Estimating Sit-to-Stand Power from a Single Smart Seat Recording: Validation Against the Supervised Chair-Stand Test in 122 Adults

Nick Mattis, Staff Software Engineer II Casana Care, Inc. · 25 August 2026

Abstract

Background. Lower-limb muscle power, how fast a person can rise from a chair, predicts mobility loss in older adults better than strength or mass. It is estimated with a supervised, stopwatch-timed chair-stand test, as in the Short Physical Performance Battery, so it mostly goes unmeasured. The Casana Smart Seat, a sensor-embedded toilet seat used with no change in routine, observes a sit-to-stand at every use.

Aims. We asked whether one recording plus age, weight, and height can recover the chair test's relative sit-to-stand power, and whether the estimate ranks people within their age decade, where age says nothing.

Methods. 896 Smart Seat recordings from 124 adults aged 24 to 89, from two IRB-approved collections, were analyzed against a supervised chair-stand reference standard. The algorithm reads only the seat's four corner load cells; every result is leave-one-subject-out at one recording per subject.

Results. Estimated relative power correlated with the chair-test-derived reference at a Spearman rho of 0.54 (95% CI 0.39 to 0.67, n = 122), median absolute error 20.0%; the raw seat stand-up interval without the model reached 0.39 and 30.5%. Against age alone, the estimate kept ranking within single decades: 0.40 in the 70 to 79 decade where age fell to 0.07, with its largest advantage at 80 and over (+0.62, 95% CI +0.26 to +0.96). Decade medians tracked the published normative distribution, and a returning cohort declined across 4.7 years at the published rate.

Conclusion. One Smart Seat recording places a person's leg power against their age peers on the clinically validated scale and, unlike age, distinguishes stronger from weaker people within a decade. Each reading is a placement, not an exact measurement, made several times a day with no visit, stopwatch, or effort. These findings support further evaluation as a passive monitor of lower-limb power in older adults.

1. Introduction

Sarcopenia, the age-related loss of muscle mass and function, has a global prevalence of 10% to 27% in older adults [2]. More than one in four US adults aged 65 and over reports a fall each year [3], and the death rate from unintentional falls in this group reached 69.9 per 100,000 in 2023, having risen across every older age group for two decades [4]. Behind both numbers is the loss of lower-limb muscle power, the product of force and velocity, expressed most simply as how quickly someone can stand up. Power declines faster with age than strength, at roughly 3.5% per year against 1% to 2% for isometric strength [5], and in mobility-limited older adults leg power explains more of the variance in physical performance than strength on every measure tested [6]. Below the published power cut-offs, the age-adjusted odds of mobility limitation rise roughly ten to fourteen-fold [1].

One clinical instrument for lower-limb power is a timed chair-stand test: a trained observer times five rises from a chair, and that time is converted to relative power in watts per kilogram of body mass [1, 7]. The test is cheap and well validated, but it needs a visit, a stopwatch, and maximal effort on command, so a measure that predicts independence is taken at best once or twice a year, and for most older adults never. That is the gap this work addresses.

The Casana Smart Seat is a sensor-embedded toilet seat that passively acquires physiological measurements during ordinary use, with no wearable hardware, no user-initiated measurement, and no change in routine behavior (casanacare.com (https://casanacare.com)). Sensors in the seat surface and base capture a single-lead electrocardiogram, a ballistocardiogram, photoplethysmography at the posterior thigh, and body weight from load cells at each corner of the seat. This paper concerns one algorithm inside that system, which reads only the load cells: every time a person sits and stands, the Smart Seat records the forces under them, unprompted.

This paper has three aims. It describes how the algorithm works, including where each of its signal measures comes from in the recorded force of a real stand. It evaluates whether the algorithm recovers the chair test's relative power from one Smart Seat recording plus basic demographics. And, because age and weight are inputs to the estimate, it tests whether the estimate carries information age cannot supply: within a single age decade, age says almost nothing about who rises quickly, so any within-decade ranking must come from the Smart Seat's signal.

2. Background and related work

2.1 Current measurement methods for this physiology

Laboratory instrumentation is the criterion method. A force plate measures ground reaction force directly during a rise, giving power without assumptions about body geometry; it is the instrument against which the chair-test power equation was validated, reaching a correlation of 0.95 and a mean difference of 6.4 W in 50 older adults [8]. The equipment is fixed and the measurement requires a dedicated visit, so nothing about it scales to routine monitoring.

The supervised chair-stand test is the clinical compromise, and it is the reference this paper compares against. In the Short Physical Performance Battery (SPPB), an observer times five rises from a standard chair with arms folded [9]. Alcazar et al. showed that this time, with the subject's stature and the chair height, yields a relative muscle power that carries the clinical signal, and established normative percentiles and functionally relevant cut-off points in a European cohort of 9,320 older adults [1, 7]. The test carries real measurement error (standard error of measurement 0.9 s in older women, minimal detectable change 2.5 s, about 17.5% of the mean [10]), and it is frequently not done: in a primary-care implementation of a national fall-prevention protocol, screening reached 79% of eligible patients in the first year, fell to 49% in the second, and the timed physical test within it was reported as time-consuming and often skipped [11].

Wearable accelerometry is the closest passive alternative. A thigh-mounted inertial sensor reproduced total chair-stand time almost perfectly (r = 0.99), yet the power it computed by integrating its own acceleration signal into velocity and scaling by body mass ran roughly 58 W below a laboratory reference [12]: timing fidelity did not carry into its power estimate. Wearables also only measure while worn and charged: in a national study of 2,208 adults aged 65 and over, only 44% produced seven valid wear days, and the odds of a valid day fell with age [13].

The pivot this paper rests on is that a toilet seat is used by essentially everyone, several times a day, without being remembered, charged, or worn. Coverage is a property of the form factor rather than of the user's motivation.

2.2 Related algorithmic approaches

The relative sit-to-stand power equation used here is the one introduced and validated by Alcazar et al., in which mean concentric power normalized to body mass is a function of stature, chair height, and the time to complete the rising phase of a chair stand [1, 7]. Because power is normalized to body mass, the relative measure depends only on body height, the assumed chair height, and the rise time.

What separates the present work from instrumented chairs and wearable approaches is the quantity the Smart Seat can observe. A seat that senses force under its own surface loses the subject at seat-off, part-way through the rising phase the power equation is defined over, so the design question is not how to measure the rise more precisely but how to recover the clinically defined rise time from the fragment that precedes seat-off. The algorithm answers it with three measures of the seat's unloading, each an established construct in the force-plate sit-to-stand literature: the duration of the unloading phase, the seat-to-foot weight-transfer interval that structures the whole rise [14, 15]; the rate of force development over a fixed brief window, the canonical measure of explosive muscle capability [16, 17]; and the peak unloading rate, the strongest instantaneous force transfer of the push-off [15].

3. Data

3.1 Data collection

Both cohorts were collected prospectively, under IRB-approved protocols designed before collection began, for the purpose of relating Smart Seat signals to validated physical performance measures.

The **clinical stability cohort** came from Casana's stability study (protocol C034), reviewed and approved by the WCG Institutional Review Board in January 2026 (IRB tracking number 20255150); every subject gave written informed consent. In a supervised session of roughly 90 minutes, research staff measured each subject's height, weight, and demographics, then collected 6 to 8 dedicated sit-and-stand recordings on the Smart Seat, interleaved with the components of the SPPB, which was administered separately on its own chair. The chair-stand assessment was observed and stopwatch-timed by staff; it was not itself recorded by the Smart Seat. The study enrolled adults spanning a wide range of age, body size, and function, including a small number of younger internal reference subjects who completed the same protocol as everyone else.

The **retirement community cohort** was collected in May 2026 at a retirement community whose research program ran its subjects under its own institutional review board approval (protocol G003, IRB tracking number 202400472), matching the clinical stability study's protocol: the same measured demographics, the same supervised SPPB, and the same dedicated Smart Seat sit-and-stand recordings. Thirteen of these subjects had also used Smart Seats installed in their own homes during an earlier Casana at-home collection run in two IRB-approved phases (protocol C014, tracking numbers IRB202100290 and IRB202100869; August to December 2021), with height and weight clinically measured at that study's enrollment; their 2021 recordings form the longitudinal follow-up corpus described in Section 3.3.

Table 1. Collections that produced the analyzed data.

Cohort	Protocol	Setting and collection	Data collected
Clinical stability cohort	WCG IRB-approved, protocol C034, IRB tracking number 20255150	Supervised sessions at Casana research facilities and community sites, from January 2026	6 to 8 dedicated Smart Seat sit-and-stand recordings per subject (load cells sampled on the device at 128 Hz); supervised SPPB with stopwatch-timed chair stands on a separate chair; measured height, weight, demographics; fall-risk questionnaire
Retirement community cohort	IRB-approved under the community's own review board, protocol G003, IRB tracking number 202400472	Supervised sessions at a retirement community, May 2026	The same inventory: dedicated Smart Seat recordings, supervised SPPB, measured demographics
Returning-subject at-home corpus	IRB-approved in two phases, protocol C014, tracking numbers IRB202100290 and IRB202100869	Subjects' own homes, August to December 2021	Smart Seat recordings of ordinary daily use; height and weight clinically measured at enrollment; age recorded at the time

3.2 Training and development data

The algorithm's chair-rise model was calibrated on the same population it was evaluated on here, a constraint of the domain: the reference standard is a supervised chair test, and the cohorts that have both a supervised chair test and Smart Seat recordings are the ones described below. Every accuracy figure in this paper was therefore computed leave-one-subject-out: for each subject the model was refit on all other subjects and then applied to that subject's recordings, so no subject informed its own estimate.

The model's three signal measures are constructs established in force-plate studies of the sit-to-stand (Section 4.2); this cohort's data informed which of those constructs entered the final model. The population the model has seen also bounds where it should be expected to generalize: predominantly older adults with generally preserved function.

3.3 Evaluation data

The evaluation population was 124 adults who completed both a supervised chair-stand assessment and at least one Smart Seat recording. One record is one seat recording, that is, one sit and stand captured by the Smart Seat, labeled with the subject's demographics and, at the subject level, the measured chair-stand result from their supervised assessment.

The clinical stability cohort contributed 79 subjects and 577 recordings, spanning ages 24 to 89; the retirement community cohort contributed 45 subjects and 319 recordings, spanning ages 63 to 88. Pooling them gave the validation a wider functional range than either alone.

Table 2. Composition of the evaluation sample.

Characteristic	Value
Subjects	124
Seat recordings analyzed	896
Recordings yielding a usable stand	886 (98.9%)
Subjects with a measured chair-stand reference	122 of 124
Recordings contributing to accuracy figures	873, over 122 subjects
Recordings per subject contributing to accuracy figures	median 7 [IQR 6 to 8], range 3 to 14
Cohorts	clinical stability 577; retirement community 319

Table 3. Baseline characteristics of the study population (n = 124).

Characteristic	Distribution
Age (years)	median 72 [IQR 65 to 79], range 24 to 89
Sex	70 female, 54 male
Height (cm)	mean 167.9 [SD 9.1], range 144.8 to 188.0
Weight (kg)	mean 76.5 [SD 19.1], range 44.0 to 156.7
Body mass index (kg/m^2)	mean 26.9 [SD 5.2], range 18.1 to 44.3
Chair-test relative power (W/kg)	median 2.84, range 0.54 to 8.25 (n = 122)

The gap between the 896 recordings and the 873 analyzed is three recordings with no stored source, seven in which no stand was detected, and the sits of the two subjects without a chair-stand reference; every record passed schema validation before analysis, and the full validation reports are packaged with the run. One subject's reference was recorded as the study's censored code for over one minute or unable; it was retained, holding the slowest rank in the rank-based analyses either way.

The **longitudinal follow-up corpus** supplements this sample: 1,549 at-home recordings made in 2021 by returning retirement community subjects. Of these, 288 belonged to four additional returning users with no recorded 2021 demographics and were not scored; of the remaining 1,261, 1,122 yielded a usable stand and 958 passed the clean-stand gate defined in Section 4.4, over the 13 subjects with recorded 2021 demographics. Both eras were reprocessed end to end by the same algorithm version, each with its own era's recorded age and clinically measured height and weight.

Body size spanned a wide range, which matters because the power equation's height term and the model's weight input both operate over it. Figure 1 shows the distribution of the three quantities the claim depends on most: age, body size, and the chair-test power the estimate is compared against.

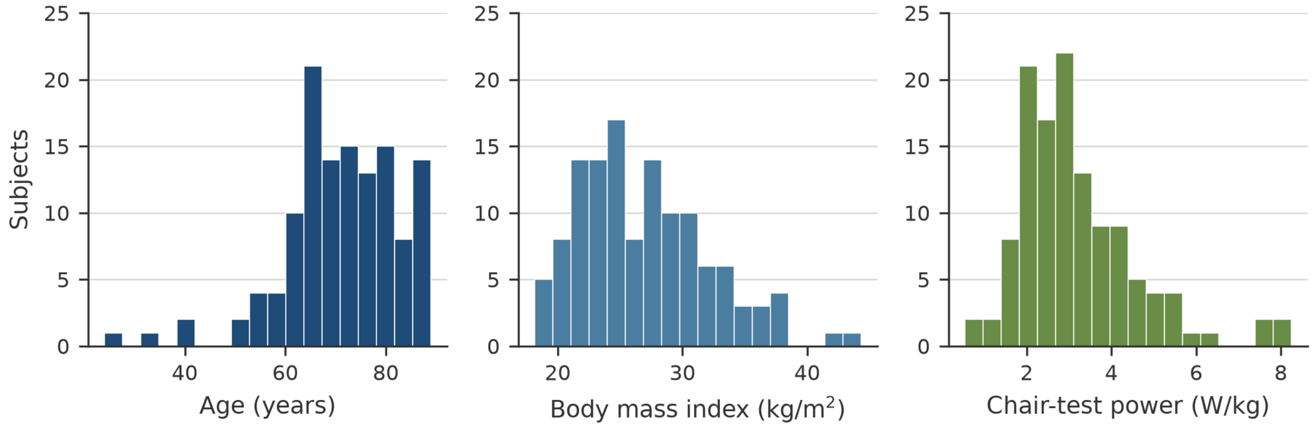

Figure 1. Composition of the study population: age and body size, which the power equation and the rise-time model both depend on, and the chair-test relative power the estimate is compared against. The cohort is predominantly older adults but spans a wide functional and body-size range.

4. Methods

4.1 Hardware and preprocessing

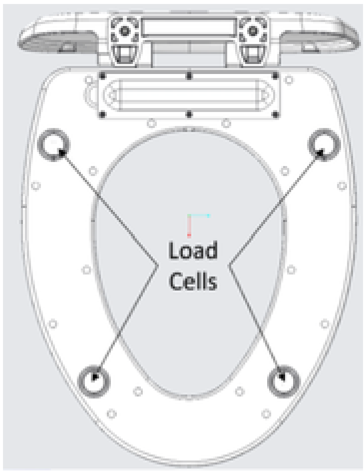

Figure 2. The Casana Smart Seat and its four corner load cells. Each load cell measures the weight carried at its corner, sampled on the device at 128 Hz; the four together give the total load on the seat. These load cells are the only sensors the algorithm reads.

The Smart Seat carries four load cells, one at each corner of the seat (Figure 2), each sampled on the device at 128 Hz and reported in pounds; during preprocessing every channel is resampled onto a common 256 Hz analysis grid. Total load is the sum across the four corners. This algorithm reads only these load cells; it does not use the electrocardiogram, the ballistocardiogram, or the photoplethysmogram.

Preprocessing establishes what the subject weighs while seated (the median load over the settled middle portion of the recording) and then locates the stand relative to the end of the recording. The recording must end with the seat unloaded, which is what confirms the subject actually stood and left; a recording that does not is rejected without an estimate. From that end point the algorithm works backward through a fixed trailing window, finding first the moment the load leaves the seat and then the point at which the load first departs from the seated baseline, so the detected event is the subject's final departure from the seat rather than an arbitrary movement earlier in the sit.

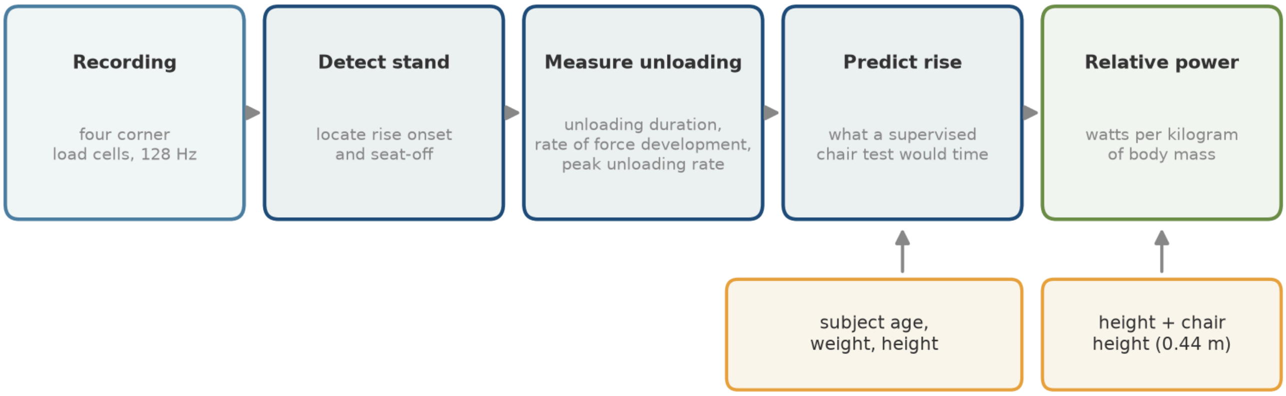

Figure 3. How one Smart Seat recording becomes a relative sit-to-stand power estimate. The three unloading measures and the subject's age, weight, and height feed the rise-time prediction; the power equation then converts the predicted rise time using only the subject's height and the chair height, with body weight entering only if absolute watts are requested.

4.2 Algorithm

The algorithm takes one seat recording plus the subject's age, weight, and height, and returns an estimate of relative sit-to-stand muscle power in watts per kilogram. It proceeds in three stages, shown in Figure 3: it detects the stand in the Smart Seat's load signal; it predicts, from three measures of how the seat unloads plus the demographics, the rise time a supervised chair-stand test would have measured; and it puts that predicted rise time through the Alcazar relative power equation [1, 7], which converts it using the subject's stature and the chair height. The interval the Smart Seat can measure directly, from the onset of rising force to seat-off, is far shorter than the rise time the power equation was defined for; fed directly into the equation it is both mis-scaled and poorly behaved on exactly the slow risers the measure should flag, and Section 5.1 quantifies what the prediction stage buys over it.

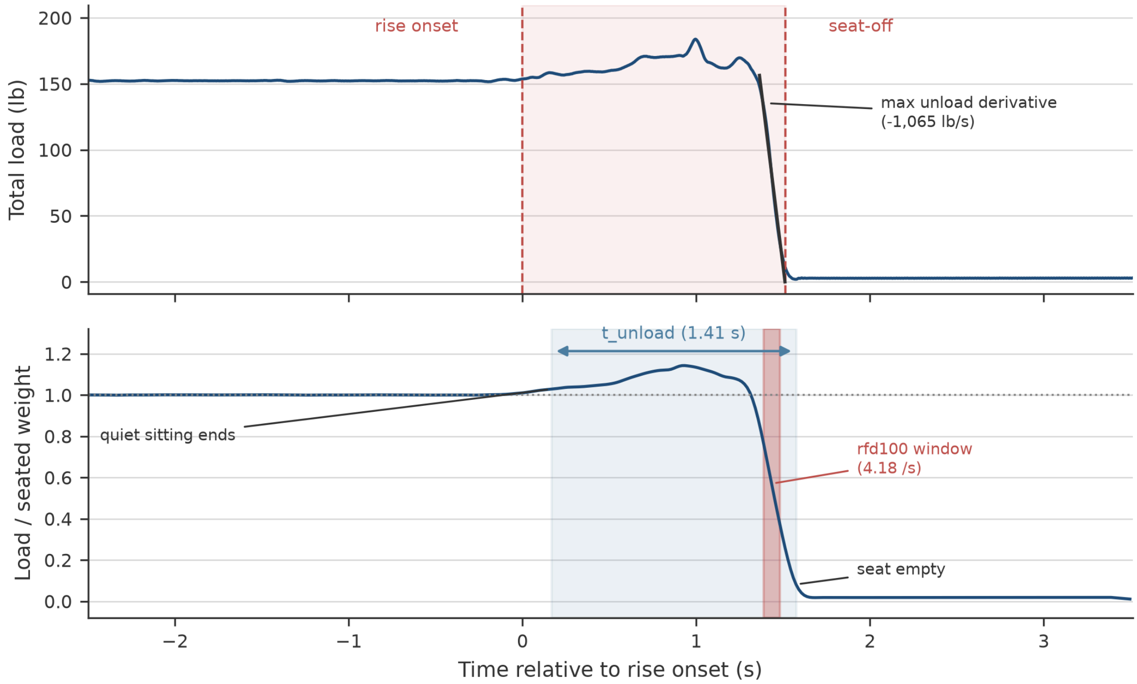

Figure 4. Where each of the algorithm's three signal measures comes from, on one real stand-up from the retirement community cohort. Top: total load across the Smart Seat's four corner load cells. Load is steady while the subject sits, rises briefly as they lean forward and load their feet, then falls sharply as they push off, reaching zero at seat-off; the solid tangent at the steepest fall is the maximum unload derivative, and the shaded interval from rise onset to seat-off is all the Smart Seat can observe, because the rest of the rise happens in the air. Bottom: the same stand on the algorithm's normalized unloading curve, the last 8 seconds of load divided by the subject's quiet-sitting weight on a fixed 64 Hz grid. The unloading duration t_unload spans from the end of quiet sitting to the empty seat, and rfd100 is the steepest fall of this curve across a fixed window of roughly 100 ms (the narrow band inside the descent).

Figure 4 shows where every signal measure the model consumes comes from, on one real stand. On the raw total load, the **maximum unload derivative** is the slope of the steepest sustained fall as the body leaves the seat, the strongest instantaneous force transfer of the push-off [15]. The other two are read from the **unloading curve**, the final stretch of the recording normalized by the subject's quiet-sitting weight, so that it runs from about 1.0 while seated to 0 when the seat is empty. The **unloading duration** (t_unload) spans from the end of quiet sitting to the empty seat, the seat-to-foot weight-transfer phase that force-plate studies identify as the interval structuring the rise [14, 15]. The **rate of force development** (rfd100) is the steepest fall of the normalized curve across a fixed window of roughly 100 ms, chosen to match the canonical early-phase windows over which explosive force production is defined and reported in the strength literature [16, 17]. Because the curve is normalized by seated weight, both curve measures are independent of body mass.

The stand is detected from the load signal by a three-cue onset (unloading rate, forward lean in the center of pressure, and countermovement loading, each biomechanically bounded), with seat-off refined by a tangent-line fit from the steepest unloading peak; the maximum unload derivative is that tangent's slope in pounds per second. The unloading curve is a fixed window of total load spanning seat-off, resampled to a common grid, lightly smoothed, and divided by the median load over the window's lead-in. On that curve, t_unload runs from the end of the last sustained quiet-sitting run to the first sample at or below a near-empty-seat fraction of seated weight, and rfd100 is the largest fall across a fixed window of roughly 100 ms. The three signal measures plus age, weight, and height are median-imputed, standardized, and fed to a Ridge regression that predicts the logarithm of the chair-test rise time. The algorithm withholds its output when the resulting relative power exceeds a plausibility ceiling of 8 W/kg, roughly twice the older-adult male 75th percentile of the published reference distribution (4.17 W/kg [1]) and above every value observed in this cohort's older-adult range, on the reasoning that such a value indicates a detection failure rather than an exceptional subject.

4.3 Configuration

Table 4. Algorithm configuration affecting interpretation.

Parameter	Value	Rationale
Assumed chair height	0.44 m	Standard toilet rim height plus the Smart Seat's own thickness. It sets the vertical travel the power equation integrates over, so the output is not interpretable without it
Reference standard	Chair-stand-derived relative power	The algorithm's calibrated target, measured under supervision

4.4 Analysis methods

The reference standard. The comparator was the relative sit-to-stand power computed from each subject's supervised SPPB chair-stand assessment: the observed result was converted to a per-repetition rise time under the published convention that the rise occupies half of each stand-and-sit cycle [7], and that rise time went through the same Alcazar power equation the algorithm's final stage uses. The assessment was timed over five repetitions where available (72 of the 122 subjects with a reference) and otherwise recorded as a 30-second repetition count (50 subjects); the two forms agreed at a correlation of 0.85 in the subjects who had both. The reference is reported in aggregate throughout, and the mixed instrument is discussed in Section 7. Because both sides share the power equation, what is actually under test is whether the Smart Seat recovers the rise time; a criterion validation against a force plate remains undone.

Metrics. Agreement was summarized at the subject level by **Spearman's rank correlation (rho)**, how well the two measures rank subjects in the same order (the Pearson correlation of the ranks), and by the **median absolute percentage error**, the median across subjects of |estimate minus reference| divided by the reference. Rank correlation is the headline statistic because the intended use is placing a person against a distribution. Confidence intervals came from 10,000 bootstrap resamples of subjects, and comparisons between two estimators on the same subjects from 5,000 paired bootstrap resamples of the difference in rho.

The production contract. The deployed algorithm reports one estimate per recording, so every primary result uses a single recording per subject (the first analyzed recording, denoted k = 1). Aggregating all of a subject's recordings by their median (k = all) is reported once, as a labeled secondary sensitivity.

Analyses, in the order Section 5 reports them:

1. Agreement between the single-recording estimate and the chair-test reference, with two anchors: the raw detected seat interval fed directly into the power equation (what the Smart Seat could report without the prediction stage), and a floor computed by predicting every subject as the population median of the others (the error of knowing nothing about the individual). How much of the true between-subject range the estimate recovers was also quantified.
2. Ranking within age bands (<60, 60 to 69, 70 to 79, 80 and over, following the published normative structure [1]): the estimate's correlation with the measured chair rise in each band, against an age-only

model fit by the identical leave-one-subject-out procedure, with paired bootstrap comparisons.

3. Placement on the published scale: band medians against the published older-adult reference distribution [1], and the estimate's overall median against the reference's.
4. Transfer between cohorts: the model calibrated on one cohort with its configuration fixed, then evaluated on the other cohort untouched, in both directions.
5. Change over time in the returning subjects: each subject's 2021 estimate (per-day median predicted rise over the clean at-home sits, converted to daily power, median across days) against their 2026 estimate (median rise over the clean clinic sits), each era using its own recorded demographics. A clean sit is a single stand attempt with no detector warnings and the two independent onset definitions agreeing within half a second. There is no 2021 chair-test reference, so the observed per-year change is read against the published age-related decline for relative sit-to-stand power [1].

4.5 Glossary

- **Sit-to-stand (STS) power**: the mechanical power generated while rising from a chair, the product of force and velocity.
- **Relative power**: power divided by body mass, in watts per kilogram, the scale the clinical cut-off points are defined on.
- **Seat-off**: the instant the body's weight leaves the seat, part-way through the rise.
- **Unloading curve**: the end of the recording's total load divided by the subject's quiet-sitting weight, running from about 1.0 (seated) to 0 (seat empty).
- **Rate of force development (RFD)**: how quickly force changes, a measure of explosive rather than maximal capability [17].
- **Leave-one-subject-out (LOSO)**: an evaluation in which each subject's estimate comes from a model refit without that subject's data.
- **Short Physical Performance Battery (SPPB)**: a standard clinical assessment of lower-extremity function combining balance, gait, and chair-stand components [9], scored here 0 to 12 under the standard EPESE scheme (from the Established Populations for Epidemiologic Studies of the Elderly).

5. Results

5.1 The single-recording estimate correlates moderately with the chair test, and its error is real skill over the floor

From a single recording per subject, estimated relative sit-to-stand power correlated with the chair-test-derived power at a Spearman correlation of 0.54 (95% CI 0.39 to 0.67, $p < 0.001$) with a median absolute error of 20.0% ($n = 122$). Aggregating all of a subject's recordings by their median, the secondary sensitivity, gave 0.57 (95% CI 0.42 to 0.69) and 18.6%, so nearly all of the achievable agreement is already present in one sit.

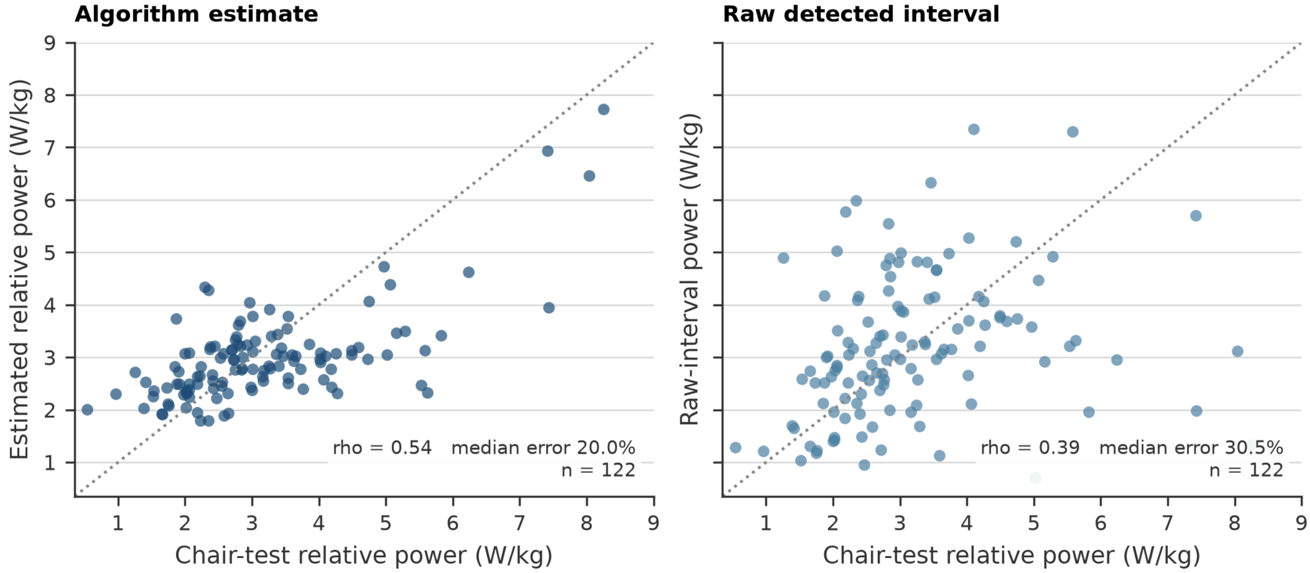

Figure 5. Single-recording estimates against chair-test-derived relative sit-to-stand power, one point per subject, every estimate held out from its own subject; the dotted line is identity. Left: the algorithm's estimate. Right: the raw detected seat interval fed directly into the power equation, the alternative available without the rise-time model.

Two anchors bound the result (Figure 5). Feeding the raw detected seat interval straight into the power equation ranked subjects at 0.39 with a 30.5% median error, so the prediction stage adds a third more rank information and removes a third of the error. At the other end, predicting every subject as the population median of the others gave a 25.0% median error; the estimate's 20.0% is therefore five points of genuine skill over knowing nothing, a bound worth stating because a median error alone reads as more absolute accuracy than it is.

The estimate is compressed toward the center of the cohort: it spanned 1.79 to 7.73 W/kg where the reference spanned 0.54 to 8.25, and its predicted rise times spanned 0.45 to 1.95 seconds against a measured 0.42 to 6.10. The top of that measured range is the single censored-code subject; the next slowest measured rise was 3.78 seconds. On the log scale the estimate recovers about a third of the true between-subject spread (slope 0.34), and in the slowest tenth of the cohort by measured rise (13 subjects, measured rises 2.02 to 6.10 s) it over-read power by a median factor of 1.50. In plain terms: the estimate reliably places a person within the population, but it pulls everyone toward the middle.

5.2 Within age decades, age collapses and the estimate keeps ranking

Pooled across all ages, an age-only model fit by the identical procedure ranked subjects against the measured chair rise at 0.39 and the algorithm at 0.54, a paired difference of +0.15 (95% CI +0.03 to +0.28). Restricting to single age bands is the sharper test, because it removes the between-decade ordering that age supplies for free.

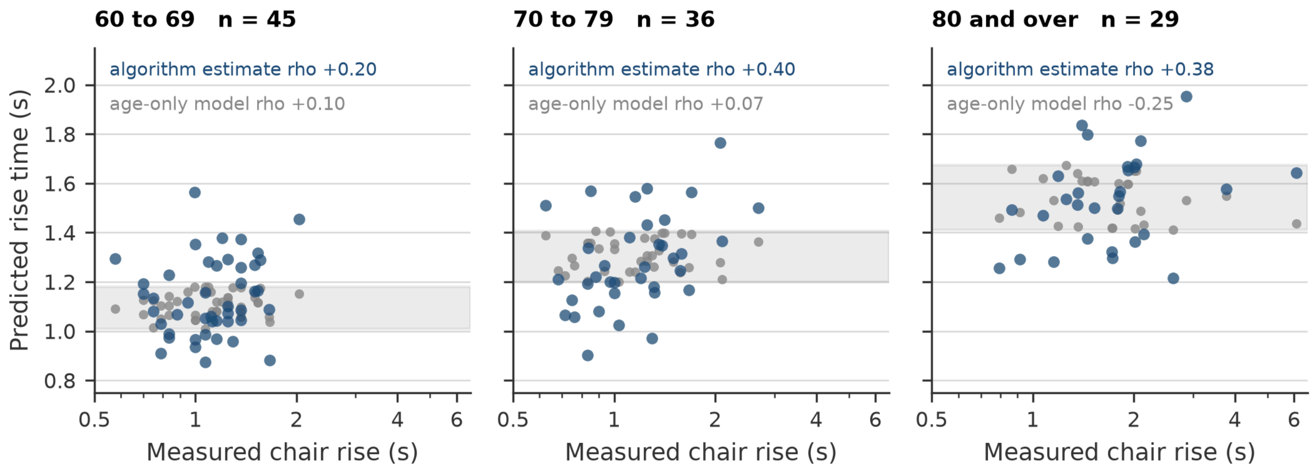

Figure 6. Held-out predicted rise time against the measured chair-test rise within each older-adult decade, one point per subject: the algorithm's estimate in blue, an age-only model fit by the same leave-one-subject-out procedure in gray, with each series' rank correlation with the measured rise shown per panel. Within a decade the age-only model predicts nearly the same rise for everyone (the flat gray stripe), so it cannot separate stronger from weaker subjects of the same age, and at 80 and over its ordering turns negative (-0.25); the algorithm's estimate spreads subjects along the measured rise in every decade. The measured axis is logarithmic; the 6.1 s value at 80 and over is the censored reference described in Section 3.3. The below-60 band (12 subjects) is reported in the text.

Figure 6 shows the three older-adult decades directly, one point per subject. Within each decade, the age-only model assigns nearly the same predicted rise to everyone, a flat stripe that carries no information about which of two same-age subjects is stronger, while the estimate spreads subjects along the measured chair rise and stays positively correlated with it in every band. In the 70 to 79 band ($n = 36$), the estimate held 0.40 against age's 0.07, a paired difference of +0.33 (95% CI -0.02 to +0.70) whose interval narrowly includes zero at one recording; the aggregated sensitivity cleared it decisively (0.58 against 0.07, paired difference +0.50, 95% CI +0.20 to +0.82). Below 60, where the band spans ages 24 to 59 in this cohort and age therefore retains real ordering information, both age (0.49) and the estimate (0.76) ranked subjects well; the within-decade comparison matters in the decades where age has nothing left to add.

The oldest band is where the advantage over age is largest. At 80 and over ($n = 29$), age's correlation with the measured chair rise turns negative (-0.25): the oldest subjects are not reliably the slowest, so an age lookup actively misleads there. The estimate's paired advantage over age in that band was the largest of any band (+0.62, 95% CI +0.26 to +0.96). The absolute within-band correlation is imprecise at that sample size (0.38 at a single recording against 0.11 aggregated), so the supported claim is relative, and it is the one that matters: where age says the least, the estimate's margin over it is greatest.

5.3 The estimate lands on the published scale

Binned by age band, single-recording estimated power declined from a median of 4.16 W/kg below 60 to 3.15 at 60 to 69, 2.78 at 70 to 79, and 2.30 W/kg at 80 and over, with the independently measured chair-test medians tracking the same descent (4.49 to 2.06 W/kg) (Figure 7). The older-adult band medians lay within or at the edges of the published 60-and-over interquartile range, with the 80 and over band at its lower boundary, where the oldest band of a distribution pooled over ages 60 to 103 belongs [1]. Across the whole cohort the estimate's median was 2.84 W/kg against the reference's 2.84, so the scale is centered, not merely correlated. The published five-repetition cut-offs for low relative power (1.9 W/kg for women, 2.5 for men [18]) are marked in the figure for clinical context.

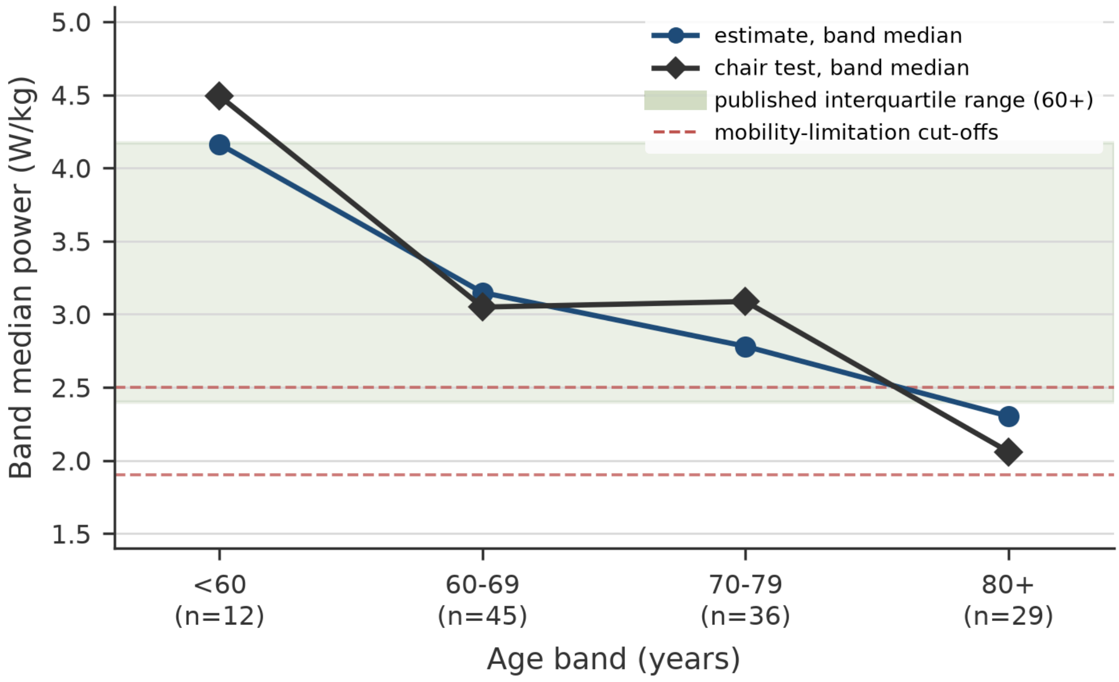

Figure 7. The estimate lands on the published scale: band medians of single-recording estimated power and of the measured chair test descend together across age bands, through the published interquartile range for adults 60 and over (women 2.40 to 3.39, men 2.99 to 4.17 W/kg) and toward the sex-specific mobility-limitation cut-offs derived from the five-repetition chair stand (1.9 women, 2.5 men). Per-subject spread is shown in the agreement figure; this figure carries the scale placement.

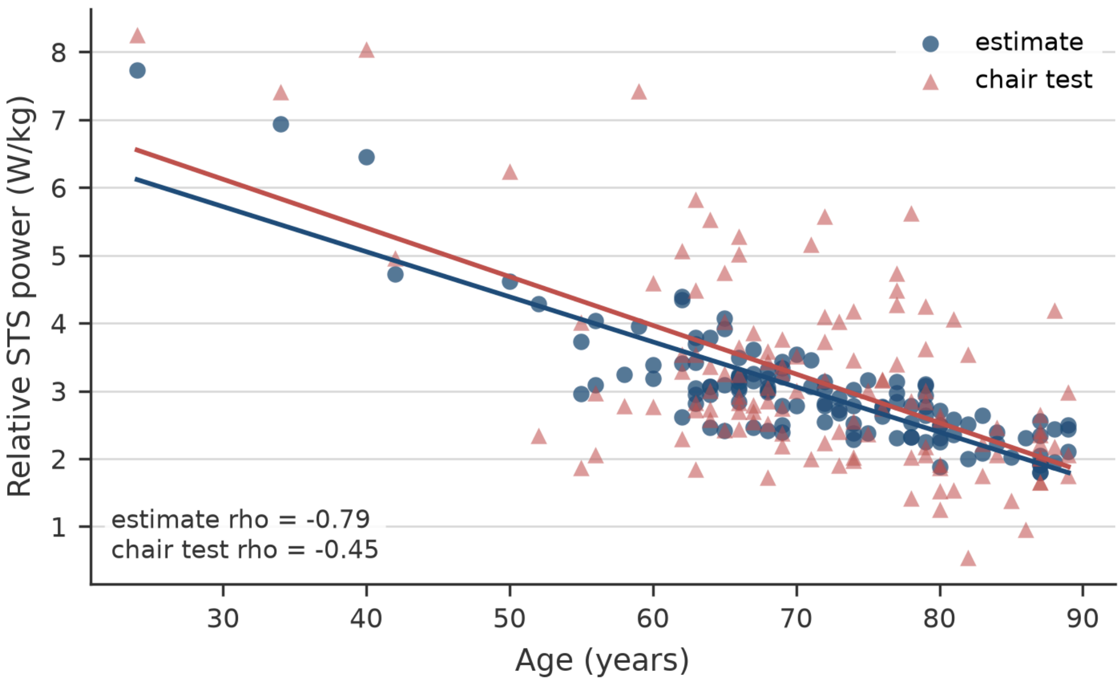

Figure 8. Relative sit-to-stand power against age, one point per subject at a single recording, for both the estimate and the measured chair test. Both fall with age. The estimate's gradient is steeper partly because age is one of its inputs; the chair test is the independent check that the cohort really does lose power with age.

Estimated power fell with age at a correlation of -0.79 (95% CI -0.85 to -0.70). Age is one of the model's inputs, so that gradient is partly by design; the independent check is the measured chair test, which fell with age at -0.45 (95% CI -0.60 to -0.28), confirming the cohort really does lose power with age (Figure 8).

5.4 The estimate transfers between cohorts

Calibrated on the clinical stability cohort with its configuration fixed and evaluated untouched on the retirement community cohort, the single-recording estimate ranked subjects at 0.48 (n = 44); in the reverse direction it ranked at 0.54 (n = 78). Both directions land near the pooled leave-one-subject-out result, so the estimate's ranking does not depend on any one collection's protocol or population.

5.5 Across 4.7 years, the estimate declines at the published rate

Estimated relative power declined in 11 of the 13 returning subjects between their 2021 at-home recordings and their 2026 clinic recordings (Figure 9), by a median of 0.32 W/kg over a median 4.7 years, a rate of 0.069 W/kg per year. The published age-related decline for relative sit-to-stand power is 0.10 to 0.13 W/kg per year for ages 50 to 80 and 0.07 to 0.08 above 80 [1]; at a median follow-up age of 79 (range 68 to 88, with 5 of 13 aged 80 or over), the cohort straddles those bands and the observed rate lands at their boundary.

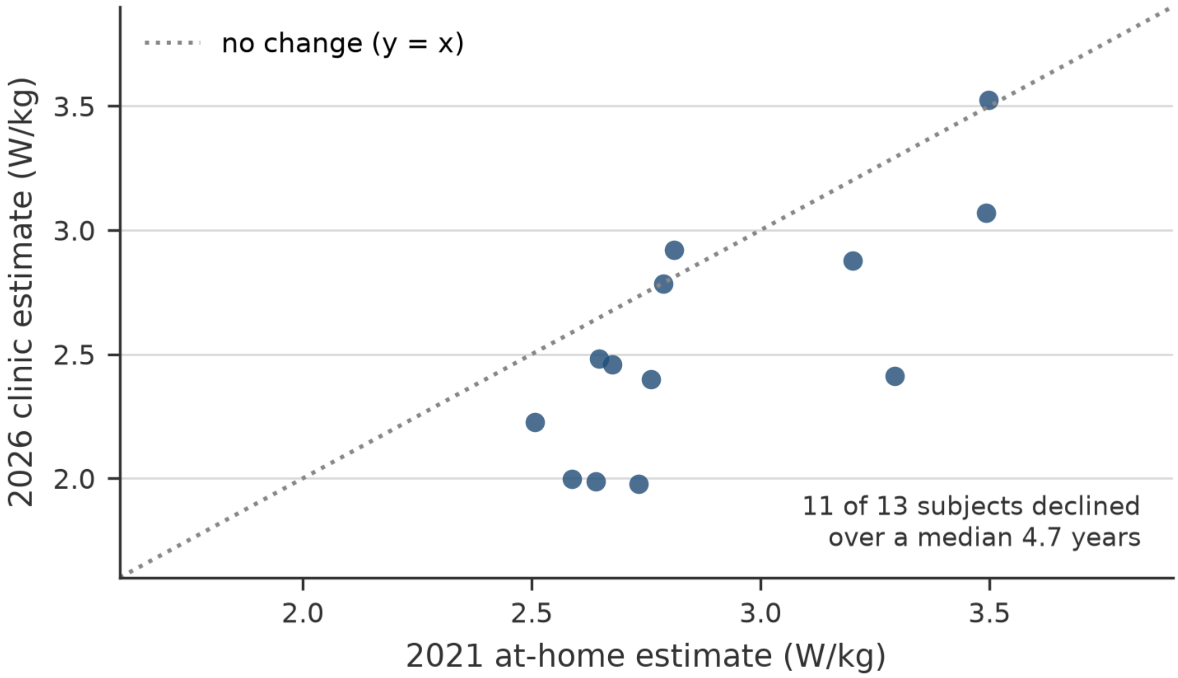

Figure 9. Estimated relative sit-to-stand power for the 13 returning retirement community subjects, from their 2021 at-home recordings against their 2026 clinic recordings, both reprocessed end to end by the same algorithm version with each era's recorded demographics. Points below the identity line declined. The two occasions differ in setting as well as in time, so the figure shows the direction and size of change rather than isolating age-related loss.

The comparison depends on the model's signal measures behaving consistently across eras and hardware revisions. The rate-of-force-development feature's distribution was near-identical between the 2021 and 2026 recordings (median 3.84 against 3.83 per second), despite the 2021 seats sampling weight at a lower native rate.

6. Discussion

From a single passive Smart Seat recording plus three demographic values, the algorithm produces a relative sit-to-stand power on the clinically validated scale that correlates moderately with a supervised chair test (rho 0.54, 20.0% median error), transfers between two independently collected cohorts, and keeps ranking people within their own age decade, where age itself collapses.

The demographic question has a quantitative answer. Age and weight are model inputs, so a fair question is whether the output is a demographic prediction restated. Within a decade the age input has almost nothing left to contribute: age ranked the 70 to 79 band's chair-rise times at 0.07, essentially a coin flip, while the estimate stayed positive in every band and its advantage over age cleared zero on the pooled chair rise. Age tells you which shelf someone is on; the Smart Seat's unloading measures are what keep ranking people on the shelf. The demographic inputs earn their place for the absolute scale, and the signal earns its place for the ranking.

Compression at the extremes is the main behavior to plan around. The estimate regresses toward the population mean because the Smart Seat only observes the fragment of the rise that precedes seat-off; a very slow riser's slowness lives mostly in the part of the movement that happens in the air, where the seat records nothing. This is why the estimate recovers about a third of the true log-scale spread, why it over-reads the slowest tenth of the cohort by roughly half again, and why it is stronger for placing a person against the published distribution and tracking central tendency than for pinpointing an individual at the tail. Any clinical-facing presentation should carry that slow-tail over-read explicitly, because at the tail the error is in the reassuring direction.

The returning cohort corroborates the trending story without proving it. Eleven of thirteen subjects declined across 4.7 years, at a median rate inside the published trajectory band for their ages. Three caveats keep this corroboration rather than proof: the sample is thirteen people; the setting changed from unsupervised home use to a supervised clinic; and the age input advanced 3 to 5 years per subject, which pushes the estimate downward on its own. What the result does establish is that the estimate's longitudinal behavior is consistent with known aging physiology rather than drifting arbitrarily.

The value concentrates where age says the least. Screening matters most in the oldest band, where functional loss accelerates, and that is exactly where chronological age stops carrying information: past 80, everyone is old, the spread in actual capacity is at its widest, and age's correlation with the measured chair rise turns negative. The estimate keeps ranking there because it is built from the movement itself, the force a person actually produces leaving the seat, rather than from a birth year, and its margin over age is the largest of any band. For a passive monitor aimed at the people most likely to be losing function, that is the property that matters.

Compared to the alternatives, the trade is accuracy for coverage. A force plate and a supervised chair test are more accurate and require a visit; a wearable is passive but only while worn, with 44% of older adults producing a full week of valid wear in a national sample [13]. The Smart Seat is less accurate than all three and is used several times a day by essentially everyone: in this evaluation 98.9% of recordings yielded a usable stand, so coverage is a property of the device rather than of adherence.

In practice this supports reading the estimate as a norm-referenced placement and trend: where a person's lower-limb power sits against their age peers, and which way it is moving across many effortless recordings. The specific capability the decade result adds is that two people of the same age with genuinely different chair-stand capacity will tend to be ordered correctly, which is exactly what an age lookup cannot do. A criterion validation against a force plate, which would test the power equation itself rather than the rise-time recovery, is the natural next study.

7. Limitations

- **The reference standard is itself an estimate, and shares an equation with the output.** The comparator is a supervised chair-stand assessment converted to a per-repetition rise time and then through the same power equation the algorithm uses, so what is under test is rise-time recovery, not absolute wattage. No claim about absolute accuracy in watts is supported until a force-plate validation is done.

- **The reference instrument was mixed.** For 50 of 122 subjects, all in the clinical stability cohort's first enrollment group, the chair-stand assessment was recorded only as a 30-second repetition count rather than a timed five-repetition trial; the count-derived rise is quantized where the timed form is continuous. The two forms agreed at 0.85 where both existed, and results are reported in aggregate, but the reference is not one uniform instrument.
- **Absolute within-band correlations are imprecise in the smallest bands.** With 29 subjects at 80 and over and 12 below 60, within-band correlations carry wide intervals and shift with the choice of recording (0.38 at one recording against 0.11 aggregated, at 80 and over). The robust within-band claim is relative: the estimate's advantage over age clears zero pooled and at 80 and over. A precise absolute correlation for the oldest band would need a larger sample.
- **Conservative at the extremes.** The estimate spans 1.79 to 7.73 W/kg where the reference spans 0.54 to 8.25, recovers about a third of the true log-scale spread, and over-reads the slowest tenth by a median factor of 1.50. It should not be the sole basis for an individual clinical decision.
- **The assumed chair height is a fixed constant, not a measurement.** The 0.44 m value is a standard toilet rim height plus the Smart Seat's own thickness rather than an arbitrary choice, but a 2 cm error in it moves every estimate by about 5.0% at this cohort's median stature, a scale offset that shifts placement against the reference distribution without reordering subjects.
- **Within-subject tracking is corroborated, not established.** The longitudinal comparison had no 2021 reference standard, its setting changed from home to clinic, its sample is 13 subjects, and age is a model input that advanced 3 to 5 years over the interval, pushing the estimate downward on its own. It shows the estimate declining at a physiologically plausible rate; validated change-tracking requires repeated supervised references on the same subjects.

8. Conclusion

A single Smart Seat recording, combined with a subject's age, weight, and height, recovers the relative sit-to-stand muscle power of a supervised chair-stand test to within a median error of 20.0%, ranks subjects moderately well against the reference (Spearman correlation 0.54, or 0.57 aggregating all of a subject's recordings; n = 122), lands centered on the clinically validated scale, and, unlike age itself, keeps ranking people within their own decade, where age collapses to a coin flip.

What a user gets is their lower-limb power placed against their age peers, produced from a device they already use several times a day, with no visit, no stopwatch, and no effort. Read within an age group it reflects measured chair-stand capacity rather than a demographic lookup, and read across recordings it points at the quantity that predicts losing independence: in the 13 subjects observed 4.7 years apart, the estimate declined at a rate consistent with published aging trajectories. Because a single reading is compressed toward the population average, it complements rather than replaces a supervised assessment, and it is not a sufficient basis on its own for an individual clinical decision. These findings support further evaluation as a passive screening and longitudinal monitoring tool for lower-limb power in older adults.

Declaration of AI assistance

This paper was prepared with substantial assistance from an AI coding agent (Claude, Anthropic), directed and reviewed by the named authors. What was and was not automated is set out here so a reader can weigh it.

What the AI did. It wrote the analysis code that produced the results reported here, generated the figures, retrieved and checked the literature cited, and drafted the prose of every section.

What it did not do. It did not generate any result. Every number in this paper comes from executing the evaluated algorithm and the packaged analysis code over the data recorded under the IRB-approved collection protocols, and the algorithm under evaluation is production software that predates this paper. No statistic, and no citation, was produced from a language model's recall.

Safeguards applied. Each quantitative claim in the prose is checked mechanically against the computed statistics, so a number cannot drift from its source during editing. Every reference was retrieved from the publisher or index record rather than recalled. The evaluation runs end to end from the raw recordings in a single reproducible command, and its accuracy statistics reproduce the algorithm's own stored calibration validation exactly.

Human accountability. The named authors defined the evaluation scope and the cohort, reviewed the full draft, corrected it, and are responsible for its claims. Domain review by the authors caught errors that no automated check in the pipeline could detect, and those corrections are reflected in this version.

References

1. Alcazar J, Alegre LM, Van Roie E, 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 to 932. DOI: 10.1002/jcsm.12737 (https://doi.org/10.1002/jcsm.12737).
2. Petermann-Rocha F, Balntzi V, Gray SR, et al. Global prevalence of sarcopenia and severe sarcopenia: a systematic review and meta-analysis. *Journal of Cachexia, Sarcopenia and Muscle*. 2022;13(1):86 to 99. DOI: 10.1002/jcsm.12783 (https://doi.org/10.1002/jcsm.12783).
3. Kakara R, Bergen G, Burns E, Stevens M. Nonfatal and fatal falls among adults aged 65 years and over: United States, 2020 to 2021. *MMWR Morbidity and Mortality Weekly Report*. 2023;72(35):938 to 943. DOI: 10.15585/mmwr.mm7235a1 (https://doi.org/10.15585/mmwr.mm7235a1).
4. Garnett MF, Weeks JD, Zehner AM. Unintentional fall deaths in adults age 65 and older: United States, 2023. *NCHS Data Brief* No. 532. June 2025. cdc.gov/nchs/products/databriefs/db532.htm (https://www.cdc.gov/nchs/products/databriefs/db532.htm).
5. Skelton DA, Greig CA, Davies JM, Young A. Strength, power and related functional ability of healthy people aged 65 to 89 years. *Age and Ageing*. 1994;23(5):371 to 377. DOI: 10.1093/ageing/23.5.371 (https://doi.org/10.1093/ageing/23.5.371).
6. Bean JF, Kiely DK, Herman S, et al. The relationship between leg power and physical performance in mobility-limited older people. *Journal of the American Geriatrics Society*. 2002;50(3):461 to 467. DOI: 10.1046/j.1532-5415.2002.50111.x (https://doi.org/10.1046/j.1532-5415.2002.50111.x).
7. Alcazar J, Losa-Reyna J, Rodriguez-Lopez C, et al. The sit-to-stand muscle power test: an easy, inexpensive and portable procedure to assess muscle power in older people. *Experimental Gerontology*. 2018;112:38 to 43. DOI: 10.1016/j.exger.2018.08.006 (https://doi.org/10.1016/j.exger.2018.08.006).
8. Baltasar-Fernandez I, Alcazar J, Rodriguez-Lopez C, et al. Sit-to-stand muscle power test: comparison between estimated and force plate-derived mechanical power and their association with physical function in older adults.

Experimental Gerontology. 2021;145:111213. DOI: 10.1016/j.exger.2020.111213 (https://doi.org/10.1016/j.exger.2020.111213).

9. Guralnik JM, Simonsick EM, Ferrucci L, et al. A short physical performance battery assessing lower extremity function: association with self-reported disability and prediction of mortality and nursing home admission. *Journal of Gerontology*. 1994;49(2):M85 to M94. DOI: 10.1093/geronj/49.2.m85 (https://doi.org/10.1093/geronj/49.2.m85).
10. Goldberg A, Chavis M, Watkins J, et al. The five-times-sit-to-stand test: validity, reliability and detectable change in older females. *Aging Clinical and Experimental Research*. 2012;24(4):339 to 344. DOI: 10.1007/BF03325265 (https://doi.org/10.1007/BF03325265).
11. Johnston YA, Reome-Nedlik C, Parker EM, et al. Preventing falls among older adults in primary care: a mixed methods process evaluation using the RE-AIM framework. *The Gerontologist*. 2023;63(3):511 to 522. DOI: 10.1093/geront/gnac111 (https://doi.org/10.1093/geront/gnac111).
12. Bochicchio G, Ferrari L, Bottari A, et al. Temporal, kinematic and kinetic variables derived from a wearable 3D inertial sensor to estimate muscle power during the 5 sit to stand test in older individuals: a validation study. *Sensors*. 2023;23(10):4802. DOI: 10.3390/s23104802 (https://doi.org/10.3390/s23104802).
13. Kocherginsky M, Huisingh-Scheetz M, Dale W, et al. Measuring physical activity with hip accelerometry among U.S. older adults: how many days are enough? *PLoS One*. 2017;12(1):e0170082. DOI: 10.1371/journal.pone.0170082 (https://doi.org/10.1371/journal.pone.0170082).
14. Hirschfeld H, Thorsteinsdottir M, Olsson E. Coordinated ground forces exerted by buttocks and feet are adequately programmed for weight transfer during sit-to-stand. *Journal of Neurophysiology*. 1999;82(6):3021 to 3029. DOI: 10.1152/jn.1999.82.6.3021 (https://doi.org/10.1152/jn.1999.82.6.3021).
15. Lindemann U, Claus H, Stuber M, et al. Measuring power during the sit-to-stand transfer. *European Journal of Applied Physiology*. 2003;89(5):466 to 470. DOI: 10.1007/s00421-003-0837-z (https://doi.org/10.1007/s00421-003-0837-z).
16. Aagaard P, Simonsen EB, Andersen JL, Magnusson P, Dyhre-Poulsen P. Increased rate of force development and neural drive of human skeletal muscle following resistance training. *Journal of Applied Physiology*. 2002;93(4):1318 to 1326. DOI: 10.1152/japplphysiol.00283.2002 (https://doi.org/10.1152/japplphysiol.00283.2002).
17. Maffiuletti NA, Aagaard P, Blazevich AJ, Folland J, Tillin N, Duchateau J. Rate of force development: physiological and methodological considerations. *European Journal of Applied Physiology*. 2016;116(6):1091 to 1116. DOI: 10.1007/s00421-016-3346-6 (https://doi.org/10.1007/s00421-016-3346-6).
18. Baltasar-Fernandez I, Alcazar J, Losa-Reyna J, et al. Comparison of available equations to estimate sit-to-stand muscle power and their association with gait speed and frailty in older people: practical applications for the 5-rep sit-to-stand test. *Experimental Gerontology*. 2021;156:111619. DOI: 10.1016/j.exger.2021.111619 (https://doi.org/10.1016/j.exger.2021.111619).