PHASE-REGISTERED ASSESSMENT OF EXERTIONAL DYSPNOEA IN ATHLETES: TIMING, SPIROMETRY AND PAIRED AIRWAY CLASSIFICATION

Ansgar Schwirtz

Technical University of Munich (TUM), Germany

Received: 13 January 2024 **Revised:** 22 April 2024 **Accepted:** 28 April 2024 **Published:** 20 June 2024

Correspondence: ansgar.schwirtz@tum.de

ABSTRACT

Exercise-induced laryngeal obstruction (EILO) and exercise-induced bronchoconstriction (EIB) require different observation windows. Combining symptoms, technically incomplete tests and measurements from separate provocations can obscure what an evaluation actually establishes. To specify a challenge-specific respiratory record that retains measurement timing, acquisition quality and the evidential status of each airway assessment. Clinical investigations and respiratory standards dated January 2011 through February 2024 were assessed for exercise-stage laryngeal observation, bronchial provocation, spirometric quality and observer agreement. Each challenge receives its own identifier and clock origin. Laryngeal video is linked to workload and symptom timing; recovery spirometry retains the highest acceptable value in each measurement set before selecting the lowest interval value. Nielsen and colleagues identified EILO in 31/88 athletes; Hammer and colleagues identified EILO in 73/101 elite-athlete referrals and asthma-supporting findings in 22/55 athletes with EILO and identifiable asthma investigations. These denominators describe separate referral and testing processes. The record retains three statuses per attempted test—positive, negative under recorded conditions and indeterminate—and all nine paired combinations. An unperformed test retains its reason. The bronchial response is the percentage fall from pre-challenge FEV_1 to the lowest valid interval value; an exercise response of at least 10% meets the ATS criterion when challenge and acquisition requirements are satisfied. Separate challenge clocks and explicit test statuses distinguish observed obstruction, an adequately acquired negative result and an unavailable observation. The method retains established diagnostic thresholds. Its clinical value requires prospective assessment of completion rates, observer agreement and interpretation errors.

Keywords: exertional dyspnoea; athlete; continuous laryngoscopy during exercise; bronchoprovocation; spirometry; temporal registration; measurement quality.

CLINICAL PROBLEM AND MEASUREMENT OBJECTIVE

Exertional breathlessness is a symptom that can arise from more than one anatomical compartment. An athlete may report respiratory difficulty during a high-demand task while the decisive upper-airway observation is acquired at another stage, or may have unremarkable pre-exercise spirometry despite a subsequent lower-airway response. Continuous laryngoscopy during exercise (CLE) and post-challenge spirometry therefore address complementary questions. The 2022 IOC subgroup review describes CLE as the preferred objective investigation for EILO; the ATS EIB guideline requires challenge-related lung-function change rather than symptoms alone (Clemm et al., 2022; Parsons et al., 2013).

The distinction has empirical support. Johansson and colleagues estimated EIB and EILO prevalence at 19.2% and 5.7%, respectively, in a general adolescent population (Johansson et al., 2015). In an asthma-referral cohort, Nielsen and colleagues identified EILO in 31 of 88 athletes completing assessment (Walsted Nielsen et al., 2013). Hammer and

Copyright © 2024 The Author(s). This article is published under the Creative Commons Attribution 4.0 International (CC BY 4.0) license.

colleagues diagnosed EILO in 73 of 101 referred elite athletes; among 55 athletes with EILO and identifiable objective asthma investigations, 22 had findings supporting asthma (Jansrud Hammer et al., 2022). These proportions describe different sampling processes. The last finding concerns asthma-supporting investigations and must not be relabelled as a confirmed EIB proportion.

Olin and colleagues synchronised laryngoscopy with metabolic recordings and assessed exercise and early-recovery windows in 71 patients (Olin et al., 2016). Phase registration links four items: *challenge identity, observation time, technical adequacy and airway-specific test status*. An examination acquired during peak demand, a recovery spirometry set and a test not performed occupy different entries. The objective is to retain the observation and acquisition conditions on which each classification depends.

LITERATURE ASSESSMENT AND MEASUREMENT METHOD

RETRIEVAL SCOPE

A targeted narrative appraisal was undertaken on 10 October 2026, with publication eligibility restricted to 1 January 2011 through 29 February 2024. Web-indexed PubMed records, journal pages and accessible articles were consulted using the terms "exercise-induced laryngeal obstruction recovery", "exercise-induced laryngeal obstruction reliability", "athletes bronchoconstriction diagnostic approach" and "indirect airway challenge testing". Exact-title and DOI lookups were used for the studies and statements discussed here. Reference lists in the IOC review were also consulted to identify directly relevant earlier measurement work. Publication dates were checked individually; search-engine rankings were not treated as an eligibility criterion.

Eligibility required direct assessment of exercise-stage laryngeal observation, bronchial challenge acquisition, grading agreement, symptoms in relation to objective findings, or athlete follow-up. Diagnostic statements and technical standards were considered separately from observational investigations, intervention case series and reviews. The assessment was non-systematic: it did not enumerate every eligible article or use independent duplicate screening. No pooled prevalence or treatment effect was calculated. The 20 references address these measurement questions; Table 2 gives the numerical findings and their denominators.

MEASUREMENT VARIABLES AND DECISION ORDER

The analysis compared anatomical target, provocation timing, observation timing and technical adequacy. Diagnostic criteria follow the cited clinical standards. Clock correction, one-second event storage and unperformed-test codes define the record's operation; they are documentation requirements rather than physiological diagnostic thresholds. Table 1 links each clinical measurement to the items retained in the record.

TABLE 1. CLINICAL MEASUREMENTS AND THE ITEMS RETAINED FOR INTERPRETATION

Element	Clinical measurement	Stored items and decision rule
Laryngeal assessment	CLE observes exercise-provoked glottic and supraglottic behaviour (Clemm et al., 2022; Halvorsen et al., 2017)	Attach interpretable video windows, symptoms and workload to one identified challenge.
Exercise/recovery timing	Synchronous laryngeal and metabolic acquisition has been demonstrated (Olin et al., 2016)	Record timing offsets, residual error and unobservable intervals; prevent joining traces from separate challenges.
Bronchial response	Serial post-challenge FEV_1 and established response criteria (Parsons et al., 2013; Hallstrand et al., 2018)	Retain manoeuvre-selection rules, actual set times, missing sets and earliest occurrence of a tied observed minimum.
Technical validity	Grading and spirometry have recognised quality limitations (Walsted et al., 2017; Graham et al., 2019)	Keep result status separate from acquisition completeness and test disposition.
Combined assessment	EILO and lower-airway disease can coexist (Jansrud Hammer et al., 2022)	Preserve positive evidence in one domain when the other is indeterminate or unperformed.

EVIDENCE RELEVANT TO MEASUREMENT VALIDITY

EILO is assessed by viewing glottic and supraglottic structures under the provoking exercise demand. The ERS/ELS statement places direct laryngeal observation at the centre of assessment (Halvorsen et al., 2017) Conversely, Christensen

and colleagues found no significant association between exercise laryngoscopy findings and pre-/post-exercise flow-volume loop assessments in 100 consecutive tests (Christensen et al., 2013) A loop obtained outside the relevant event is not a substitute for an adequate laryngeal view during that event.

Table 2 retains each investigation's denominator and design. Symptoms, technical quality and observed anatomy require separate fields. In the longitudinal CLE investigation, 29 of 98 invited athletes returned (29.6%); most scores remained stable, although 3 of 18 athletes with initially absent or mild obstruction later had moderate obstruction (Irewall et al., 2023) Selective return limits inference about all invited athletes. Fraction percentages are rounded to one decimal place.

TABLE 2. AIRWAY FINDINGS BY INVESTIGATION, DENOMINATOR AND MEASUREMENT CONSEQUENCE. EACH ROW RETAINS ITS OWN POPULATION; NO ESTIMATES ARE POOLED.

Investigation	Numerical finding	Interpretive limitation	Record requirement
Johansson et al.(2015); population study	Questionnaire invitation population 3,838; exercise-tested subsample 146; estimated EIB 19.2%, EILO 5.7%.	Population estimates from a subsample, not athlete-referral frequencies.	Retain sampling population and test-specific denominator.
Nielsen et al (Walsted Nielsen et al., 2013); retrospective referral cohort	EILO in 31/88 (35.2%); inspiratory symptoms did not differentiate groups.	Symptomatic referral selection.	Keep symptom description separate from anatomy.
Hammer et al (Jansrud Hammer et al., 2022); elite-athlete referral cohort	EILO 73/101 (72.3%); asthma-supporting findings 22/55 (40.0%) with identifiable asthma investigations.	Available-test denominator differs from the full EILO group.	Record each test and the condition its findings support.
Olin et al.(2016); retrospective time-window analysis	71 included patients; greatest observed obstruction at peak workload, with rapid recovery.	Selected young patients; not a reporting-method trial.	Observe peak demand and document early resolution.
Walsted et al.(2017); repeat-testing study	20 subjects completed four CLE tests; inter-rater weighted kappa 0.16–0.45.	Small case/control sample; grade agreement is not diagnostic sensitivity.	Preserve raters, grade components and adjudication.
Irewall et al.(2023); longitudinal follow-up	29/98 returned (29.6%); 3/18 (16.7%) with initially absent/mild obstruction later had moderate obstruction without dyspnoea.	Selective return; most scores remained stable.	Date observations and record symptoms independently.

Eucapnic voluntary hyperpnoea (EVH) has limitations in diagnostic performance and repeatability; lower-airway investigations also differ in performance (Hull et al., 2016; Reier-Nilsen et al., 2023) Interpretation of a negative result retains the named challenge, achieved stimulus, medications and environment.

CHALLENGE-SPECIFIC ACQUISITION

INDICATION, UNIT OF RECORDING AND SAFETY

Before testing, a clinician records the indication for each airway investigation, competing diagnoses and the intended test. The record does not mandate both tests for every breathless athlete. Initial respiratory assessment and further cardiopulmonary investigation follow clinical indications; the IOC consensus addresses the differential of non-infective respiratory illness (Schwellnus et al., 2022)

One acquisition record represents one provocation. An athlete identifier links records, while a separate challenge identifier distinguishes date, modality and session. For CLE, $t = 0$ is cessation of the externally imposed exercise workload, marked by the operator. For a lower-airway exercise challenge it is cessation of exercise; for EVH it is cessation of the prescribed hyperpnoea stimulus. An incremental challenge outside these two lower-airway modalities requires its own standard-specific anchors and interpretation and is not forced into the sampling schedule below.

Sequential provocations retain separate identifiers even if performed during one visit. Two exhaustive respiratory provocations should not be scheduled together simply to obtain a complete record (Clemm et al., 2022) A test stopped for safety retains its acquired evidence and termination reason. Additional effort or delayed rescue treatment is never justified by a documentation requirement.

TIMING, SYNCHRONISATION AND OBSERVABLE INTERVALS

The operator uses one monotonic elapsed-time clock for all streams within a challenge and separately stores the session date/time and time zone. Synchronisation cues visible or audible in the laryngoscopic recording are entered in the workload/event log at acquisition start and recording end. A third, non-disruptive cue during exercise provides an independent check. Device-native timestamps are retained. If a common clock is unavailable, the start/end anchors define a linear correction $t_{\mathrm{master}} = a + b\,t_{\mathrm{native}}$, where b is the ratio of the master-clock anchor interval to the native-clock anchor interval and a aligns the start anchors. Store the anchors, coefficients, native times and corrected times.

Event times are recorded to the nearest second. Clock acceptance requires a residual difference of no more than 1 s at the independent check cue after correction; fitting the two anchors alone cannot verify alignment accuracy. This one-second limit is an operational rule, not a validated physiological threshold. If the check is absent or fails, retain the measured error or failure reason and mark fine temporal ordering indeterminate. Failure of timing alignment need not invalidate an adequate anatomical observation or valid FEV_1 response; it prevents claims about their precise timing relationship.

Symptom onset means the athlete's first contemporaneous indication that the usual abnormal breathing has begun, rather than the first moment of expected exertional breathlessness. Before exercise, the athlete selects a brief hand signal or phrase that does not impede the task. The operator logs it and distinguishes the athlete's signal from an observed sound. Peak-symptom time is entered contemporaneously if possible; retrospective recollection is labelled as such and is not assigned artificial second-level precision. Symptom resolution can be recorded as an interval between the last positive and first negative check.

Every video interval is marked interpretable, partly obscured or unavailable. Workload and symptom events outside interpretable intervals remain in the record. Figure 1 locates the acquisition windows on separate challenge clocks. The bars mark recording periods, and the circles mark spirometry sets.

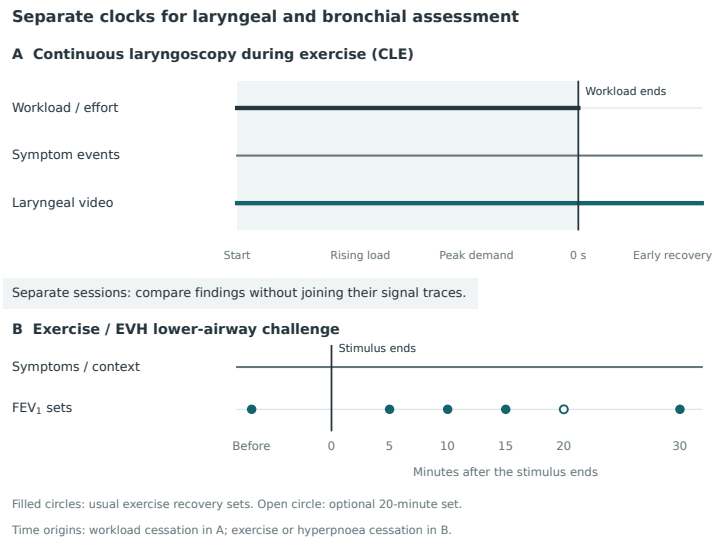


FIGURE 1. CHALLENGE-SPECIFIC ACQUISITION WINDOWS. PANEL A LINKS LARYNGEAL VIDEO, SYMPTOM EVENTS AND WORKLOAD WITHIN A CLE CHALLENGE, WITH EARLY RECOVERY RETAINED WHEN SAFELY OBSERVABLE. PANEL B SHOWS PRE-CHALLENGE AND RECOVERY SPIROMETRY FOR A SEPARATELY IDENTIFIED EXERCISE OR EVH CHALLENGE. FILLED CIRCLES INDICATE THE USUAL 5-, 10-, 15- AND 30-MINUTE RECOVERY SETS; THE OPEN CIRCLE INDICATES AN OPTIONAL 20-MINUTE SET. SEPARATE CHALLENGES ARE NOT EVIDENCE OF SIMULTANEOUS PATHOPHYSIOLOGY. PANEL A IS SCHEMATIC AND NOT TO SCALE.

LARYNGEAL CHANNEL

The exercise task addresses the athlete's usual trigger and is recorded as treadmill, cycle or a specified sport task. Store stage duration, speed/grade or power, achieved intensity, heart rate or ventilation if acquired, termination reason and the athlete's judgement of task similarity. CLE requires specialist equipment and trained personnel (Clemm et al., 2022)

Store continuous video from the available pre-exercise period through peak effort and, when safe and tolerated, early recovery until symptom resolution or termination of recording. This is intended to capture onset and resolution rather than to restrict laryngeal observation to negative clock times. Record the final interpretable frame time and the reason for ending the video.

The named CLE visual grading system assigns separate glottic and supraglottic scores from 0 to 3 at moderate and maximal effort. Retain all four scores, their video windows, rater and uncertainty rather than a total alone (Walsted et al., 2017) A grade of at least 2 at either level is commonly used to identify clinically relevant obstruction; retain anatomical configuration, technical quality and clinical context (Clemm et al., 2022) A negative laryngeal classification requires an interpretable peak or another adequate provoking view.

For independent grading, two trained observers score the selected windows while masked to lower-airway findings. Disagreement about threshold status or anatomical level goes to a third observer; both initial scores and the final adjudication remain available. Training uses a shared image atlas and practice recordings. Agreement is calculated from the independent scores before adjudication, preserving the observer variation identified by Walsted and colleagues (Walsted et al., 2017)

LOWER-AIRWAY CHANNEL

The responsible respiratory laboratory follows a named exercise or EVH standard. Store the intended stimulus, achieved ventilation or other standard-specific adequacy measure, exposure duration, inspired-air temperature/humidity where measured and recent exercise. Medication names, doses and actual last-use times are recorded alongside the laboratory's preparation instructions and any departures. Medication withholding is prescribed by the supervising clinician; a universal withholding interval is not introduced here. These requirements preserve the context in which an indirect challenge can be interpreted (Hallstrand et al., 2018)

Pre-challenge and recovery manoeuvres require documented equipment checks, acceptability and repeatability. Pre-challenge spirometry aims for at least three acceptable FEV_1 manoeuvres; for athletes aged 12 years or older, the two largest acceptable values should differ by no more than 0.150 L. The applicable ATS/ERS standard and set quality are recorded; usable but unacceptable manoeuvres are flagged rather than silently included (Graham et al., 2019) After an exercise challenge, the usual recovery schedule is 5, 10, 15 and 30 min, with an optional 20-min set or additional earlier sets if specified by the laboratory. At least two repeatable FEV_1 manoeuvres are obtained per interval, with the highest acceptable value retained for that interval (Parsons et al., 2013) EVH uses its named protocol-specific schedule rather than assuming that every preparation and sampling detail is interchangeable with exercise.

Store scheduled time, actual start/end times for the set and each manoeuvre timestamp. The timestamp of the selected manoeuvre is the time assigned to the retained interval value. A set that falls outside its laboratory-defined timing tolerance remains visible with the deviation and its interpretation. Rescue treatment, termination or technical failure can change the available sampling window; the record states when this occurred and which sets remain interpretable.

VALUE SELECTION AND THE OBSERVED MINIMUM

Let $B > 0$ be the highest acceptable pre-challenge FEV_1 in a set satisfying the applicable repeatability requirement. At scheduled interval i, let $\mathcal{A}_i$ contain the acceptable manoeuvres in a repeatable set, with measured values F_{ij} and timestamps t_{ij}. Define

$$R_i = \max_{j \in \mathcal{A}_i} F_{ij},$$
$$t_i = t_{i,j_i}, \quad j_i \in \arg\max_{j \in \mathcal{A}_i} F_{ij}, \tag{1}$$
$$D = 100\,\frac{B - \min_{i \in \mathcal{V}} R_i}{B},$$
$$\mathcal{V} = \{i : \text{valid post-challenge set within the prescribed window}\}. \tag{2}$$

The earliest manoeuvre is retained when acceptable values tie at stored instrument precision. The observed minimum time is t_{i^*}, where i^* indexes the lowest retained interval value; all tied intervals are also stored. For the exercise protocol, the prescribed analysis window extends through 30 min. A fall of at least 10% follows the ATS EIB criterion when challenge and acquisition requirements are met (Parsons et al., 2013) A laboratory using another prespecified age- or challenge-specific criterion records it explicitly; thresholds are not changed after inspecting the result.

The calculation is undefined when B or all valid recovery sets are absent. Missing sets are not imputed. If all recovery values exceed B, a negative calculated fall is retained as an increase in measured function, not clipped to manufacture a zero. A qualifying valid reduction can support a positive classification even if later sets are missing; the incomplete acquisition is still flagged. An early safety interruption before an adequate stimulus ordinarily prevents a negative classification.

The minimum in Equation 2 is a *sampled minimum*. It does not locate the physiological nadir between manoeuvres. Actual sampling intervals and omissions accompany D, especially when comparing observed minimum times across visits.

EVIDENCE STATUS, TEST DISPOSITION AND COMBINED REPORTING

THREE STATUSES FOR ATTEMPTED TESTS

Apply the following order separately to each domain:

1. **Positive (P):** qualifying objective evidence is interpretable and meets the named criterion under an accepted provocation. A missing later observation does not erase this evidence.
2. **Negative under recorded test conditions (N):** no criterion is met, and the required stimulus, observation window and technical quality are adequate. This is an episode-specific result, not a permanent exclusion of disease.
3. **Indeterminate (I):** no qualifying positive evidence is established and a critical requirement for a negative interpretation is unmet. Record the reason: inadequate stimulus, absent peak view, invalid spirometry, missing required sets, timing failure affecting the claim, or safety interruption.

Temporal adequacy and diagnostic adequacy are separate fields. A video can support an anatomical finding while its exact relation to symptom onset is uncertain. Similarly, one technically valid positive bronchial response can coexist with incomplete characterisation of its subsequent time course.

Whether the challenge elicits the usual symptoms is recorded as yes, no, uncertain or not assessed, with the athlete's account of task similarity. Symptoms do not override valid positive evidence. A CLE challenge that does not elicit the usual trigger limits inference from a negative view. For a standardised lower-airway challenge, symptom occurrence does not replace the objective stimulus-adequacy criterion.

UNPERFORMED TESTS AND PAIRED OUTCOMES

A test not attempted receives a disposition, not an attempted-test result: `NI` (not clinically indicated), `SD` (safety-deferred), `DC` (declined), `UA` (unavailable) or `PN` (planned, not completed). No such code is translated into N. When both domains are attempted, all nine status combinations are retained (Table 3). When one is unperformed, report the measured domain and the other domain's disposition in the same way.

TABLE 3. PAIRED EVIDENCE STATES FOR ATTEMPTED UPPER- AND LOWER-AIRWAY TESTS. "DETECTED" REFERS ONLY TO THE NAMED CHALLENGE CRITERION.

	EIB P	**EIB N**	**EIB I**
EILO P	Both criteria detected	EILO detected; EIB not detected under tested conditions	EILO detected; EIB unresolved
EILO N	EIB detected; EILO not detected under tested conditions	Neither criterion detected under tested conditions	EILO not detected under tested conditions; EIB unresolved
EILO I	EIB detected; EILO unresolved	EIB not detected under tested conditions; EILO unresolved	Both domains unresolved

Separate positive challenges establish that both conditions were detected during the evaluation period. They do not demonstrate simultaneous obstruction within one exercise episode. When both domains are negative, the report identifies the tested conditions and returns unresolved symptoms to the broader differential rather than assigning an airway explanation by exclusion.

ACQUISITION COMPLETENESS AND CLINICAL INFERENCE

Completeness is a separate field: complete, incomplete with preserved interpretable findings, or no interpretable acquisition. The summary contains the criterion, value/grade, actual observation time, adequacy decision, missing requirements and any unperformed-test disposition. A P/I pair means "EILO detected; lower-airway response unresolved", with the reason for I. It retains the positive laryngeal finding without assigning a bronchial diagnosis.

Direct laryngeal grading, symptom severity and activity restriction are retained separately during follow-up. The same visual grade can have a different clinical meaning when workload, symptoms or task exposure changes. Ersson and colleagues provide longitudinal evidence that test-defined EIB and EILO can change over time (Ersson et al., 2023) Task and exposure comparability must therefore accompany repeat assessments.

PROSPECTIVE VALIDATION PROTOCOL

CONSECUTIVE COHORT AND PRIMARY ESTIMAND

Clinical validation requires a prospective cohort; the protocol in this section has not been conducted. A three-centre cohort would enrol consecutive competitive athletes aged 12 years or older referred with unresolved exertional respiratory symptoms for whom both airway investigations are clinically indicated before study interpretation. Competitive participation is defined as organised club-level competition or higher within the preceding 12 months. Athletes outside this dual-indication group can enter a parallel registry but do not enter its primary denominator. Adolescents and adults are analysed separately.

The planned evaluation episode extends for 42 days from the initial testing decision. This interval is an administrative study rule, not a disease threshold. Contraindications, patient decisions, lack of availability, technical failures and unfinished appointments remain in the enrolled denominator. Research participation does not require an unsafe challenge. Ethical approval, consent and adolescent assent would be obtained before any prospective recruitment or secondary use of identifiable video.

The primary estimand is the proportion of enrolled dual-indication episodes with an interpretable P or N result for each domain by day 42:

$$C = \frac{\#\{\text{episodes with both domain statuses in } \{P, N\}\}}{N_{\text{enrolled}}}. \tag{3}$$

Report C with a 95% Wilson confidence interval, alongside domain-specific P/N/I frequencies, unperformed-test dispositions and acquisition-completeness categories. Because valid positive findings can survive incomplete acquisition, report a second, stricter proportion with every required acquisition item complete. Do not remove technical failures or safety deferrals to increase either proportion.

A conservative precision calculation uses $p = 0.5$, desired half-width $h = 0.05$ and $z = 1.96$: $n \approx z^2 p(1-p)/h^2 = 384.16$, rounded to 385 enrolled dual-indication episodes. At $n = 385$, the worst-case Wilson half-width is approximately 0.0497. This is a design calculation, not an observed sample or a power claim. The precision pertains to the pooled proportion; centre-specific estimates will be less precise. All 385 episodes remain in the denominator, so incomplete tests are outcomes rather than attrition excluded from analysis.

Secondary implementation outcomes are staff minutes for entry and adjudication, the proportion of correctly linked recordings, synchronisation residual error, missing-field frequency, and adverse events. The analysis compares these outcomes across centres, modality and age groups and reports their distributions without assuming a clinical benefit.

REPORT-INTERPRETATION EXPERIMENT

A nested experiment would use 80 deidentified cohort episodes sampled by paired evidence state, with deliberate representation of partial and unperformed evaluations. Each episode generates two reports containing the same factual observations: a field-structured report and a chronological prose report. Both retain technical failures and unavailable results; the contrast tests organisation and explicit status representation rather than additional measurements.

Twelve respiratory/sports-medicine readers each interpret 40 different cases, 20 per format, in a balanced allocation giving each case six readings, three per format. Each reader sees an episode only once. Format and order are randomised within reader, and reporting labels are removed where possible. A separate three-member panel reviews the full deidentified files, masked to reader responses; its compartment-specific judgement follows the prespecified rules. Irreconcilable panel decisions remain unresolved rather than becoming a manufactured reference diagnosis.

The experiment's primary outcome is an *unqualified negative interpretation*: a reader labels an indeterminate or unperformed required domain negative without acknowledging its limitation. Report the absolute difference in error proportions between formats with a 95% interval from a logistic model with crossed random intercepts for case and reader, using marginal standardisation. Record unresolved panel cases separately. A sensitivity analysis uses case- and reader-resampling appropriate to the crossed design. Secondary outcomes are interpretation time, correct preservation of a positive result in a partial evaluation and correct identification of an unperformed test.

The 80-case, 12-reader experiment is a prespecified feasibility design with 480 readings; these are not 480 independent patients. No formal power is claimed. Interval width and variance components would inform a later confirmatory reader study. The experiment measures understanding of available evidence, not the sensitivity or specificity of CLE or bronchoprovocation against an independent disease reference.

AGREEMENT, TIMING AND FOLLOW-UP ANALYSIS

For laryngeal grading, report glottic and supraglottic ordinal agreement separately using weighted kappa with linear weights $w_{uv} = 1 - |u - v|/3$ for grades $u, v \in \{0, 1, 2, 3\}$, raw agreement and confidence intervals from 2,000 athlete-level bootstrap resamples. Report threshold-status agreement separately; a high percentage agreement does not replace inspection of the marginal grade distribution. In a volunteer subset of 50 athletes without an intervening treatment change, repeat matched-task tests only when clinically justified and safe within 7–21 days. Describe eligibility, refusals and protocol departures. This subset size is for exploratory repeatability estimation, not a claim of definitive precision.

For continuous timing measurements, use paired differences and agreement limits, with separate analysis for mode or workload changes. Symptom onsets expressed as intervals are analysed as intervals, not exact events. Observed FEV_1 minimum times are compared only when sampling schedules and valid sets are comparable; their discrete distribution and differences retain the sampling resolution. Missing measurement values are not replaced with normal values. A missing day-42 outcome remains incomplete for the primary estimand, with its reason retained.

At three months, record attendance and training/competition restriction, symptom burden and treatment use independently of test classification. These observational outcomes describe participation; they cannot establish that the reporting specification caused improvement. Intervention effects require a separate controlled study.

REHABILITATION RELEVANCE AND LIMITATIONS

Treatment follow-up separates symptoms, physiology and participation. Kolnes and colleagues identified symptom improvement in a four-athlete physiotherapy case series, whose design does not support broad causal inference (Kolnes et al., 2019) In Fujiki and colleagues' speech-language treatment cohort, questionnaire participation declined from 59 initially to 27 at six months, corresponding to 45.8% retention (Fujiki et al., 2023) Treatment effects and follow-up participation therefore require separate interpretation. The diagnostic and management review by Liyanagedara and colleagues also identifies variation in treatment findings across interventions and study designs (Liyanagedara et al., 2017)

Acquisition depends on specialist access, safe performance of the provoking task, cooperation and instrument quality. A one-second synchronisation target can be technically unattainable in some systems and does not establish one-second biological resolution. Separately acquired tests may occur under different medication and environmental conditions; those conditions remain attached to each challenge.

The literature assessment was non-systematic and restricted to January 2011–February 2024. Referral fractions do not estimate prevalence in unselected athletes. Christensen and colleagues' general-public investigation has a different sampling population from specialist athlete cohorts (Christensen et al., 2011) Observer variation, selective follow-up and challenge-specific limitations constrain interpretation. A complete record does not establish the cause of dyspnoea when both airway investigations are negative or another disease is responsible.

The validation protocol tests whether explicit status entries reduce unqualified negative interpretations of incomplete examinations. Interpretation errors, acquisition completion and participation outcomes are separate endpoints. Clinical adoption requires evidence that these changes justify staff time and access costs.

CONCLUSION

A phase-registered airway record links each finding to its challenge, observation time and acquisition quality. The bronchial calculation selects the highest acceptable manoeuvre within each repeatable set and the lowest valid interval value across recovery. Paired classification retains all nine P/N/I combinations, while unperformed tests retain their disposition. These rules preserve positive findings and prevent unavailable observations from becoming negative results. Completion rates, observer agreement and clinical benefit require prospective evaluation.

DECLARATIONS

CONFLICT OF INTEREST

The author declares that there are no conflicts of interest related to this study.

DATA AVAILABILITY

The data supporting the findings of this study are available from the corresponding author upon reasonable request.

FUNDING

This research received no external funding.

REFERENCES

1. Christensen PM, Thomsen SF, Rasmussen N, Backer V. Exercise-induced laryngeal obstructions: prevalence and symptoms in the general public. *Eur Arch Otorhinolaryngol*. 2011;268(9):1313–1319. doi: 10.1007/s00405-011-1612-0.

2. Christensen PM, Maltbæk N, Jørgensen IM, Nielsen KG. Can flow-volume loops be used to diagnose exercise-induced laryngeal obstructions? A comparison study examining the accuracy and inter-rater agreement of flow-volume loops as a diagnostic tool. *Prim Care Respir J*. 2013;22(3):306–311. doi: 10.4104/pcrj.2013.00067.

3. Clemm HH, Olin JT, McIntosh C, et al. Exercise-induced laryngeal obstruction (EILO) in athletes: a narrative review by a subgroup of the IOC Consensus on 'acute respiratory illness in the athlete'. *Br J Sports Med*. 2022;56(11):622–629. doi: 10.1136/bjsports-2021-104704.

4. Ersson K, Mallmin E, Nordang L, Malinovschi A, Johansson H. A longitudinal study of exercise-induced bronchoconstriction and laryngeal obstruction in high school athletes. *Scand J Med Sci Sports*. 2023;33(8):1509–1518. doi: 10.1111/sms.14373.

5. Fujiki RB, Olson-Greb B, Braden M, Thibeault SL. Therapy outcomes for teenage athletes with exercise-induced laryngeal obstruction. *Am J Speech Lang Pathol*. 2023;32(4):1517–1531. doi: 10.1044/2023_AJSLP-22-00359.

6. Graham BL, Steenbruggen I, Miller MR, et al. Standardization of spirometry 2019 update. An official American Thoracic Society and European Respiratory Society technical statement. *Am J Respir Crit Care Med*. 2019;200(8):e70–e88. doi: 10.1164/rccm.201908-1590ST.

7. Hallstrand TS, Leuppi JD, Joos G, et al. ERS technical standard on bronchial challenge testing: pathophysiology and methodology of indirect airway challenge testing. *Eur Respir J*. 2018;52(5):1801033. doi: 10.1183/13993003.01033-2018.

8. Halvorsen T, Walsted ES, Bucca C, et al. Inducible laryngeal obstruction: an official joint European Respiratory Society and European Laryngological Society statement. *Eur Respir J*. 2017;50(3):1602221. doi: 10.1183/13993003.02221-2016.

9. Hull JH, Ansley L, Price OJ, Dickinson JW, Bonini M. Eucapnic voluntary hyperpnea: gold standard for diagnosing exercise-induced bronchoconstriction in athletes? *Sports Med*. 2016;46(8):1083–1093. doi: 10.1007/s40279-016-0491-3.

10. Irewall T, Bäcklund C, Naumburg E, Ryding M, Stenfors N. A longitudinal follow-up of continuous laryngoscopy during exercise test scores in athletes irrespective of laryngeal obstruction, respiratory symptoms, and intervention. *BMC Sports Sci Med Rehabil*. 2023;15:87. doi: 10.1186/s13102-023-00681-9.

11. Jansrud Hammer I, Halvorsen T, Vollsæter M, et al. Conundrums in the breathless athlete; exercise-induced laryngeal obstruction or asthma? *Scand J Med Sci Sports*. 2022;32(6):1041–1049. doi: 10.1111/sms.14137.

12. Johansson H, Norlander K, Berglund L, et al. Prevalence of exercise-induced bronchoconstriction and exercise-induced laryngeal obstruction in a general adolescent population. *Thorax*. 2015;70(1):57–63. doi: 10.1136/thoraxjnl-2014-205738.

13. Kolnes LJ, Vollsæter M, Røksund OD, Stensrud T. Physiotherapy improves symptoms of exercise-induced laryngeal obstruction in young elite athletes: a case series. *BMJ Open Sport Exerc Med*. 2019;5(1):e000487. doi: 10.1136/bmjsem-2018-000487.

14. Liyanagedara S, McLeod R, Elhassan HA. Exercise induced laryngeal obstruction: a review of diagnosis and management. *Eur Arch Otorhinolaryngol*. 2017;274(4):1781–1789. doi: 10.1007/s00405-016-4338-1.

15. Olin JT, Clary MS, Fan EM, et al. Continuous laryngoscopy quantitates laryngeal behaviour in exercise and recovery. *Eur Respir J*. 2016;48(4):1192–1200. doi: 10.1183/13993003.00160-2016.

16. Parsons JP, Hallstrand TS, Mastronarde JG, et al. An official American Thoracic Society clinical practice guideline: exercise-induced bronchoconstriction. *Am J Respir Crit Care Med*. 2013;187(9):1016–1027. doi: 10.1164/rccm.201303-0437ST.

17. Reier-Nilsen T, Sewry N, Chenuel B, et al. Diagnostic approach to lower airway dysfunction in athletes: a systematic review and meta-analysis by a subgroup of the IOC consensus on 'acute respiratory illness in the athlete'. *Br J Sports Med*. 2023;57(8):481–489. doi: 10.1136/bjsports-2022-106059.

18. Schwellnus M, Adami PE, Bougault V, et al. International Olympic Committee (IOC) consensus statement on acute respiratory illness in athletes part 2: non-infective acute respiratory illness. *Br J Sports Med*. 2022;56(19):1089–1103. doi: 10.1136/bjsports-2022-105567.

19. Walsted ES, Hull JH, Hvedstrup J, Maat RC, Backer V. Validity and reliability of grade scoring in the diagnosis of exercise-induced laryngeal obstruction. *ERJ Open Res*. 2017;3(3):00070-2017. doi: 10.1183/23120541.00070-2017.

20. Walsted Nielsen E, Hull JH, Backer V. High prevalence of exercise-induced laryngeal obstruction in athletes. *Med Sci Sports Exerc*. 2013;45(11):2030–2035. doi: 10.1249/MSS.0b013e318298b19a.