Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Sex differences in repeated handgrip across adulthood depend on measurement scale and protocol

Peng Li
Beijing Hospital
Yun Dong
Beijing Anzhen Hospital
Jiqiu Wang
Beijing Anzhen Hospital
Xi Yang
yangxi720@126.com
Beijing Hospital
Guiying Liu
Beijing Hospital

Article

Keywords: handgrip strength, sex differences, ageing, within-session variability, functional limitation, measurement protocol, coefficient of variation, NHANES

Posted Date: November 23rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11198914/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Sex differences in repeated handgrip across adulthood depend on measurement scale and protocol

Peng Li MD, PhD1#, Yun Dong MD, PhD2#, Jiqiu Wang MD, PhD3#, Xi Yang MD, PhD1, Guiying Liu MD, PhD2

1Beijing Institute of Geriatgrics, Beijing Hospital, National Center for Gerontology, Beijing 100730, China

2 Department of Pediatrics, Beijing Anzhen Hospital, Capital Medical University, Beijing , China

# These authors are co-first author

Correspondence: Guiying Liu, Email: liugvying@126.com; Xi Yang, Email: yangxi720@126.com

Short title: Sex-sensitive interpretation of repeated grip

Abstract

Sex differences in handgrip are usually characterized by strength maxima, while repeated-trial variability and protocol dependence remain poorly defined. We analysed six bilateral maximal-grip trials from NHANES 2011–2014 using complete complex-survey domain methods. Analyses included 9,040 adults aged 18–79 years; 3,854 participants aged 6–17 years formed a developmental extension, and 8,487 adults contributed physical-function outcomes. Female participants had lower mean grip but smaller absolute within-session variability. After normalization by mean strength, relative variability was higher in later adulthood, whereas trial-level location-scale models continued to show lower female absolute variability. Female participants had 1.55–2.06-fold higher adjusted prevalence of grip-relevant upper-limb difficulty across age groups without age heterogeneity, but only mean-grip deficit—not strength-adjusted variability—was independently associated with functional difficulty. Applying maximum-derived cut-points to mean summaries increased low-grip prevalence; prevalence-equivalent calibration removed aggregate shifts but not all person-level discordance. Sex comparisons therefore require explicit separation of force magnitude, variability, and testing protocol.

Keywords: handgrip strength; sex differences; ageing; within-session variability; functional limitation; measurement protocol; coefficient of variation; NHANES

Introduction

Ageing is accompanied by loss of skeletal-muscle mass, strength and power, although the decline in force-generating capacity is not fully explained by loss of muscle tissue alone.[1,2] Handgrip strength is widely used in clinical and epidemiological research because it is inexpensive, reproducible under standardised conditions and associated with disability and mortality.[3-7] Sex is central to interpretation: females generally have lower absolute grip strength, live longer, and experience a greater burden of frailty and functional limitation, whereas males have shorter average survival despite lower frailty levels.[8,9] Reporting sex as a biological variable and distinguishing it from gendered social exposures are therefore necessary for equitable ageing research.[10] Most population studies reduce repeated maximal-grip trials to one maximum value or to a summary of hand-specific maxima. In NHANES 2011–2012 and 2013–2014, each hand was tested three times, yet the commonly released and analysed summary retains only hand-specific maxima.[11] This convention discards information about within-session trial-to-trial variation and embeds trial order and extreme-value selection into the reported maximum. Earlier work showed that the coefficient of variation (CoV) from repeated grip trials is unstable at the individual level, and recent data indicate that percentage and absolute repeatability indices can produce different sex comparisons.[12,13]

The distinction between measurement scale and physiological mechanism is important. Laboratory studies have documented sex differences in motor-unit discharge, motoneuronal excitability and muscle morphology, but these effects depend on contraction intensity, recruitment threshold and muscle group.[14] Force steadiness is usually quantified from rapid fluctuations around a sustained submaximal target and reflects motor-unit and contractile processes.[15] A recent npj Aging study combined continuous torque signals, high-density surface electromyography, muscle morphology and free-living activity, reporting sex-dependent motor-unit behaviour and partial convergence at older ages.[16] Separated maximal-effort trials in NHANES measure a different construct: the consistency with which maximal performance is reproduced within one examination. Clinical interpretation adds a second layer. Low grip strength contributes to sarcopenia case definitions, but current consensus groups use different sex-specific thresholds and testing conventions.[17-19] A threshold developed for a maximum value cannot be transferred unchanged to a mean of repeated trials without changing prevalence. The effect may also differ by sex because male and female grip distributions differ in location and spread. Protocol calibration is therefore relevant both to disease surveillance and to fair comparison of ageing phenotypes.

We used all six bilateral grip trials from NHANES 2011–2014 to address three questions. First, we examined whether adult sex differences in mean grip, absolute within-session variability, raw CoV and trial-level residual scale followed the same age pattern, with ages 6–17 years analysed separately as a developmental extension. Second, we tested whether lower mean grip, mean-strength-adjusted variability and right-left asymmetry were independently associated with functional difficulty and whether these associations differed by sex. Third, we quantified protocol-dependent low-grip classification and used prevalence-equivalent cut-points to distinguish aggregate scale shifts from person-level discordance.

Results

Study population and survey-domain design

The complete positive-MEC-weight survey design contained 19,151 participants and 32 design degrees of freedom. The primary adult domain included 9,040 participants aged 18–79 years (4,549 female and 4,491 male),

and the developmental extension included 3,854 participants aged 6–17 years (Fig. 1a and Supplementary Table S1). Recalculation of the two hand-specific maxima exactly reproduced the released NHANES combined-grip variable, supporting correct reconstruction of hand and trial order. Physical-function analyses included 8,487 adults aged 20–79 years with the primary upper-limb outcome; 7,806 contributed to the fully adjusted functional model. Survey-weighted participant characteristics are shown in Table 1.

Adult sex differences depended on the variability metric

Across adulthood, female mean repeated grip was approximately 62–65% of male grip (Fig. 2a and Table 2). At age 25 years, model-standardised mean grip was 28.88 kg in females and 45.47 kg in males, corresponding to a female-to-male ratio of 0.635 (95% simultaneous confidence interval [CI], 0.621-0.649). At age 75 years, the corresponding values were 21.60 and 34.64 kg, with a ratio of 0.624 (0.587-0.662). The global sex-by-age interaction was supported for mean grip ($P=1.1 \times 10{-}5$).

Absolute within-session variability followed a different pattern. The female-to-male pooled within-hand SD ratio was 0.592 (95% simultaneous CI, 0.550-0.638) at age 25 years, 0.692 (0.602-0.795) at age 55 years and 0.747 (0.655-0.853) at age 75 years (Fig. 2b). Thus, females did not show greater absolute trial-to-trial deviations. By contrast, the raw CoV ratio rose with age: 0.928 (0.863-0.998) at age 25 years, 0.998 (0.905-1.101) at age 40 years, 1.132 (1.017-1.261) at age 65 years and 1.204 (1.060-1.368) at age 75 years (Fig. 2c). The adult sex-by-age interaction was supported for both absolute SD (P=0.001) and raw CoV (P=0.0002). These findings show a metric-dependent reversal: absolute variability remained lower in females, while division by the lower female mean produced higher relative variability in later adulthood.

Trial-level location-scale models confirmed lower female absolute variability

A trial-level location-scale model used all six observations per participant and a leave-one-trial-out variance proxy, reducing reliance on individual SD estimates calculated from only three trials per hand. Sex-by-age terms were supported in the location model ($P=5.3 \times 10{-}5$) and scale model (P=0.007), and the scale relation also differed by mean strength (sex-by-mean-strength P=0.0018). The female-to-male residual scale ratio was 0.751 (95% simultaneous CI, 0.674-0.836) at age 25 years, 0.817 (0.732-0.912) at age 55 years and 0.820 (0.719-0.935) at age 75 years (Fig. 2d and Table 2). The trial-level analysis therefore confirmed lower female absolute variability despite higher female raw CoV in later adulthood.

Mean-strength-matched estimates were constrained by limited overlap

When absolute variability was modelled at a common age-specific grip level, the female-to-male SD ratio increased from 0.764 (95% simultaneous CI, 0.688-0.849) at age 25 years to 0.929 (0.789-1.095) at age 75 years, indicating attenuation of the difference with age rather than emergence of universal female instability. However, weighted density overlap between female and male mean grip was only 0.19-0.25 across adult age bands (Fig. 3d and Supplementary Table S4). Common-support restrictions retained 13.6% of male participants and 18.5% of female participants. Trimming and overlap-weighted analyses preserved the overall direction but widened late-life CIs (Supplementary Table S5), so mean-strength-matched estimates were treated as secondary model-based comparisons rather than direct population contrasts.

Developmental extension and cycle-specific reproducibility

In the developmental extension, male and female mean grip were similar in childhood and separated rapidly during adolescence (Fig. 3a). Sex-by-age interactions were supported for mean grip ($P<1 \times 10^{-14}$) and absolute SD ($P=4.6 \times 10^{-6}$), but not for raw CoV (P=0.160) or mean-strength-matched SD (P=0.124; Fig. 3b and Supplementary Table S2). These findings were interpreted as developmental context rather than adult ageing. Adult estimates were directionally consistent across 2011–2012 and 2013–2014 survey samples (Fig. 3c). Cycle-by-sex-by-age heterogeneity tests were non-significant for mean grip (P=0.363), absolute SD (P=0.753), raw CoV (P=0.775) and mean-strength-matched SD (P=0.808).

Female participants had a persistently greater burden of functional difficulty

Grip-relevant upper-limb difficulty increased with age in both sexes and was consistently more prevalent among females (Fig. 4a). Survey-weighted prevalence was 7.8% in females and 3.8% in males at ages 20–44 years, 19.3% and 12.7% at ages 45–59 years, 37.4% and 21.4% at ages 60–69 years, and 47.4% and 24.9% at ages 70–79 years. The adjusted female-to-male prevalence ratios were 2.063 (95% CI, 1.424-2.989), 1.553 (1.205-2.003), 1.746 (1.346-2.265) and 1.914 (1.502-2.438), respectively (Fig. 4b and Table 3). There was no evidence that this disparity varied materially across age groups (sex-by-age-group P=0.680; false-discovery-rate [FDR] q=0.808). Additional functional outcomes showed the same general female excess (Supplementary Table S6).

Mean grip, but not residual variability, was independently associated with function

In the joint functional model, each adverse SD difference in mean grip was associated with a 49% higher prevalence of grip-relevant upper-limb difficulty after adjustment for age, sex, their interaction, survey cycle, race and ethnicity, education, income-to-poverty ratio, height and body mass index (prevalence ratio [PR], 1.486; 95% CI, 1.398-1.580; Fig. 4c and Table 3). The association was also evident for difficulty grasping small objects (PR, 1.720; 1.491-1.982), mobility difficulty, self-care difficulty and instrumental or social activity difficulty (Supplementary Table S7).

Mean-strength-adjusted variability was not independently associated with the primary upper-limb outcome (PR per adverse SD, 0.966; 95% CI, 0.900-1.037), and age- and sex-adjusted asymmetry was also not clearly associated (PR, 1.041; 0.990-1.095). Neither dimension showed consistent associations across the other functional outcomes. The functional association of mean-grip deficit was reproducible across cycles (PR, 1.567 in 2011–2012 and 1.458 in 2013–2014), whereas residual variability remained close to the null; corrected cycle-heterogeneity P values were 0.447, 0.910 and 0.414 for mean-grip deficit, residual variability and residual asymmetry, respectively.

Sex-specific grip-function interactions were exploratory

Three of 24 sex-by-grip-dimension tests had unadjusted P values below 0.05, including mean-grip deficit and asymmetry for difficulty grasping small objects, and residual variability for instrumental or social difficulty. None remained significant after Benjamini-Hochberg correction across all 24 tests; the minimum q value was 0.199 (Fig. 4d and Supplementary Table S8). Restriction to nine pre-specified key tests yielded a minimum q value of approximately 0.075. The data therefore supported a persistent female burden of functional difficulty but did not provide multiplicity-adjusted evidence that grip-function associations differed by sex.

Low-grip classification depended on summary protocol and calibration

Among adults aged 60–79 years, use of identical nominal cut-points across grip summaries produced substantial prevalence differences (Fig. 5a). Under EWGSOP2 thresholds, low-grip prevalence was 2.42% using the maximum of all six trials, 5.28% using the six-trial mean and 7.18% using the first bilateral-trial mean. Corresponding values were 1.93%, 4.24% and 6.53% under FNIH thresholds and 3.63%, 9.07% and 12.08% under AWGS 2019 thresholds.

Prevalence-equivalent calibration lowered the cut-points for mean-based summaries (Fig. 5b and Table 4). Under EWGSOP2, a maximum-of-six threshold of 27 kg in males and 16 kg in females corresponded to six-trial-mean cut-points of 24.71 kg and 13.88 kg, respectively. Calibration returned aggregate prevalence to approximately the reference prevalence and improved agreement for the six-trial mean from kappa 0.617 to 0.799 and for the first bilateral-trial mean from 0.486 to 0.754 (Fig. 5c). Agreement was not perfect, indicating remaining person-level ranking differences. Calibrated low-grip definitions remained associated with upper-limb difficulty (EWGSOP2 PRs, 1.82-1.92; Fig. 5d). Equivalent cut-points were regarded as exploratory translations between protocols, not new clinical thresholds.

Discussion

In this nationally representative study, apparent sex differences in repeated maximal handgrip depended on the metric used to express variability and on the protocol used to summarise repeated trials. Females had markedly lower mean grip and smaller absolute trial-to-trial deviations. Nevertheless, their raw CoV became higher in later adulthood because the female-to-male difference in mean strength was larger than the difference in absolute SD. Trial-level location-scale models confirmed lower female residual scale, and mean-strength-matched analyses suggested attenuation of the absolute difference with age. Female participants simultaneously had a greater burden of functional difficulty, but that burden was associated with lower mean grip rather than with residual variability. Finally, maximum-derived thresholds transferred unchanged to mean-based summaries increased low-grip prevalence; protocol-specific calibration removed the aggregate shift but not all person-level discordance.

The divergence between mean grip and absolute variability is consistent with the distinction between muscle force-generating capacity and the broader construct of dynapenia.[20,21] Mean strength reflects muscle mass, muscle quality, neural activation, anthropometry, health conditions and task execution. Absolute trial-to-trial variation also scales with force magnitude and instrument resolution. Consequently, smaller absolute variability in females does not imply superior neuromuscular ageing, just as lower mean strength does not by itself identify the causal level at which function is constrained. Our findings should not be interpreted as a population replication of laboratory force-steadiness or motor-unit studies. Sex differences in motor-unit firing and motoneuronal excitability vary across task intensity and muscle group.[22-24] Recent work has also identified sex-related differences in age-associated neuromuscular decline, adaptations among highly trained older adults and firing behaviour during development.[25-27] These studies involve sustained submaximal contractions, electrophysiological decomposition or carefully strength-matched experimental groups. Continuous force steadiness is influenced by common synaptic input, discharge variability and contractile filtering,[28-32] whereas NHANES recorded separated maximal efforts. The two paradigms can inform each other but should not be assigned the same physiological meaning.

The metric-dependent reversal has a straightforward statistical basis. Reliability and measurement-error frameworks distinguish absolute error from relative indices because they answer different questions.[33,34] CoV is

useful when variability is proportional to the mean, but it is not automatically free of scale effects, especially when group means differ substantially.[35,36] In the present data, females could have lower absolute SD and higher CoV at the same age without any contradiction. Statements that one sex is more variable are therefore incomplete unless the numerator, denominator, task and reference population are specified. The trial-level analysis strengthened this interpretation. Participant-specific SDs estimated from three trials per hand have low precision and can be coupled to the participant mean. The leave-one-trial-out scale model retained all observations, adjusted for trial order and hand, and modelled residual variation conditional on age and mean strength. Its estimates continued to show lower female absolute scale across adulthood. The increasing ratio with age indicates narrowing of the difference, but the simultaneous confidence bands did not support a claim that female absolute variability became greater.

The functional analyses place measurement findings in a clinically relevant context. Female participants had approximately 1.5- to 2.1-fold higher adjusted prevalence of grip-relevant upper-limb difficulty across adult age groups, consistent with the broader female morbidity-frailty disadvantage despite longer survival. This disparity was already present in early adulthood and did not show clear age heterogeneity. However, the public-use NHANES sex variable cannot distinguish biological, reproductive, behavioural and social pathways. The observed female excess should therefore be interpreted as a population-level disparity, not as evidence for one specific mechanism.

Mean-grip deficit had consistent associations with upper-limb, fine-motor, mobility, self-care and instrumental or social difficulty. Residual trial-to-trial variability did not add independent functional information, and no sex interaction survived multiplicity correction. These null results are informative: the repeated-trial metrics revealed how the examination behaves, but they did not establish an independent functional ageing biomarker. Laboratory studies have linked force steadiness and perceived fatigability to mobility,[37] yet those measures arise from different tasks. Similarly, wrist-accelerometry studies demonstrate age-related variation in free-living movement,[38,39] but our earlier accelerometry analysis did not validate residual maximal-grip variability. Repeated-trial variation should therefore remain a measurement characteristic unless longitudinal and mechanistic evidence demonstrates otherwise.

Protocol dependence has practical implications for sarcopenia research. Standard clinical grip protocols typically obtain one or more maximal efforts and retain the greatest value.[40-43] A mean of repeated trials is necessarily no greater than the maximum, so use of the same numeric cut-point increases low-grip prevalence by construction. Prevalence-equivalent calibration separated this aggregate scale translation from residual person-level discordance. The large improvement in kappa for the six-trial and first-trial means indicates that much of the nominal disagreement was due to a shift in measurement scale, but incomplete agreement showed that trial order and within-person ranking also mattered. Calibration must also be sex specific. Male and female grip distributions differ in location and spread, so the translation between a maximum-based and mean-based threshold was not identical across sexes. Applying one protocol's cut-point to another protocol can therefore change sex-specific case counts and potentially distort comparisons of sarcopenia prevalence. The equivalent values derived here are internal translations for analytic interpretation; they are not replacement clinical thresholds. External cohorts with standardised dynamometry, functional outcomes and longitudinal events are required before any protocol-specific cut-point is considered for practice.

Right-left asymmetry represented a further potential dimension. Population studies show that between-side grip differences vary across age and should not be reduced to an invariant threshold.[44] In our data, residual asymmetry was not independently associated with self-reported functional difficulty, despite an earlier association with wrist

activity. This discrepancy may reflect different constructs, hand dominance, unilateral disorders, self-report limitations or chance. It argues against presenting asymmetry as a confirmed clinical marker in this study.

The study has several strengths. It retained all six original measurements in a nationally representative sample; separated adulthood from developmental growth; implemented complete complex-survey domain inference; used simultaneous confidence bands rather than selecting a data-driven crossing age; added trial-level location-scale modelling; quantified common support for strength-matched comparisons; examined reproducibility across independent survey samples; linked measurement dimensions to multiple functional outcomes; controlled the false discovery rate for sex interactions; and distinguished nominal protocol transfer from prevalence-equivalent calibration. The findings are also consistent with the wider observation that habitual activity and lifelong exercise influence neuromuscular function,[38,39,45] while remaining cautious about mechanisms not measured in NHANES.

Several limitations require emphasis. NHANES is cross-sectional, so the age curves do not represent within-person ageing trajectories, and cohort effects or selective survival may contribute. The public-use variable records binary sex and lacks gender identity, reproductive stage, menopause, sex hormones and gendered exposures. Each hand contributed only three trials, the examination measured maximal peaks rather than continuous force, and age 80 years or older was top-coded. Functional outcomes were self-reported, and questionnaire skip patterns required explicit assumptions. Mean-strength-matched estimates were model dependent because male and female strength distributions had limited overlap. Protocol calibration was performed internally and did not assess longitudinal disability, falls, institutionalisation or mortality. Finally, some exploratory sex interactions had nominal P values below 0.05 but none survived FDR control.

In conclusion, sex differences in repeated handgrip were contingent on measurement scale and testing protocol. Female participants showed smaller absolute trial-to-trial variability, but their lower mean strength generated higher relative variability in later adulthood. Female participants also had a consistently greater burden of functional difficulty, yet this burden was associated with lower mean grip rather than strength-adjusted variability. Clinical low-grip prevalence changed when maximum-derived cut-points were transferred to mean-based summaries; prevalence calibration removed the aggregate shift but did not eliminate person-level discordance. Studies of sex differences in neuromuscular ageing should distinguish force magnitude from variability and interpret clinical thresholds within the protocol in which they were developed.

Methods

Study design and data source

We conducted a secondary cross-sectional analysis of the public-use 2011–2012 and 2013–2014 National Health and Nutrition Examination Survey (NHANES) cycles. NHANES uses a stratified, multistage probability design to represent the non-institutionalised civilian US population.[46,47] The study was reported in accordance with STROBE and SAGER recommendations.[48,10]

Study populations and analysis domains

The complete survey design retained all participants with positive MEC examination weights before defining analytic subpopulations. The primary adult domain comprised participants aged 18–79 years who completed both hands, had six positive non-missing grip values and had all six trials classified by the examiner as maximal effort.

Participants aged 6–17 years meeting the same grip criteria formed a developmental extension. Adults aged 20–79 years were eligible for physical-function analyses. Public-use age was top-coded at 80 years; these participants were excluded from continuous age models. Recorded sex was obtained from RIAGENDR and coded as male or female. Exact examination age in months was used below age 20 when available.

Handgrip examination and trial reconstruction

Grip strength was measured with the NHANES digital dynamometer protocol. Participants completed three trials with each hand in an alternating sequence, so approximately 60 seconds separated repeated trials of the same hand. The files record practice hand, starting hand, hand dominance, testing status and examiner assessment of effort. We reconstructed six trial-level observations (H1-T1, H2-T1, H1-T2, H2-T2, H1-T3 and H2-T3) and mapped hand order to physical right and left hands using the starting-hand variable. The recalculated sum of hand-specific maxima was compared with the released combined-grip variable as an integrity check.

Mean grip, variability and asymmetry

Mean repeated grip was the arithmetic mean of all six trials. For each hand, the SD of its three trials was calculated; pooled absolute variability was the square root of the degrees-of-freedom-weighted mean of the two hand-specific variances. Raw CoV was 100 x pooled SD divided by the six-trial mean. Right-left asymmetry was the absolute log ratio of the mean right-hand to mean left-hand value. The best-value summary was the average of the maximum trial from each hand, and best-value inflation was 100 x (best-value summary / six-trial mean - 1). These measures quantify within-session performance and do not represent continuous force steadiness or motor-unit firing variability.

Complex survey design and adult age models

Four-year MEC weights were calculated as WTMEC2YR/2. Strata and primary sampling units were nested within survey cycle. Survey objects were constructed using the full positive-weight sample, and analytic groups were specified as domains rather than by deleting observations before design construction.[49] Adult mean grip, log absolute SD and log raw CoV were modelled using sex, fixed natural-cubic-spline terms for age, sex-by-age interactions and survey cycle. Fixed knots avoided data-dependent re-selection when analyses were repeated in subdomains.[50] Adult global interactions were evaluated with design-based Wald tests.

Simultaneous confidence bands

Point estimates were generated for each integer age from 18 to 79 years. Ninety-five per cent simultaneous confidence bands were obtained by simulating 20,000 coefficient vectors from the design-based model covariance matrix, calculating the maximum absolute studentised deviation across the full age grid and using its 95th percentile as the curve-wide critical value.[51] The manuscript therefore emphasised pre-specified ages rather than a data-derived crossing age. The developmental extension used a separately specified spline over ages 6–17 years.

Trial-level location-scale analysis

To use all repeated observations and reduce dependence on an SD estimated from three values, we fitted trial-level location and scale models. The location model included sex-by-age terms, trial, hand, dominance and cycle. For trial t within a participant-hand, the reference mean was the average of the other two trials. Under independent equal-variance errors, (2/3) x (Y_t - mean[Y_-t])^2 estimates the within-hand variance. This positive variance proxy

was modelled using a log-link quasi-Gamma survey model including sex-by-age, sex-by-leave-one-out-mean, trial, hand, dominance and cycle. Each trial row received one-sixth of the participant MEC weight.

Mean-strength matching and common support

Participant-level log SD was modelled as a function of sex-by-age and sex-by-log-mean-grip spline terms. Both sexes were predicted at the geometric mean of their age-specific modelled grip. We quantified sex-specific weighted percentiles, P5–P95 common-support intervals, density-overlap coefficients and weighted effective sample sizes. Sensitivity analyses restricted observations to common support and used survey weight multiplied by overlap weight derived from a weighted sex-propensity model.[52] These analyses were descriptive standardisations and were not interpreted as causal sex effects.

Physical-function outcomes

Physical-function outcomes were obtained from PFQ. The primary outcome combined difficulty lifting or carrying 10 pounds, grasping or holding small objects and pushing or pulling large objects. Secondary outcomes captured extended upper-limb difficulty, fine-motor difficulty, mobility, self-care, instrumental or social activity difficulty, any PFQ061 difficulty and severe difficulty or inability. For participants aged 20-59 years who followed the official no-limitation skip pattern, structurally missing PFQ061 items were coded as no difficulty. “Do not do this activity” was missing in the primary analysis and coded as difficulty in sensitivity analysis.

Residualised grip dimensions and functional models

Three continuous grip dimensions were evaluated jointly: lower age-sex-expected mean grip, excess absolute variability after conditioning on age, sex and mean grip, and excess right-left asymmetry after conditioning on age and sex. Each was standardised by its survey-weighted SD, with higher values denoting a less favourable profile. Binary functional outcomes were analysed using survey-weighted Poisson regression with a log link and design-based standard errors to estimate PRs.[53] Base models adjusted for sex-by-age and cycle; fully adjusted models additionally included race and ethnicity, education, income-to-poverty ratio, height and body mass index.

Sex interactions, multiplicity and cycle-specific reproducibility

Sex-by-grip-dimension interactions were estimated in joint models. Benjamini-Hochberg FDR control was applied across all 24 interaction tests and separately to nine pre-specified key tests.[54] Female-to-male functional PRs were estimated within age groups, with a global sex-by-age-group test. Cycle-specific reproducibility analyses fitted the same models separately in 2011–2012 and 2013–2014 using the original two-year MEC weights. Pooled exposure-by-cycle interactions provided formal heterogeneity tests.

Low-grip thresholds and protocol calibration

Low-grip classifications used EWGSOP2 (<27 kg in males, <16 kg in females), FNIH (<26 kg and <16 kg) and AWGS 2019 (<28 kg and <18 kg) thresholds.[17-19] Thresholds were applied to the maximum of all six trials, dominant-hand maximum, mean of hand-specific maxima, six-trial mean and mean of the first bilateral trials. We reported survey-weighted prevalence, directional discordance, agreement and weighted kappa.[55] Because the thresholds were developed for maximum-based measurements, mean-based summaries were also assigned sex-specific prevalence-equivalent cut-points: within the same available-case sample, the alternative-summary cut-point

was the weighted quantile reproducing the reference protocol's low-grip prevalence. Calibration results were treated as protocol translations rather than new diagnostic thresholds.

Missing data and sensitivity analyses

The main grip analysis was complete-case by design. Functional models used available PFQ outcomes and covariates; no imputation was performed because substantial missingness arose from component eligibility, age restrictions and questionnaire routing. Sensitivity analyses included all complete grip trials regardless of examiner effort classification, exclusion of pain, prior hand or wrist surgery, seated testing or grip-angle limitation, alternative spline complexity, alternative variability metrics, common-support restriction, overlap weighting, alternative PFQ coding and multiple low-grip threshold systems.

Software

Analyses were performed in R version 4.6.1 using the survey, dplyr, tidyr, ggplot2 and related packages. Custom scripts generated the survey-domain models, simultaneous bands, trial-level location-scale estimates, functional models and protocol calibration.

Ethics

NHANES protocols were approved by the National Center for Health Statistics Research Ethics Review Board, and participants or their proxies provided informed consent. This analysis used publicly available, de-identified data.

Data availability

The public-use NHANES 2011–2012 and 2013–2014 data files used in this study are available from the US National Center for Health Statistics. The exact component files, variable map and derived analytic definitions will accompany the archived analysis code. No restricted individual-level data were used.

Acknowledgements

The authors received no specific funding for this work.

Competing interests

The authors declare no competing interests.

References

1. Larsson, L. et al. Sarcopenia: aging-related loss of muscle mass and function. Physiol. Rev. 99, 427-511 (2019).
2. Tieland, M., Trouwborst, I. & Clark, B. C. Skeletal muscle performance and ageing. J. Cachexia Sarcopenia Muscle 9, 3-19 (2018).
3. Roberts, H. C. et al. A review of the measurement of grip strength in clinical and epidemiological studies: towards a standardised approach. Age Ageing 40, 423-429 (2011).
4. Dodds, R. M. et al. Grip strength across the life course: normative data from twelve British studies. PLoS One 9, e113637 (2014).
5. Bohannon, R. W. Grip strength: an indispensable biomarker for older adults. Clin. Interv. Aging 14, 1681-1691 (2019).
6. Rantanen, T. et al. Midlife hand grip strength as a predictor of old age disability. JAMA 281, 558-560 (1999).
7. Leong, D. P. et al. Prognostic value of grip strength: findings from the Prospective Urban Rural Epidemiology study. Lancet 386, 266-273 (2015).
8. Gordon, E. H. et al. Sex differences in frailty: a systematic review and meta-analysis. Exp. Gerontol. 89, 30-40 (2017).
9. Austad, S. N. & Fischer, K. E. Sex differences in lifespan. Cell Metab. 23, 1022-1033 (2016).
10. Heidari, S., Babor, T. F., De Castro, P., Tort, S. & Curno, M. Sex and Gender Equity in Research: rationale for the SAGER guidelines and recommended use. Res. Integr. Peer Rev. 1, 2 (2016).
11. Wang, Y. C. et al. Summary of grip strength measurements obtained in the 2011-2012 and 2013-2014 National Health and Nutrition Examination Surveys. J. Hand Ther. 32, 489-496 (2019).
12. Fairfax, A. H., Balnave, R. & Adams, R. D. Variability of grip strength during isometric contraction. Ergonomics 38, 1819-1830 (1995).

13. Abe, T. et al. Sex differences in the reliability of maximal handgrip strength expressed using absolute and relative indices. J. Sports Sci. Med. 24, 707-711 (2025).

14. Lulic-Kuryllo, T. & Inglis, J. G. Sex differences in motor unit behaviour: a review. J. Electromyogr. Kinesiol. 66, 102689 (2022).

15. Enoka, R. M. & Farina, D. Force steadiness: from motor units to voluntary actions. Physiology 36, 114-130 (2021).

16. Carbonaro, M. et al. Motor unit behavior adaptations across lifespan: sex differences in young, middle-aged and old adults. npj Aging; 10.1038/s41514-026-00427-0 (2026).

17. Cruz-Jentoft, A. J. et al. Sarcopenia: revised European consensus on definition and diagnosis. Age Ageing 48, 16-31 (2019).

18. Studenski, S. A. et al. The FNIH sarcopenia project: rationale, study description, conference recommendations, and final estimates. J. Gerontol. A Biol. Sci. Med. Sci. 69, 547-558 (2014).

19. Chen, L. K. et al. Asian Working Group for Sarcopenia: 2019 consensus update on sarcopenia diagnosis and treatment. J. Am. Med. Dir. Assoc. 21, 300-307.e2 (2020).

20. Clark, B. C. & Manini, T. M. Sarcopenia is not dynapenia. J. Gerontol. A Biol. Sci. Med. Sci. 63, 829-834 (2008).

21. Mitchell, W. K. et al. Sarcopenia, dynapenia, and the impact of advancing age on human skeletal muscle size and strength: a quantitative review. Front. Physiol. 3, 260 (2012).

22. Inglis, J. G. & Gabriel, D. A. Sex differences in motor unit discharge rates at maximal and submaximal levels of force output. Appl. Physiol. Nutr. Metab. 45, 1197-1207 (2020).

23. Inglis, J. G. & Gabriel, D. A. Sex differences in the modulation of the motor unit discharge rate leads to reduced force steadiness. Appl. Physiol. Nutr. Metab. 46, 1065-1072 (2021).

24. Jenz, S. T. et al. Estimates of persistent inward currents in lower limb motoneurons are larger in females than in males. J. Neurophysiol. 129, 1322-1333 (2023).

25. Guo, Y. et al. Sex disparities of human neuromuscular decline in older humans. J. Physiol. 603, 151-165 (2025).

26. Piasecki, J. et al. Influence of sex on the age-related adaptations of neuromuscular function and motor unit properties in elite masters athletes. J. Physiol. 599, 193-205 (2021).

27. Yacyshyn, A. F. et al. Sex-related differences in motoneuron firing behavior during typical development. J. Neurophysiol. 133, 1307-1319 (2025).

28. Castronovo, A. M. et al. Decrease in force steadiness with aging is associated with increased power of the common but not independent input to motor neurons. J. Neurophysiol. 120, 1616-1624 (2018).

29. Galganski, M. E., Fuglevand, A. J. & Enoka, R. M. Reduced control of motor output in a human hand muscle of elderly subjects during submaximal contractions. J. Neurophysiol. 69, 2108-2115 (1993).

30. Watanabe, K. et al. Age-related changes in motor unit firing pattern of vastus lateralis muscle during low-moderate contraction. AGE 38, 48 (2016).

31. Orssatto, L. B. R., Blazevich, A. J. & Trajano, G. S. Ageing reduces persistent inward current contribution to motor neurone firing: potential mechanisms and the role of exercise. J. Physiol. 601, 3705-3716 (2023).

32. Hepple, R. T. & Rice, C. L. Innervation and neuromuscular control in ageing skeletal muscle. J. Physiol. 594, 1965-1978 (2016).

33. Atkinson, G. & Nevill, A. M. Statistical methods for assessing measurement error in variables relevant to sports medicine. Sports Med. 26, 217-238 (1998).

34. Hopkins, W. G. Measures of reliability in sports medicine and science. Sports Med. 30, 1-15 (2000).

35. Reed, G. F., Lynn, F. & Meade, B. D. Use of coefficient of variation in assessing variability of quantitative assays. Clin. Diagn. Lab. Immunol. 9, 1235-1239 (2002).

36. Curran-Everett, D. Explorations in statistics: the log transformation. Adv. Physiol. Educ. 42, 343-347 (2018).

37. Mani, D. et al. Motor unit activity, force steadiness, and perceived fatigability are correlated with mobility in older adults. J. Neurophysiol. 120, 1988-1997 (2018).

38. Wanigatunga, A. A. et al. Wrist-worn accelerometry, aging, and gait speed in the Baltimore Longitudinal Study of Aging. J. Aging Phys. Act. 31, 408-416 (2023).

39. Belcher, B. R. et al. US population-referenced percentiles for wrist-worn accelerometer-derived activity. Med. Sci. Sports Exerc. 53, 2455-2464 (2021).

40. Mathiowetz, V. et al. Grip and pinch strength: normative data for adults. Arch. Phys. Med. Rehabil. 66, 69-74 (1985).

41. Innes, E. Handgrip strength testing: a review of the literature. Aust. Occup. Ther. J. 46, 120-140 (1999).

42. Hamilton, A., Balnave, R. & Adams, R. Grip strength testing reliability. J. Hand Ther. 7, 163-170 (1994).

43. Coldham, F., Lewis, J. & Lee, H. The reliability of one versus three grip trials in symptomatic and asymptomatic subjects. J. Hand Ther. 19, 318-327 (2006).

44. Wang, Y. C. et al. Between-side differences in hand-grip strength across the age span: findings from 2011-2014 NHANES and 2011 NIH Toolbox studies. Laterality 24, 697-706 (2019).

45. Jones, E. J. et al. Lifelong exercise is associated with more homogeneous motor unit potential features across deep and superficial areas of vastus lateralis. GeroScience 43, 1555-1565 (2021).

46. National Center for Health Statistics. National Health and Nutrition Examination Survey: Plan and Operations, 1999-2010. Vital Health Stat. 1(56), 1-37 (2013).

47. Johnson, C. L. et al. National Health and Nutrition Examination Survey: Analytic Guidelines, 1999-2010. Vital Health Stat. 2(161), 1-24 (2013).

48. von Elm, E. et al. The Strengthening the Reporting of Observational Studies in Epidemiology statement: guidelines for reporting observational studies. Ann. Intern. Med. 147, 573-577 (2007).

49. Lumley, T. Complex Surveys: A Guide to Analysis Using R (Wiley, 2010).

50. Harrell, F. E. Jr. Regression Modeling Strategies 2nd edn (Springer, 2015).

51. Efron, B. & Tibshirani, R. J. An Introduction to the Bootstrap (Chapman & Hall/CRC, 1993).

52. Li, F., Morgan, K. L. & Zaslavsky, A. M. Balancing covariates via propensity score weighting. J. Am. Stat. Assoc. 113, 390-400 (2018).

53. Barros, A. J. D. & Hirakata, V. N. Alternatives for logistic regression in cross-sectional studies: an empirical comparison of models that directly estimate the prevalence ratio. BMC Med. Res. Methodol. 3, 21 (2003).

54. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. B 57, 289-300 (1995).

55. Cohen, J. A coefficient of agreement for nominal scales. Educ. Psychol. Meas. 20, 37-46 (1960).

Figure legends

Figure 1 | Study design and analytical framework. a, Complete positive-MEC survey design and the adult, developmental and functional analysis domains. b, Alternating bilateral grip protocol with three trials per hand. c, Pre-specified measurement dimensions: six-trial mean, pooled within-hand absolute SD, raw CoV and trial-level residual scale. d, Integrated metric, robustness, function and protocol-calibration framework. Sample sizes are unweighted. PFQ, Physical Functioning Questionnaire.

Figure 2 | Metric-dependent sex differences across adulthood. a, Model-standardised mean of all six maximal-grip trials by sex. b, Female-to-male ratio of pooled within-hand absolute SD. c, Female-to-male ratio of raw CoV. d, Female-to-male residual scale ratio from the trial-level location-scale model. Shaded areas in b-d are 95% simultaneous confidence bands across the full adult age grid; dashed horizontal lines indicate a ratio of 1. The adult primary domain comprised 9,040 participants aged 18-79 years.

Figure 3 | Developmental context, cycle reproducibility and common support. a, Sex-specific mean-grip curves in the developmental extension aged 6-17 years. b, Female-to-male raw-CoV and mean-strength-matched SD ratios in the developmental extension. c, Cycle-specific adult raw-CoV ratios at pre-specified ages; error bars are design-based 95% confidence intervals. d, Weighted density-overlap coefficients for female and male mean grip across adult age bands. The developmental extension was analysed separately from the adult ageing analysis.

Figure 4 | Sex differences in functional burden and functional relevance. a, Survey-weighted prevalence of grip-relevant upper-limb difficulty by age group and sex; error bars are 95% confidence intervals. b, Adjusted female-to-male PRs for selected functional outcomes. c, Fully adjusted PRs for grip-relevant upper-limb difficulty per adverse survey-weighted SD in mean-grip deficit, mean-strength-adjusted variability and age-sex-adjusted asymmetry. d, Unadjusted P values for the nine smallest sex-by-grip interactions; the dashed line marks P=0.05. None of the 24 interaction tests remained significant after FDR correction (all q>=0.199).

Figure 5 | Testing protocol changes low-grip classification and agreement. a, Survey-weighted low-grip prevalence among adults aged 60-79 years when the same nominal sex-specific threshold was applied to five grip summaries. b, Sex-specific prevalence-equivalent cut-points for three alternative summaries, using the maximum of all six trials as the reference. c, Weighted kappa before and after prevalence-equivalent calibration. d, Adjusted PRs for grip-relevant upper-limb difficulty under calibrated low-grip definitions. Error bars are 95% confidence intervals. Equivalent cut-points are internal protocol translations and are not proposed as clinical thresholds.

Table 1 | Survey-weighted characteristics of the adult primary domain.

Values are survey-weighted means or percentages. No between-sex hypothesis tests were performed in this descriptive table. Arm lean mass is available only in the DXA-eligible age range; the functional-difficulty percentage is based on the age 20–79 PFQ domain.

Characteristic	All (n=9,040)	Male (n=4,491)	Female (n=4,549)
Age, years	44.8	44.5	45.1
Height, cm	169.1	176.0	162.3
Body mass index, kg/m2	28.9	28.7	29.1
Bilateral arm lean mass, kg	6.28	7.9	4.51
Mean of all six grip trials, kg	35.28	43.35	27.44
Pooled within-hand absolute SD, kg	2.02	2.49	1.57
Raw coefficient of variation, %	5.8	5.77	5.83
Absolute right-left log ratio	0.094	0.089	0.098
Best-value inflation over six-trial mean, %	5.0	4.97	5.03
Mexican American, %	8.4	9.0	7.9
Other Hispanic, %	5.6	5.8	5.5
Non-Hispanic White, %	66.8	66.9	66.8
Non-Hispanic Black, %	11.4	10.5	12.2
Non-Hispanic Asian, %	5.0	4.9	5.2
Other/multiracial, %	2.7	2.9	2.5
Grip-relevant upper-limb difficulty, %	14.9	10.5	19.0

Table 2 | Adult metric-specific sex differences at pre-specified ages.

Ratios are female/male. Confidence intervals are simultaneous across the adult age grid. CoV, coefficient of variation; SD, standard deviation.

Age, years	Female mean grip, kg	Male mean grip, kg	Female/male mean-grip ratio (95% simultaneous CI)	Female/male absolute-SD ratio (95% simultaneous CI)	Female/male raw-CoV ratio (95% simultaneous CI)	Mean-strength-matched SD ratio (95% simultaneous CI)	Trial-level scale ratio (95% simultaneous CI)
18	27.54	42.31	0.651 (0.619-0.685)	0.611 (0.538-0.694)	0.938 (0.821-1.071)	0.749 (0.645-0.871)	0.699 (0.586-0.834)
25	28.88	45.47	0.635 (0.621-0.649)	0.592 (0.550-0.638)	0.928 (0.863-0.998)	0.764 (0.688-0.849)	0.751 (0.674-0.836)
40	29.35	46.1	0.637 (0.622-0.652)	0.635 (0.576-0.700)	0.998 (0.905-1.101)	0.829 (0.735-0.935)	0.803 (0.698-0.923)
55	27.05	42.53	0.636 (0.614-0.658)	0.692 (0.602-0.795)	1.086 (0.944-1.249)	0.859 (0.748-0.985)	0.817 (0.732-0.912)
65	24.65	39.29	0.627 (0.602-0.654)	0.712 (0.643-0.787)	1.132 (1.017-1.261)	0.876 (0.787-0.976)	0.812 (0.740-0.890)
75	21.6	34.64	0.624 (0.587-0.662)	0.747 (0.655-0.853)	1.204 (1.060-1.368)	0.929 (0.789-1.095)	0.820 (0.719-0.935)

Table 3 | Female functional burden and independent functional relevance of grip dimensions.

Panel A gives adjusted female/male prevalence ratios for grip-relevant upper-limb difficulty. Panel B gives fully adjusted prevalence ratios per adverse survey-weighted SD, with all three grip dimensions entered jointly.

Panel A. Female-to-male prevalence ratios

Age group, years	Female/male adjusted PR (95% CI)	P value
20-44	2.063 (1.424-2.989)	0.0003742
45-59	1.553 (1.205-2.003)	0.001279
60-69	1.746 (1.346-2.265)	0.0001242
70-79	1.914 (1.502-2.438)	5.269e-06

Panel B. Joint functional model

Grip dimension (per adverse 1-SD difference)	Adjusted PR (95% CI)	P value
Lower age-sex-expected mean grip, per SD	1.486 (1.398-1.580)	1.798e-14
Higher mean-strength-adjusted variability, per SD	0.966 (0.900-1.037)	0.3272
Higher age-sex-adjusted asymmetry, per SD	1.041 (0.990-1.095)	0.1142

Table 4 | Protocol calibration among adults aged 60-79 years.

The reference is the maximum of all six trials. Equivalent cut-points reproduce the reference prevalence within the same available-case sample. Values are internal protocol translations, not proposed clinical thresholds.

Threshold	Alternative summary	Nominal low-grip prevalence, %	Male equivalent cut, kg	Female equivalent cut, kg	Calibrated prevalence, %	Nominal kappa	Calibrated kappa	Calibrated low-grip functional PR (95% CI)
EWGSOP2	Mean of hand-specific maxima	3.55	25.85	15.02	2.4	0.806	0.803	1.818 (1.430-2.311)
EWGSOP2	Mean of all six trials	5.28	24.71	13.88	2.42	0.617	0.799	1.917 (1.527-2.407)
EWGSOP2	Mean of first bilateral trials	7.18	22.9	13.2	2.3	0.486	0.754	1.848 (1.428-2.393)
FNIH	Mean of hand-specific maxima	2.71	24.32	15.02	1.91	0.828	0.848	1.914 (1.507-2.432)
FNIH	Mean of all six trials	4.24	22.75	13.88	1.87	0.615	0.833	1.994 (1.571-2.530)
FNIH	Mean of first bilateral trials	6.53	21.56	13.2	1.81	0.44	0.801	1.966 (1.512-2.555)
AWGS2019	Mean of hand-specific maxima	6.08	26.25	16.86	3.61	0.736	0.86	1.956 (1.612-2.373)
AWGS2019	Mean of all six trials	9.07	25.29	16.01	3.61	0.548	0.765	1.755 (1.404-2.194)
AWGS2019	Mean of first bilateral trials	12.08	23.9	15.28	3.51	0.43	0.715	1.686 (1.342-2.119)

Figure 1 | Study design and analytical framework. a, Complete positive-MEC survey design and the adult, developmental and functional analysis domains. b, Alternating bilateral grip protocol with three trials per hand. c, Pre-specified measurement dimensions: six-trial mean, pooled within-hand absolute SD, raw CoV and trial-level residual scale. d, Integrated metric, robustness, function and protocol-calibration framework. Sample sizes are unweighted. PFQ, Physical Functioning Questionnaire.

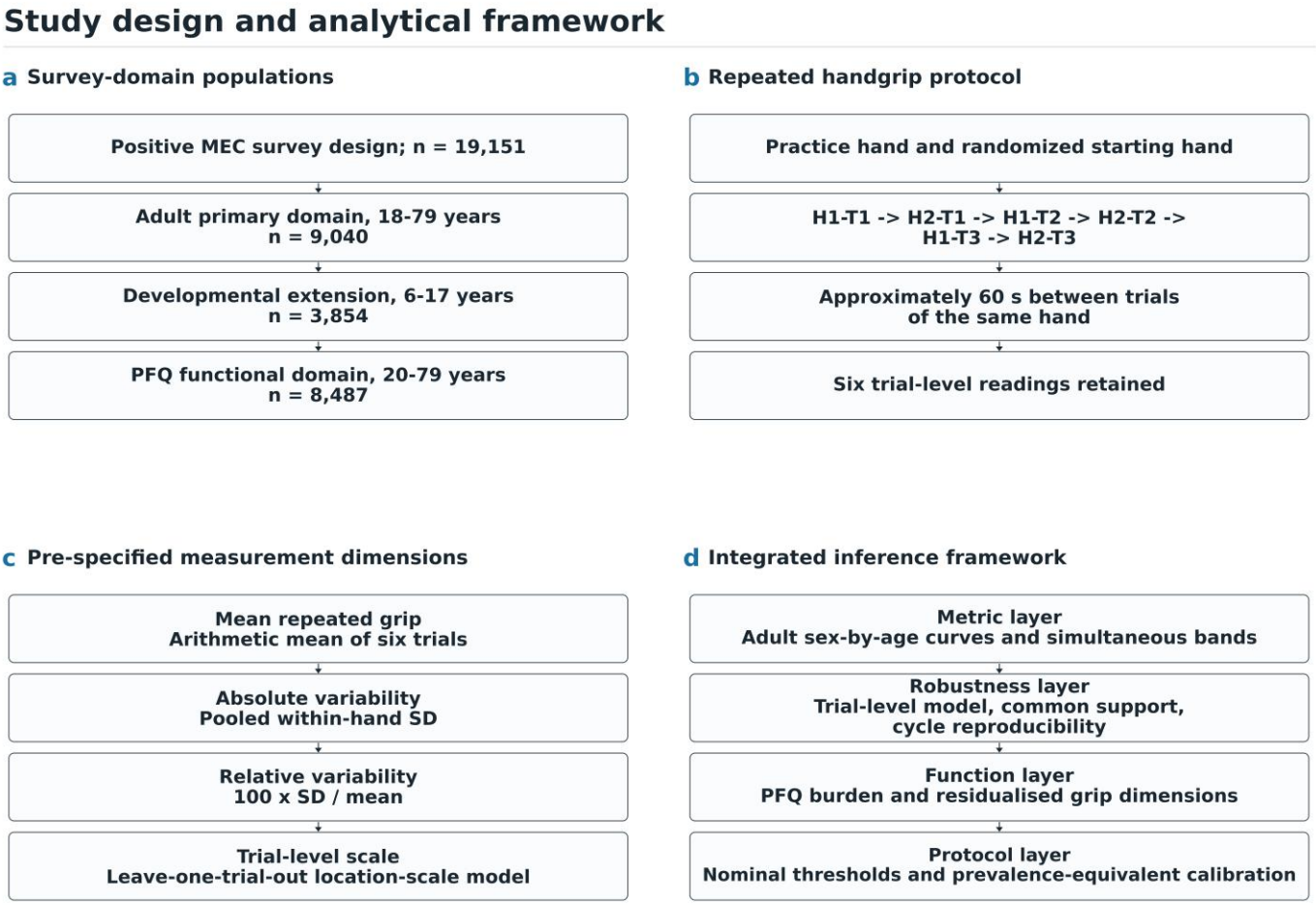


Figure 2 | Metric-dependent sex differences across adulthood. a, Model-standardised mean of all six maximal-grip trials by sex. b, Female-to-male ratio of pooled within-hand absolute SD. c, Female-to-male ratio of raw CoV. d, Female-to-male residual scale ratio from the trial-level location-scale model. Shaded areas in b-d are 95% simultaneous confidence bands across the full adult age grid; dashed horizontal lines indicate a ratio of 1. The adult primary domain comprised 9,040 participants aged 18-79 years.

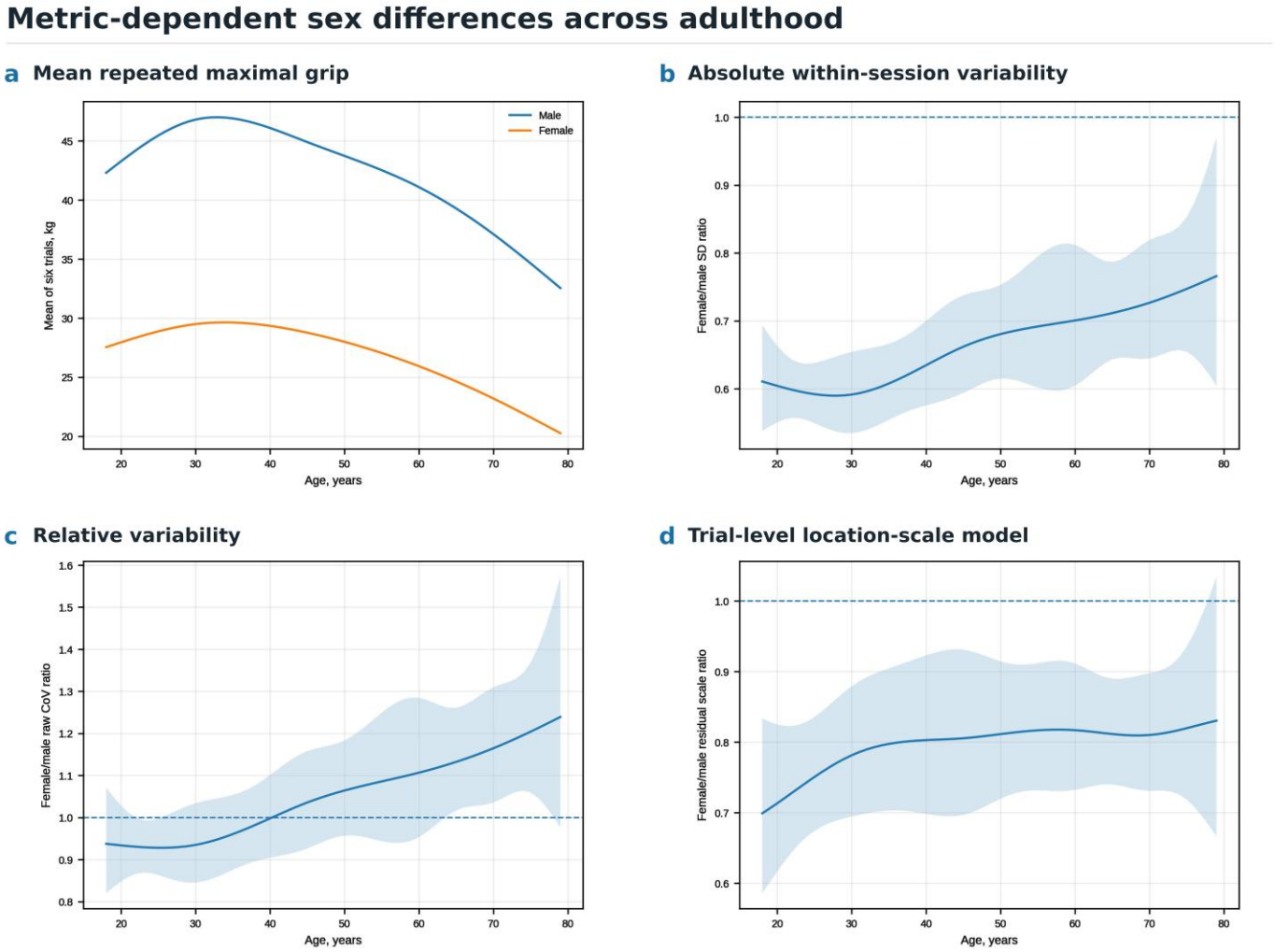


Figure 3 | Developmental context, cycle reproducibility and common support. a, Sex-specific mean-grip curves in the developmental extension aged 6-17 years. b, Female-to-male raw-CoV and mean-strength-matched SD ratios in the developmental extension. c, Cycle-specific adult raw-CoV ratios at pre-specified ages; error bars are design-based 95% confidence intervals. d, Weighted density-overlap coefficients for female and male mean grip across adult age bands. The developmental extension was analysed separately from the adult ageing analysis.

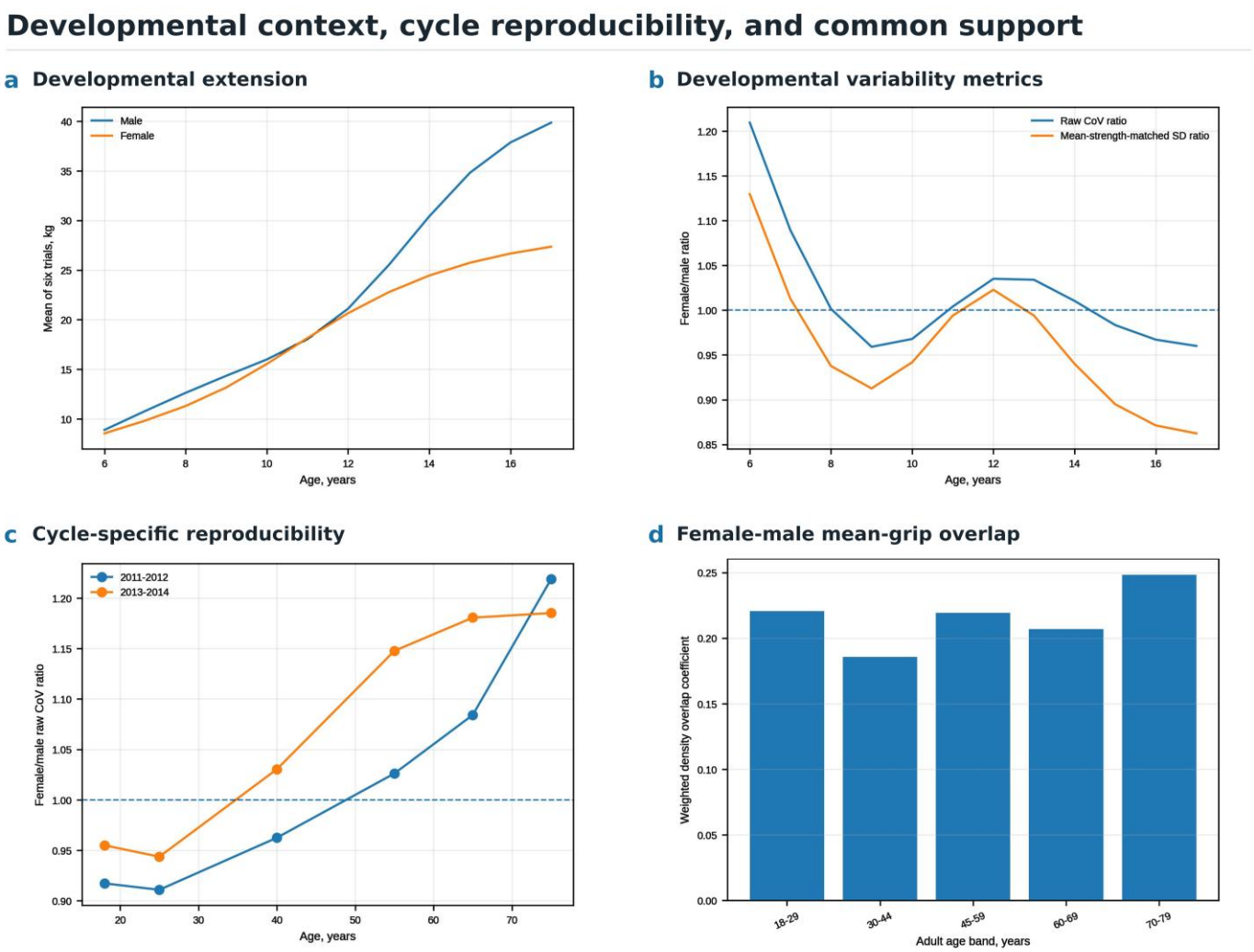


Figure 4 | Sex differences in functional burden and functional relevance. a, Survey-weighted prevalence of grip-relevant upper-limb difficulty by age group and sex; error bars are 95% confidence intervals. b, Adjusted female-to-male PRs for selected functional outcomes. c, Fully adjusted PRs for grip-relevant upper-limb difficulty per adverse survey-weighted SD in mean-grip deficit, mean-strength-adjusted variability and age-sex-adjusted asymmetry. d, Unadjusted P values for the nine smallest sex-by-grip interactions; the dashed line marks P=0.05. None of the 24 interaction tests remained significant after FDR correction (all q>=0.199).

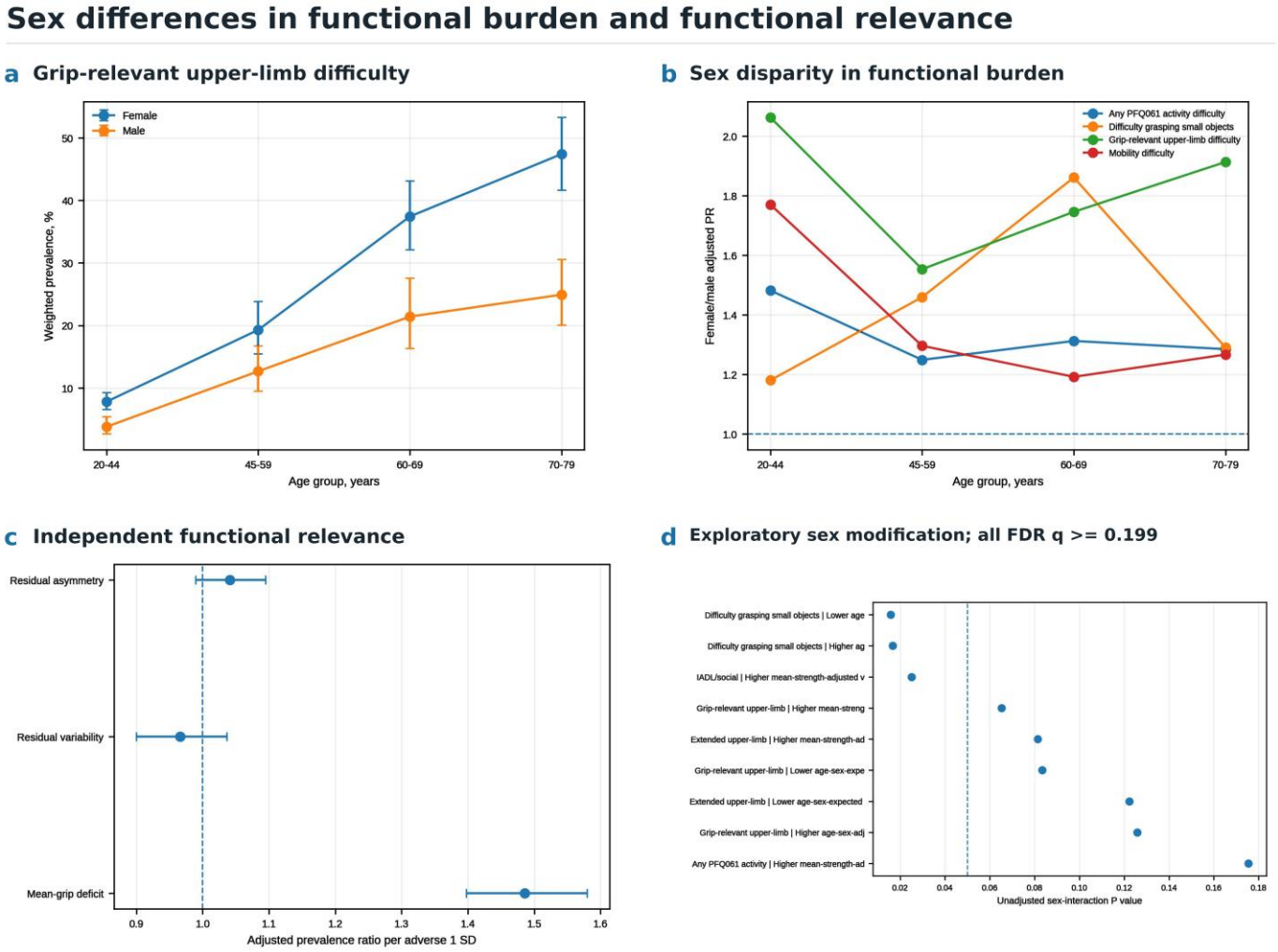
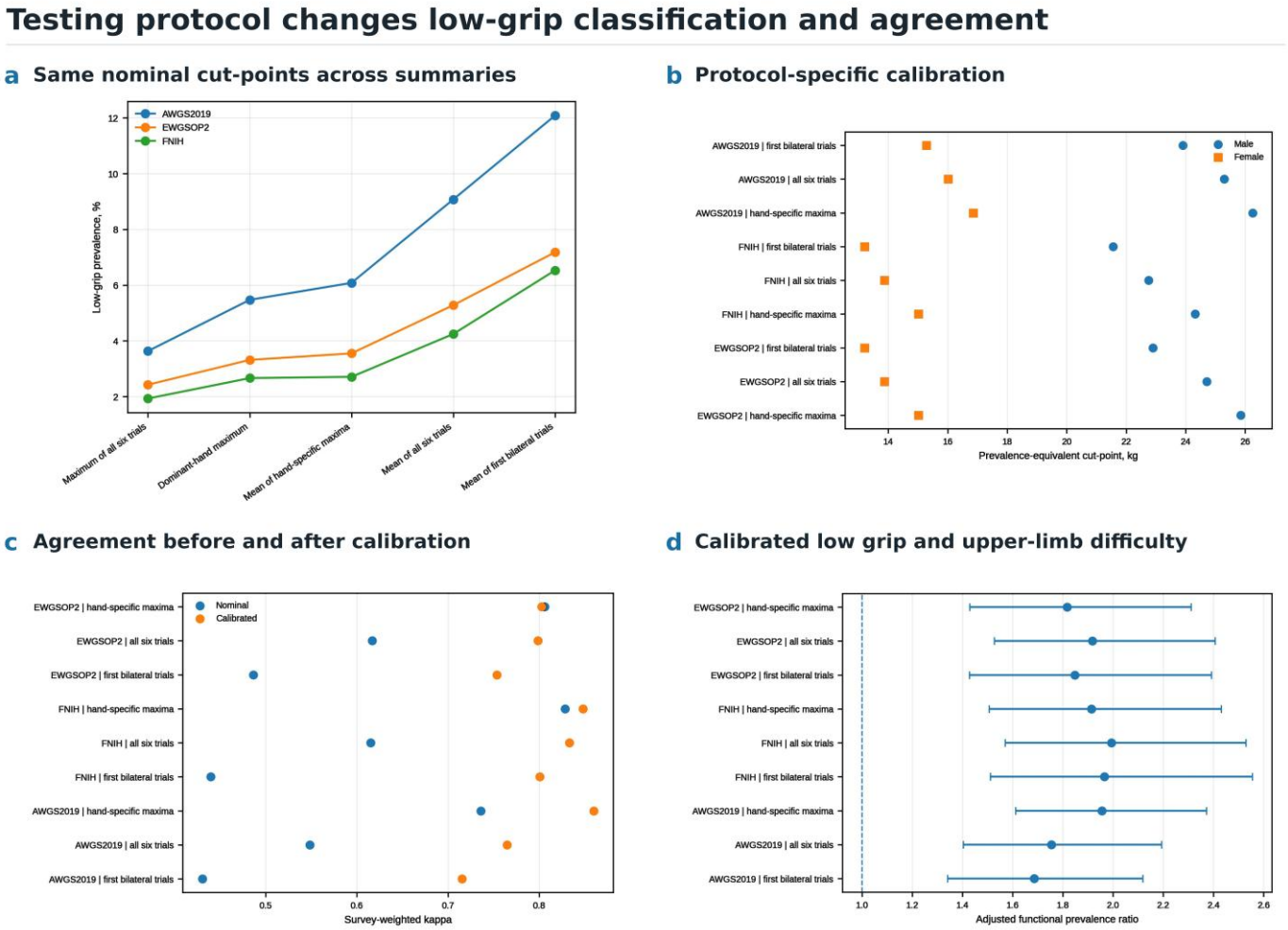


Figure 5 | Testing protocol changes low-grip classification and agreement. a, Survey-weighted low-grip prevalence among adults aged 60-79 years when the same nominal sex-specific threshold was applied to five grip summaries. b, Sex-specific prevalence-equivalent cut-points for three alternative summaries, using the maximum of all six trials as the reference. c, Weighted kappa before and after prevalence-equivalent calibration. d, Adjusted PRs for grip-relevant upper-limb difficulty under calibrated low-grip definitions. Error bars are 95% confidence intervals. Equivalent cut-points are internal protocol translations and are not proposed as clinical thresholds.



Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- npjAgingSupplementaryFiguresTables9.27.docx