Risk

The statistical analysis is appropriate for the design of the study, limiting potential for presentation of invalid or spurious results

Overall comments regarding statistical analysis & reporting _____

eResults. Sensitivity Analyses and GRADE Assessment

Post-hoc Sensitivity Analyses

Although peak VO_2 was higher among non-hospitalized individuals, subgroup analysis suggested that the mean difference by symptom status did not vary by the proportion hospitalized (more hospitalized: -4.7; 95%CI -6.5 to -3.0 versus fewer hospitalized: -4.6; 95%CI -7.3 to -2.0; p=0.95). Similarly, subgroup analysis comparing studies by median time after SARS-CoV-2 infection suggested that time since infection was not a major cause of heterogeneity (<6 months: -5.0; 95%CI -7.1 to -3.0; ≥6 months: -4.5; 95%CI -6.4 to -3.4; p=0.73).

Leave One Out Analysis

Omitted study	**Mean diff (95%CI)**
Abdallah	-4.94 (-6.58 – -3.30)
Aparisi	-4.90 (-6.62 – -3.17)
Barbagelata	-5.18 (-6.81 – -3.56)
Brown	-5.01 (-6.71 – -3.30)
Durstenfeld	-4.69 (-6.22 – -3.16)
Ladlow	-5.44 (5.93 – -3.07)
Margalit	-5.18 (-6.81 – -3.56)
Skjørten	-4.25 (-5.44 – -3.05)
Szekely	-5.18(-6.76 – -3.60)
Overall	**-4.87 (-6.36 – -3.39)**

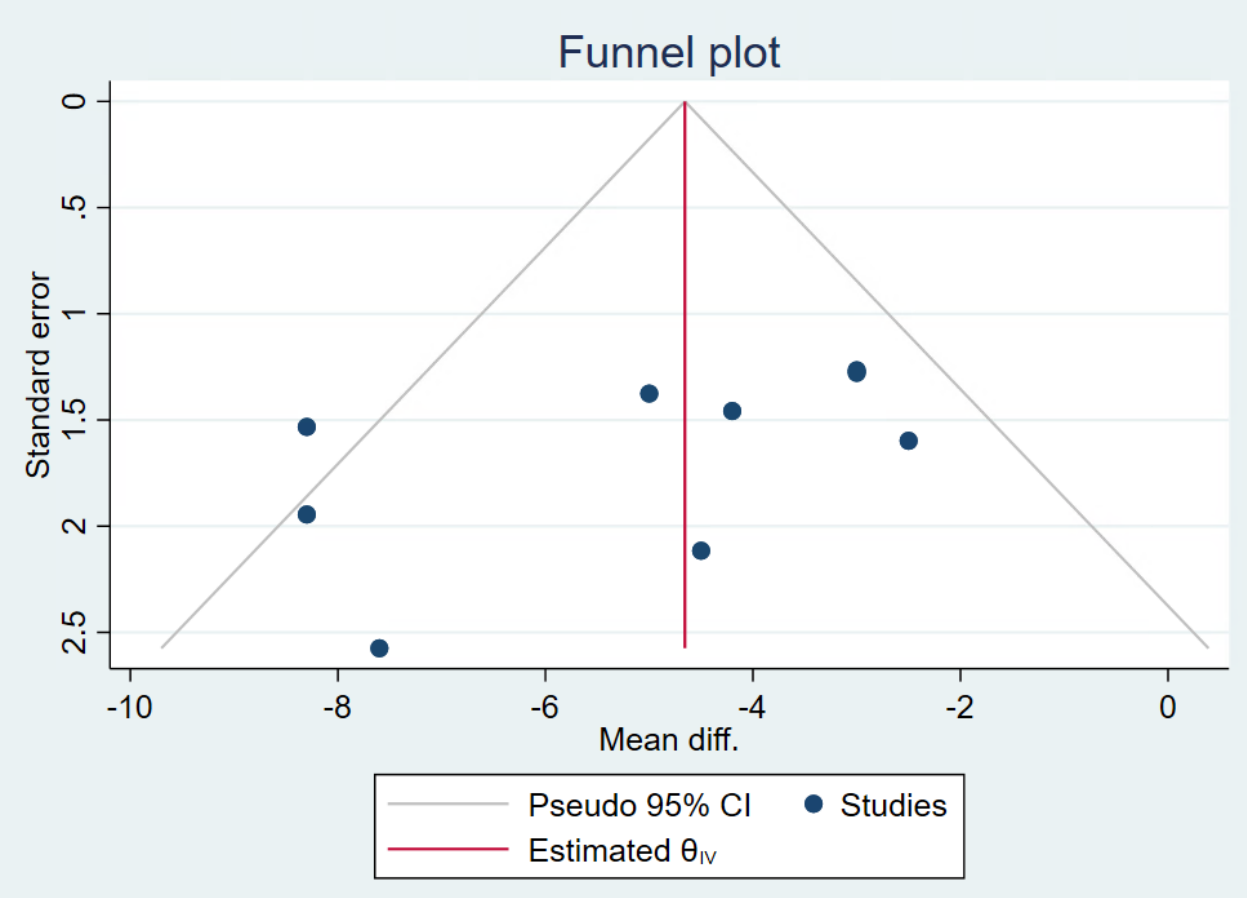
Summary of GRADE Assessment Discussion for Aim 1

Starting for Observational Data: Low
Risk of bias: Downgrade for issues with selection bias and confounding
Imprecision: No change for precision; whether the average effect is -6 ml/kg/min or -3 ml/kg/min would not dramatically change our interpretation (although the greater estimate suggests a higher prevalence, which we were not able to estimate).
Inconsistency: Upgrade for consistency: in the subgroup analyses and leave one out analyses the effects were fairly consistent.
Indirectness: Downgrade for indirectness in measuring Long COVID symptoms.
Publication bias: Uncertain. Two studies (Clavario et al & Cassar et al) that did find a statistically significant result and therefore did not report peak VO_2 by symptom status, so it is possible that there are other negative studies that have not been published. We attempted to find these through preprints or conference abstracts in case they are having a difficult time being published.
Overall team impression: Low Certainty

Summary of GRADE Assessment Discussion for Aim 2

Starting for Observational Data: Low
Risk of bias: Downgrade for issues with selection bias and confounding
Imprecision: Downgrade for lack of precision especially with regards to classification of deconditioning vs muscular/peripheral issues, issues with not excluding submaximal tests.
Inconsistency: Downgrade for inconsistency; the patterns observed across different studies are not at all consistent, and some studies report negative findings that are the most common pattern observed in other studies.
Indirectness: Downgrade for indirectness in measuring "Long COVID"
Publication bias: Uncertain
Overall team impression: Very Low Certainty

eFigure. Funnel Plot of Studies Comparing Peak $\dot{V}O_2$ Among People With and Without Symptoms



eFigure 1 Legend: Funnel plot of studies included for Aim 1 (With vs without LC Symptoms) using inverse variance.