



Evidence Annex: retest windows, reliability sources and references

Companion to the SIIM Core, version 11

Last reviewed September 2026

Next scheduled review September 2027

This Annex holds every dated value and source the protocol relies on. It is reviewed at least once a year, and each revision is logged.

Purpose

The Core document sets out rules: name the reference frame, compare against measurement error, report process and outcome separately. Those rules change rarely. The values that make them operable, such as a reliability figure, a retest window or a reference cut-off, change whenever better evidence is published, independently of the rules themselves.

This Annex holds those values, so that updating a number never requires changing the Core. Each entry states its source and a status.

Status labels

- **Physiological:** set by known biology, such as red cell lifespan.
- **Convention:** a reasoned working figure, to be replaced when better data appear.
- **Structural:** a property of the measure itself, which a longer interval does not resolve.

Table 1. Retest windows and reliability evidence

Measure	Earliest meaningful retest	Basis for the window	Status	Reliability and reference evidence
Static single-leg balance	4 to 6 weeks	Mainly neural and proprioceptive adaptation, which is fast	Convention	Normative values by age, with excellent inter-rater reliability for the best of three trials (A2). A test-retest MDC is to be established locally.
Sit-to-stand power	6 to 8 weeks	Neural drive and rate of force development adapt before muscle size	Convention	Normative data and functional cut-offs from 9,906 adults, with a minimal clinically important difference reported (A3)
Maximal gait speed	8 to 12 weeks	Depends on lower-limb power, balance and neural control, so it follows its components	Convention	Norms from 1,320 adults; fair test-retest ICCs and high MDCs, with limited value for judging individual change (A1)
Estimated aerobic capacity	8 to 12 weeks	Central adaptations such as stroke volume, plus peripheral mitochondrial change	Convention	Device validity and reliability reported, with mean absolute percentage error above 10% against laboratory measurement (A5)
Resting heart rate variability	8 weeks, read as a multi-week aggregate at both ends	Responds to training load within weeks, with large day-to-day variation	Structural	Normative ranges from a systematic review of 44 studies (A4). A repeatability threshold is to be established locally.
Grip strength	12 weeks minimum, treated as a long-horizon measure	Small transfer effects from exercise training (A8)	Convention	European norms from 27 countries (A6); robust outcome associations in large cohorts (A7)
Lean mass by DXA	12 weeks minimum; 16 to 24 weeks more realistic	Lean mass accrues slowly relative to scanner precision	Convention	European DXA reference values from 10,894 adults (A9)
HbA1c	About 12 weeks	Red cell lifespan of roughly 120 days sets a floor	Physiological	Reference change value approach (A14, A15)
Lipid panel	6 to 8 weeks	Effects of diet and training usually stabilise within this window	Convention	Reference change value approach (A14, A15)
hs-CRP	A single reading may be uninterpretable at any interval	Large within-person variation; acute illness confounds single values	Structural	Repeat sampling required
Cognitive measures	6 to 12 months, with alternate test forms	Practice effects dominate shorter intervals	Structural	Instrument-specific

Notes on Table 1

Local calibration. For several measures, the published literature provides reference values and inter-rater reliability, while a test-retest minimum detectable change for the operator's own equipment and population is still missing. Core section 12 anticipates this: until thresholds are calibrated on the operator's own clients, verdicts carry that uncertainty, and any threshold used as a working convention should be labelled as one.

Gait speed. The largest reliability study of four-metre gait speed (A1) is useful in two ways. Its norms support Interpretation, and its authors' conclusion, that fair ICCs and high MDCs limit the measure's use in judging change, supports caution in Monitoring. A measure can serve one step well and another poorly.

Grip strength. Grip carries robust outcome associations in large cohorts (A7). A meta-analysis of exercise training in adults aged 60 and over found only small transfer effects on grip strength (standardised mean difference 0.28), and its authors concluded that grip cannot be recommended as a general marker of functional change for all types of exercise programme (A8). Predictive strength and responsiveness are separate properties.

Estimation method. Different estimation equations applied to the same raw test can produce different absolute values of aerobic capacity. The method is held constant across a client's assessments.

Table 2. One construct, two sources, two thresholds

This table illustrates Core section 6, Rule 1 (name the reference frame).

Source	Context	Grip threshold, men	Grip threshold, women	Purpose
Swiss framework (Core reference 3)	Clinical risk assessment	Below 26 kg	Below 18 kg	Associated with a twofold increase in all-cause hospital admissions
Delphi consensus (Core reference 4)	Locomotion within intrinsic capacity	Below 27 kg	Below 16 kg	Normative cut-off for a practical clinical assessment

Both thresholds are defensible and current. They come from different sources and serve different purposes, so a cited grip threshold should always name its source.

Table 3. Accuracy and repeatability

Property	Meaning	What it governs
Accuracy	Agreement with a criterion measurement	Trust in the absolute reported value
Repeatability	Consistency of the device's own reading in the same person under the same conditions	Trust in a change between two readings

A validation study of a seismocardiography-based estimate of maximal oxygen uptake reported a mean absolute percentage error above 10% against laboratory measurement, and concluded that the method needs further development before clinical use (A5). Much of that error is systematic and largely cancels when the same device measures the same person twice. This is why repeatability governs the Monitoring verdict in Core section 8, Rule 4, and why the absolute value is always reported as an estimate.

Evidence references

Functional measures

- A1. Bohannon RW, Wang YC. Four-meter gait speed: normative values and reliability determined for adults participating in the NIH Toolbox study. *Archives of Physical Medicine and Rehabilitation*. 2019;100(3):509-513. doi:10.1016/j.apmr.2018.06.031
- A2. Springer BA, Marin R, Cyhan T, Roberts H, Gill NW. Normative values for the unipedal stance test with eyes open and closed. *Journal of Geriatric Physical Therapy*. 2007;30(1):8-15. doi:10.1519/00139143-200704000-00003
- A3. Alcazar J, Alegre LM, Van Roie E, Magalhães JP, Nielsen BR, González-Gross M, et al. Relative sit-to-stand power: aging trajectories, functionally relevant cut-off points, and normative data in a large European cohort. *Journal of Cachexia, Sarcopenia and Muscle*. 2021;12(4):921-932. doi:10.1002/jcsm.12737
- A4. Nunan D, Sandercock GRH, Brodie DA. A quantitative systematic review of normal values for short-term heart rate variability in healthy adults. *Pacing and Clinical Electrophysiology*. 2010;33(11):1407-1417. doi:10.1111/j.1540-8159.2010.02841.x
- A5. Hansen MT, Rømer T, Højgaard A, Husted K, Sørensen K, Schmidt SE, et al. Validity and reliability of seismocardiography for the estimation of cardiorespiratory fitness. *Cardiovascular Digital Health Journal*. 2023;4(5):155-163. doi:10.1016/j.cvdhj.2023.08.020
- A6. Grgic J, Wazny VK, Maier AB, Schoenfeld BJ, Pedisic Z. Reference values for handgrip strength in Europe: analysis of individual participant data from 27 countries. *GeroScience*. 2026;48(3):4349-4371. doi:10.1007/s11357-025-01919-9
- A7. Leong DP, Teo KK, Rangarajan S, Lopez-Jaramillo P, Avezum A, Orlandini A, et al. Prognostic value of grip strength: findings from the Prospective Urban Rural Epidemiology (PURE) study. *The Lancet*. 2015;386(9990):266-273. doi:10.1016/S0140-6736(14)62000-6
- A8. Labott BK, Bucht H, Morat M, Morat T, Donath L. Effects of exercise training on handgrip strength in older adults: a meta-analytical review. *Gerontology*. 2019;65(6):686-698. doi:10.1159/000501203
- A9. Ofenheimer A, Breyer-Kohansal R, Hartl S, Burghuber OC, Krach F, Schrott A, et al. Reference values of body composition parameters and visceral adipose tissue (VAT) by DXA in adults aged 18-81 years: results from the LEAD cohort. *European Journal of Clinical Nutrition*. 2020. doi:10.1038/s41430-020-0596-5

Intervention and activity evidence

A10. Sherrington C, Fairhall NJ, Wallbank GK, Tiedemann A, Michaleff ZA, Howard K, et al. Exercise for preventing falls in older people living in the community. *Cochrane Database of Systematic Reviews*. 2019;(1):CD012424. doi:10.1002/14651858.CD012424.pub2

A11. Stamatakis E, Ahmadi MN, Gill JMR, Thøgersen-Ntoumani C, Gibala MJ, Doherty A, et al. Association of wearable device-measured vigorous intermittent lifestyle physical activity with mortality. *Nature Medicine*. 2022;28(12):2521-2529. doi:10.1038/s41591-022-02100-x

A12. Currier BS, D'Souza AC, Fiatarone Singh MA, Lowisz CV, Rawson ES, Schoenfeld BJ, et al. American College of Sports Medicine position stand: resistance training prescription for muscle function, hypertrophy, and physical performance in healthy adults: an overview of reviews. *Medicine & Science in Sports & Exercise*. 2026;58(4):851-872.

A13. Nuzzo JL, Pinto MD, Kirk BJC, Nosaka K. Resistance exercise minimal dose strategies for increasing muscle strength in the general population: an overview. *Sports Medicine*. 2024;54(5):1139-1162. doi:10.1007/s40279-024-02009-0

Blood biomarkers

A14. Harris EK, Yasaka T. On the calculation of a "reference change" for comparing two consecutive measurements. *Clinical Chemistry*. 1983;29(1):25-30.

A15. Fraser CG, Harris EK. Generation and application of data on biological variation in clinical chemistry. *Critical Reviews in Clinical Laboratory Sciences*. 1989;27(5):409-437. doi:10.3109/10408368909106595

Revision log

Date	Change
September 2026	Annex version 1 created, holding all dated values and evidence citations separately from the Core.
September 2026	Annex version 2. Every reference checked against its primary source. Gait speed reliability source replaced with the NIH Toolbox norming study (A1). Grip responsiveness evidence added (A8). Reference change value sources added (A14, A15). Reliability values that could not be traced to their cited sources removed, pending local calibration.

Maintenance

This Annex is reviewed at least once a year. Between reviews, any entry superseded by a larger cohort, a newer position stand or a locally calibrated value is updated, dated and logged above. The Core changes only when one of the conditions in Core section 12 is met.