Journal of Experimental Zoology Part A: Ecological and Integrative Physiology

WILEY

JEZ-A ECOLOGICAL AND INTEGRATIVE PHYSIOLOGY

REVIEW

Assessing Aerobic Capacity in Rat and Mouse Models: A Critical Narrative Review

Dursun Alper Yılmaz[1,2] | Fadıl Özyener[1,3]

[1]Department of Physiology, Institute of Health Sciences, Bursa Uludağ University, Bursa, Türkiye | [2]Department of Nursing, Faculty of Health Sciences, Ağrı İbrahim Çeçen University, Ağrı, Türkiye | [3]Department of Physiology, Faculty of Medicine, Bursa Uludağ University, Bursa, Türkiye

Correspondence: Fadıl Özyener (fozyener@uludag.edu.tr)

Received: 18 August 2026 | **Revised:** 28 September 2026 | **Accepted:** 12 October 2026

Keywords: aerobic capacity | exercise testing | O_2peak | rodent models | $\dot{V}O_2$max$\dot{V}$

ABSTRACT

Aerobic capacity reflects pulmonary ventilation, oxygen transport, cardiovascular performance, skeletal-muscle perfusion, and mitochondrial oxygen utilization. In rodents, $\dot{V}O_2$max, $\dot{V}O_2$peak, running speed, time to exhaustion, critical speed, lactate-derived thresholds, and voluntary activity are often grouped as aerobic capacity despite representing distinct constructs. This review aims to critically examine the physiological foundations, experimental modalities, outcome measures, methodological determinants, and translational challenges of aerobic-capacity assessment in rats and mice. A structured search of PubMed, Web of Science Core Collection, and Scopus was conducted from database inception to August 6, 2026. Studies addressing aerobic and endurance capacity, oxygen uptake, exercise performance, treadmill running, voluntary wheel running, swimming, indirect calorimetry, and methodological issues were synthesized narratively. Directly measured $\dot{V}O_2$max, or $\dot{V}O_2$peak when maximal criteria are not fulfilled, provides the strongest index of aerobic capacity. Running speed, distance, and time to exhaustion are performance outcomes influenced by running economy, anaerobic capacity, neuromuscular function, motivation, and test tolerance. Lactate-derived thresholds and critical speed provide complementary information on sustainable exercise intensity. Treadmill running with indirect calorimetry offers control for maximal oxygen-uptake assessment, whereas voluntary wheel running reflects self-selected activity and swimming is useful for lactate-based models. Validity depends on biological, procedural, environmental, and technical factors. Reliable assessment requires alignment among physiological construct, exercise modality, outcome measure, and experimental population. Model-specific validation, objective maximality and exhaustion criteria, appropriate oxygen-uptake normalization, calibrated systems, refined motivational procedures, and transparent reporting are essential for reproducibility, animal welfare, and translational interpretation.

1 | Introduction

Aerobic capacity is an integrated physiological phenotype determined by the coordinated uptake, transport, distribution, and mitochondrial utilization of oxygen. Rat and mouse models enable controlled investigation of its cardiovascular, respiratory, metabolic, neuromuscular, and molecular determinants. Rodent experiments have contributed foundational concepts to exercise physiology rather than merely serving as reduced-scale implementations of concepts established in humans. Classic studies in rats demonstrated that endurance training markedly increases skeletal-muscle mitochondrial respiratory capacity and respiratory-enzyme activity (Holloszy 1967), and subsequently linked mitochondrial markers, including citrate synthase and cytochrome c, to endurance performance (Fitts et al. 1975). Rodent models have also provided mechanistic insight into the cardiovascular and peripheral determinants of exercise tolerance; for example, regional-perfusion experiments demonstrated that endurance training alters the distribution of skeletal-muscle blood flow during exercise after myocardial

All rights reserved, including rights for text and data mining and training of artificial intelligence technologies or similar technologies.

© 2026 Wiley Periodicals LLC.

Highlights

- Rodent aerobic capacity requires construct-specific outcome interpretation.
- Treadmill calorimetry best measures maximal whole-body oxygen uptake.
- Wheel running and swimming capture distinct, non-equivalent exercise phenotypes.
- Protocol design, motivation, and exhaustion criteria can alter test validity.
- A minimum reporting framework is proposed to improve reproducibility.

infarction while increasing maximal oxygen uptake and endurance capacity (Musch et al. 1992). These studies illustrate the ability of animal models to integrate whole-body exercise responses with invasive tissue, metabolic, and circulatory measurements that are difficult or impossible to obtain simultaneously in humans. In addition, experimentally controlled disease models, standardized training exposures, and direct physiological measurements allow specific components of oxygen transport and utilization to be interrogated under conditions in which disease severity, exercise dose, and measurement timing can be tightly defined. However, rodent studies often group maximal oxygen uptake ($\dot{V}O_2$max), peak oxygen uptake ($\dot{V}O_2$peak), maximal or peak running speed, time to exhaustion, running distance, critical speed, lactate threshold, and maximal lactate steady state under the general term aerobic capacity, although these outcomes represent related but distinct constructs. $\dot{V}O_2$max primarily reflects the upper limit of whole-body oxygen transport and utilization, whereas performance-based outcomes are also influenced by running economy, anaerobic capacity, substrate availability, neuromuscular function, thermoregulation, motivation, and tolerance of the testing procedure (Bedford et al. 1979; Høydal et al. 2007; Marcaletti et al. 2011; Massett et al. 2021; Petrosino et al. 2016).

The distinction between $\dot{V}O_2$max and $\dot{V}O_2$peak is therefore essential. A plateau in oxygen uptake despite a further increase in workload provides strong evidence that $\dot{V}O_2$max has been attained; however, the absence of a plateau during a single incremental test does not necessarily exclude attainment of the physiological maximum. In human exercise testing, a subsequent verification bout has been advocated as an additional means of confirming maximality when the primary test does not demonstrate a clear plateau (Poole et al. 2008; Poole and Jones 2017). Because such verification approaches have not been established uniformly in rodent models, their application to rats and mice requires model-specific validation. When maximality cannot be adequately demonstrated, the highest observed oxygen uptake should be reported as $\dot{V}O_2$peak. Test termination may reflect behavioral refusal, abnormal gait, inadequate treadmill compliance, or locomotor rather than cardiorespiratory limitation, and the recorded peak is influenced by treadmill inclination, workload progression, stage duration, familiarization, gas-analysis response time, and exhaustion criteria (Ayachi et al. 2016; Marcaletti et al. 2011; Petrosino et al. 2016; Qin et al. 2020). Foundational studies demonstrated that progressive treadmill testing can detect physiological differences and support individualized exercise prescription. Bedford et al. (1979) established a multistage rat protocol; Wisløff et al. (2001) identified a clear plateau at a 25° grade in Sprague–Dawley rats; Kemi et al. (2002) reported high test–retest reliability in C57BL/6 J mice; and Leandro et al. (2007) used repeated $\dot{V}O_2$max assessment to adjust training intensity in Wistar rats. No single treadmill configuration is optimal for all rodent models. A slowly accelerating level ramp may be appropriate in some mouse strains, whereas combined increases in speed and inclination or steeper grades may be more effective in other populations. Protocol performance varies with species, strain, age, body size, gait, training status, chamber characteristics, and workload progression; consequently, protocols should be validated within the target population and measurement system (Ayachi et al. 2016; Kemi et al. 2002; Petrosino et al. 2016; Qin et al. 2020; Wisløff et al. 2001). This model specificity reflects the integrative nature of aerobic capacity, which depends on pulmonary gas exchange, hemoglobin-mediated transport, cardiac output, regional blood flow, muscle oxygen diffusion, and mitochondrial utilization. The relative contribution of these components changes with disease, aging, genetic background, and training, and animals with similar $\dot{V}O_2$max values may differ substantially in sustainable exercise capacity and fatigue profiles (Copp et al. 2010; Høydal et al. 2007; Marcaletti et al. 2011; Petrosino et al. 2016; Rezende et al. 2006). Outcome interpretation should therefore match the physiological level being assessed. Direct measurement of exercise $\dot{V}O_2$ provides the closest in vivo estimate of maximal whole-body oxygen transport and utilization. Lactate-derived indices instead characterize the transition between sustainable and non-steady-state exercise and are sensitive to species, modality, workload progression, sampling procedures, mathematical method, and threshold definition (Abreu et al. 2016; Almeida et al. 2013; Gobatto et al. 2001; Potolitsyna et al. 2023). Critical speed similarly defines a functional boundary between sustainable and non-sustainable exercise but is not equivalent to maximal oxygen-transport capacity (Copp et al. 2010). Maximal speed, distance, and time to exhaustion integrate aerobic and anaerobic metabolism with biomechanics, muscle recruitment, motivation, and behavioral tolerance and should not be labeled as $\dot{V}O_2$max without gas-exchange measurement. Running speed remains useful when interpreted as a performance or prescription variable. $\dot{V}O_2$ generally rises with treadmill speed during submaximal exercise, and maximal speed is associated with oxygen consumption in healthy and diabetic rats (Høydal et al. 2007; Rodrigues et al. 2007). Peak velocity may also provide a practical basis for training prescription in mice, producing adaptations similar to those obtained when training is prescribed from the velocity associated with $\dot{V}O_2$max (Picoli et al. 2018). Nevertheless, the same absolute speed may represent a lower relative intensity after training because running economy and $\dot{V}O_2$max can improve. Direct reassessment is therefore preferable when precise individualized intensity control is required (Høydal et al. 2007; Picoli et al. 2018; Rodrigues et al. 2007).

Biological variability further limits the transferability of reference values. Age, sex, strain, substrain, genetic background, body mass, disease status, baseline activity, training history, and willingness to run influence both oxygen consumption and

exercise performance. Differences have been documented among rat strains, between adult and aged mice, and between sexes, indicating that values obtained in one population should not be generalized uncritically to another (Bedford et al. 1979; Kemi et al. 2002; Schefer and Talan 1996; Wisløff et al. 2001). Treadmill performance is also shaped by strain-specific behavioral compliance. Marked differences in exercise duration, forced treadmill performance, voluntary wheel running, and training adaptation have been reported across mouse strains, demonstrating that physiological capacity interacts with locomotor behavior, stress responsiveness, and genetic background (Gibb et al. 2016; Lerman et al. 2002; Lightfoot et al. 2001; Rezende et al. 2006). Body-size normalization is another important analytical issue. Expressing $\dot{V}O_2$ only as milliliters per kilogram per minute assumes linear scaling with body mass and may bias comparisons involving animals that differ in size or body composition. Allometric approaches using exponents such as 0.67 or 0.75, logarithmic modeling, or statistical adjustment for body mass or lean mass may be preferable in some settings, although no universal method is suitable for all populations. Studies should therefore report absolute and mass-normalized $\dot{V}O_2$ values, describe any allometric or statistical adjustment, and justify the selected approach (Bedford et al. 1979; Høydal et al. 2007; Speakman 2013). Methodological standardization must also address animal behavior and welfare. Electrical grids and other aversive stimuli can influence stress physiology, motivation, pain tolerance, and the operational definition of exhaustion. Standardized familiarization, prospective stopping criteria, consistent handling, and transparent reporting of motivational procedures are essential for distinguishing physiological limitation from behavioral withdrawal (Booth et al. 2010; Marcaletti et al. 2011; Rezaee 2025). Non-noxious strategies, such as the rubber-ball stimulus described by de Assis et al. (2024), may improve welfare and reduce measurement bias. Standardization, however, should not imply a universal external workload; it should mean transparent, reproducible, and model-specific validation. The substantial heterogeneity in mouse endurance-training studies confirms that protocol design and reporting quality can materially affect observed adaptations (Massett et al. 2021). Finally, aerobic capacity, endurance capacity, and training adaptation should remain conceptually distinct. Aerobic capacity refers primarily to the upper limit of aerobic energy production, endurance capacity to the ability to sustain a workload, and training adaptation to changes in physiological, functional, or molecular outcomes after repeated exercise. Improvements in time to exhaustion, maximal speed, running economy, or oxidative markers may occur without a corresponding increase in $\dot{V}O_2$max, and changes in $\dot{V}O_2$max may not produce proportional improvements in every endurance outcome. Tissue-level markers should therefore be interpreted as mechanistic correlates rather than direct measures of whole-body aerobic capacity (Leandro et al. 2007; Marcaletti et al. 2011; Massett et al. 2021). Accordingly, this narrative review treats aerobic capacity as a multidimensional construct. Directly measured $\dot{V}O_2$max, or $\dot{V}O_2$peak when maximal criteria are not fulfilled, is considered the primary whole-body physiological index; lactate-derived thresholds and critical speed are viewed as complementary indicators of sustainable metabolic performance; maximal speed, distance, and time to exhaustion are interpreted as integrated functional outcomes; and tissue-level oxidative and molecular markers are considered mechanistic correlates. The review critically examines exercise modalities, direct and indirect indices, protocol validity, biological and technical sources of variability, animal-welfare considerations, reporting standards, and translational relevance in rat and mouse models.

1.1 | Literature Search and Narrative Synthesis

This article was designed as a comprehensive narrative review integrating the physiological foundations, experimental models, measurement technologies, testing protocols, interpretative limitations, and translational challenges associated with the assessment of aerobic capacity in rats and mice. The review was not intended to quantify the effect of a single intervention or to generate a pooled estimate across methodologically homogeneous studies. Instead, it aimed to determine how aerobic capacity has been conceptualized and measured in rodent exercise research, to distinguish direct physiological measurements from indirect performance-based indicators, and to identify the methodological factors that limit comparability across laboratories, strains, species, exercise modalities, and experimental conditions.

A structured, iterative, and deliberately sensitive literature search was conducted in PubMed/MEDLINE, Web of Science Core Collection, and Scopus from database inception to August 6, 2026. No lower publication-year limit was imposed because several early investigations established the physiological and technical foundations of maximal oxygen-uptake measurement, treadmill exercise testing, swimming-based endurance assessment, and gas-exchange analysis in small laboratory animals. Excluding older studies would therefore have removed historically important protocol-development and validation evidence. At the same time, recent studies were specifically examined to determine whether advances in indirect calorimetry, automated metabolic systems, exercise-test standardization, animal handling, welfare-oriented motivation, and transparent reporting had addressed the limitations identified in the earlier literature.

The search process was conducted through several partially overlapping search sets rather than through a single restrictive query. This approach was selected because the terminology used to describe rodent aerobic capacity is highly heterogeneous. Studies evaluating essentially similar physiological constructs may use terms such as maximal oxygen uptake, peak oxygen uptake, aerobic power, cardiorespiratory fitness, aerobic performance, exercise capacity, endurance capacity, running performance, maximal running speed, peak treadmill velocity, critical speed, lactate threshold, the historically used term “anaerobic threshold,” running economy, work capacity, fatigue resistance, or time to exhaustion. Conversely, some studies use the expression “aerobic exercise” only to describe the intervention and do not actually measure aerobic capacity. Multiple overlapping searches were therefore necessary to maximize retrieval sensitivity while allowing conceptual relevance to be determined during title, abstract, and full-text evaluation.

The search strategy was structured around four complementary concept domains. The first domain identified the experimental

species and included terms corresponding to rats, mice, rodents, *Rattus norvegicus*, and *Mus musculus*. Species terms were searched using controlled vocabulary and free-text variants, including singular, plural, abbreviated, and taxonomic forms where appropriate. Broader rodent terminology was retained in selected searches to capture studies in which the species was specified only in the abstract or indexing terms rather than in the title. However, studies were retained for the synthesis only when the relevant findings were obtained from rats or mice or when species-specific rodent data could be clearly separated from findings obtained in other animals.

The second domain represented the physiological construct of aerobic capacity. Search terms included maximal oxygen uptake, maximum oxygen consumption, peak oxygen uptake, VO_2max, VO_2peak, aerobic capacity, aerobic power, aerobic fitness, cardiorespiratory fitness, exercise capacity, maximal exercise capacity, endurance capacity, endurance performance, running capacity, running performance, exercise tolerance, fatigue resistance, work capacity, and time to exhaustion. Because maximal aerobic capacity may also be estimated using external workload or velocity rather than directly measured oxygen uptake, the searches additionally included maximal running speed, peak running speed, maximal treadmill velocity, peak treadmill velocity, maximal aerobic speed, critical speed, critical velocity, and related workload-based indices. Terms related to submaximal physiological transition points, including lactate threshold, blood lactate, lactate turn point, the historical term "anaerobic threshold," ventilatory threshold, gas-exchange threshold, and exercise intensity domains, were also incorporated when they could contribute to the characterization of aerobic performance.

The third domain addressed exercise-testing modalities and measurement systems. Relevant terms included treadmill running, motorized treadmill exercise, graded treadmill testing, incremental exercise testing, progressive exercise testing, maximal exercise testing, exhaustion testing, swimming, forced swimming, swim-to-exhaustion testing, load-adjusted swimming, voluntary wheel running, running wheels, spontaneous locomotor activity, indirect calorimetry, open-circuit calorimetry, closed-circuit calorimetry, respirometry, metabolic chambers, metabolic treadmills, gas-exchange analysis, oxygen-consumption measurement, carbon-dioxide production, respiratory exchange ratio, respiratory quotient, and whole-body metabolic assessment. Searches were not limited to studies explicitly using the term "aerobic-capacity test," because many technically relevant studies were indexed according to the apparatus, exercise modality, or physiological variable measured rather than the broader construct of aerobic capacity.

The fourth domain was used to identify methodological and interpretative evidence. This domain included terms related to validity, criterion validity, construct validity, reliability, reproducibility, repeatability, test–retest reliability, standardization, familiarization, acclimatization, protocol comparison, exercise-test termination, exhaustion criteria, motivation, aversive stimulation, electrical stimulation, shock grids, handling, stress, animal welfare, circadian timing, sex differences, age differences, strain differences, body mass, thermoregulation, environmental temperature, treadmill inclination, stage duration, speed increments, chamber dimensions, gas-flow rate, analyzer calibration, sampling interval, and signal averaging. These terms were combined selectively with species, outcome, and modality terms to identify investigations focused on the technical and biological determinants of test performance rather than on the effects of a particular training or pharmacological intervention.

Controlled vocabulary terms were combined with free-text keywords whenever supported by the database. In PubMed, relevant Medical Subject Headings were combined with title and abstract terms to capture both formally indexed records and recently published or online-first articles that had not yet received complete indexing. In Scopus, searches were adapted primarily to title, abstract, and keyword fields. In Web of Science Core Collection, the corresponding topic fields were used to search titles, abstracts, author keywords, and database-generated keywords. Database-specific syntax, truncation, phrase searching, field tags, and Boolean operators were adapted as required, while the underlying conceptual structure was preserved across platforms.

The general Boolean logic can be summarized as follows:

(rat OR rats OR mouse OR mice OR rodent OR *Rattus norvegicus* OR *Mus musculus*) AND (maximal oxygen uptake OR maximum oxygen consumption OR VO_2max OR VO_2peak OR aerobic capacity OR cardiorespiratory fitness OR exercise capacity OR endurance capacity OR maximal running speed OR critical speed OR lactate threshold OR running economy OR time to exhaustion) AND (treadmill OR running OR swimming OR voluntary wheel running OR incremental exercise test OR graded exercise test OR indirect calorimetry OR respirometry OR gas-exchange analysis OR metabolic chamber).

This core structure was broadened or narrowed in successive search sets. Broad searches were used to identify terminology, landmark studies, and potentially relevant methodological evidence, whereas focused searches combined specific outcomes with particular modalities, such as VO_2max with treadmill testing, critical speed with swimming, lactate threshold with graded running, or voluntary wheel running with exercise performance. The use of multiple overlapping search sets was intentional because preliminary examination showed that no single combination of indexing terms adequately retrieved all major methodological traditions in the field.

Search results were exported in comma-separated-value format and consolidated into a working bibliographic library. The available search files comprised several overlapping PubMed search sets, a detailed Scopus export, and a supplementary evidence-discovery search intended to identify recent or semantically related records that might not use conventional aerobic-capacity terminology. Supplementary discovery records were not accepted solely on the basis of automatically generated relevance statements. Their titles, abstracts, bibliographic information, and, where required, original journal records were examined before they were considered for inclusion in the narrative evidence base.

Duplicate records were identified hierarchically using digital object identifiers, PubMed identifiers or other database-specific identifiers, and normalized article titles. Title-based matching

was necessary when older studies lacked a DOI, when the same article appeared with minor punctuation or capitalization differences, or when online-first and final versions were exported as separate records. Apparent duplicates were checked against author names, publication year, journal, volume, issue, and page or article number before removal. Multiple reports from the same experimental series were not automatically treated as duplicates when they contained distinct protocols, samples, physiological outcomes, or methodological analyses.

The search was restricted to English-language journal articles and reviews involving rat or mouse models. When database-level language filters could not be applied consistently without reducing retrieval sensitivity, language eligibility was assessed during screening. Both final publications and sufficiently documented online-first or in-press journal articles available by August 6, 2026, were eligible. Reviews were used to identify established concepts, historical developments, methodological controversies, and additional primary studies. However, conclusions regarding protocol performance, measurement validity, physiological responses, and reproducibility were based preferentially on the original experimental reports.

Studies were considered potentially eligible when they addressed at least one of the following areas: direct measurement of oxygen uptake during exercise; determination or interpretation of VO_2max or VO_2peak; maximal or peak treadmill speed; critical-speed or critical-velocity modeling; lactate-derived exercise thresholds; ventilatory or gas-exchange thresholds; running economy; endurance time or time to exhaustion; swimming-based exercise capacity; voluntary wheel-running performance; calibration or validation of indirect-calorimetry systems; reliability or repeatability of exercise tests; criteria used to define maximal effort or exhaustion; effects of familiarization and motivation; biological determinants of test variability; or animal-welfare considerations relevant to exercise testing.

Studies involving healthy, trained, untrained, sedentary, aged, genetically modified, or disease-model animals were eligible when their methods or findings contributed to understanding the measurement or interpretation of aerobic capacity. Disease-model studies were not included merely because exercise was used as an intervention. They were considered relevant only when they provided transferable methodological information, directly assessed an aerobic-capacity outcome, compared testing approaches, or demonstrated how pathological conditions altered the validity or interpretation of the selected outcome. Similarly, training studies were prioritized when they reported sufficiently detailed testing protocols or provided evidence regarding responsiveness, construct validity, or the physiological meaning of the measured outcome.

Human-only studies, studies involving non-rodent species without separable rat or mouse data, and studies limited exclusively to cellular or ex vivo experiments were excluded from the core synthesis. Studies reporting only molecular, histological, oxidative, inflammatory, or metabolic adaptations after an intervention were also excluded when no exercise-capacity test or methodologically relevant performance outcome was reported. General locomotor-activity studies were not considered equivalent to aerobic-capacity assessments unless wheel-running or spontaneous activity data were explicitly interpreted in relation to endurance, aerobic phenotype, or exercise performance. Studies concerned exclusively with resistance exercise, isolated muscle force, anaerobic sprint performance, or motor coordination were excluded unless they included an aerobic-capacity measure relevant to the objectives of the review. Conference abstracts, editorials, brief commentaries, and other reports lacking adequate methodological detail were not used as principal evidence, although they could assist in locating a subsequent full publication.

Screening proceeded from broad conceptual relevance to detailed methodological relevance. Titles and abstracts were first examined to identify records involving rats or mice and at least one relevant exercise modality, physiological outcome, or measurement issue. Full texts were then evaluated when the abstract indicated direct measurement of aerobic capacity, use of an established surrogate, comparison of testing protocols, or discussion of technical or biological determinants of test performance. Broadly retrieved records that merely mentioned exercise, endurance, metabolism, or aerobic pathways without evaluating whole-animal exercise capacity were excluded at this stage. Because the review was narrative rather than systematic, screening was directed toward achieving conceptual completeness across the major methodological domains rather than establishing a narrowly defined intervention–comparator–outcome evidence set.

Information from eligible studies was charted in a structured manner. Bibliographic and design characteristics included publication year, study purpose, experimental design, and whether the article represented a protocol-development study, validation study, reliability study, training experiment, disease-model investigation, methodological comparison, or review. Animal characteristics included species, strain or stock, sex, age, developmental stage, body mass, training status, disease or genetic model, and sample size when reported.

Exercise-test characteristics included the modality used; treadmill model or swimming apparatus; chamber dimensions; treadmill inclination; initial speed; stage duration; speed or workload increments; warm-up procedures; total test duration; familiarization schedule; interval between repeated tests; load attached during swimming; water depth and temperature; wheel dimensions and resistance; and whether testing was voluntary or externally imposed. Environmental and procedural characteristics included room temperature, humidity, time of day, light–dark cycle, feeding or fasting status, handling procedures, investigator presence, and the method used to motivate continued exercise.

For studies using indirect calorimetry, attention was given to the type of respirometry system, airflow characteristics, oxygen and carbon-dioxide analyzers, calibration procedures, chamber equilibration, sampling frequency, data-averaging interval, correction for delay between exercise and gas measurement, and expression of oxygen uptake. Oxygen uptake reported in absolute units was distinguished from values normalized to body mass, lean body mass, or allometrically scaled body size. This distinction was important because simple ratio normalization may distort comparisons among animals that differ substantially in body mass, adiposity, age, sex, strain, or disease status.

The criteria used to identify maximal effort were recorded whenever available. These included the presence or absence of an oxygen-uptake plateau despite increasing workload, achievement of a predefined respiratory exchange ratio, blood-lactate concentration, failure to maintain treadmill position, repeated contact with the rear of the treadmill, inability to continue swimming, loss of coordinated movement, or failure to respond to motivational stimuli. Particular attention was paid to whether the reported outcome should be interpreted as VO_2max, VO_2peak, maximal tolerated oxygen uptake, or simply the highest oxygen-uptake value observed during the protocol. Studies that used the terms VO_2max and VO_2peak interchangeably were interpreted according to the actual measurement and verification procedures rather than according to the terminology selected by the authors.

Performance-based outcomes were also differentiated according to the construct they most directly represented. Maximal or peak treadmill speed was treated primarily as an integrated performance outcome influenced by aerobic capacity, neuromuscular function, coordination, motivation, body mass, treadmill familiarity, and aversive-stimulus tolerance. Time to exhaustion was interpreted as an endurance-performance measure rather than as a direct equivalent of maximal aerobic capacity because it is sensitive to protocol intensity, investigator-defined termination criteria, motivation, stress, substrate availability, thermoregulation, and pacing opportunities. Critical speed was considered a model-derived estimate of the boundary between sustainable and non-sustainable exercise domains, provided that the model was based on multiple appropriately selected exhaustive trials and demonstrated an acceptable fit. Lactate-derived thresholds and running economy were evaluated as complementary submaximal indices that describe different dimensions of aerobic function and cannot be assumed to substitute directly for VO_2max.

The evidence was subsequently synthesized according to a predefined thematic framework. The principal domains were: conceptual definitions of aerobic capacity; direct measurement of VO_2max and VO_2peak; treadmill-based exercise testing; swimming-based capacity and endurance tests; voluntary wheel-running models; maximal speed and time-to-exhaustion outcomes; critical-speed approaches; lactate-derived thresholds; running economy and submaximal efficiency; indirect-calorimetry and respirometry methodology; biological sources of variability; protocol standardization and reproducibility; motivational procedures and exhaustion criteria; animal welfare; and translation from rodent models to human exercise physiology.

Within each domain, evidence was organized chronologically and methodologically rather than solely according to the direction of the reported findings. Foundational studies were used to explain the origins of commonly applied protocols and assumptions. Later validation and reliability studies were used to assess whether those assumptions had been empirically supported. Recent investigations, methodological reviews, and reporting recommendations were used to evaluate contemporary practice, technological developments, reproducibility, and welfare. When multiple studies used apparently similar outcome labels but substantially different protocols, the differences were described explicitly rather than combining the findings under a single outcome category.

Greater interpretative weight was given to studies that directly measured gas exchange, described calibration and sampling procedures, reported animal and protocol characteristics in sufficient detail, used explicit and reproducible exhaustion criteria, evaluated test–retest reliability, or compared a performance surrogate with a direct physiological measure. Studies providing limited information about the exercise apparatus, familiarization, motivation, environmental conditions, or test termination were retained when historically or conceptually important but were interpreted more cautiously. Accordingly, the narrative synthesis distinguished between evidence that was physiologically direct, methodologically validated, reproducible across repeated testing, or merely conventional within a specific research tradition.

Formal meta-analysis was not undertaken because the literature was characterised by substantial heterogeneity in species, strain, sex, age, body size, disease status, training background, exercise modality, treadmill slope, speed progression, stage duration, swimming load, familiarization, motivational procedures, gas-analysis systems, normalization methods, maximality criteria, and outcome definitions. These differences would have made a pooled numerical estimate difficult to interpret and potentially misleading. Likewise, a formal risk-of-bias assessment and certainty-of-evidence grading were not performed because the purpose of the review was not to estimate a single causal treatment effect. Methodological credibility was instead considered narratively according to measurement directness, protocol transparency, validation, reliability, reproducibility, biological plausibility, reporting completeness, and susceptibility to stress- or motivation-related confounding.

The synthesis therefore emphasized patterns of convergence, sources of inconsistency, and conditions under which a given outcome can be interpreted validly. Agreement across independent laboratories, strains, exercise modalities, or measurement systems was noted when present. Conflicting findings were examined in relation to differences in protocol design, animal characteristics, normalization strategy, familiarization, test termination, and technical measurement procedures rather than being reduced to a simple count of positive and negative studies. Particular care was taken not to equate statistically significant training responses with validation of the measurement method itself.

Overall, this structured narrative approach was intended to provide a physiologically coherent and methodologically transparent account of rodent aerobic-capacity assessment. It enabled direct physiological measurements and indirect performance surrogates to be considered within the same conceptual framework while preserving the distinctions among maximal aerobic power, submaximal efficiency, sustainable exercise intensity, endurance tolerance, spontaneous activity, and motivation-dependent performance. The resulting synthesis was used to identify established practices, unresolved methodological problems, minimum reporting requirements, and priorities for improving reproducibility and cross-species translation in future rat and mouse exercise studies.

2 | Experimental Modalities and Test Protocols

The validity of aerobic-capacity assessment in rodents depends on both the physiological outcome and the exercise modality used to elicit it. Motorized treadmill running, voluntary wheel running, and swimming differ in workload control, locomotor pattern, muscle recruitment, behavioral and motivational demands, thermoregulatory stress, and compatibility with respiratory gas analysis. They should therefore not be treated as interchangeable methods. Modality selection should be based on whether the objective is to measure maximal oxygen uptake, sustainable exercise intensity, endurance tolerance, training intensity, habitual physical activity, or long-term adaptation (Marcaletti et al. 2011; Massett et al. 2021; Rezaee 2025). Disease phenotype should also guide both test selection and interpretation. A protocol appropriate for healthy animals may become limited by disease-specific factors such as impaired skeletal-muscle perfusion, locomotor dysfunction, weakness, altered gait, pain, or reduced tolerance of imposed exercise before maximal cardiorespiratory capacity is reached. Consequently, lower running speed or shorter time to exhaustion in a disease model should not automatically be interpreted as a proportional reduction in $\dot{V}O_2$max. Where the research question concerns maximal oxygen transport, direct gas-exchange measurement is preferable; where the pathology is expected primarily to affect peripheral perfusion, muscle metabolism, neuromuscular function, or fatigue tolerance, complementary physiological and functional outcomes should be selected accordingly. Accordingly, disease-model testing should distinguish between reduced maximal oxygen transport and exercise limitation arising from altered perfusion, muscle metabolism, locomotor function, or fatigue tolerance (Musch and Terrell 1992; Musch et al. 1992).

2.1 | Treadmill Running

Motorized treadmill running is the most standardized modality for assessing maximal aerobic capacity in rats and mice because speed, inclination, stage duration, acceleration, and total exercise time can be controlled. When enclosed within a metabolic chamber, it also permits simultaneous measurement of $\dot{V}O_2$, $\dot{V}CO_2$, and respiratory exchange ratio. Incremental or ramp protocols progressively increase speed, inclination, or both and are most appropriate for determining $\dot{V}O_2$max, $\dot{V}O_2$peak, maximal running speed, and related cardiorespiratory responses. Constant-load protocols maintain a predefined workload and are better suited to endurance tolerance, running economy, metabolic stability, and fatigue resistance. Improved time to exhaustion during a constant-load test should not automatically be interpreted as increased $\dot{V}O_2$max because it may also reflect changes in substrate utilization, thermoregulation, anaerobic reserve, running economy, or peripheral fatigue resistance (Marcaletti et al. 2011; Petrosino et al. 2016; Qin et al. 2020). Foundational studies established the methodological basis of treadmill-based aerobic testing. Bedford et al. (1979) developed a continuous 10-stage protocol that detected differences associated with strain, sex, age, hypertension, and training status, while also demonstrating that treadmill competence may represent an eligibility issue. In Sprague–Dawley rats, Wisløff et al. (2001) used a 15-min warm-up followed by speed increments of 0.03 m/s every 2 min and found that a 25° inclination elicited the highest oxygen uptake and a distinct plateau accompanied by a respiratory exchange ratio above 1.05 and blood lactate above 6 mmol/L. Leandro et al. (2007) subsequently combined a 10-stage treadmill test with open-circuit calorimetry in Wistar rats and used repeated $\dot{V}O_2$max measurements to adjust training intensity, demonstrating that the same absolute speed may represent a lower relative intensity after adaptation (Høydal et al. 2007; Leandro et al. 2007). Kemi et al. (2002) adapted a similar intensity-controlled approach to C57BL/6 J mice. Testing across several inclinations produced a test–retest correlation of 0.98 at 25°, while peak oxygen uptake occurred over an inclination range of approximately 15°–35°. Repeated testing enabled individualized training prescription and detected improvements in $\dot{V}O_2$max and running economy in both sexes. However, subsequent studies showed that steep inclination is not universally required. Ayachi et al. (2016) found that a slowly accelerating ramp at 0° elicited the highest $\dot{V}O_2$peak in FVB/N mice, together with a plateau and elevated lactate and respiratory exchange ratio. Petrosino et al. (2016) used staged increases in speed and grade to derive maximal oxygen uptake, a submaximal metabolic transition point reported by the authors as an "anaerobic threshold," substrate crossover, running economy, maximal workload, and recovery characteristics from a single test. The term "anaerobic threshold," however, should be interpreted cautiously because lactate accumulation does not indicate the onset of muscle anoxia; more specific terms such as lactate threshold or gas-exchange threshold are preferable when supported by the measurement used (Poole et al. 2021). Qin et al. (2020) further demonstrated that the oxygen cost of speed and inclination varied with age in Wistar rats. Collectively, these findings indicate that optimal treadmill architecture is species-, strain-, age-, and system-dependent. Peak running speed may provide a practical performance or prescription variable when direct gas analysis is unavailable. Picoli et al. (2018) reported similar adaptations when training was prescribed from peak velocity or from the velocity associated with $\dot{V}O_2$max. Similarly, Rodrigues et al. (2007) found significant relationships between maximal treadmill speed and oxygen consumption in control and diabetic rats. Nevertheless, peak speed remains an indirect outcome and should not be presented as a direct measurement of maximal oxygen uptake. Its relationship with $\dot{V}O_2$ should be validated within the specific disease model because gait, muscle function, autonomic regulation, and motivation may alter performance independently of cardiorespiratory capacity.

Protocol duration and workload progression must allow the physiological response to develop without causing premature mechanical, peripheral, thermal, or behavioral fatigue. Very rapid increments may exceed the animal's locomotor capacity before maximal oxygen transport is reached, whereas excessively long stages may increase fatigue and exposure to motivational stimuli. Chamber mixing, analyser response time, starting workload, acceleration, stage duration, inclination, and total test duration should therefore be considered together (Ayachi et al. 2016; Marcaletti et al. 2011; Petrosino et al. 2016).

The magnitude and nature of training adaptation are strongly protocol- and outcome-dependent. In male rats, Lambert and Noakes (1989) reported an increase in $\dot{V}O_2$max from 75.5 ± 1.9 to 86.0 ± 2.7 $mL \cdot kg^{-1} \cdot min^{-1}$ after 12 weeks of treadmill training, corresponding to an increase of approximately 14%, whereas submaximal running endurance increased by 120%, 320%, and 372% after 4, 8, and 12 weeks, respectively. In male Swiss mice, Picoli et al. (2018) reported increases in $\dot{V}O_2$max of 49.1% and 56.2% after 4 weeks of treadmill training when

exercise intensity was prescribed from v$\dot{V}O_2$max or peak running velocity, respectively. In contrast, Teixeira-Coelho et al. (2017) found that 8 weeks of treadmill training improved running performance and economy in male Wistar rats without a corresponding increase in $\dot{V}O_2$peak. These findings illustrate that training-induced improvements in endurance performance may be considerably larger than, or occur independently of, changes in maximal oxygen uptake. The substantial heterogeneity reported across mouse endurance-training studies in strain, sex, inclination, session duration, familiarization, testing procedures, and outcome selection confirms that protocol design can influence the apparent magnitude of adaptation and limit between-study comparison (Massett et al. 2021).

2.2 | Familiarization, Motivation, and Exhaustion Criteria

Treadmill familiarization is necessary to distinguish physiological limitation from novelty, fear, and inadequate running skill. Animals must learn to orient toward the front of the lane, maintain belt speed, recover after minor disturbances, and respond consistently to the selected motivational procedure. In C57BL/6 J mice, one or two short familiarization sessions may be sufficient to reduce aversive-stimulus exposure, although the required familiarization period should be validated for each strain, disease model, age group, and experimental condition (Marcaletti et al. 2011). Familiarization should reduce learning effects without itself becoming a training stimulus. Protocols should therefore report the number and duration of familiarization sessions, treadmill speed and inclination, rest intervals, chamber exposure, handler involvement, and motivational procedures. The same approach should be applied across groups. When treadmill competence is an eligibility requirement, criteria should be defined before randomization because post-randomization exclusion of "poor runners" may introduce attrition bias.

Motivational procedures influence both animal welfare and measurement validity. Electrical-grid stimulation has been widely used to maintain treadmill running, but stimulus exposure can differ substantially among animals and may introduce stress-, avoidance-, and pain-related variability into the measured performance endpoint (Marcaletti et al. 2011; Rezaee 2025). Escalating stimulation after an animal is no longer able to maintain the required workload does not improve discrimination between physiological exhaustion and behavioral or stimulus-related responses. Less aversive alternatives, including gentle touching, brushing, gas stimulation, visual cues, and strategically positioned objects, may therefore be preferable when they are effective and standardized. de Assis et al. (2024), for example, used a non-noxious rubber-ball stimulus during treadmill exercise in rats and detected training-related improvements in treadmill performance without electrical shock.

Running to exhaustion should be distinguished from the motivational method used to elicit continued running; an exhaustion test is not, by itself, evidence of an unacceptable welfare endpoint. Rather, interpretation depends on how exhaustion is operationally defined, how motivation is applied, and whether the test is terminated promptly once the predefined endpoint is reached. Objective criteria may include persistent inability to maintain belt speed, failure to re-engage after standardized prompting, and reproducible alterations in gait or posture. In male Sprague–Dawley rats, Copp et al. (2009) defined fatigue using inability or unwillingness to maintain treadmill pace together with characteristic gait changes and attenuation of the righting reflex. Fatigued rats required several seconds or longer to right themselves when placed supine, whereas non-fatigued rats righted themselves immediately. Thus, attenuation of the righting response can provide corroborative evidence of profound exercise fatigue, but it should not be interpreted as a criterion for attainment of $\dot{V}O_2$max.

Published stimulation-count criteria should also be interpreted cautiously. Marcaletti et al. (2011), for example, described a stimulation rate exceeding 20 electrical stimulations per minute or failure to re-engage for more than 15 s as operational criteria for terminating a particular mouse treadmill protocol. These values describe that experimental procedure and should not be interpreted as universal welfare recommendations or acceptable exposure limits. A numerical stimulation threshold alone cannot establish the welfare acceptability of a protocol; stimulus intensity, duration, cumulative exposure, animal behavior, and recovery should also be considered. Once the predefined termination criterion is met, treadmill running and the motivational stimulus should cease, and both stimulus exposure and the specific reason for termination should be reported.

Strain-specific behavior further complicates exhaustion-based testing. Lightfoot et al. (2001) found marked differences among ten inbred mouse strains: some demonstrated prolonged running, some repeatedly interacted with the shock grid and refused to continue, and others attempted to run but could not maintain relatively low speeds. Lerman et al. (2002) likewise observed substantial genetic differences in forced treadmill and voluntary wheel-running performance, with limited correspondence between the two modalities. Gibb et al. (2016) also reported strain-specific differences in treadmill compliance and training adaptation between FVB/NJ and C57BL/6 J mice. Strain selection and behavioral compliance should therefore be treated as components of protocol validity rather than incidental characteristics (Gibb et al. 2016; Lerman et al. 2002; Lightfoot et al. 2001).

2.3 | Voluntary Wheel Running

Voluntary wheel running allows mice to exercise according to their natural circadian pattern with minimal investigator intervention and generally without aversive stimulation. It is particularly useful for studying spontaneous physical activity, chronic exercise exposure, behavioral phenotypes, aging, cardiometabolic adaptation, and interactions between genetic background and exercise preference (Lerman et al. 2002; Manzanares et al. 2019; Rezaee 2025). Mice typically accumulate substantial distances through numerous short bouts, mainly during the dark cycle. This pattern resembles spontaneous murine locomotion more closely than continuous treadmill running and can induce endurance-like adaptations, although the self-selected dose varies across individuals, strains, sexes, ages, and disease states (Manzanares et al. 2019). The principal limitation is the lack of precise control over exercise intensity, duration, frequency, and progression. Daily distance alone is insufficient because the same distance may be completed through different combinations of speed, bout duration, and recovery.

Reports should therefore include daily distance, average and peak speed, activity time, bout number and duration, dark–light distribution, and longitudinal changes. When physiological exercise capacity is of interest, wheel-running exposure should be complemented by a separate standardized capacity test, such as treadmill gas-exchange testing or a validated exercise-tolerance protocol. Wheel characteristics may also alter behavior. Diameter, width, orientation, resistance, inertia, surface properties, and accessibility can affect revolutions, speed, and willingness to run. Mice show preferences for particular wheel designs, and angled and upright wheels may produce different activity levels (Banjanin and Mrosovsky 2000; Manzanares et al. 2019). Housing conditions introduce additional limitations: individual housing permits animal-level measurement but may alter stress and metabolism, whereas group housing with one wheel prevents reliable attribution of activity. Electronically identified systems may reduce this problem. Mechanical resistance, bedding contamination, sensor resolution, logging intervals, and false revolutions caused by climbing or wheel manipulation should also be considered. Voluntary wheel-running activity should therefore be interpreted as habitual physical activity, exercise behavior, or self-selected exercise exposure rather than as a measure of exercise capacity. Distance, duration, or average wheel-running speed reflect how much and how an animal chooses to run under the specific housing and wheel conditions; they do not directly quantify maximal or sustainable physiological exercise capacity. This distinction is supported by the poor correspondence between voluntary-wheel and forced-treadmill performance across mouse strains. Lerman et al. (2002), for example, found no significant overall relationship between the two paradigms; notably, C57BL/6 J mice showed among the poorest treadmill performance while displaying the greatest voluntary wheel-running distance and high average wheel speed. Accordingly, greater voluntary wheel activity should not be interpreted as evidence of higher $\dot{V}O_2$max or greater exercise capacity, and low wheel activity should not be interpreted as evidence of impaired cardiorespiratory function. When exercise capacity is the outcome of interest, wheel-running exposure should be complemented by a standardized physiological or performance-capacity test (Lerman et al. 2002).

2.4 | Swimming

Swimming permits continuous whole-body exercise without treadmill-associated impact loading and may be useful when running is limited by limb injury, gait impairment, poor compliance, or apparatus availability. Exercise intensity can be manipulated through duration, water current, swimming force, or body-mass-relative external loading. It has been used extensively to assess lactate threshold, lactate minimum, and maximal lactate steady state in rats (Abreu et al. 2016; Gobatto et al. 2001; Manchado et al. 2006). However, swimming presents greater methodological challenges for direct aerobic-capacity assessment because continuous gas-exchange measurement is difficult and workload is affected by technique, floating, bobbing, climbing, diving, body position, and limb use. A load expressed as a percentage of body mass therefore does not guarantee equivalent mechanical work across animals (dos Reis et al. 2018; Rezaee 2025).

Gobatto et al. (2001) established a swimming-based maximal lactate steady-state protocol in Wistar rats using loads of 5%–10% of body mass and showed that training increased the highest workload at which lactate remained stable. Manchado et al. (2006) proposed a non-exhaustive double-bout test in which the change in lactate between efforts was used to estimate a critical load corresponding closely to maximal lactate steady state. Cunha et al. (2009) found that repeated lactate-threshold tests in aged rats produced similar values and estimates that closely approximated the steady-state workload, supporting their use when protocols are validated for the target population. Potolitsyna et al. (2023) reported that a 4% body-mass load best differentiated endurance performance in untrained Wistar rats and that threshold workload differed between animals with long and short swimming times. Thus, the same external load may represent different metabolic intensities according to intrinsic capacity.

Swimming behavior can also reduce reliability. Using a tethered force-measurement system, dos Reis et al. (2018) distinguished continuous surface swimming from bobbing, climbing, and diving. Mechanical force was reproducible at maximal lactate steady-state intensity, whereas blood lactate was less reproducible and appeared sensitive to apnea, stress, and submersion risk. Behavioral observation and direct workload measurement are therefore important components of swimming-based testing. Water temperature, depth, tank dimensions, loading method, habituation, drying, recovery, and the number of animals tested simultaneously should be standardized. Cold water increases thermogenic demand and hypothermia risk, whereas excessively warm water may promote hyperthermia and cardiovascular stress (Rezaee 2025). Group swimming may also allow animals to climb on one another, use the tank wall for support, or interfere with other animals. Individual lanes can reduce these effects. Because inability to remain at the surface can rapidly lead to submersion, swimming tests require continuous observation and conservative, predefined humane stopping criteria.

Swimming additionally imposes endocrine, thermal, and environmental stressors that may be partly dissociated from the intended locomotor workload. In one of the most direct exercise-physiology comparisons, Contarteze et al. (2008) studied exercise-adapted Wistar rats at maximal lactate steady-state intensity and during exhaustive exercise at 25% above that intensity. Serum ACTH was significantly higher during swimming than during treadmill running at both intensities, while corticosterone increased with both exercise modalities but showed a greater response to swimming, particularly at the higher intensity. Thus, even when swimming and treadmill workloads are anchored to a physiological intensity domain, their neuroendocrine responses cannot be assumed to be equivalent (Contarteze et al. 2008). Previous training history also modifies this response. Watanabe et al. (1992) reported that the ACTH response to an acute swimming stress was reduced by 62.4% after 4 weeks of swim training, whereas four or ten weeks of running training reduced the response by only 13.8%–16.4%, indicating substantial mode-specific adaptation as well as more limited cross-adaptation.

Water temperature, immersion, and swimming mechanics represent additional sources of physiological variability. During repeated forced-swim exposure, Rabasa et al. (2013) found that

ACTH adaptation occurred by day 4 in rats swimming at 36°C but was delayed until day 13 at 25°C, where the thermal disturbance was also greater, demonstrating that water temperature materially modifies adaptation to the stressor. Drugan et al. (2005) similarly found moderate-to-severe hypothermia after swim exposure at 20°C–25°C, whereas core temperature decreased by only approximately 1°C at 30°C; corticosterone nevertheless increased after swimming, and confined controls showed an intermediate response, indicating that environmental exposure itself contributes to the measured stress phenotype. Acute forced-swim experiments have also demonstrated duration- and temperature-dependent increases in corticosterone together with changes in glucose, lactate, carbon dioxide, electrolytes, and acid–base balance (Abel 1993). Water depth further modifies these responses: compared with shallower-water exposure, deeper-water testing produced higher serum lactate, lower serum carbon dioxide, and greater metabolic acidosis (Abel 1994). Importantly, the studies by Rabasa et al. (2013), Drugan et al. (2005), and Abel (1993, 1994) used forced-swim stress paradigms rather than aerobic-capacity tests. Their findings should therefore be interpreted as evidence of potential physiological confounding mechanisms rather than as direct validation of swimming-based exercise-capacity assessment. Collectively, the evidence indicates that swimming protocols should report and standardize water temperature and depth, habituation, prior exercise exposure, immersion duration, external loading or tethering, handling, sampling time, submersion behavior, and post-exercise recovery. The absence of treadmill-associated aversive stimulation should not be taken to imply the absence of a substantial stress response.

2.5 | Selection of the Experimental Modality

No modality is universally optimal. Treadmill running combined with indirect calorimetry is the most appropriate approach when the objective is to quantify maximal whole-body oxygen uptake. Incremental or ramp protocols are suitable for $\dot{V}O_2$max, $\dot{V}O_2$peak, and maximal-speed assessment, whereas constant-load tests are more appropriate for endurance tolerance and submaximal metabolic responses. Voluntary wheel running is better suited to habitual activity, self-selected exercise, behavioral phenotypes, and long-term adaptation but does not provide precise control of exercise intensity. Swimming is useful for low-impact exercise and lactate-based assessment but requires strict control of external loading, water conditions, swimming behavior, and humane endpoints. Modalities may also be combined. Voluntary wheel running can serve as a chronic training exposure followed by standardized metabolic treadmill testing of $\dot{V}O_2$max or running economy. Swimming-derived lactate thresholds can be combined with muscle oxidative markers when sustainable metabolic intensity is the primary question. In all cases, modality, workload progression, familiarization, motivational procedures, exhaustion criteria, and physiological outcomes should be aligned with the intended construct and translational objective. Representative protocols and validation studies across treadmill running, voluntary wheel running, and swimming are summarized in Table 1, illustrating the substantial variation in protocol architecture, physiological endpoints, and validation criteria used to characterize aerobic capacity and related exercise phenotypes in rodents.

3 | Direct and Indirect Indices of Aerobic Capacity

Aerobic capacity in rat and mouse models can be evaluated using physiological, performance-based, metabolic, cardiovascular, and tissue-level indices. These outcomes are related but should not be used interchangeably. $\dot{V}O_2$max and $\dot{V}O_2$peak characterize whole-body oxygen transport and utilization; maximal speed, distance, and time to exhaustion reflect integrated exercise performance; lactate-derived thresholds and critical speed indicate sustainable exercise intensity; running economy and heart rate provide information about submaximal efficiency and internal load; and skeletal-muscle markers identify peripheral adaptation. Improved treadmill or swimming performance should therefore not be interpreted as increased maximal aerobic capacity without direct gas-exchange measurement or other validated physiological evidence (Høydal et al. 2007; Marcaletti et al. 2011; Massett et al. 2021; Petrosino et al. 2016).

3.1 | Direct Measurement of Oxygen Uptake

Exercise $\dot{V}O_2$, although measured by indirect calorimetry, is the most direct non-invasive whole-body index of aerobic metabolism in living rodents. Open-circuit calorimetry calculates $\dot{V}O_2$ and $\dot{V}CO_2$ from the differences between inflowing and outflowing gas concentrations and measured airflow. Coupling this system to a motorized treadmill permits simultaneous recording of $\dot{V}O_2$, $\dot{V}CO_2$, and respiratory exchange ratio during progressive changes in speed or inclination (Marcaletti et al. 2011; Petrosino et al. 2016; Speakman 2013). Measurement accuracy depends on chamber volume, airflow, humidity, gas mixing, analyser precision, calibration, sampling frequency, response time, and washout characteristics. Delayed chamber responses may assign oxygen uptake generated during one stage to a subsequent workload, particularly when stages are short. Analysers should therefore be stabilized and calibrated, airflow validated, and system-response time considered when determining stage duration and averaging intervals (Bedford et al. 1979; Marcaletti et al. 2011; Speakman 2013). Baseline metabolism and chamber-related stress may also affect the recorded response. Animals should be acclimatized to the metabolic treadmill, and baseline $\dot{V}O_2$ should be recorded under standardized conditions. Ambient temperature, feeding status, circadian phase, handling, and recent exercise should be controlled because small rodents' high mass-specific metabolic rate makes measurements sensitive to environmental and behavioral variation (Rezaee 2025; Speakman 2013).

3.2 | $\dot{V}O_2$max and $\dot{V}O_2$Peak

$\dot{V}O_2$max is the highest oxygen uptake attained when a further increase in workload produces no meaningful increase in $\dot{V}O_2$. A plateau despite increasing treadmill speed or inclination therefore provides the strongest primary evidence that $\dot{V}O_2$max has been attained. In Sprague–Dawley rats, Wisløff et al. (2001) showed that treadmill inclination markedly affected the highest oxygen uptake achieved; a 25° inclination produced the clearest plateau, together with a respiratory exchange ratio above 1.05 and blood lactate above 6 mmol/L. Similar intensity-controlled

TABLE 1 | Representative protocols and validation studies for the assessment of aerobic capacity and related exercise phenotypes in rats and mice.

Study	Animal model	Primary construct/ outcome	Familiarization/ preparation	Exercise-test protocol	Maximality/endpoint criteria	Main methodological contribution
Bedford et al. (1979)	Rats; Sprague–Dawley, Wistar–Kyoto and spontaneously hypertensive strains; different sexes, ages and training states	$\dot{V}O_2$max	Chamber equilibration before testing; initial stage functioned as low-intensity familiarization/ warm-up	Continuous 10-stage treadmill protocol; speed progressively increased from approximately 8.2 to 43.8 m/min with grade increasing from 0° to 15°; stages generally 3–4 min	$\dot{V}O_2$ increase < 5% despite further increase in workload; inability/ unwillingness to continue	Foundational standardized rat $\dot{V}O_2$max protocol; demonstrated sensitivity to strain, sex, age and training status and high test–retest reliability
Wisløff et al. (2001)	Sprague–Dawley rats	$\dot{V}O_2$max, running economy and exercise intensity	15-min warm-up at approximately 40%–50% $\dot{V}O_2$max	Speed increased by 0.03 m/s every 2 min; several inclinations tested from 0° to 30°	$\dot{V}O_2$ plateau supported by RER > 1.05 and blood lactate > 6 mmol/L	A 25° inclination produced the clearest maximal response in this model; established an intensity-controlled approach for repeated testing and training prescription
Kemi et al. (2002)	C57BL/6 J mice; males and females	$\dot{V}O_2$max and running economy	Two treadmill familiarization exposures; warm-up at approximately 40%–50% $\dot{V}O_2$max	Speed increased by 0.03 m/s every 2 min; inclinations of 0°, 15°, 25°, 35° and 45° were examined	$\dot{V}O_2$ plateau with RER > 1.0 as a supportive criterion	High test–retest reproducibility; $\dot{V}O_2$ responses indicated an optimal inclination range of approximately 15°–35° rather than one universally optimal grade
Leandro et al. (2007)	Male Wistar rats	$\dot{V}O_2$max and individualized training intensity	Preliminary protocol adaptation to the strain and treadmill system	Initial speed approximately 0.3 km/h with increments of 0.3 km/h every 3 min in a multistage protocol; speed and inclination progressively manipulated	Plateau defined by ≤ 5% additional increase in $\dot{V}O_2$ despite greater workload; inability to maintain treadmill speed also considered	Demonstrated the importance of strain-specific protocol optimization and repeated $\dot{V}O_2$max assessment for adjustment of training intensity
Copp et al. (2009)	Male Sprague–Dawley rats	$\dot{V}O_2$peak and endurance-capacity reproducibility	5–10 min/day for 5 days at 25 m/min and 10% grade	$\dot{V}O_2$peak test: 25 m/min for 2–3 min, then 40 m/min, followed by approximately 5–10 m/min increases each minute. Separate progressive endurance test also performed	Inability/unwillingness to maintain belt speed; gait change and/or absence of further $\dot{V}O_2$ increase used to support successful maximal testing; RER was supportive but not sufficient alone	Demonstrated excellent within-rat reproducibility: overall $\dot{V}O_2$peak CV ≈ 0.06 and consecutive-test CV ≈ 0.03–0.04 across repeated measurements

(Continues)

TABLE 1 | (Continued)

Study	Animal model	Primary construct/outcome	Familiarization/preparation	Exercise-test protocol	Maximality/endpoint criteria	Main methodological contribution
Ayachi et al. (2016)	FVB mice, $n = 8$	$\dot{V}O_2$max versus $\dot{V}O_2$peak; protocol dependency	Four 10-min familiarization sessions with progressively increasing speeds up to 9 m/min and 48-h intervals	Incremental protocol: 10 m/min +3 m/min every 3 min; Ramp3: start 3 m/min, acceleration 3 m/min^2; Ramp12: acceleration 12 m/min^2; all tested at 0° and 25°	$\dot{V}O_2$ plateau maintained for ≥ 30 s with lactate and RER used as secondary evidence; exhaustion based on inability to maintain speed despite standardized stimulation	Demonstrated strong protocol dependence of $\dot{V}O_2$peak. The slow 3 m/min^2 ramp at 0° elicited the highest and clearest maximal response. The source paper contains an internal discrepancy regarding sex between its abstract and Methods, so sex is intentionally not specified here.
Petrosino et al. (2016)	C57BL/6 J and FVB/NJ mice; healthy and cardio-metabolic dysfunction models	Integrated cardiometabolic phenotype: $\dot{V}O_2$max, a submaximal metabolic transition point reported by the authors as "anaerobic threshold," substrate crossover, and maximal performance	Three treadmill acclimation sessions separated by approximately 60 h, followed by 1 week of rest	Graded mouse exercise test with simultaneous progression of speed and grade: approximately 6 m/min/0°, 9/5°, 12/10°, 15/15°, followed by higher speeds at 15°	Maximal exertion, RER ≥ 1.0, plateau/decrease after peak $\dot{V}O_2$, post-exercise lactate increase and inability to continue; 5-s continuous grid contact used as an operational endpoint	Demonstrated that a staged speed-plus-inclination GXT can provide a broader cardio-metabolic phenotype than conventional progressive treadmill testing
Picoli et al. (2018)	Male Swiss mice	Peak running velocity (Vpeak), v$\dot{V}O_2$max and training prescription	Three days of treadmill adaptation with progressive increases in speed and duration	5-min warm-up at 10 cm/s; initial test speed 19 cm/s followed by +9 cm/s every 3 min at 0° until exhaustion	Inability to remain running in the final third of the treadmill for > 10 s; incomplete final stage incorporated into adjusted Vpeak calculation	Demonstrated that Vpeak can be a practical, low-cost variable for exercise prescription, but it should be interpreted as a performance measure rather than a direct measurement of $\dot{V}O_2$max
Qin et al. (2020)	Male Wistar rats of different ages	Age-specific $\dot{V}O_2$max and relative exercise intensity (%$\dot{V}O_2$max)	One week of treadmill familiarization; 5-min warm-up at approximately 40%–50% estimated $\dot{V}O_2$max	Three protocols compared: combined speed/grade progression (P1), speed-only progression (P2) and speed followed by increasing grade (P3); stages generally 3 min	Any two of the following: $\dot{V}O_2$ increase < 5% despite increased workload, RER > 1.00 and observable exhaustion characteristics	Demonstrated that optimal protocol architecture changes with age. Speed-dominant protocols were more suitable for young rats, whereas slope-based progression reduced stress/injury and improved maximal testing in older rats.

(Continues)

TABLE 1 | (Continued)

Study	Animal model	Primary construct/ outcome	Familiarization/ preparation	Exercise-test protocol	Maximality/endpoint criteria	Main methodological contribution
Copp et al. (2010)	Male Sprague–Dawley rats	Critical speed (CS)	$\dot{V}O_2$max and the corresponding running speed determined before CS trials	Five randomized constant-speed trials at 95%, 100%, 105%, 110% and 115% of the speed eliciting $\dot{V}O_2$max; recovery between trials	Inability to maintain treadmill pace despite evident effort; gait alteration and other behavioral indicators used to validate exhaustion	Validated CS as a metabolic/ performance boundary in rats. Below CS, $\dot{V}O_2$ stabilized at a submaximal value; above CS, $\dot{V}O_2$ increased to $\dot{V}O_2$max. CS was approximately 48.6 m/min.
Ferreira et al. (2007)	Five-month-old male C57BL/6 J mice	Running maximal lactate steady state (MLSS)	Incremental test first used to determine peak workload	Incremental test: 6–33 m/ min with +3 m/min every 3 min at 0%. MLSS: 10-min warm-up at 6 m/ min followed by 28-min constant-load trials; workload adjusted by 3 m/min between sessions	MLSSw defined as the highest workload with < 1 mmol/L rise in blood lactate between min 7 and 28	MLSSw = 15.1 ± 0.7 m/min, corresponding to approximately 60 ± 2% of maximal running speed; intra- and interobserver lactate measurements were highly reproducible.
Lightfoot et al. (2001)	Female mice from 10 inbred strains	Maximal treadmill performance/genetic variability	Two treadmill-orientation exposures separated by ≥ 2 days	Start at 22 m/min; speed increased to 28.4, 36.6 and 42.4 m/min every 4 min; thereafter grade increased by 2% every 4 min	Inability or refusal to run forward from the shock grid	Demonstrated substantial strain-dependent variation in maximal treadmill performance and showed that behavioral compliance may strongly affect apparently physiological exercise outcomes
Lerman et al. (2002)	Male mice from seven strains	Forced treadmill performance versus voluntary wheel-running performance	Three 15-min treadmill acclimation sessions over 2 weeks	Endurance: 30 min at 20 m/min and 7°. Graded stress test: start 20 m/min and 7°, +1.5 m/min every 2 min. Voluntary wheel activity recorded separately over 2 weeks	Treadmill failure defined as inability to continue regular running despite standardized additional stimulation	Demonstrated marked strain effects and, importantly, no significant correlation between forced treadmill and voluntary wheel performance, indicating that the two paradigms represent different exercise phenotypes.
Gobatto et al. (2001)	Adult male Wistar rats	Swimming MLSS	Progressive water adaptation	20-min swimming trials with external loads corresponding to 5%, 6%, 7%, 8%, 9% or 10% of body	Highest workload at which blood lactate remained stable	Established swimming MLSS in rats; sedentary animals showed MLSS at approximately 5–6% body-

(Continues)

TABLE 1 | (Continued)

Study	Animal model	Primary construct/ outcome	Familiarization/ preparation	Exercise-test protocol	Maximality/endpoint criteria	Main methodological contribution
				mass; lactate measured every 5 min		mass load and approximately 5.5 mmol/L blood lactate; training shifted the sustainable workload upward
Manchado et al. (2006)	Adult male Wistar rats	Critical load and non-exhaustive estimation of MLSS	15-day progressive water familiarization	Randomized swimming loads of 4%, 6%, 7% and 8% body mass; two 5-min bouts separated by 2 min; subsequent 25-min validation trials at 90%, 100% and 110% of estimated critical load	MLSS defined as ≤ 1 mmol/L increase in blood lactate between min 10 and 25	Estimated critical load ≈ 4.79% body mass and closely corresponded to MLSS, supporting a non-exhaustive method for identifying the aerobic–anaerobic transition
dos Reis et al. (2018)	Male Wistar rats, approximately 120 days old	Swimming MLSS and mechanical workload reproducibility	Extended forced-swimming familiarization	Tethered swimming; 3–4 constant-force trials of approximately 30 min at individually adjusted workloads, with 48-h intervals	MLSS defined by < 1 mmol/L change in lactate between min 10 and 30; swimming-force and behavioral patterns continuously assessed	Mechanical exercise work was highly reproducible (ICC ≈ 0.90), whereas blood lactate showed substantially poorer test–retest reproducibility, highlighting the influence of swimming behavior, apnea and stress on physiological endpoints.

Abbreviations: CS, critical speed; CV, coefficient of variation; GXT, graded exercise test; ICC, intraclass correlation coefficient; MLSS, maximal lactate steady state; MLSSw, maximal lactate steady-state workload; RER, respiratory exchange ratio; $\dot{V}O_2$, oxygen uptake; $\dot{V}O_2$max, maximal oxygen uptake; $\dot{V}O_2$peak, peak oxygen uptake; Vpeak, peak running velocity; v$\dot{V}O_2$max, velocity associated with $\dot{V}O_2$max.

methods demonstrated high test–retest reliability in C57BL/6 J mice (Kemi et al. 2002; Wisløff et al. 2001).

Representative values illustrate both the magnitude of rodent maximal oxygen uptake and its dependence on species, protocol, training status, and normalization method. In Sprague–Dawley rats, Wisløff et al. (2001) reported peak oxygen uptake values ranging from 64.5 ± 1.9 at 0° to 79.9 ± 3.1 $mL \cdot kg^{-1} \cdot min^{-1}$ at 25°, with the latter inclination producing the clearest maximal response. In untrained C57BL/6 J mice tested at 25°, Kemi et al. (2002) reported $\dot{V}O_2$peak values of 48.1 ± 3.8 and 51.7 ± 3.1 $mL \cdot kg^{-0.75} \cdot min^{-1}$ in females and males, respectively; after endurance training, corresponding values increased to 75.4 ± 3.2 and 65.0 ± 2.8 $mL \cdot kg^{-0.75} \cdot min^{-1}$. In FVB/N mice, Ayachi et al. (2016) reported a highest $\dot{V}O_2$peak of 59.0 $mL \cdot kg^{-0.75} \cdot min^{-1}$ during a slowly accelerating 0° ramp protocol. These values should be regarded as protocol- and population-specific examples rather than universal reference ranges, particularly because conventional body-mass normalization and allometric scaling are not numerically interchangeable.

A plateau is not always observed because exercise may terminate through behavioral refusal, inability to maintain belt speed, abnormal gait, repeated contact with the motivational device, or locomotor limitation. The absence of a plateau during a single incremental test, however, does not necessarily establish that the physiological maximum was not attained. Where an appropriately validated verification procedure is available, it may provide additional evidence of maximality; otherwise, when maximality cannot be demonstrated, the highest observed oxygen uptake should be reported as $\dot{V}O_2$peak. $\dot{V}O_2$peak is particularly sensitive to treadmill grade, acceleration, stage duration, familiarization, environmental conditions, and exhaustion criteria (Ayachi et al. 2016; Marcaletti et al. 2011; Petrosino et al. 2016).

Protocol-comparison studies further demonstrate this dependency. Ayachi et al. (2016) found that a slowly accelerating ramp elicited a higher and more clearly defined maximal response than alternative stepwise protocols in mice. Petrosino et al. (2016) showed that progressive changes in speed and inclination could produce a broader physiological phenotype, including $\dot{V}O_2$max, maximal workload, running economy, a submaximal metabolic transition point, and substrate crossover. Qin et al. (2020) further demonstrated that age and treadmill grade modified the workload–$\dot{V}O_2$ relationship in rats. Failure to observe a plateau may therefore reflect unsuitable protocol architecture rather than the absence of a physiological maximum.

In the absence of a demonstrable $\dot{V}O_2$–work-rate plateau, fixed secondary criteria based on respiratory exchange ratio, blood lactate concentration, or maximal heart rate should not be used to validate $\dot{V}O_2$max. Human studies have demonstrated substantial interindividual variability in these responses and shown that arbitrary cut-off values may both accept a submaximal $\dot{V}O_2$ as maximal and reject tests in which a true maximum has been attained (Poole et al. 2008; Poole and Jones 2017). RER and blood lactate nevertheless remain physiologically informative variables because they characterize substrate use, acid–base and metabolic responses, and the severity of the exercise challenge; their role should therefore be descriptive and complementary rather than confirmatory of maximal oxygen uptake. In human exercise testing, a subsequent constant-work-rate verification bout has been proposed as a more objective means of establishing whether $\dot{V}O_2$max has been reached (Poole and Jones 2017). Similar verification strategies, however, cannot be assumed to be valid in rats and mice without model-specific evaluation. Rodent studies should therefore report the complete $\dot{V}O_2$ response, the operational plateau criterion, test-termination procedures, and any secondary physiological responses separately rather than using RER or blood lactate thresholds alone to classify an observed peak as $\dot{V}O_2$max. $\dot{V}O_2$max and $\dot{V}O_2$peak also depend on data processing. A single instantaneous value may reflect analyser noise, whereas an excessively long averaging interval may obscure a brief maximum. Studies should therefore report the averaging method and explicitly define the maximum permitted increase in $\dot{V}O_2$ used to identify a plateau.

3.3 | Normalization of Oxygen Uptake

$\dot{V}O_2$ is commonly expressed in milliliters per kilogram per minute, although this assumes a proportional relationship between oxygen uptake and body mass. Ratio normalization may bias comparisons involving different ages, sexes, strains, disease phenotypes, or body-composition profiles because larger animals may appear to have lower aerobic capacity even when $\dot{V}O_2$ does not scale linearly with mass (Bedford et al. 1979; Høydal et al. 2007; Speakman 2013). Allometric scaling has therefore been proposed. Høydal et al. (2007) used an exponent of 0.75, whereas values closer to 0.67 have also been supported. Bedford et al. (1979) recommended examining the logarithmic relationship between body mass and $\dot{V}O_2$max rather than relying solely on simple ratios. However, no exponent is universally appropriate. This issue is especially relevant in obesity, cachexia, growth, aging, and disease, where total mass and metabolically active tissue may change disproportionately. Statistical adjustment for body or lean mass may therefore be preferable in some designs. Studies should report absolute $\dot{V}O_2$, conventional mass-specific values, any allometric or statistical adjustment, and the rationale for the selected approach (Bedford et al. 1979; Høydal et al. 2007; Speakman 2013).

3.4 | Running Economy and Submaximal Oxygen Cost

Running economy describes the $\dot{V}O_2$ required to maintain a standardized submaximal speed and grade. Lower $\dot{V}O_2$ at the same absolute workload indicates improved economy when mechanical and environmental conditions are comparable. Running economy is distinct from $\dot{V}O_2$max. Training may reduce the oxygen cost of running without altering maximal oxygen transport, allowing longer exercise or higher speed at the same metabolic demand. Improvements following intensity-controlled training have been demonstrated in rats and mice, together with changes in the speed–$\dot{V}O_2$ relationship (Høydal et al. 2007; Kemi et al. 2002; Wisløff et al. 2001). Consequently, a treadmill speed prescribed at a given relative intensity at baseline may represent a progressively lower relative

physiological stimulus as aerobic capacity or running economy improves. In longitudinal training studies designed to maintain a constant relative exercise intensity, $\dot{V}O_2$max, $\dot{V}O_2$peak, or the relevant submaximal physiological response should therefore be reassessed periodically and the treadmill workload adjusted accordingly. By contrast, when the objective is to determine whether running economy has changed over time, speed and grade should remain fixed so that $\dot{V}O_2$ can be compared at the same absolute external workload. Thus, workload adjustment is appropriate for maintaining relative training intensity, whereas workload standardization is required for longitudinal assessment of exercise economy (Høydal et al. 2007; Kemi et al. 2002; Wisløff et al. 2001).

3.5 | Maximal and Peak Running Speed

Maximal running speed is generally defined as the highest completed treadmill speed during an incremental test. The velocity at which $\dot{V}O_2$max is first attained and the peak speed reached before termination may differ because an animal can continue running after $\dot{V}O_2$ has plateaued. At workloads capable of eliciting $\dot{V}O_2$max, maximal oxygen uptake may therefore be attained before exhaustion when time to limitation (Tlim) exceeds approximately three $\dot{V}O_2$ time constants ($3\tau\dot{V}O_2$), after accounting for any response delay; the physiological and methodological basis of this relationship is discussed in Section 4.4 (Behnke et al. 2002; Poole et al. 2020). During submaximal exercise, $\dot{V}O_2$ usually increases approximately linearly with speed. Significant associations between maximal speed and oxygen uptake have been reported in healthy and diabetic rats, supporting speed as a useful functional index when gas analysis is unavailable (Høydal et al. 2007; Rodrigues et al. 2007). Nevertheless, maximal speed is not a direct substitute for $\dot{V}O_2$max. Running beyond the $\dot{V}O_2$ plateau may be supported by anaerobic reserves and influenced by neuromuscular function, economy, gait, motivation, and treadmill tolerance. Similar $\dot{V}O_2$max values may therefore coexist with different peak speeds, and vice versa. In models with locomotor or neuromuscular impairment, the attainable workload may be limited by motor dysfunction; in such cases, peak speed should be interpreted as a performance outcome constrained by locomotor capacity rather than as a proxy for oxygen-transport capacity.

Peak running speed should also be distinguished from critical speed (CS). Peak speed is the highest velocity attained during an incremental test, whereas CS is derived from the speed–duration relationship across multiple constant-speed trials and represents a functional boundary between exercise that can be sustained for substantially longer durations and exercise associated with progressive metabolic instability and eventual exhaustion. Rodent studies have demonstrated the applicability of the CS construct in both mice and rats (Billat et al. 2005; Copp et al. 2010). Thus, peak speed and CS provide complementary but non-interchangeable information: peak speed characterizes maximal treadmill performance, whereas CS provides information about sustainable exercise intensity. The physiological and methodological interpretation of CS is discussed in detail in Section 3.8. Peak velocity can support training prescription, as similar adaptations have been obtained using peak speed and the velocity associated with $\dot{V}O_2$max, but it should remain classified as a performance-derived measure validated for the relevant species, strain, and protocol (Høydal et al. 2007; Picoli et al. 2018; Rodrigues et al. 2007).

3.6 | Time to Exhaustion and Total Distance

Time to exhaustion and total distance are inexpensive and functionally meaningful measures of whole-animal exercise tolerance, but their physiological interpretation is protocol dependent. Rapid high-speed tests emphasize locomotor power and anaerobic metabolism, whereas prolonged constant-load tests place greater demands on oxidative metabolism, substrate availability, thermoregulation, and fatigue resistance. Speed, grade, stage duration, expected test length, and exhaustion criteria should therefore be fully reported (Marcaletti et al. 2011; Massett et al. 2021). These outcomes are also influenced by motivation, stress, pain sensitivity, familiarization, gait, body mass, glycogen availability, temperature, and observer decisions. Protocol heterogeneity consequently contributes to variation in reported endurance adaptations (Massett et al. 2021). Strain comparisons illustrate the behavioral contribution. Lightfoot et al. (2001) found substantial differences in treadmill duration among inbred strains, but low performance reflected behavioral refusal in some strains and inability to maintain speed in others. Lerman et al. (2002) similarly demonstrated genetic variation in forced treadmill and voluntary wheel-running performance, with different rankings between modalities. Time and distance should therefore be reported as endurance or performance outcomes unless direct physiological evidence confirms aerobic limitation (Lerman et al. 2002; Lightfoot et al. 2001).

3.7 | Lactate-Derived Indices

Blood lactate measurements characterize the transition from workloads at which lactate appearance and removal remain balanced to those associated with progressive accumulation. They are useful for identifying sustainable intensity and prescribing training but do not directly measure maximal whole-body oxygen transport (Abreu et al. 2016). Maximal lactate steady state is generally defined as the highest constant workload at which blood lactate remains stable or increases only minimally. It is usually determined through repeated constant-load trials conducted on separate days. In swimming rats, maximal lactate steady state occurred at approximately 6% of body mass and at concentrations higher than those often reported during treadmill exercise. Differences between swimming and treadmill protocols indicate that fixed concentrations such as 4 mmol/L should not be universally applied (Abreu et al. 2016; Gobatto et al. 2001; Manchado et al. 2006). Lactate-based assessment has also been applied to aged rats, obese Zucker rats, and animals with different intrinsic endurance levels. These findings indicate that thresholds must be established within the relevant age, disease phenotype, performance level, and exercise modality rather than transferred directly between models (Almeida et al. 2013; Cunha et al. 2009; Potolitsyna et al. 2023). Incremental lactate tests are less time-consuming but depend on stage duration, workload increments,

sampling time, and the visual or mathematical method used to determine the breakpoint. Similarly, the lactate-minimum test is influenced by the method used to induce hyperlactatemia, recovery duration, workload progression, sampling schedule, and curve-fitting procedure. Identically labeled thresholds may therefore reflect different physiological and mathematical constructs (Abreu et al. 2016). Repeated blood sampling may increase stress, alter behavior, and reduce hemoglobin concentration if total sampling volume is not controlled. These issues are especially important in mice because of their small circulating blood volume. Sampling site, timing, and volume should therefore be standardized and minimized (Abreu et al. 2016; Rezaee 2025).

3.8 | Critical Speed

Critical speed models the hyperbolic relationship between treadmill speed and exercise tolerance using several constant-speed trials to exhaustion. As trial duration increases, tolerable speed approaches an asymptote defined as critical speed; the model may also estimate a finite distance or work capacity available above this boundary. Critical speed represents the transition between sustainable and non-sustainable exercise rather than maximal oxygen transport. Below critical speed, $\dot{V}O_2$ and other metabolic variables may stabilize; above it, metabolic disturbance and oxygen uptake progressively increase until exhaustion. The applicability of this construct to mice was demonstrated by Billat et al. (2005), who derived critical running speed from four exhaustive treadmill trials in CD1, C57BL/6 J, and FVB/N mice. The distance–time relationships showed coefficients of determination greater than 0.98 across strains and sexes, supporting the mathematical applicability of the model in mice. Critical speed differed substantially among strains, averaging 24.1 ± 4.6 m/min in male CD1, 22.1 ± 2.8 m/min in male FVB/N, and 18.0 ± 1.5 m/min in male C57BL/6 J mice, whereas no sex difference was detected within the C57BL/6 J strain. Critical speed was also closely related to lactate-threshold velocity, with mean values of 19.6 ± 3.6 and 19.7 ± 2.9 m/min, respectively, across the subgroup in which both were measured. These findings demonstrate that critical speed can provide a reproducible functional index in mice while also emphasizing substantial strain- and individual-dependent variability (Billat et al. 2005). In rats, exercise above critical speed alters hindlimb blood-flow distribution and increases recruitment of highly glycolytic muscle fibers (Copp et al. 2010). Critical speed may be more closely related to endurance tolerance than $\dot{V}O_2$max because animals with similar maximal uptake may sustain different fractions of that capacity. However, estimates depend on trial selection, exhaustion-time range, recovery, mathematical modeling, motivation, and treadmill competence. Critical speed should therefore be interpreted as an integrated functional threshold rather than a substitute for $\dot{V}O_2$max (Burnley 2023; Copp et al. 2010; Rezaee 2025).

3.9 | Heart Rate and Cardiovascular Indices

Heart rate increases with treadmill workload and oxygen consumption and can help characterize the relationship between external speed, metabolic demand, and cardiovascular strain (Brooks and White 1978). Under standardized conditions, it may provide a reasonable estimate of relative exercise intensity (Høydal et al. 2007). Continuous measurement generally requires implanted telemetry or transducers. Implantation introduces surgical stress and recovery requirements, while device mass is particularly relevant in mice. Motion, electrode displacement, and electromagnetic interference may also affect signal quality (Høydal et al. 2007; Rezaee 2025). Heart rate should therefore be considered a complementary internal-load measure rather than a direct index of oxygen consumption. Its relationship with $\dot{V}O_2$ may be modified by autonomic dysfunction, medication, disease, temperature, stress, and training and should be validated within each model. Cardiac output and regional or microvascular skeletal-muscle blood flow can help identify central and peripheral limitations to oxygen delivery, particularly in disease models, but they are not standalone measures of aerobic capacity (Musch and Terrell 1992; Kindig et al. 1999; Diederich et al. 2002).

3.10 | Skeletal-Muscle Oxidative and Mitochondrial Markers

Citrate synthase, succinate dehydrogenase, cytochrome-c oxidase, mitochondrial respiratory capacity, capillary characteristics, and fiber-type composition provide mechanistic evidence of peripheral aerobic adaptation (Holloszy 1967; Fitts et al. 1975; Musch et al. 1986). Classic rat experiments demonstrated large training-induced increases in mitochondrial respiratory capacity and respiratory-enzyme activity, while subsequent work linked citrate synthase, cytochrome c, and respiratory capacity with endurance performance (Holloszy 1967; Fitts et al. 1975). Training-induced increases in oxidative-enzyme activity have also been reported alongside increases in $\dot{V}O_2$max in both healthy and disease-model rodents (Kemi et al. 2002; Leandro et al. 2007; Musch et al. 1986). However, a sample obtained from one muscle does not necessarily represent the total musculature recruited during whole-body exercise. $\dot{V}O_2$max also depends on pulmonary gas exchange, hemoglobin concentration, cardiac output, vascular conductance, blood-flow distribution, neuromuscular recruitment, and motivation. Mitochondrial adaptations may therefore occur without increased $\dot{V}O_2$max when another component remains limiting, and improved oxygen delivery may increase $\dot{V}O_2$max without equivalent changes in every muscle. Tissue markers should be described as mechanistic correlates rather than substitutes for $\dot{V}O_2$max or $\dot{V}O_2$peak (Leandro et al. 2007; Marcaletti et al. 2011; Massett et al. 2021; Rezaee 2025).

3.11 | Integrated Interpretation

No single outcome fully characterizes aerobic function. Directly measured $\dot{V}O_2$max or $\dot{V}O_2$peak provides the strongest whole-body physiological endpoint but does not describe sustainable intensity, running economy, behavioral tolerance, or the underlying mechanisms. Maximal speed and time to exhaustion provide functional information but are influenced by anaerobic capacity, neuromuscular performance, and motivation. Lactate-derived thresholds and critical speed identify sustainable metabolic boundaries, heart rate reflects cardiovascular strain,

and tissue markers characterize peripheral adaptation. The most informative designs combine complementary outcomes across physiological levels. Direct gas exchange may be evaluated alongside performance, sustainable-intensity, and mechanistic measures. When only indirect measures are available, conclusions should be restricted to the specific performance or endurance outcome assessed rather than generalized to maximal aerobic capacity. The physiological meaning, principal methodological strengths and limitations, and recommended terminology for the major outcomes used to characterize aerobic function in rodent models are summarized in Table 2.

4 | Methodological Validity, Standardization, and Sources of Variability

The validity of rodent aerobic-capacity assessment depends on whether the selected protocol measures the intended physiological construct rather than a combination of aerobic function, locomotor competence, behavioral compliance, stress tolerance, and familiarity with the apparatus. This distinction is particularly important in forced treadmill testing, where termination may reflect physiological limitation, inability to meet the mechanical demands of the belt, behavioral refusal, or intolerance of the motivational procedure. A methodologically valid protocol should therefore elicit the intended response, distinguish cardiorespiratory limitation from non-metabolic termination, and produce reproducible outcomes under standardized conditions (Booth et al. 2010; Høydal et al. 2007; Poole et al. 2020).

4.1 | Construct and Criterion Validity

As discussed in Section 3, the terminology used should correspond to the outcome actually measured. Direct gas-exchange measurements are required when the objective is to quantify maximal whole-body oxygen uptake, whereas maximal speed, work, distance, and time to exhaustion should be described as performance outcomes unless their relationship with $\dot{V}O_2$ has been validated in the relevant species, strain, age group, disease model, and protocol. Although running speed is associated with $\dot{V}O_2$ during submaximal exercise, the relationship becomes less precise near maximal intensity because animals may continue running after oxygen uptake has plateaued through contributions from anaerobic reserves, running economy, neuromuscular function, gait, and motivation (Ayachi et al. 2016; Høydal et al. 2007; Marcaletti et al. 2011; Poole et al. 2020). Criterion validity is strengthened when performance outcomes are evaluated alongside direct gas-exchange measurements. Simultaneous or parallel assessment of $\dot{V}O_2$, $\dot{V}CO_2$, respiratory exchange ratio, maximal speed, running time, and reason for termination helps distinguish physiological adaptation from behavioral or locomotor influences. A protocol that differentiates trained from sedentary animals on the basis of running time may demonstrate discriminatory validity, but this alone does not confirm a change in maximal aerobic capacity (Marcaletti et al. 2011; Petrosino et al. 2016). Verification testing may provide additional support for maximal oxygen-uptake assessment. A progressive ramp test can be followed by a constant-speed or constant-load bout performed above the maximal workload achieved during the initial test. Failure to obtain a higher $\dot{V}O_2$ during verification supports the interpretation that the initial value represented $\dot{V}O_2$max. However, verification workload, recovery interval, and duration should be adapted to the animal model, and the additional burden should be balanced against incomplete recovery and conditioning effects (Poole et al. 2020).

4.2 | Reliability and Measurement Error

Test–retest reliability is essential for interpreting longitudinal changes and between-group differences. Reproducible oxygen-uptake measurements have been demonstrated in C57BL/6 J mice and Sprague–Dawley rats when treadmill workload, familiarization, warm-up, environmental conditions, motivational procedures, and gas-analysis methods were standardized. Kemi et al. (2002) reported a test–retest correlation of 0.98 and a coefficient of variation of 9.1% for submaximal oxygen uptake in mice, while Wisløff et al. (2001) demonstrated reproducible $\dot{V}O_2$max measurements in rats. Repeated measurements reviewed by Poole et al. (2020) also produced within-animal variation close to the expected technical error of the measurement system. Reliability should nevertheless be established within each laboratory. Differences in treadmill dimensions, belt mechanics, chamber volume, airflow, analyser technology, calibration, animal handling, motivational systems, and termination criteria may alter measurement error. Pilot test–retest studies are particularly important when introducing a new species, strain, age group, disease model, apparatus, or gas-analysis system. Reliability should be reported using appropriate agreement statistics, including intraclass or test–retest correlations, coefficients of variation, typical error, and, where possible, limits of agreement. A non-significant difference between repeated tests does not demonstrate acceptable agreement. The expected intervention effect should also be compared with measurement error; when the anticipated change is similar to the typical error, modest differences may not represent genuine biological adaptation. Repeated maximal tests may themselves improve treadmill competence and induce a minor conditioning effect. The number of baseline tests, intervals between tests, and total pre-intervention exercise exposure should therefore be identical across groups and reported explicitly (Poole et al. 2020).

4.3 | Familiarization and Learning Effects

The principles of familiarization were described in Section 2; methodologically, its primary purpose is to stabilize running behavior without producing a training adaptation. The required duration may vary according to strain, age, disease status, previous handling, and apparatus complexity. One or two sessions may be sufficient for some C57BL/6 J mice, whereas other populations may require longer exposure (Marcaletti et al. 2011). Familiarization should be conducted using the apparatus intended for formal testing because behavior may differ between an open multi-lane treadmill and an enclosed metabolic treadmill. Kemi et al. (2002) familiarized mice before reliability assessment, while Wisløff et al. (2001) observed rapid learning of treadmill and rear-grid avoidance in rats. The

TABLE 2 | Physiological interpretation, methodological strengths, limitations, and recommended terminology of outcomes used to characterize aerobic function in rodent models.

Outcome	Primary physiological construct	Typical assessment approach	Simple and inexpensive; associated with $\dot{V}O_2$ in validated models (Høydal et al. 2007; Rodrigues et al. 2007).	Principal limitations/ confounders	Report as maximal/peak running speed; not as $\dot{V}O_2$max (Høydal et al. 2007; Picoli et al. 2018).
$\dot{V}O_2$max	Upper limit of integrated whole-body oxygen uptake, transport, delivery, and utilization during maximal dynamic exercise	Progressive or ramp treadmill exercise combined with indirect calorimetry	Most direct in vivo whole-body physiological index of maximal aerobic capacity; integrates pulmonary gas exchange, cardiovascular function, oxygen transport, peripheral perfusion, and muscle oxygen utilization	Influenced by treadmill grade, acceleration, stage duration, test duration, familiarization, metabolic-chamber response characteristics, environmental conditions, locomotor competence, motivation, and maximality criteria	Use $\dot{V}O_2$max when physiologically defensible evidence indicates that maximal oxygen uptake has been reached, preferably including a $\dot{V}O_2$ plateau despite further workload increase. Secondary criteria such as RER, lactate, HR, or gait should support rather than independently define $\dot{V}O_2$max
$\dot{V}O_2$peak	Highest oxygen uptake actually attained during a specific exercise test	Progressive or ramp exercise with indirect calorimetry when formal maximality cannot be established	Provides an objective whole-body gas-exchange endpoint when a clear maximal response cannot be demonstrated	Highly protocol dependent; premature locomotor failure, behavioral refusal, inadequate treadmill compliance, unsuitable grade or acceleration, and test termination may prevent attainment of a physiological maximum	Use $\dot{V}O_2$peak when the highest measured $\dot{V}O_2$ cannot be confidently classified as $\dot{V}O_2$max. $\dot{V}O_2$peak and $\dot{V}O_2$max should not be used interchangeably
Running economy/ submaximal oxygen cost	Metabolic efficiency of locomotion at a standardized submaximal external workload	$\dot{V}O_2$ measured at a fixed treadmill speed and grade	Sensitive to training adaptation independent of $\dot{V}O_2$max; useful for longitudinal assessment and for interpreting changes in relative exercise intensity	Requires highly standardized speed, grade, body size, gait, environmental conditions, and testing procedures; changes in economy alter the relationship between speed and physiological intensity over time	Report as running economy or submaximal oxygen cost. A lower $\dot{V}O_2$ at the same standardized workload indicates improved economy but does not by itself indicate increased maximal aerobic capacity
Maximal/peak running speed	Integrated maximal locomotor exercise performance	Progressive treadmill exercise until the highest completed or tolerated running speed	Simple, inexpensive, useful for functional phenotyping and exercise prescription; can correlate strongly with $\dot{V}O_2$ within a validated model	Influenced by anaerobic reserve, economy, neuromuscular function, gait, coordination, motivation, body mass, treadmill competence, and aversive-stimulus tolerance. Animals may continue running after $\dot{V}O_2$ has plateaued. Speed–$\dot{V}O_2$ relationships differ with disease, training, age, and model	Report as maximal running speed, peak running speed, or peak velocity. Treat it as a performance-derived outcome. It may serve as a validated surrogate or prescription variable within a specific model but should not be labeled $\dot{V}O_2$max without direct gas-exchange measurement

(Continues)

TABLE 2 | (Continued)

Outcome	Primary physiological construct	Typical assessment approach	Simple and inexpensive; associated with $\dot{V}O_2$ in validated models (Høydal et al. 2007; Rodrigues et al. 2007).	Principal limitations/ confounders	Report as maximal/peak running speed; not as $\dot{V}O_2$max (Høydal et al. 2007; Picoli et al. 2018).
Time to exhaustion (TTE)	Whole-animal endurance tolerance at a defined workload or during a specified progressive protocol	Constant-load or progressive treadmill/swimming exercise continued to predefined exhaustion	Inexpensive, functionally meaningful, sensitive to training or disease effects, and feasible without metabolic equipment	Strongly protocol dependent; affected by workload, test duration, anaerobic contribution, substrate availability, thermoregulation, gait, familiarization, motivation, stress, pain sensitivity, motivational procedures, and investigator-defined stopping criteria	Report explicitly as time to exhaustion, exercise tolerance, or endurance performance. Improvement in TTE alone should not be described as increased $\dot{V}O_2$max
Total running/ swimming distance	Integrated exercise-performance or endurance capacity	Distance accumulated during treadmill, swimming, exhaustion, or voluntary-exercise protocols	Simple global functional endpoint that is useful for phenotype screening and intervention studies	Depends on speed profile, duration, workload progression, biomechanical efficiency, modality, motivation, behavior, and termination criteria	Report as exercise-performance distance or endurance distance. Do not interpret distance alone as maximal aerobic capacity
Critical speed (CS)	Functional metabolic boundary separating sustainable from non-sustainable high-intensity exercise domains	Multiple constant-speed trials to exhaustion followed by mathematical modeling of speed–time or distance–time relationships	Provides a physiologically meaningful estimate of sustainable exercise capacity; distinguishes endurance characteristics not captured by $\dot{V}O_2$max alone	Sensitive to selection and number of trials, exhaustion-time range, recovery interval, mathematical model, motivation, treadmill competence, and exhaustion criteria	Report as critical speed. Interpret CS as a sustainable-intensity/metabolic boundary and integrated endurance parameter rather than a substitute for $\dot{V}O_2$max
Lactate threshold/ lactate-derived threshold	Exercise intensity associated with a defined change in the blood-lactate response and transition toward progressively greater lactate accumulation	Incremental exercise with serial blood-lactate measurements and a predefined physiological, fixed-concentration, or mathematical criterion	Useful for individualized submaximal intensity prescription and characterization of the transition toward progressively greater lactate accumulation	Highly sensitive to sampling site and timing, workload increments, stage duration, exercise modality, species, strain, training status, mathematical method, stress, and threshold definition. Different methods do not necessarily identify the same physiological point	Report the specific lactate-threshold method and definition used. Prefer terms such as lactate threshold or lactate-derived threshold when applicable; avoid an unspecified "anaerobic threshold." A universal fixed lactate concentration should not be assumed

(Continues)

TABLE 2 | (Continued)

Outcome	Primary physiological construct	Typical assessment approach	Simple and inexpensive; associated with $\dot{V}O_2$ in validated models (Høydal et al. 2007; Rodrigues et al. 2007).	Principal limitations/ confounders	Report as maximal/peak running speed; not as $\dot{V}O_2$max (Høydal et al. 2007; Picoli et al. 2018).
Maximal lactate steady state (MLSS)	Highest constant workload at which blood lactate can remain approximately stable over time	Multiple constant-load trials, usually on separate days, with serial blood-lactate measurements	Strong physiological index of the upper boundary of sustainable metabolic exercise; useful for individualized training prescription and validation of lactate-derived methods	Labor intensive and invasive; affected by exercise modality, motor pattern, workload increments, test duration, sampling procedures, stress, training status, and MLSS definition	Report separately as MLSS workload (MLSSw) and, where relevant, MLSS blood-lactate concentration (MLSSc). A criterion such as ≤ 1 mmol/L increase over a specified interval must be explicitly stated. Do not assume that MLSS corresponds universally to 4 mmol/L or to $\dot{V}O_2$max
Heart rate (HR)	Cardiovascular internal load and autonomic/ hemodynamic response to exercise	Telemetry, implanted transducers, or ECG during standardized exercise	Can track relative exercise intensity and provide mechanistic cardiovascular information; HR and $\dot{V}O_2$ may show a near-linear association across submaximal workloads	Continuous measurement usually requires implanted instrumentation; surgery, recovery, device mass, motion artifact, autonomic dysfunction, medication, disease, stress, temperature, and training alter the HR–$\dot{V}O_2$ relationship	Report HR as a complementary cardiovascular/internal-load measure, not as a direct measurement of aerobic capacity. HR-based intensity prescription requires validation in the specific experimental model
Other cardiovascular indices	Central and peripheral cardiovascular mechanisms contributing to exercise capacity	Blood pressure, cardiac output, stroke volume, oxygen pulse, regional blood flow, vascular conductance	Help identify cardiovascular determinants and limitations of exercise performance and $\dot{V}O_2$max	Technically demanding and often invasive; individual indices represent only one component of an integrated physiological system	Describe as mechanistic cardiovascular outcomes rather than standalone measures of maximal aerobic capacity
Voluntary wheel-running activity	Self-selected habitual physical activity, exercise behavior, and chronic voluntary exercise exposure	Instrumented running wheels measuring distance, duration, frequency, and sometimes speed	Minimal investigator intervention; avoids forced treadmill stimulation; useful for genetic, behavioral, circadian, aging, and chronic-exercise studies	Exercise intensity is not standardized; marked effects of strain, sex, age, circadian behavior, motivation, and individual preference. Wheel distance may correlate poorly with forced treadmill performance and $\dot{V}O_2$max	Report as voluntary wheel-running activity, habitual physical activity, or self-selected exercise exposure. It should not be reported as a measure or direct surrogate of exercise capacity or maximal aerobic capacity.
Skeletal-muscle oxidative and	Peripheral cellular and molecular	Citrate synthase, succinate dehydrogenase, cytochrome-c	Provide mechanistic evidence of peripheral adaptation and	A single muscle or molecular marker cannot represent whole-	Report as mechanistic correlates of aerobic adaptation, peripheral

(Continues)

TABLE 2 | (Continued)

Outcome	Primary physiological construct	Typical assessment approach	Simple and inexpensive; associated with $\dot{V}O_2$ in validated models (Høydal et al. 2007; Rodrigues et al. 2007).	Principal limitations/ confounders	Report as maximal/peak running speed; not as $\dot{V}O_2$max (Høydal et al. 2007; Picoli et al. 2018).
mitochondrial markers	adaptations supporting oxidative metabolism	oxidase, OXPHOS proteins, mitochondrial respiratory capacity, mitochondrial DNA, capillary density, and fiber-type characterization	are useful for confirming biological responses to endurance training	body oxygen transport and utilization; adaptations may occur despite unchanged $\dot{V}O_2$max, while $\dot{V}O_2$max may improve without parallel changes in every tissue marker	oxidative adaptations, or the specific measured molecular/ enzymatic outcome. They should not be presented as substitutes for $\dot{V}O_2$max or $\dot{V}O_2$peak

Abbreviations: CS, critical speed; HR, heart rate; MLSSc, maximal lactate steady-state concentration; MLSSw, maximal lactate steady-state workload; OXPHOS, oxidative phosphorylation; RER, respiratory exchange ratio; TTE, time to exhaustion; $\dot{V}O_2$, oxygen uptake; $\dot{V}O_2$max, maximal oxygen uptake; $\dot{V}O_2$peak, peak oxygen uptake.

protocol should report the number, frequency, duration, speed, inclination, rest intervals, chamber exposure, and motivational procedures used. Successful familiarization should be evaluated through observable criteria such as stable gait, reduced time at the rear of the treadmill, fewer motivational contacts, completion of predefined low-intensity workloads, and stabilization of performance. Animals should not be classified as "poor runners" solely because they require repeated stimulation. When treadmill competence is an eligibility criterion, it should be defined before randomization, and all exclusions, their timing, and their reasons should be reported to avoid attrition bias (Percie du Sert et al. 2020).

4.4 | Protocol Architecture

Incremental-test architecture influences whether the intended physiological response can be elicited. Starting workload, increment size, stage duration, acceleration rate, inclination, warm-up, and total test duration must balance adequate physiological response with the risk of premature peripheral, thermal, or behavioral fatigue. Stages that are too short may not allow either the animal's oxygen-uptake kinetics or the chamber–analyser system to respond adequately, whereas excessively long stages may cause fatigue before maximal oxygen transport is reached (Ayachi et al. 2016; Petrosino et al. 2016; Poole et al. 2020). The temporal characteristics of the $\dot{V}O_2$ response are therefore important when stage duration or time to exercise limitation (Tlim) is used to interpret aerobic function. For a monoexponential response with amplitude A, time delay (TD), and time constant $\tau\dot{V}O_2$, approximately 63% of the response amplitude is attained after one time constant and approximately 95% after three time constants $[1 - \exp(-3)]$. Consequently, when the imposed workload is sufficiently high to elicit $\dot{V}O_2$max, a Tlim exceeding approximately $TD + 3\tau\dot{V}O_2$ provides sufficient time for $\dot{V}O_2$ to approach its maximal asymptote before exhaustion. Thus, $\dot{V}O_2$max may be attained before task failure rather than coinciding with the point of exhaustion. Importantly, however, Tlim $> 3\tau\dot{V}O_2$ is a temporal condition rather than an independent criterion of maximality: exercise performed at a workload below that required to elicit $\dot{V}O_2$max may continue for many time constants without reaching maximal oxygen uptake, and a slow component during severe exercise may further complicate a simple monoexponential interpretation (Behnke et al. 2002; Poole et al. 2020). As a quantitative rodent example, Behnke et al. (2002) reported a $\tau\dot{V}O_2$ of 23.2 ± 1.0 s in contracting rat spinotrapezius muscle, corresponding to approximately 70 s for 3τ; this in situ muscle estimate should not be interpreted as a universal pulmonary or treadmill $\tau\dot{V}O_2$. Directly modeled pulmonary $\dot{V}O_2$ kinetics remain relatively sparsely characterized in exercising rodents, and quantitative values established in humans should not be transferred directly to rats or mice without model-specific validation. Nevertheless, available rodent studies emphasize the importance of allowing sufficient time for the physiological response to develop. Qin et al. (2020), for example, used 3-min stages when characterizing treadmill $\dot{V}O_2$ responses in Wistar rats, whereas Orsi et al. (2023) reported a Tlim of 28.4 ± 8.3 min during running at critical velocity in C57BL/6 J mice, illustrating the substantially longer exercise durations encountered in endurance-oriented protocols. Disease may

additionally alter the kinetics of local O_2 delivery–utilization matching. In rats with myocardial-infarction-induced chronic heart failure, Diederich et al. (2002) reported primary microvascular $P_{mv}O_2$ time constants of 19.0 ± 1.5 s in sham-operated animals, 13.2 ± 1.9 s in moderate heart failure, and 28.2 ± 3.4 s in severe heart failure. These microvascular measurements should not be interpreted as pulmonary $\tau\dot{V}O_2$; rather, they demonstrate that the dynamic matching of muscle O_2 utilization to O_2 delivery varies with disease severity. Stage duration should therefore reflect both the test objective and the expected temporal characteristics of the physiological response. Wisløff et al. (2001) used longer stages for submaximal assessment to permit $\dot{V}O_2$ stabilization and shorter increments during maximal testing. Submaximal economy tests require sufficient time to approach a steady state, whereas maximal protocols must progress rapidly enough to limit prolonged fatigue but slowly enough for the metabolic response to follow the workload. No inclination or acceleration pattern is universally optimal. Ayachi et al. (2016) obtained the highest $\dot{V}O_2$ in mice using a slowly accelerating level ramp, whereas Wisløff et al. (2001) identified the clearest plateau in Sprague–Dawley rats at a 25° inclination. Kemi et al. (2002) also demonstrated reliable testing across inclined treadmill protocols in mice. These differences indicate that optimal protocol architecture is population- and system-specific rather than fixed. Age further modifies the relationship among treadmill speed, grade, and oxygen consumption. Qin et al. (2020) showed that protocols developed for young Wistar rats did not impose the same relative physiological demand in older animals. Starting workloads and increments should therefore reflect expected functional capacity while retaining a sufficient progression to reach the intended physiological endpoint. Inclination should also be justified because steep grades increase metabolic demand at lower speeds but may alter gait, muscle recruitment, mechanical efficiency, and willingness to run (Ayachi et al. 2016; Kemi et al. 2002; Qin et al. 2020; Wisløff et al. 2001).

4.5 | Exhaustion, Motivation, and Refinement

Termination criteria should be objective, prospectively specified, and applicable without escalating aversive stimulation. Common criteria include persistent inability to maintain belt speed, failure to re-engage after standardized prompting, prolonged time at the rear of the treadmill, loss of coordinated gait, or a predefined number of motivational contacts. These criteria are not equivalent and may produce different performance outcomes. Marcaletti et al. (2011) proposed operational criteria such as more than 20 stimulations within 1 min or failure to resume running for longer than 15 s. Brief loss of position should be distinguished from persistent inability to continue, and stimulus exposure should be documented. When subjective judgment remains necessary, the observer should be blinded to allocation, and treadmill running and the motivational stimulus should be stopped immediately once the predefined criterion is met.

A marked slowing or attenuation of the righting reflex may provide corroborative evidence of fatigue after treadmill exercise, particularly when accompanied by inability to maintain pace or an altered gait (Copp et al. 2009). However, it should not be used as an independent criterion for $\dot{V}O_2$max. The specific reason for termination "such as physiological inability, behavioral refusal, abnormal gait, injury, equipment failure, or completion of a predefined maximum test duration" should be reported separately from the numerical outcome.

Motivational procedures affect both animal welfare and measurement validity. Repeated electrical-grid contact can make performance dependent on avoidance behavior, stress responsiveness, or pain tolerance rather than aerobic physiology (Booth et al. 2010; Marcaletti et al. 2011; Rezaee 2025). Increased stimulation may also reflect exhaustion, poor familiarization, pain, locomotor dysfunction, disease severity, or strain-specific behavior rather than inadequate motivation. Strain-comparison studies illustrate this problem. Lightfoot et al. (2001) showed that low treadmill performance could reflect active refusal in A/J and AKR/J mice but physical inability to maintain speed in DBA/2 J mice. Lerman et al. (2002) found limited correspondence between forced treadmill and voluntary wheel-running performance, and Gibb et al. (2016) reported differences in treadmill compliance between FVB/NJ and C57BL/6 J mice. Behavioral compliance should therefore be treated as a model characteristic rather than an incidental procedural issue.

Alternative motivational approaches, including gentle touching, brushing, visual cues, soft air stimulation, and strategically positioned objects, may reduce reliance on electrical stimulation but must still be standardized. de Assis et al. (2024) used a rubber ball as a non-electrical motivational stimulus during treadmill training in rats and reported improved exercise performance. However, alternative procedures may also affect measurement: injected gases can disturb chamber gas mixing, touching may vary between operators, and food reinforcement may alter nutritional status. The selected procedure should therefore be described in detail and evaluated for potential interference with the physiological outcome.

4.6 | Technical Validity of Indirect Calorimetry

Technical validity requires control of the complete measurement chain. Oxygen and carbon dioxide analysers should be warmed, calibrated using certified gas mixtures, and checked against reference gases. Reports should specify chamber volume, airflow, sampling interval, humidity and pressure correction, purge time, and analyser response characteristics (Marcaletti et al. 2011; Speakman 2013). Wisløff et al. (2001) and Kemi et al. (2002) reported analyser calibration, accuracy, and chamber airflow, while Marcaletti et al. (2011) emphasized repeated reference-gas measurement and frequent gas sampling. These details are critical because small errors in oxygen concentration or airflow can produce substantial errors in calculated $\dot{V}O_2$ in small animals. System delay is especially important when chamber volume is large relative to the animal's respiratory gas exchange. Gas mixing and washout may delay or smooth rapid changes in $\dot{V}O_2$, causing the apparent peak to be assigned to the wrong treadmill stage. Laboratories should characterize analyser lag, chamber washout, airflow accuracy, and recovery of a known gas infusion and align the gas trace with workload after correcting for delay (Speakman 2013). Data-processing methods should also be reported, including whether the peak represents a single sample,

moving average, or mean of consecutive values; the averaging interval; artifact management; and the plateau criterion. A single value may reflect analyser noise, whereas excessive averaging may suppress a genuine short-duration maximum. Operational characteristics of the apparatus are as important as its manufacturer and model. Belt dimensions, lane width, enclosure, airflow, noise, belt surface, inclination accuracy, and rear-grid configuration may alter locomotion and behavior. When training and testing are performed on different systems, familiarization with both should be reported. Customized apparatus should be described in sufficient detail to permit replication.

4.7 | Biological, Environmental, and Pre-Analytical Variability

Biological variability should be distinguished from technical error. Species, strain, substrain, sex, age, genotype, body size, body composition, disease severity, baseline activity, and previous training are genuine determinants of aerobic phenotype. Bedford et al. (1979) demonstrated differences in $\dot{V}O_2$max among rat strains, while Lightfoot et al. (2001) reported a 3.6-fold difference in treadmill duration across inbred mouse strains, with genetic background accounting for a substantial proportion of variance. Lerman et al. (2002) and Gibb et al. (2016) further identified strain-specific variation in compliance, voluntary activity, and training adaptation. Reference values should therefore not be transferred directly across strains or substrains. Genetically modified animals should be compared with appropriately matched controls, and breeder, substrain, and genetic background should be reported because even minor differences may influence body mass, activity, stress responses, and performance. Age affects both aerobic physiology and test completion. Schefer and Talan (1996) reported lower resting, submaximal, and maximal oxygen consumption and lower termination speeds in aged C57BL/6 J mice. Protocols for aged or diseased rodents may require lower starting workloads, smaller increments, longer familiarization, and more conservative endpoints while preserving the intended physiological construct. Sex should also be incorporated into study design. Kemi et al. (2002) demonstrated reliable measurements in male and female mice but identified sex-related differences in training responses, while Wisløff et al. (2001) reported sex differences after body-mass scaling. Studies including both sexes should be adequately designed and report the corresponding analysis strategy. Body size influences treadmill mechanics and $\dot{V}O_2$ interpretation. As detailed in Section 3.3, absolute, conventional mass-specific, and allometrically adjusted values should be presented when groups differ substantially in size or composition (Bedford et al. 1979; Høydal et al. 2007; Speakman 2013). Disease models require similar caution because reduced performance may arise from impaired oxygen delivery, metabolic dysfunction, weakness, pain, neurological impairment, altered gait, or treatment effects. Circadian phase, feeding status, temperature, humidity, handling, and recent procedures influence baseline metabolism and exercise tolerance (Poole et al. 2020; Rezaee 2025; Speakman 2013). Testing time should be standardized or counterbalanced. Kemi et al. (2002) tested animals during the dark phase with a consistent handler, whereas Gibb et al. (2016) demonstrated that circadian activity and treadmill performance differed between strains. Testing order should be randomized to prevent time of day, chamber warming, analyser drift, or operator fatigue from becoming confounded with group allocation. Feeding and water access should be standardized because fasting alters glycogen availability and substrate metabolism, whereas recent feeding may affect respiratory exchange ratio and gastrointestinal comfort. Room and chamber temperature should be monitored because small rodents are particularly sensitive to heat loss and thermogenic demand. Recent anesthesia, surgery, catheterization, blood sampling, drug administration, or telemetry implantation may independently affect performance. Recovery should be confirmed using predefined criteria, and the interval between invasive procedures and exercise testing should be standardized across groups (Poole et al. 2020; Rezaee 2025).

4.8 | Experimental Design and Minimal Reporting

Randomization and blinding prevent routine procedural variation from becoming systematic bias. Animals should be randomized to testing order, operator, treadmill lane, and test day where relevant. Each testing day should include animals from all groups whenever feasible. Observers applying exhaustion criteria and investigators processing gas traces, identifying plateaus, removing artifacts, or excluding tests should be blinded to allocation and apply prespecified rules. The experimental unit should be defined correctly. Cage, litter, treadmill lane, testing day, and batch effects may compromise independence when animals are housed, trained, or assessed collectively. Sample-size calculations should be based on the primary outcome and realistic estimates of biological variability and measurement error. Missing data and exclusions should be anticipated prospectively rather than addressed through post hoc removal. ARRIVE 2.0 identifies study design, sample size, eligibility criteria, randomization, blinding, outcome measures, statistical methods, animal characteristics, procedures, and results as essential reporting elements. These principles are directly relevant to exercise studies because poor-runner exclusions, exhaustion criteria, and gas-trace processing can materially affect outcomes (Percie du Sert et al. 2020). The PREPARE guidelines complement these requirements by emphasizing prospective planning, pilot testing, risk assessment, humane endpoints, equipment evaluation, and collaboration with animal-facility personnel (Smith et al. 2018). Standardization should not be interpreted as imposing a universal protocol. A protocol is standardized when it is explicit, internally consistent, validated in the target population, and reported in sufficient detail for replication. Model-specific adaptations are appropriate, but their rationale should be explained. The heterogeneity identified by Massett et al. (2021) in strain, sex, treadmill inclination, session duration, frequency, familiarization, outcome selection, and reporting demonstrates how protocol variation limits cross-study comparison. To facilitate reproducibility and cross-study comparison, the minimum information that should be reported in rodent aerobic-capacity and exercise-performance studies is summarized in Table 3. The checklist integrates general experimental-design principles with exercise-specific requirements related to animal characteristics, familiarization, workload architecture, motivational procedures, exhaustion criteria, metabolic-system performance, data processing, and outcome interpretation. These recommendations are intended to standardize reporting rather than impose a single universal testing protocol, because appropriate workloads and validation criteria remain model- and system-specific (Table 3). Raw or minimally processed gas-exchange traces should be retained and, where feasible, made available so that

TABLE 3 | Minimal reporting checklist for rodent aerobic-capacity and exercise-performance studies.

Reporting domain	Minimum information that should be reported	Methodological rationale/potential source of bias
1. Study design and experimental unit	Primary and secondary outcomes; experimental unit; group allocation; sample size and its justification; inclusion and exclusion criteria; number of animals initially enrolled, tested, excluded, and analyzed	Prevents selective outcome reporting, pseudoreplication, post hoc exclusions, and ambiguity regarding the unit of analysis
2. Randomization and blinding	Method of group allocation; randomization of testing order, test day, treadmill lane, operator, or metabolic chamber when applicable; whether investigators applying exhaustion criteria or processing gas-exchange data were blinded	Time of day, analyser drift, operator behavior, lane effects, and subjective test termination may otherwise become systematically associated with group allocation
3. Animal characteristics	Species; strain and substrain; sex; age; genotype; breeder/source; body mass; relevant body-composition information; health status or disease model and severity; previous training or habitual-activity status	Aerobic phenotype, treadmill compliance, $\dot{V}O_2$, running performance, and training responsiveness vary substantially with biological background
4. Housing and pre-test environmental conditions	Housing density; light–dark cycle; phase/time of testing; room and chamber temperature; humidity when relevant; feeding/fasting status; water access; handling conditions	Circadian phase, thermoregulation, substrate availability, and handling stress directly affect resting metabolism and exercise tolerance
5. Recent procedures and recovery status	Recent anesthesia, surgery, catheterization, telemetry implantation, drug administration, blood sampling, or other invasive procedures; recovery interval; predefined recovery criteria	Residual pain, anesthesia, blood loss, surgical stress, or instrumentation may independently alter exercise performance and physiological measurements
6. Exercise apparatus and modality	Exercise modality; manufacturer/model or detailed custom-apparatus description; treadmill belt dimensions, number and width of lanes, belt surface, inclination and enclosure; wheel dimensions/resistance when relevant; swimming tank dimensions, water depth and temperature, loading or tethering method when applicable	Apparatus characteristics influence gait, mechanical workload, behavior, heat exchange, stimulus exposure, and comparability across laboratories
7. Familiarization/ acclimatization	Number and frequency of sessions; duration; speed/workload; inclination; chamber exposure; rest intervals; handler involvement; motivational procedure; criteria used to judge successful familiarization	Familiarization must reduce novelty, fear, poor locomotor skill, and learning effects without becoming an unintended training intervention
8. Baseline rest and warm-up	Pre-test resting/equilibration period; baseline gas-exchange collection when applicable; warm-up duration, speed, inclination, and intended relative intensity	Baseline conditions and warm-up modify metabolic starting conditions and may influence $\dot{V}O_2$ kinetics and maximal performance
9. Exercise-test architecture	Starting workload; speed; grade/inclination; increment size; stage duration; acceleration rate for ramp protocols; constant-load intensity when applicable; total test duration; transitions between stages	Protocol architecture determines whether the intended physiological endpoint is reached before locomotor, thermal, behavioral, or peripheral fatigue occurs
10. Exercise-intensity prescription	Whether intensity was based on absolute speed, $\%\dot{V}O_2$max, $\%\dot{V}O_2$peak, %maximal speed, critical speed, lactate-derived threshold, MLSS, HR, or another criterion; timing and frequency of reassessment during longitudinal studies	The same absolute workload may represent different relative intensities between animals and after training, aging, or disease progression

(Continues)

TABLE 3 | (Continued)

Reporting domain	Minimum information that should be reported	Methodological rationale/potential source of bias
11. Motivational procedures and animal welfare	Exact motivational method; electrical-grid settings if used; stimulus intensity, duration, and frequency; manual prompting, brushing, gas stimulation, or other alternatives; number of motivational contacts where feasible	Motivational procedures affect stress, pain tolerance, cardiovascular physiology, willingness to run, and the operational definition of exhaustion
12. Exhaustion and termination criteria	Prospectively defined operational definition of exhaustion; duration of failure to maintain belt position or speed; gait criteria; permitted motivational contacts; humane stopping criteria; actual reason for termination for each test when feasible	Distinguishes physiological exhaustion from behavioral refusal, poor treadmill competence, pain, stress reactivity, or locomotor failure
13. Indirect-calorimetry system	Chamber volume; airflow; analyser manufacturer/model; O_2 and CO_2 analyser type; calibration gases and calibration procedure; reference-gas procedures; gas drying or humidity correction; barometric/temperature corrections when applicable	Small errors in airflow or gas concentration can substantially distort calculated $\dot{V}O_2$ and $\dot{V}CO_2$ in small rodents
14. Gas sampling and system dynamics	Sampling frequency; reference sampling frequency; analyser response time; chamber washout characteristics; system delay/lag; purge or equilibration time; method used to align gas-exchange data with treadmill workload	Chamber mixing and analyser delay can shift measured $\dot{V}O_2$ into the wrong treadmill stage and alter identification of peak or plateau responses
15. Gas-exchange data processing	Formula/equation used to calculate $\dot{V}O_2$ and $\dot{V}CO_2$; averaging interval; whether peak values are single samples, moving averages, or consecutive means; artifact handling; baseline correction; treatment of missing values; software and relevant settings	Different averaging and artifact-processing strategies can materially alter $\dot{V}O_2$peak and apparent plateau detection
16. $\dot{V}O_2$ normalization	Absolute $\dot{V}O_2$ and the normalized value used; body-mass normalization; lean-mass normalization where appropriate; allometric exponent or statistical adjustment and its rationale	Simple ratio normalization may bias comparisons among animals differing in body mass, age, sex, adiposity, genotype, or disease state
17. $\dot{V}O_2$max/$\dot{V}O_2$peak classification	Explicit plateau definition; magnitude of acceptable $\dot{V}O_2$ change despite increasing workload; duration over which plateau was assessed; secondary maximality criteria such as RER, blood lactate, HR, or gait; terminology used for the final endpoint	Prevents unverified peak oxygen uptake from being incorrectly labeled as $\dot{V}O_2$max
18. Verification testing, when used	Verification-bout workload; recovery interval; bout duration; stopping criterion; comparison between verification $\dot{V}O_2$ and initial maximal/peak value	Provides additional evidence that the initial progressive test reached maximal oxygen uptake but also adds physiological burden and possible recovery effects
19. Performance outcomes	Exact definition and units of maximal/peak speed, TTE, distance, work, power, or other performance variable; whether incomplete final stages were incorporated; precise test-termination rule	Performance endpoints are highly protocol dependent and should not be interpreted as direct measures of maximal oxygen uptake
20. Lactate-derived outcomes	Blood-sampling site; sample volume; sampling timing; analyser/assay; incremental-test architecture; threshold-detection method; fixed or individualized criterion; for MLSS,	Lactate thresholds and MLSS are strongly dependent on methodology and cannot be compared validly without explicit definitions

(Continues)

TABLE 3 | (Continued)

Reporting domain	Minimum information that should be reported	Methodological rationale/potential source of bias
	constant-load duration, workload increments, and acceptable lactate change	
21. Critical-speed testing	Number and order of exhaustive trials; selected speeds; range of TTE values; recovery between tests; mathematical model and goodness-of-fit; criteria for accepting the CS estimate	Critical-speed estimates depend on trial selection and modeling assumptions and may become unstable when trial durations or workloads are inappropriate
22. Reliability and repeat testing	Number of baseline/repeat tests; intervals between tests; total familiarization/pre-intervention exposure; test–retest correlation or ICC; coefficient of variation; typical error and limits of agreement when possible	Distinguishes biological change from technical error, learning effects, and day-to-day variability
23. Missing data, exclusions, and failed tests	Number of incomplete or technically invalid tests; reasons for exclusion; timing of exclusion; handling of artifacts and missing data; whether exclusion rules were prespecified	Post hoc removal of "poor runners," atypical gas traces, or failed tests may introduce substantial selection and attrition bias
24. Data transparency and reproducibility	Availability of raw or minimally processed gas traces where feasible; individual-animal outcome data; protocol deviations; analysis rules; software/version; sufficient information for independent replication	Enables assessment of workload–gas alignment, plateau identification, artifact handling, inter-individual variability, and reproducibility

Abbreviations: CS, critical speed; HR, heart rate; ICC, intraclass correlation coefficient; MLSS, maximal lactate steady state; RER, respiratory exchange ratio; TTE, time to exhaustion; $\dot{V}CO_2$, carbon-dioxide production; $\dot{V}O_2$, oxygen uptake; $\dot{V}O_2max$, maximal oxygen uptake; $\dot{V}O_2peak$, peak oxygen uptake.

workload alignment, plateau identification, artifacts, and averaging decisions can be evaluated (Percie du Sert et al. 2020). Cross-laboratory harmonization should focus on common validation and reporting principles rather than identical external workloads. Shared calibration procedures and reference protocols may improve comparability, but suitability should still be confirmed within each model and laboratory (Poole et al. 2020).

5 | Translational Relevance, Limitations, and Future Directions

Rat and mouse models bridge reductionist cellular research and human exercise studies by allowing whole-body aerobic phenotypes to be linked with cardiac, vascular, muscular, mitochondrial, and molecular mechanisms under controlled conditions. Invasive measurements of skeletal-muscle oxidative capacity, regional blood flow, capillary hemodynamics, and microvascular oxygenation can be integrated with functional exercise outcomes in ways that are generally impractical in humans (Holloszy 1967; Musch et al. 1986; Musch and Terrell 1992; Kindig et al. 1999; Diederich et al. 2002). Intensity-controlled endurance training in rodents reproduces several adaptations observed in humans, including increased maximal oxygen uptake, improved exercise economy, lower heart rate at a given workload, physiological cardiac remodeling, and enhanced skeletal-muscle oxidative capacity. These similarities support the use of rodents to identify mechanisms regulating cardiovascular and muscular adaptation (Høydal et al. 2007; Kemi et al. 2002; Wisløff et al. 2001). Nevertheless, similarity in the direction of adaptation does not establish direct equivalence. Rats and mice differ from humans in body size, mass-specific metabolism, heart rate, posture, gait, thermoregulation, circadian activity, muscle recruitment, and locomotor behavior. Absolute treadmill speed, inclination, and duration therefore cannot be converted directly into equivalent human exercise doses. Cross-species comparisons should instead be based on relative physiological domains, such as proximity to maximal oxygen uptake, sustainable metabolic boundaries, substrate disturbance, and homeostatic strain. The translational question should determine the outcome hierarchy. Whole-body oxygen uptake is central when maximal cardiorespiratory function is the target, whereas submaximal oxygen cost, sustainable exercise intensity, muscle oxygen extraction, or mitochondrial function may be more informative for fatigue resistance and disease-related exercise intolerance. Each model should therefore represent a defined physiological process rather than be expected to reproduce every dimension of human exercise.

5.1 | Intrinsic Aerobic Capacity and Disease Susceptibility

Selective breeding for low- and high-capacity running established intrinsic aerobic fitness as an experimentally tractable biological trait. Koch and Britton (2001) developed low-capacity runner and high-capacity runner rat lines using treadmill endurance as the selection phenotype. Because these differences were inherited rather than training-induced, the model enabled investigation of intrinsic aerobic capacity independently of habitual exercise. High-capacity runners subsequently demonstrated greater $\dot{V}O_2max$ and more effective peripheral oxygen transfer, confirming that inherited endurance capacity reflects coordinated cardiovascular, vascular, muscular, and metabolic characteristics (Henderson et al. 2002). Low-capacity animals also developed greater adiposity, insulin resistance, dyslipidaemia, hypertension, and cardiovascular

dysfunction, supporting a biological association between low intrinsic aerobic capacity and disease susceptibility (Wisløff et al. 2005). The model should not, however, be interpreted as a direct representation of physically active and sedentary humans. Human cardiorespiratory fitness reflects complex interactions among genetics, development, habitual activity, disease, environmental exposure, and socioeconomic conditions. The principal value of the HCR/LCR model lies in identifying biological networks linking intrinsic aerobic capacity with metabolic resilience. High intrinsic capacity also does not confer universal protection. Hjortbak et al. (2020) found that aerobic-capacity phenotype influenced myocardial sensitivity to ischemia–reperfusion injury but did not determine the effectiveness of ischemic preconditioning. Alsahly et al. (2021) similarly reported greater baseline cardiac mitochondrial respiratory capacity in high-capacity runners, although this advantage disappeared following severe ischemia–reperfusion injury. Aerobic capacity should therefore be considered one component of physiological reserve whose protective effects depend on the organ, pathological insult, severity, and time course.

5.2 | Disease Models and Clinical Interpretation

Rodent disease models can clarify whether reduced exercise capacity arises primarily from impaired oxygen delivery, peripheral metabolic dysfunction, neuromuscular impairment, or systemic disease burden. Their translational value is greatest when the modeled pathology, expected mechanism, and selected exercise outcome are causally aligned. Chronic heart failure provides a clear example of how rodent models can dissect the pathophysiological basis of exercise intolerance. In rats with myocardial-infarction-induced heart failure, endurance training increased $\dot{V}O_2$max and skeletal-muscle succinate dehydrogenase activity while reducing the blood-lactate response to submaximal exercise, demonstrating that changes in whole-body exercise capacity can coexist with peripheral oxidative and hemodynamic adaptations (Musch et al. 1986). Regional-perfusion studies subsequently showed that skeletal-muscle blood flow during a standardized treadmill workload is reduced after myocardial infarction and that the magnitude of this deficit increases with infarct size and left-ventricular dysfunction (Musch and Terrell 1992). At the microvascular level, chronic heart failure reduced the proportion of capillaries supporting continuous red-blood-cell flow from 87% to 66%, while red-blood-cell velocity decreased from 252 ± 20 to $144 \pm 9\ \mu m \cdot s^{-1}$ and flux from 21.4 ± 2.4 to 9.4 ± 1.1 cells$\cdot s^{-1}$, thereby reducing the potential for muscle O_2 delivery (Kindig et al. 1999). Moreover, the dynamics of microvascular O_2 pressure during muscle contractions varied with heart-failure severity: the primary response time constant was 19.0 ± 1.5 s in sham-operated rats, 13.2 ± 1.9 s in moderate heart failure, and 28.2 ± 3.4 s in severe heart failure, indicating that disease severity alters the dynamic matching between muscle O_2 utilization and O_2 delivery (Diederich et al. 2002). Together, these findings demonstrate why reduced treadmill performance in a disease model should not be interpreted as a unitary reduction in "aerobic capacity"; the underlying limitation may reside at central, regional-perfusion, microvascular, or skeletal-muscle metabolic levels.

Feng et al. (2017), for example, used low-capacity rats to investigate interactions among intrinsic fitness, surgical stress, neuroinflammation, cognition, and the gut microbiome. Preoperative exercise attenuated postoperative neuroinflammation and cognitive decline and was accompanied by microbial changes. This model illustrated how rodents can connect metabolic, neuroimmune, behavioral, and microbial pathways that are difficult to isolate in clinical populations. Fiuza-Luces et al. (2019) used Harlequin mice with apoptosis-inducing factor deficiency and mitochondrial complex I dysfunction to examine exercise adaptation in a genetically defined mitochondrial disorder. Combined aerobic and resistance training improved functional performance and induced skeletal-muscle molecular adaptations, demonstrating that exercise may recruit compensatory pathways despite a specific respiratory-chain defect. Experimentally induced or genetically defined disorders nevertheless reproduce only selected components of human disease. Clinical populations commonly have advanced age, multimorbidity, polypharmacy, psychosocial limitations, variable disease duration, and prolonged inactivity, whereas rodent studies often involve young animals with one dominant pathology. Restoration of exercise performance in an animal model may therefore demonstrate biological efficacy without establishing clinical effectiveness.

Translational claims should identify the specific disease component represented. Increased running time in a neurological model may arise from improved motor coordination, while better performance in obesity may partly reflect reduced body mass and mechanical loading. Similarly, preserved peak oxygen uptake in mitochondrial disease may coexist with abnormal substrate use or fatigue resistance. Interpretation should therefore remain aligned with the proposed mechanism of benefit.

5.3 | Forced and Voluntary Exercise as Translational Constructs

As outlined previously, forced treadmill exercise and voluntary wheel running represent different translational constructs. Treadmill protocols most closely model prescribed training, rehabilitation, and laboratory-based cardiopulmonary testing, whereas wheel running reflects self-selected physical activity and behavioral engagement. Lerman et al. (2002) demonstrated substantial strain variation in both treadmill endurance and voluntary wheel behavior, with limited correspondence between performance rankings across modalities. This indicates that willingness to exercise spontaneously, tolerance of imposed exercise, and maximal physiological capacity have overlapping but distinct determinants. Wheel-running volume should consequently not be interpreted as a direct equivalent of cardiorespiratory fitness. It is more informative for questions concerning habitual activity, behavioral motivation, reward processing, self-selected exercise dose, and long-term interactions between activity and disease. Conversely, treadmill testing is preferable when a controlled physiological stimulus is required, although its imposed nature limits ecological equivalence to voluntary human activity.

5.4 | Integration of Whole-Body and Mechanistic Outcomes

A major strength of rodent research is the ability to connect whole-body phenotypes with direct organ and tissue analysis. Cardiac imaging and histology, isolated-heart experiments, regional perfusion measurements, skeletal-muscle morphology, mitochondrial respirometry, and molecular analyses can reveal

why interventions producing similar functional changes may operate through different mechanisms. Equivalent improvements in treadmill performance may result from increased cardiac output, enhanced muscle oxygen extraction, altered substrate use, reduced body mass, or improved locomotor efficiency. Conversely, an isolated increase in mitochondrial protein abundance, citrate-synthase activity, or capillary density does not establish functional benefit. Tissue-level adaptations should therefore be interpreted as components of a causal pathway connecting the intervention to integrated physiological function.

5.5 | Emerging Systemic Mechanisms

Exercise adaptation involves communication among skeletal muscle, heart, liver, adipose tissue, nervous and immune systems, endocrine pathways, and the intestinal microbiome. These systems may influence aerobic function through metabolites, inflammatory mediators, neural regulation, and inter-organ signaling. Uchida et al. (2023) reported that aerobic training altered the gut microbiota in mice, while antibiotic treatment attenuated improvements in endurance performance, muscle citrate-synthase activity, and PGC-1α. Transfer of microbiota from trained mice also improved endurance-related outcomes in recipient animals, suggesting that microbial changes may contribute causally to exercise adaptation. Direct taxonomic translation to humans remains difficult because rodent microbiota are strongly influenced by diet, housing, vendor, cage effects, and laboratory environment, whereas human microbial communities are shaped by broader dietary, pharmacological, cultural, and environmental exposures. Translational research may therefore benefit more from studying conserved microbial functions and metabolites, including short-chain fatty-acid production, bile-acid metabolism, redox regulation, inflammatory signaling, and substrate availability.

5.6 | Future Directions

Future studies should prioritize biologically diverse and clinically relevant models, including aging, comorbidities, medication exposure, and genetic variation. Longitudinal, minimally invasive phenotyping should combine repeated exercise testing with imaging, telemetry, body-composition measures, circulating biomarkers, and selected omics approaches to clarify the temporal and mechanistic basis of adaptation. Prospectively aligned rodent–human studies, supported by causal experiments and automated behavioral monitoring, may further improve translation by linking whole-body aerobic phenotypes with molecular mechanisms and daily functional outcomes. Rodent models cannot reproduce the full biological, behavioral, and social complexity of human exercise, but they are valuable for isolating mechanisms and linking whole-body performance with organ-, cellular-, and molecular-level adaptations. Their translational relevance is greatest when each model addresses a clearly defined physiological question and conclusions are restricted to the processes represented. Future progress will depend on diverse and clinically relevant models, longitudinal multimodal phenotyping, causal mechanistic studies, and bidirectional rodent–human research. Within this framework, aerobic-capacity assessment provides an integrative physiological anchor connecting exercise adaptation with health and disease.

Author Contributions

Dursun Alper Yılmaz: conceptualization, data curation, investigation, methodology, visualization, writing – original draft, writing – review and editing. **Fadıl Özyener:** conceptualization, methodology, supervision, validation, writing – review and editing.

Acknowledgments

The authors have nothing to report.

Funding

The authors have nothing to report.

Ethics Statement

The authors have nothing to report. This article is a narrative review of previously published studies and did not involve new experiments involving human participants or animals.

Conflicts of Interest

The authors declare no conflicts of interest.

Policy on Using ChatGPT and Similar AI Tools

During the preparation of this manuscript, the authors used ChatGPT (OpenAI) solely to assist with English-language editing, grammatical refinement, and improvements in clarity and readability. The tool was not used to generate, analyze, or interpret research data, nor to make scientific decisions or conclusions. All AI-assisted text was critically reviewed, verified, and revised by the authors, who take full responsibility for the accuracy, integrity, and final content of the manuscript.

Data Availability Statement

No new data were generated or analyzed in this narrative review. All information synthesized in this article was derived from previously published studies cited in the manuscript.

References

Abel, E. L. 1993. "Physiological Correlates of the Forced Swim Test in Rats." *Physiology & Behavior* 54, no. 2: 309–317. https://doi.org/10.1016/0031-9384(93)90116-W.

Abel, E. L. 1994. "Behavioral and Physiological Effects of Different Water Depths in the Forced Swim Test." *Physiology & Behavior* 56, no. 2: 411–414. https://doi.org/10.1016/0031-9384(94)90215-1.

Abreu, P., S. V. D. Mendes, J. H. Leal-Cardoso, and V. M. Ceccatto. 2016. "Anaerobic Threshold Employed on Exercise Training Prescription and Performance Assessment for Laboratory Rodents: A Short Review." *Life Sciences* 151: 1–6. https://doi.org/10.1016/j.lfs.2016.02.016.

Almeida, J., B. Petriz, C. P. Gomes, L. A. Rocha, R. Pereira, and O. Franco. 2013. "Determination of the Maximal Lactate Steady State in Obese Zucker Rats." *International Journal of Sports Medicine* 34, no. 3: 214–217. https://doi.org/10.1055/s-0032-1316360.

Alsahly, M. B., M. O. Zakari, L. G. Koch, et al. 2021. "Augmented Cardiac Mitochondrial Capacity in High Capacity Aerobic Running "Disease-Resistant" Phenotype at Rest Is Lost Following Ischemia Reperfusion." *Frontiers in Cardiovascular Medicine* 8, Article: 752640. https://doi.org/10.3389/fcvm.2021.752640.

de Assis, G. G., E. O. N. de Souza, P. F. de Almeida-Neto, H. İ. Ceylan, and N. L. Bragazzi. 2024. "A Proposal for a Noxious Stimuli-Free, Moderate-Intensity Treadmill Running Protocol to Improve Aerobic

Performance in Experimental Research on Rats." *Metabolites* 14, no. 10, Article: 534. https://doi.org/10.3390/metabo14100534.

Ayachi, M., R. Niel, I. Momken, V. L. Billat, and L. Mille-Hamard. 2016. "Validation of a Ramp Running Protocol for Determination of the True VO2max in Mice." *Frontiers in Physiology* 7: 372, Article. 372. https://doi.org/10.3389/fphys.2016.00372.

Banjanin, S., and N. Mrosovsky. 2000. "Preferences of Mice, *Mus musculus*, for Different Types of Running Wheel." *Laboratory Animals* 34, no. 3: 313–318. https://doi.org/10.1258/002367700780384681.

Bedford, T. G., C. M. Tipton, N. C. Wilson, R. A. Oppliger, and C. V. Gisolfi. 1979. "Maximum Oxygen Consumption of Rats and Its Changes With Various Experimental Procedures." *Journal of Applied Physiology* 47, no. 6: 1278–1283. https://doi.org/10.1152/jappl.1979.47.6.1278.

Behnke, B. J., T. J. Barstow, C. A. Kindig, P. McDonough, T. I. Musch, and D. C. Poole. 2002. "Dynamics of Oxygen Uptake Following Exercise Onset in Rat Skeletal Muscle." *Respiratory Physiology & Neurobiology* 133: 229–239. https://doi.org/10.1016/S1569-9048(02)00183-0.

Billat, V. L., E. Mouisel, N. Roblot, and J. Melki. 2005. "Inter- and Intrastrain Variation in Mouse Critical Running Speed." *Journal of Applied Physiology* 98, no. 4: 1258–1263. https://doi.org/10.1152/japplphysiol.00991.2004.

Booth, F. W., M. J. Laye, and E. E. Spangenburg. 2010. "Gold Standards for Scientists Who Are Conducting Animal-Based Exercise Studies." *Journal of Applied Physiology* 108, no. 1: 219–221. https://doi.org/10.1152/japplphysiol.00125.2009.

Brooks, G. A., and T. P. White. 1978. "Determination of Metabolic and Heart Rate Responses of Rats to Treadmill Exercise." *Journal of Applied Physiology* 45, no. 6: 1009–1015. https://doi.org/10.1152/jappl.1978.45.6.1009.

Burnley, M. 2023. "Invited Review: The Speed-Duration Relationship Across the Animal Kingdom." *Comparative Biochemistry and Physiology. Part A, Molecular & Integrative Physiology* 279, Article: 111387. https://doi.org/10.1016/j.cbpa.2023.111387.

Contarteze, R., F. B. Manchado, C. A. Gobatto, and M. De Mello. 2008. "Stress Biomarkers in Rats Submitted to Swimming and Treadmill Running Exercises." *Comparative Biochemistry and Physiology. Part A, Molecular & Integrative Physiology* 151, no. 3: 415–422. https://doi.org/10.1016/j.cbpa.2007.03.005.

Copp, S. W., R. T. Davis, D. C. Poole, and T. I. Musch. 2009. "Reproducibility of Endurance Capacity and VO2peak in Male Sprague-Dawley Rats." *Journal of Applied Physiology* 106, no. 4: 1072–1078. https://doi.org/10.1152/japplphysiol.91566.2008.

Copp, S. W., D. M. Hirai, T. I. Musch, and D. C. Poole. 2010. "Critical Speed in the Rat: Implications for Hindlimb Muscle Blood Flow Distribution and Fibre Recruitment." *Journal of Physiology* 588, no. 24: 5077–5087. https://doi.org/10.1113/jphysiol.2010.198382.

Cunha, R. R., V. N. de Carvalho Cunha, P. R. Segundo, et al. 2009. "Determination of the Lactate Threshold and Maximal Blood Lactate Steady State Intensity in Aged Rats." *Cell Biochemistry and Function* 27, no. 6: 351–357. https://doi.org/10.1002/cbf.1580.

Diederich, E. R., B. J. Behnke, P. McDonough, et al. 2002. "Dynamics of Microvascular Oxygen Partial Pressure in Contracting Skeletal Muscle of Rats With Chronic Heart Failure." *Cardiovascular Research* 56, no. 3: 479–486. https://doi.org/10.1016/S0008-6363(02)00545-X.

Drugan, R. C., S. Eren, A. Hazi, J. Silva, J. P. Christianson, and S. Kent. 2005. "Impact of Water Temperature and Stressor Controllability on Swim Stress-Induced Changes in Body Temperature, Serum Corticosterone, and Immobility in Rats." *Pharmacology, Biochemistry and Behavior* 82, no. 2: 397–403. https://doi.org/10.1016/j.pbb.2005.09.011.

Feng, X., Y. Uchida, L. Koch, et al. 2017. "Exercise Prevents Enhanced Postoperative Neuroinflammation and Cognitive Decline and Rectifies the Gut Microbiome in a Rat Model of Metabolic Syndrome." *Frontiers in Immunology* 8, Article: 1768. https://doi.org/10.3389/fimmu.2017.01768.

Ferreira, J. C., N. P. Rolim, J. B. Bartholomeu, C. A. Gobatto, E. Kokubun, and P. C. Brum. 2007. "Maximal Lactate Steady State in Running Mice: Effect of Exercise Training." *Clinical and Experimental Pharmacology & Physiology* 34, no. 8: 760–765. https://doi.org/10.1111/j.1440-1681.2007.04635.x.

Fitts, R., F. Booth, W. Winder, and J. Holloszy. 1975. "Skeletal Muscle Respiratory Capacity, Endurance, and Glycogen Utilization." *American Journal of Physiology-Legacy Content* 228, no. 4: 1029–1033. https://doi.org/10.1152/ajplegacy.1975.228.4.1029.

Fiuza-Luces, C., P. L. Valenzuela, S. Laine-Menéndez, et al. 2019. "Physical Exercise and Mitochondrial Disease: Insights From a Mouse Model." *Frontiers in Neurology* 10: Article 790. https://doi.org/10.3389/fneur.2019.00790.

Gibb, A. A., L. A. McNally, D. W. Riggs, D. J. Conklin, A. Bhatnagar, and B. G. Hill. 2016. "FVB/NJ Mice Are a Useful Model for Examining Cardiac Adaptations to Treadmill Exercise." *Frontiers in Physiology* 7, Article: 636. https://doi.org/10.3389/fphys.2016.00636.

Gobatto, C. A., M. A. de Mello, C. Y. Sibuya, J. R. de Azevedo, L. A. dos Santos, and E. Kokubun. 2001. "Maximal Lactate Steady State in Rats Submitted to Swimming Exercise." *Comparative Biochemistry and Physiology. Part A, Molecular & Integrative Physiology* 130, no. 1: 21–27. https://doi.org/10.1016/S1095-6433(01)00362-2.

Henderson, K. K., H. Wagner, F. Favret, et al. 2002. "Determinants of Maximal O2uptake in Rats Selectively Bred for Endurance Running Capacity." *Journal of Applied Physiology* 93, no. 4: 1265–1274. https://doi.org/10.1152/japplphysiol.00809.2001.

Hjortbak, M. V., T. S. Grønnebæk, N. R. Jespersen, et al. 2020. "Differences in Intrinsic Aerobic Capacity Alters Sensitivity to Ischemia–Reperfusion Injury but Not Cardioprotective Capacity by Ischemic Preconditioning in Rats." *PLoS One* 15, no. 10), Article: e0240866. https://doi.org/10.1371/journal.pone.0240866.

Holloszy, J. O. 1967. "Biochemical Adaptations in Muscle." *Journal of Biological Chemistry* 242, no. 9: 2278–2282.

Høydal, M. A., U. Wisløff, O. J. Kemi, and O. Ellingsen. 2007. "Running Speed and Maximal Oxygen Uptake in Rats and Mice: Practical Implications for Exercise Training." *European Journal of Cardiovascular Prevention and Rehabilitation: Official Journal of the European Society of Cardiology, Working Groups on Epidemiology & Prevention and Cardiac Rehabilitation and Exercise Physiology* 14, no. 6: 753–760. https://doi.org/10.1097/HJR.0b013e3281eacef1.

Kemi, O. J., J. P. Loennechen, U. Wisløff, and Ø. Ellingsen. 2002. "Intensity-Controlled Treadmill Running in Mice: Cardiac and Skeletal Muscle Hypertrophy." *Journal of Applied Physiology* 93, no. 4: 1301–1309. https://doi.org/10.1152/japplphysiol.00231.2002.

Kindig, C. A., T. I. Musch, R. J. Basaraba, and D. C. Poole. 1999. "Impaired Capillary Hemodynamics in Skeletal Muscle of Rats in Chronic Heart Failure." *Journal of Applied Physiology* 87, no. 2: 652–660. https://doi.org/10.1152/jappl.1999.87.2.652.

Koch, L. G., and S. L. Britton. 2001. "Artificial Selection for Intrinsic Aerobic Endurance Running Capacity in Rats." *Physiological Genomics* 5, no. 1: 45–52. https://doi.org/10.1152/physiolgenomics.2001.5.1.45.

Lambert, M. I., and T. D. Noakes. 1989. "Dissociation of Changes in VO2 max, Muscle QO2, and Performance With Training in Rats." *Journal of Applied Physiology* 66, no. 4: 1620–1625. https://doi.org/10.1152/jappl.1989.66.4.1620.

Leandro, C. G., A. C. Levada, S. M. Hirabara, et al. 2007. "A Program of Moderate Physical Training for Wistar Rats Based on Maximal Oxygen Consumption." *Journal of Strength and Conditioning Research* 21, no. 3: 751–756. https://doi.org/10.1519/R-20155.1.

Lerman, I., B. C. Harrison, K. Freeman, et al. 2002. "Genetic Variability in Forced and Voluntary Endurance Exercise Performance in Seven Inbred Mouse Strains." *Journal of Applied Physiology* 92, no. 6: 2245–2255. https://doi.org/10.1152/japplphysiol.01045.2001.

Lightfoot, J. T., M. J. Turner, K. A. DeBate, and S. R. Kleeberger. 2001. "Interstrain Variation in Murine Aerobic Capacity." *Medicine and Science in Sports and Exercise* 33, no. 12: 2053–2057. https://doi.org/10.1097/00005768-200112000-00012.

Manchado, F. B., C. A. Gobatto, F. A. Voltarelli, and M. A. Rostom de Mello. 2006. "Non-Exhaustive Test for Aerobic Capacity Determination in Swimming Rats." *Applied Physiology, Nutrition, and Metabolism* 31, no. 6: 731–736. https://doi.org/10.1139/h06-079.

Manzanares, G., G. Brito-da-Silva, and P. G. Gandra. 2019. "Voluntary Wheel Running: Patterns and Physiological Effects in Mice." *Brazilian Journal of Medical and Biological Research* 52, no. 1, Article: e7830. https://doi.org/10.1590/1414-431X20187830.

Marcaletti, S., C. Thomas, and J. N. Feige. 2011. "Exercise Performance Tests in Mice." *Current Protocols in Mouse Biology* 1, no. 1: 141–154. https://doi.org/10.1002/9780470942390.mo100160.

Massett, M. P., C. Matejka, and H. Kim. 2021. "Systematic Review and Meta-Analysis of Endurance Exercise Training Protocols for Mice." *Frontiers in Physiology* 12, Article: 782695. https://doi.org/10.3389/fphys.2021.782695.

Musch, T. I., R. L. Moore, D. J. Leathers, A. Bruno, and R. Zelis. 1986. "Endurance Training in Rats With Chronic Heart Failure Induced by Myocardial Infarction." *Circulation* 74, no. 2: 431–441. https://doi.org/10.1161/01.CIR.74.2.431.

Musch, T. I., C. T. Nguyen, H. V. Pham, and R. L. Moore. 1992. "Training Effects on the Regional Blood Flow Response to Exercise in Myocardial Infarcted Rats." *American Journal of Physiology-Heart and Circulatory Physiology* 262, no. 6: H1846–H1852. https://doi.org/10.1152/ajpheart.1992.262.6.H1846.

Musch, T. I., and J. A. Terrell. 1992. "Skeletal Muscle Blood Flow Abnormalities in Rats With a Chronic Myocardial Infarction: Rest and Exercise." *American Journal of Physiology-Heart and Circulatory Physiology* 262, no. 2: H411–H419. https://doi.org/10.1152/ajpheart.1992.262.2.H411.

Orsi, J. B., L. S. Araujo, P. P. M. Scariot, et al. 2023. "Critical Velocity, Maximal Lactate Steady State, and Muscle MCT1 and MCT4 After Exhaustive Running in Mice." *International Journal of Molecular Sciences* 24, no. 21: 15753. https://doi.org/10.3390/ijms242115753.

Percie du Sert, N., V. Hurst, A. Ahluwalia, et al. 2020. "The Arrive Guidelines 2.0: Updated Guidelines for Reporting Animal Research." *PLoS Biology* 18, no. 7), Article: e3000410. https://doi.org/10.1371/journal.pbio.3000410.

Petrosino, J. M., V. J. Heiss, S. K. Maurya, et al. 2016. "Graded Maximal Exercise Testing to Assess Mouse Cardio-Metabolic Phenotypes." *PLoS One* 11, no. 2), Article: e0148010. https://doi.org/10.1371/journal.pone.0148010.

Picoli, C. C., P. V. S. Romero, G. R. Gilio, et al. 2018. "Peak Velocity as an Alternative Method for Training Prescription in Mice." *Frontiers in Physiology* 9, Article: 42. https://doi.org/10.3389/fphys.2018.00042.

Poole, D. C., S. W. Copp, T. D. Colburn, et al. 2020. "Guidelines for Animal Exercise and Training Protocols for Cardiovascular Studies." *American Journal of Physiology-Heart and Circulatory Physiology* 318, no. 5: H1100–H1138. https://doi.org/10.1152/ajpheart.00697.2019.

Poole, D. C., and A. M. Jones. 2017. "Measurement of the Maximum Oxygen Uptake V̇o2max: V̇o2peakis No Longer Acceptable." *Journal of Applied Physiology* 122, no. 4: 997–1002. https://doi.org/10.1152/japplphysiol.01063.2016.

Poole, D. C., H. B. Rossiter, G. A. Brooks, and L. B. Gladden. 2021. "The Anaerobic Threshold: 50+ Years of Controversy." *Journal of Physiology* 599, no. 3: 737–767. https://doi.org/10.1113/JP279963.

Poole, D. C., D. P. Wilkerson, and A. M. Jones. 2008. "Validity of Criteria for Establishing Maximal O2 Uptake During Ramp Exercise Tests." *European Journal of Applied Physiology* 102, no. 4: 403–410. https://doi.org/10.1007/s00421-007-0596-3.

Potolitsyna, N., O. Parshukova, N. Vakhnina, et al. 2023. "Lactate Thresholds and Role of Nitric Oxide in Male Rats Performing a Test With Forced Swimming to Exhaustion." *Physiological Reports* 11, no. 17), Article: e15801. https://doi.org/10.14814/phy2.15801.

Qin, F., Y. Dong, S. Wang, et al. 2020. "Maximum Oxygen Consumption and Quantification of Exercise Intensity in Untrained Male Wistar Rats." *Scientific Reports* 10, Article: 11520. https://doi.org/10.1038/s41598-020-68455-8.

Rabasa, C., R. Delgado-Morales, A. Gómez-Román, R. Nadal, and A. Armario. 2013. "Adaptation of the Pituitary-Adrenal Axis to Daily Repeated Forced Swim Exposure in Rats Is Dependent on the Temperature of Water." *Stress (Amsterdam)* 16, no. 6: 698–705. https://doi.org/10.3109/10253890.2013.824964.

dos Reis, I. G. M., L. E. B. Martins, G. G. de Araujo, and C. A. Gobatto. 2018. "Forced Swim Reliability for Exercise Testing in Rats by a Tethered Swimming Apparatus." *Frontiers in Physiology* 9, Article: 1839. https://doi.org/10.3389/fphys.2018.01839.

Rezaee, Z. 2025. "Guidelines for Animal Models of Endurance and Resistance Exercise." *Journal of Neuroscience Methods* 416, Article: 110362. https://doi.org/10.1016/j.jneumeth.2025.110362.

Rezende, E. L., T. Garland, Jr., M. A. Chappell, J. L. Malisch, and F. R. Gomes. 2006. "Maximum Aerobic Performance in Lines of *Mus* Selected for High Wheel-Running Activity: Effects of Selection, Oxygen Availability and the Mini-Muscle Phenotype." *Journal of Experimental Biology* 209, no. 1: 115–127. https://doi.org/10.1242/jeb.01883.

Rodrigues, B., D. M. Figueroa, C. T. Mostarda, M. V. Heeren, M.-C. Irigoyen, and K. De Angelis. 2007. "Maximal Exercise Test Is a Useful Method for Physical Capacity and Oxygen Consumption Determination in Streptozotocin-Diabetic Rats." *Cardiovascular Diabetology* 6, Article: 38. https://doi.org/10.1186/1475-2840-6-38.

Schefer, V., and M. I. Talan. 1996. "Oxygen Consumption in Adult and Aged C57BL/6J Mice During Acute Treadmill Exercise of Different Intensity." *Experimental Gerontology* 31, no. 3: 387–392. https://doi.org/10.1016/0531-5565(95)02032-2.

Smith, A. J., R. E. Clutton, E. Lilley, K. E. A. Hansen, and T. Brattelid. 2018. "PREPARE: Guidelines for Planning Animal Research and Testing." *Laboratory Animals* 52, no. 2: 135–141. https://doi.org/10.1177/0023677217724823.

Speakman, J. R. 2013. "Measuring Energy Metabolism in the Mouse: Theoretical, Practical, and Analytical Considerations." *Frontiers in Physiology* 4, Article: 34. https://doi.org/10.3389/fphys.2013.00034.

Teixeira-Coelho, F., C. G. Fonseca, N. H. S. Barbosa, et al. 2017. "Effects of Manipulating the Duration and Intensity of Aerobic Training Sessions on the Physical Performance of Rats." *PLoS One* 12, no. 8: e0183763. https://doi.org/10.1371/journal.pone.0183763.

Uchida, M., S. Fujie, H. Yano, and M. Iemitsu. 2023. "Aerobic Exercise Training-Induced Alteration of Gut Microbiota Composition Affects Endurance Capacity." *Journal of Physiology* 601, no. 12: 2329–2344. https://doi.org/10.1113/JP283995.

Watanabe, T., A. Morimoto, Y. Sakata, N. Tan, K. Morimoto, and N. Murakami. 1992. "Running Training Attenuates the ACTH Responses in Rats to Swimming and Cage-Switch Stress." *Journal of Applied Physiology* 73, no. 6: 2452–2456. https://doi.org/10.1152/jappl.1992.73.6.2452.

Wisløff, U., J. Helgerud, O. J. Kemi, and Ø. Ellingsen. 2001. "Intensity-Controlled Treadmill Running in Rats: V̇O2 maxand Cardiac Hypertrophy." *American Journal of Physiology-Heart and Circulatory Physiology* 280, no. 3: H1301–H1310. https://doi.org/10.1152/ajpheart.2001.280.3.H1301.

Wisløff, U., S. M. Najjar, Ø. Ellingsen, et al. 2005. "Cardiovascular Risk Factors Emerge After Artificial Selection for Low Aerobic Capacity." *Science* 307, no. 5708: 418–420. https://doi.org/10.1126/science.1108177.