Sexual dimorphism in human muscle ageing

Thomas W. Fieldsend[1], Callum R. O'Neill[1,2], Akshita Shrivastava[1], Helen E. Ogden[3], Nick Dand[4*], and Simon M. Hughes[1*]

[1]Randall Centre for Cell and Molecular Biophysics. School of Basic and Medical Biosciences, King's College London, SE1 1UL, UK

[2]Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, OX3 7BN, UK

[3]School of Mathematical Sciences and Southampton Statistical Sciences Research Institute, University of Southampton, SO17 1BJ, UK

[4]Department of Medical & Molecular Genetics, School of Basic and Medical Biosciences, King's College London, SE1 1UL, UK

*Co-correspondence s.hughes@kcl.ac.uk, nick.dand@kcl.ac.uk. These authors contributed equally to the work.

Abstract

Primary sarcopenia is a common geriatric syndrome characterised by age-related decline in muscle strength, mass, and quality, which is associated with reduced quality of life and a significant social burden. We report results from an analysis of arm muscle ageing on 478,438 UK Biobank participants aged 40–82 yr. A clear sexual dimorphism was identified: while the sex difference in age-related strength decline was modest, muscle mass loss was considerably more pronounced in males, both in absolute and percentage terms. We also introduce a novel measure of muscle quality, which showed substantially greater age-related decline in females. These trends held across cross-sectional and longitudinal analyses, and persisted after accounting for systematic size differences between the sexes. Furthermore, sex differences were apparent irrespective of female menopause status or HRT usage history. Our findings will inform the ongoing debate around the definition and diagnosis of sarcopenia, and prompt consideration of sex-specific interventions.

Contents

Abstract 1
Results 5
Modest sexual dimorphism in age-related strength decline 5
Age-related muscle mass decline is greater in males 6
Age-related muscle quality decline is greater in females 7
Females exhibit distinct muscle ageing trajectories from males irrespective of menopause status and history of HRT use 8
Discussion 8
Online Methods 13
Data preparation 13
Hand grip strength 14
Muscle mass 14
Muscle quality 15
Participant matching 15
Calculations and analyses 16
Regression analysis 16
Data visualisation 17
Acknowledgements 17
References 17
Tables 25
Figures 29
Extended Data 34
Supplementary Information 1 Participant exclusion flowcharts 42
Supplementary Information 2 Regression analysis 46
Supplementary Information 3 Adjustment of hand grip strength readings 53
Supplementary Information 4 Muscle mass adjustment and estimation 61
Adjustment of BIA impedance values 61
Generating novel muscle mass estimates 68
Comparative analysis of muscle mass estimates 69
Supplementary Information 5 UK Biobank Data-Fields included in this study 71
Supplementary Information 6 Muscle mass of participants with no self-reported history of cancer 72
Supplementary Information 7 Ulna length and muscle quality 73
Supplementary Information 8 Muscle quality calculated with alternative estimates of ulna length 75

Supplementary Information 9 Participant matching 77

Sarcopenia is a syndrome characterised by a progressive and generalised reduction of skeletal muscle strength, size, and quality [1,2], which is associated with a host of negative outcomes including functional impairment [3], loss of independence [4], lowered quality of life [5], and increased all-cause mortality [6,7]. Although reported sarcopenia incidence depends to a large extent on the diagnostic criteria used [8], true global prevalence in those aged 60 and above is probably ~10% [8,9], which would make sarcopenia more common in this age group than other geriatric conditions such as rheumatoid arthritis (~1–2% [10,11]) and cognitive impairment (~5.4–7.3% [12,13]). The combined socio-economic and health impacts of sarcopenia therefore represent a considerable 'social burden', the weight of which will likely increase as the global population continues to age [14].

While sarcopenia can occur secondary to other factors (e.g. systemic disease), it is predominantly a geriatric syndrome ('primary sarcopenia') [2]. Primary sarcopenia evidently has a large genetic component [15], with muscle strength and size heritability generally estimated to exceed 50% [16,17]. The development and progression of primary sarcopenia may also be influenced by other factors, such as lifestyle [18-20], socio-economic status [21], ethnicity [22], and the menopause [23].

Humans are sexually dimorphic, with males being generally taller and heavier than females, and possessing more skeletal muscle and less adipose tissue for their size [24-27]. Body size and composition change substantially with age, with both sexes typically exhibiting an age-related decline in height [28,29], fat-free mass, and bone mass [30,31], whereas fat mass generally increases with age [30]. The timing and magnitude of some of these changes differ by sex (e.g. fat accretion and height loss [29,32]), as does the prevalence of some geriatric syndromes (e.g. osteoporosis [33]). It is therefore essential from a public health perspective to identify and accurately characterise any sex differences in the age-related reductions of skeletal muscle strength, size, and quality that characterise primary sarcopenia.

Muscle strength is typically defined as the maximum force or torque generated at maximal voluntary contraction [34-36], and can be measured during concentric, isometric, or eccentric contraction [e.g. 37]. There is incontrovertible evidence that average muscle strength is considerably greater in males than females at all ages, and that muscle strength declines with age in both sexes [24,38-41]. However, it remains an open question whether sexual dimorphism exists in the age-related decline of muscle strength: while a limited number of small studies have reported greater age-related strength declines in males [42-44], a comparable amount of evidence exists that rates of decline are in fact similar between the sexes [38,40,45].

Muscle size can be quantified as muscle mass, volume, or cross-sectional area [e.g. 24,46,47]. Any sex differences in the size-age relationship would be highly relevant to the definition and diagnosis of primary sarcopenia. For example, diagnoses based on low *relative* muscle mass (as compared to reference populations of young adults) assume no sex difference in the general trajectory of age-related muscle size decline [e.g. 3,48,49], despite there being only one study to our knowledge that actually reports such a trend (a two-year follow-up study of Koreans aged 70–84 yr [19]). The limited evidence scattered across the literature is more suggestive of greater age-related muscle size decline in males [24,44,46,47,50-56]. However, the low sample sizes and high variance associated with many of these studies [e.g. 24,44,46] make it difficult to draw firm conclusions.

Muscle quality is defined as a measure of function (force exerted) per unit of muscle. Given the various ways in which both muscle function and size can be quantified, the literature on muscle quality is highly heterogenous. It is well established that age-related strength decline typically exceeds that attributable to size reduction alone (reviewed in [57]), indicating that reduced quality is a general feature of aged muscle. However, evidence of sexual dimorphism

in age-related muscle quality decline is inconsistent, with studies variously reporting a greater decline in females [38,58] or males [37,42], or no sex difference [37,42,58]. Indeed, several studies provide evidence for multiple trends: for instance, Lynch et al. (1999) detected no sex difference in age-related leg muscle quality decline, whereas a greater decline in arm muscle quality was evident in males [42]. Similarly, Lindle et al. (1997) reported similar rates of age-related muscle quality decline when strength was quantified using concentric peak torque, but greater male decline when eccentric peak torque measurements were used [37]. It is therefore apparent that no clear consensus exists regarding the general relationship between age, sex, and muscle quality [40,59].

The UK Biobank (UKB) is a prospective cohort study of >500,000 UK participants aged 40–70 yr at recruitment in 2006–2010 [60]. The wealth of phenotypic and genotypic data available in the UKB dataset has facilitated several large-scale studies on sarcopenia and relevant muscle phenotypes [18,47,56,61-68]. We employed both cross-sectional and longitudinal study designs to characterise sex-specific ageing trajectories of muscle strength, size, and quality in UKB participants, with the goal of informing the ongoing debate around how sarcopenia should be defined and diagnosed [e.g. 6,8,69].

Results

Modest sexual dimorphism in age-related strength decline

We performed a cross-sectional analysis of mean hand grip strength (HGS) readings taken at the baseline assessments of 368,336 UKB participants (202,028 females and 166,308 males) aged ≥40 and <70 years ('Cross-sectional Participants'; Table 1; Supp. Info. 1). Mean HGS of males (M) was ~68% greater than that of females (F) (F 23.7 ± 0.01 kgf vs M 39.8 ± 0.02 kgf [mean ± standard error of the mean (SEM)]; P <0.001; False Positive Risk [FPR] <0.001), despite males being slightly older on average (Table 1). The mean HGS of 69-year-old females was 6.8 kgf lower (-24.9%) than that of 40-year-old females (27.4 ± 0.14 kgf vs 20.6 ± 0.07 kgf), equivalent to a linear age-related decline of -0.24 kgf·yr^{-1} (-0.98%·yr^{-1}). In comparison, 69-year-old males were 8.3 kgf (-18.9%) weaker than 40-year-old males (43.7 ± 0.2 kgf vs 35.4 ± 0.1 kgf), equivalent to a linear age-related decline of -0.28 kgf·yr^{-1} (-0.72%·yr^{-1}). Males exhibited a trajectory of gradually accelerating HGS decline, whereas females did not display a comparable acceleration beyond 60 yr (Fig. 1a), suggesting sexual dimorphism in muscle strength ageing.

To formally quantify the apparent sexual dimorphism, we fitted a generalised additive mixed model (GAMM), regressing HGS on age, with sex, self-reported ethnicity, and potential technical confounders included as covariates (Online Methods). Including sex-specific splines for age-related HGS decline as an additional covariate substantially improved model fit—as indicated by the substantially lower Akaike Information Criterion (AIC) score (a measure of model quality [70]) when sex-specific splines were included (ΔAIC = 532—HGS model 1 in Supp. Info. 2)—thereby confirming sexual dimorphism in the HGS-age trajectory in the Cross-sectional Participants. The general trends identified in the observed data were supported by the sex-specific trajectory model, with a fitted annual HGS decline between 40–69 yr of -0.24 kgf·yr^{-1} (-1.02 %·yr^{-1}) in females, and -0.30 kgf·yr^{-1} (-0.75 %·yr^{-1}) in males (Fig. 1a). Males therefore appear to lose more absolute strength than females during middle age, but less proportional strength.

Cross-sectional and longitudinal studies can produce markedly different results when used to assess age-related anthropometric change, due to phenomena such as cohort effects [28,29,31]. We therefore investigated the relationship between HGS and age in a second, independent group of 45,629 UKB participants who attended an additional ('repeat')

assessment ('Longitudinal Participants'; Table 1; Supp. Info. 1), 33,362 of whom had suitable HGS readings taken at both assessments (17,146 females and 16,216 males). Simple cross-sectional analysis of the data from each measurement on the Longitudinal Participants showed that mean HGS of 80-year-old females (19.3 ± 0.64 kgf) was 9.5 kgf (-32.8%) lower than that of 40-year-old females (28.8 ± 0.39 kgf), equivalent to a linear age-related decline of -0.24 $kgf \cdot yr^{-1}$ ($-0.99\% \cdot yr^{-1}$), whereas 80-year-old males (31.6 ± 0.69 kgf) were 13.8 kgf (-30.5%) weaker than their 40-year-old counterparts (45.4 ± 0.74 kgf), equivalent to a linear age-related change of -0.35 $kgf \cdot yr^{-1}$ ($-0.90\% \cdot yr^{-1}$). Thus, the Longitudinal Participants showed broadly similar declines to the Cross-sectional Participants. We next implemented a longitudinal GAMM that formally confirmed the sexual dimorphism (ΔAIC = 1,936—HGS model 2 in Supp. Info. 2), with fitted declines between 40–80 yr similar to those observed in the raw data, at -0.24 $kgf \cdot yr^{-1}$ for females and -0.36 $kgf \cdot yr^{-1}$ for males (Fig. 1b). As in the cross-sectional data, the HGS-age trajectories of females and males appeared similar until ~60 yr, after which an accelerating rate of HGS decline was apparent in males only.

HGS is positively and strongly correlated with body size [70], but none of the above results account for potential age or age-sex differences in body size. We therefore refitted the cross-sectional and longitudinal GAMMs with additional covariates for standing height^2 (cm^2) and body mass$^{2/3}$ ($kg^{2/3}$) under the principle of isometric scaling [e.g. 35,70], continuing to observe the same general sexual dimorphism in trajectories (ΔAICs ≥511—HGS models 3–4 in Supp. Info. 2; Extended Data 1). As a further control for body size, we also analysed two age- and size-matched subsets: 61,683 female:male pairs drawn from the Cross-sectional Participants ('Matched Cross-sectional Participants'), and 5,435 female:male pairs drawn from the Longitudinal Participants ('Matched Longitudinal Participants') (Table 1; Methods). Males remained substantially stronger than females after matching (e.g. ~46% in the Matched Cross-sectional Participants; F 25.35 ± 0.03 kgf vs M 37.10 ± 0.04 kgf; P <0.001; FPR <0.001). Sexual dimorphism in age-related HGS decline was still detected in both matched subsets (ΔAICs ≥61—HGS models 5–6 in Supp. Info. 2), and the trajectories were broadly comparable to those of the unmatched participant groups, with an acceleration between 60–70 yr in males that was not observed in females, although females showed some evidence of accelerating decline beyond 70 yr in the Matched Longitudinal Participants (Figs 1a–d). As with all of the prior analyses, age-related HGS decline across the assessed age range differed by sex, being greater in males in terms of kilogram-force, and greater in proportional terms in females. This held true across both the observed data and the model-fitted relationships (Table 2).

Age-related muscle mass decline is greater in males

We next performed similar analyses on Bioelectrical Impedance Analysis (BIA)-derived measures of mean arm muscle mass (MM) (Fig. 2; Table 3). MM of male Cross-sectional Participants was ~68% greater than that of females (F 2.1 ± 0.001 kg vs M 3.53 ± 0.001 kg; P <0.001; FPR <0.001), and remained ~42% greater in the Matched Cross-sectional Participants (F 2.29 ± 0.001 kg vs M 3.26 ± 0.002 kg; P <0.001; FPR <0.001), confirming a substantial sex difference in muscularity not explicable by differences in body size (Figs 2a,c).

The general findings of the MM analyses were consistent across the Cross-sectional Participants, Matched Cross-sectional Participants, Longitudinal Participants, and Matched Longitudinal Participants, with all four groups exhibiting a similar and striking sexual dimorphism. Observed rates of age-related MM decline were around four times greater in males than females, being equivalent to linear declines of -0.003 $kg \cdot yr^{-1}$ (F) vs -0.013 $kg \cdot yr^{-1}$ (M) in Cross-sectional Participants between 40–69 yr, and -0.004 $kg \cdot yr^{-1}$ (F) vs -0.016 $kg \cdot yr^{-1}$ (M) in Longitudinal Participants between 40–80 yr. Male declines were also over twice as great as female declines in percentage terms, equivalent to $-0.15\% \cdot yr^{-1}$ (F) vs $-0.36\% \cdot yr^{-1}$ (M) in the

Cross-sectional Participants, and -0.19%·yr^{-1} (F) vs -0.48%·yr^{-1} (M) in the Longitudinal Participants (Fig. 2a,b; Table 3). Consequently, the magnitude of the observed sex difference in muscularity was found to decrease with age: for instance, MM of male Longitudinal Participants was ~75% greater than that of their female counterparts at 40 yr, but this difference was reduced to ~56% by 80 yr (Fig. 2b; Table 3).

Sexual dimorphism in the MM-age trajectory was overwhelmingly supported by GAMMs fitted to all matched and unmatched participants groups, irrespective of whether standing height^2 and body mass $^{2/3}$ were included as covariates (ΔAICs ≥1,561—MM models 1–6 in Supp. Info. 2). Model-fitted estimates were in close accordance with the observed data, with fitted male declines at least three times as great as those of females in absolute terms, and over twice as great in percentage terms (Fig. 2; Table 3; Extended Data 2). Substantially greater male muscle mass decline with age was also evident in Dual X-ray Absorptiometry (DXA) scan data available for 4,253 Longitudinal Participants (Fig. 2e; ΔAIC = 54.3—MM model 7 in Supp. Info. 2), and across three alternative sets of BIA-derived muscle mass estimates (Extended Data 3). Intriguingly, models fitted to both the MM and DXA data predict female muscle mass to remain essentially stable with age in the absence of age-related body-size decline, whereas substantial muscle mass decline is still predicted for males (Extended Data 4), emphasising the existence of a striking sexual dimorphism in the MM-age trajectory.

Age-related muscle quality decline is greater in females

We next combined the HGS and MM observations with individual estimates of forearm (ulna) length to derive a novel measure of mean arm muscle quality (MQ). MQ measures estimated specific force, i.e. force-generating capacity per unit of forearm muscle cross-sectional area (CSA—Methods).

Age-related MQ decline was much more pronounced in females than males across all participant groups (Fig. 3; Table 4). The observed rate of female MQ decline between 40–69 yr was around twice that of males when assessed cross-sectionally (Cross-sectional Participants: F -0.97%·yr^{-1} vs M -0.45%·yr^{-1}) and after body-size matching (Matched Cross-sectional Participants: F -0.91%·yr^{-1} vs M -0.46%·yr^{-1}) (Fig. 3a,c). Sexual dimorphism was similarly pronounced between 40–80 yr in the Longitudinal Participants (F -0.87%·yr^{-1} vs M -0.48%·yr^{-1}; Fig. 3b). Model-fitted estimates generally accorded with the observed data, except for a discrepancy in the male Matched Longitudinal Participants (observed: F -0.92%·yr^{-1} vs M -0.25%·yr^{-1}; model-fitted: F -0.98%·yr^{-1} vs M -0.52%·yr^{-1}; Fig. 3d), which apparently reflects the small sample sizes amongst the very oldest participants in this participant group (e.g. 18 males at 80 yr). Sexual dimorphism in ageing trajectories was strongly supported by GAMMs fitted to all participant groups, including those in which standing height^2 and body mass$^{2/3}$ were included as covariates (ΔAICs ≥216—MQ models 1–6 in Supp. Info. 2; Extended Data 5). Cross-sectional assessment of the DXA-scanned Longitudinal Participants (N = 3,972—2,061 females and 1,911 males) provided additional support for sex-specific MQ-age trajectories (Fig. 3e; ΔAIC = 23—MQ model 7 in Supp. Info. 2), and age-related muscle quality decline between 47–76 yr was over twice as great in females as in males (observed: F -1.44%·yr^{-1} vs M -0.66%·yr^{-1}; model-fitted: F -1.10%·yr^{-1} vs M -0.50%·yr^{-1}; Fig. 3e).

Observed MQ of females and males was almost identical at 40 yr in both the Cross-sectional Participants and the Longitudinal Participants (Fig. 3a–b). In contrast, average female MQ exceeded that of males until ~60 yr in the matched participant groups (Fig. 3c–d), and also when systematic sex differences in body size were accounted for within GAMMs (Extended Data 5). This distinction was explained by a positive correlation between MQ and standing height in both sexes (Extended Data 6,7). The greater mean standing height of males

compensated for their lower MQ in the unmatched participant groups, whereas the elimination of this sex difference in the matched participant groups caused female MQ to appear relatively greater (Extended Data 7a). These complex allometric relationships between standing height, CSA, MM, and MQ suggest that changes in muscle geometry with age contribute to the observed sexual dimorphism.

Females exhibit distinct muscle ageing trajectories from males irrespective of menopause status and history of HRT use

To investigate the relationship between menopause status and muscle ageing, we used GAMMs regressing HGS, MM, and MQ on age to perform pairwise comparisons on groups of 53,840 premenopausal females, 116,261 postmenopausal females, and 195,794 males drawn from the Cross-sectional Participants ('Menopause Participants', N = 365,895; Methods; Table 1; Supp. Info. 1; Extended Data 8). GAMMs included self-reported ethnicity, technical covariates, standing height^2, and body mass$^{2/3}$. Models were fitted between 45–60 yr, the window within which menopause typically occurs [71]. Perimenopause was not included in these analyses because perimenopausal individuals cannot reliably be identified in the UKB.

Model comparison strongly supported the existence of distinct HGS-age trajectories for premenopausal females, postmenopausal females, and males (ΔAICs ≥11.4; HGS menopause models 1–3 in Supp. Info. 2) (Fig. 4a). Age-related change in female MM was minimal (Figure 2), and no difference in MM-age trajectory was found between pre- and postmenopausal females (ΔAIC = 0.076; MM menopause model 1 in Supp. Info. 2). However, postmenopausal status was associated with significantly lower MM than premenopausal status (~0.037 kg lower MM, as estimated without group-specific age splines; P <0.001; FPR <0.001; MM menopause model 1_{GLOBAL} in Supp. Info. 2). The MM-age trajectories of both female groups were distinct from that of males (ΔAICs ≥550; MM menopause models 2–3 in Supp. Info. 2) (Fig. 4b). Three distinct MQ-age trajectories were also supported (ΔAICs ≥11.2; MQ menopause models 1–3 in Supp. Info. 2; Fig. 4c) indicating that the observed difference in the HGS-age trajectory between pre- and postmenopausal females is attributable to MQ differences between these groups.

Further stratifying postmenopausal females by history of hormone replacement therapy (HRT) use (Online Methods) revealed distinct HGS-age, MM-age, and MQ-age trajectories for females with versus without a history of HRT (ΔAICs ≥13.6; HGS/MM/MQ menopause model 4 in Supp. Info. 2; Fig. 4d–f), although it was not possible to determine whether these differences were due to the effects of HRT *per se*, or to systematic pre-existing differences between females that affected their HRT usage history. Importantly, differences between males and postmenopausal females were still apparent after stratifying the latter by HRT usage history (ΔAICs ≥168; HGS/MM/MQ menopause models 5–6 in Supp. Info. 2), demonstrating that sexual dimorphism in muscle ageing persists irrespective of both female menopause status and history of HRT use.

Discussion

Our analysis of 478,438 UKB participants aged 40–82 yr identified clear sexual dimorphism in arm muscle ageing. Whereas only a modest sex difference was detected in age-related hand grip strength (HGS) decline, age-related muscle mass (MM) decline was considerably more pronounced in males. Consequently, age-related decline of muscle quality (MQ)—an estimate of specific force—was substantially greater in females. These general trends held across cross-

sectional and longitudinal analyses, persisted after accounting for systematic sex differences in body size, and were apparent irrespective of female menopause status or HRT usage history.

Robust evidence of a functionally relevant sex difference in human muscle ageing

Our study provides strong evidence that age-related MM decline is substantially greater in males than females. Although the magnitude of this dimorphism is surprising, the general finding is in broad agreement with the available literature [24,44,46,47,50-56] [24,44,46,47,50-56]. Perhaps more striking is the evidence that age-related MM decline is minimal in females after accounting for decline in overall body size, whereas male MM loss with age clearly exceeds that attributable to overall body size decrease. While this is a remarkable finding, it is nevertheless consistent with two low-powered studies, which also contain hints that age-related decline of female arm muscle size is minimal [24,46].

The pronounced sex difference in MM change is especially compelling in light of the relatively modest sex difference in age-related HGS decline. Sexual dimorphism in age-related decline of HGS and MM ultimately presented as a greater age-related MQ decline in females across all ages and participants groups. This finding is compatible with Hakkinen et al. (1996) [38], who reported significantly lower quadriceps muscle quality in older (~70 yr) than younger (~50 yr) females, but no difference in analogous groups of males. Similarly, a longitudinal study of 511 participants >50 yr detected significantly faster age-related quadriceps muscle quality decline in females, although—unlike in our study—this difference was not statistically significant after adjusting for body size (standing height) [58]. Muscle quality appears to be the sarcopenia-associated muscle phenotype for which age-sex differences are least characterised, and no consistent relationship is evident in the literature [37,38,40,42,58,59]. Nevertheless, a decline in function of remaining muscle mass is acknowledged as a key diagnostic criterion for sarcopenia in the most-recent European Working Group on Sarcopenia in Older People (EWGSOP2) guidelines [2]. Our finding of greater age-related MQ loss in females is plausible, given that the evidence supporting greater age-related muscle size decline in males is relatively strong [24,44,46,47,50-56], whereas evidence of a correspondingly greater loss of male muscle strength is equivocal in the literature [38,40,42-45].

Origins of sexual dimorphism in human muscle ageing

The distinct muscle ageing trajectories of females and males may derive from intrinsic sex differences in the composition, architecture, and neural activation of human skeletal muscle. Although sex differences in these aspects of muscle physiology have been reported [e.g. 25,39,72], their alteration with age is less clear.

To understand ageing-related changes, it is important first to consider how female and male musculature diverges during adolescence and early adulthood, and to ask whether reduction of these differences with age can explain the distinct muscle ageing trajectories reported in this study. Female and males exhibit little difference in musculature until puberty, at which point a rapid and marked divergence occurs [25,73]. Most notably, the average muscle mass of adult males is considerably greater than that of adult females, even after accounting for systematic sex differences in body size [24-27]. In this study, MM exhibited an accelerating decline between 40–70 yr in males, whereas the modest decline in female MM over this age range occurred at a relatively constant rate. Consequently, the magnitude of the sex difference in MM decreased with age. Androgens—particularly testosterone—are key promotors of muscle mass, especially in males [74]. Serum testosterone levels decline modestly between 40–70 yr in males, but change little in females [75]. This convergence of androgen levels with age bears some resemblance to the convergence of MM. However, female serum testosterone levels remain unchanged during the menopausal years [76], and thus cannot explain the lower MM of

postmenopausal females. Furthermore, testosterone replacement has no effect on male HGS [77], suggesting that testosterone decline does not exert a substantial effect on the male HGS-age trajectory. While the loss of fluctuating female hormone levels at menopause may influence the female MM-age trajectory, these changes do not correlate well with the trajectory of HGS, MM or MQ (as discussed more fully below). Thus, the rapid and marked divergence of both hormonal milieu and musculature between females and males at puberty is partially reversed in later life. However, it remains unclear whether the distinct trajectories leading to reconvergence simply reflect the partial loss of sex-specific secondary sexual characteristics, or whether they represent a continued divergence in aspects of the ageing process.

In the search for a single physiological, cytological, or biochemical difference that is consistent with all aspects of the observed sexual dimorphism in muscle ageing trajectories, one attractive possibility is that females experience an age-associated replacement of force-generating contractile material with material that does not enhance isometric force. Age-related increase in intra- and intermuscular adipose tissue (IMAT) is well established [78,79]. If the methods used to detect IMAT fail to detect a significant fat accumulation, then elevated levels of IMAT in the muscle of older females could theoretically explain why female MQ declines with age, whereas MM remains relatively stable. However, a previous longitudinal study found no significant relationship between thigh IMAT and muscle quality [58], and MRI scans of the thighs of UKB participants also show that thigh IMAT volume correlates only weakly with fat-free thigh muscle volume [47]. Thus, unless the methods used to detect IMAT fail to detect a significant fat accumulation—and barring a radical difference in IMAT accretion between the upper and lower limbs—replacement of arm muscle tissue by IMAT seems unlikely to explain our findings. The alternatives of accumulation of 1) non-IMAT fat in the form of intramyofibre lipid droplets, 2) other non-contractile tissues (e.g. fibrous connective tissue) or 3) other cytoplasmic components (such as mitochondria or non-functional myosin) replacing functional myofibrils in female arm muscle remain to be investigated.

Muscle architecture may be another mediator of age-sex differences in MQ. A positive correlation between standing height and MQ arose in both sexes because greater standing heights are typically associated with both higher HGS and lower CSA, i.e. the numerator and denominator of our MQ measure, respectively. Both predicted arm (ulna) length (UL) and MM scaled positively allometrically with standing height in our dataset; UL scaled more strongly with standing height than MM, meaning that taller individuals were predicted to have more elongate arm muscles—and thus smaller CSAs—despite possessing more MM. Positive allometric scaling between arm length and standing height—including forearm (radius) length—has been reported elsewhere [80,81]. An inverse correlation between muscle quality and CSA has also been observed in at least one study [82]. The increase in MQ with standing height and the inverse correlation between CSA and MQ observed in our study are both consistent with a pennate muscle architecture model under which elongated forearm muscles in taller individuals contain more muscle fibres, thereby increasing physiological cross-sectional area (PCSA)—i.e. the summed component of force deriving from the cross-sectional areas of all fibres within a muscle—at a given CSA [83]. Muscle fibre number may also help to explain why male MQ generally exceeds that of females at a given CSA. There is evidence that muscle fibre pennation angles are typically greater in males [84-89]: while greater pennation angles reduce the specific force of individual muscle fibres, this loss can be more-than offset by the denser muscle fibre packing that is possible at greater pennation angles [90], suggesting that denser muscle fibre packing in males may lead to a relative increase in male PCSA. Interestingly, vastus lateralis muscle fibre pennation angles have been shown to decrease with age in both females and males [84]. If this reduction in pennation angle occurs to compensate for age-related loss of muscle fibres [91,92], then it would be expected to lead to age-related MQ decline in both sexes, as

observed in our study. The extent to which age-sex differences in muscle architecture contribute to sexual dimorphism in muscle ageing trajectories remains an open question deserving of further investigation.

Sexual dimorphism in muscle ageing also likely reflects age-sex differences in gene expression. For example, age-related upregulation of stimulatory muscle growth regulatory genes is thought to occur in females, but not males discussed in [93]. A small study by Raue et al. (2007) 77 [94] also reported significant age-related upregulation of the catabolic genes *FOXO3A* and *MuRF1* in females, although we are not aware of an equivalent study performed on males. However, if catabolic genes are in fact upregulated in both sexes, then increased muscle catabolism could conceivably be offset by increased muscle growth in older females but not older males, thus helping to explain why age-related MM decline is attenuated in females, although not why female MQ decline is more rapid. We intend to perform Genome Wide Association Studies (GWASs) on our muscle phenotypes, to characterise the genetic architecture underpinning the age-sex differences described in this study. GWAS analyses should also help to explain the high degree of variation observed in these phenotypes, much of which likely has a genetic basis [16,17,64,95].

Muscle ageing is associated with a number of health and lifestyle factors that were not considered in our study, including physical activity levels [19], socio-economic circumstances [21], diet 20 [20], and acquired disease 84 [96]. Although systematic age-sex differences in these factors are unlikely to fully explain the observed sexual dimorphism, it is certainly possible that each factor exerts a systematic effect. Although the primary aim of our study was to characterise—rather than explain—sex differences in muscle ageing, it is nevertheless clear that future studies should examine how both intrinsic and extrinsic factors contribute to the sexual dimorphism that we have described, as well as the potential for intrinsic factors to influence lifestyle in 'virtuous' or 'vicious' cycles.

A potential acute effect of perimenopause on female muscle mass

One of the most intriguing findings of our study was that postmenopausal females were predicted to have ~37 g less MM than their premenopausal counterparts, despite there being no significant difference in the shape of the MM-age trajectory of these two groups. This finding strongly suggests substantial MM reduction during the brief (mean duration ~1.3 yr 80 [71]) perimenopausal phase, and is strikingly consistent with the findings of a recent literature review [97], which concluded that 1) postmenopausal females have lower muscle mass than premenopausal females, and 2) oestradiol (an oestrogen) decline influences muscle mass loss during—but not after—menopause. Female levels of serum oestrogen decline substantially both during and after the menopause [97]. Nevertheless, the parallel MM-age trajectories of pre- and postmenopausal females argue against oestrogens being key drivers of female MM decline, and the abrupt change in oestrogen levels that occurs during the menopause bears no relation to the steady age-related decline in female HGS and MQ observed between 40–70 yr. Interestingly, our data offer no support for the protective effect of HRT on female muscle reported elsewhere [e.g. 98,99]. However, we stress that it is not possible to determine whether the differences that we detected between postmenopausal females with and without a history of HRT use are directly attributable to HRT, or whether traits associated with muscle ageing are simply also associated with likelihood of undergoing HRT. The effects of perimenopause and HRT use on female muscle ageing require further investigation.

Study limitations

The UKB cohort is healthier and less socioeconomically deprived than the general UK population [100], and we found this bias to be particularly pronounced in our Longitudinal

Participants, who were—on average—even less socioeconomically deprived than the Cross-sectional Participants, and were demonstrably healthier, e.g. having lower mean body-mass index ('BMI') scores and fewer self-reported cancers than their cross-sectional counterparts (Online Methods). It is therefore apparent that the known volunteer bias in the UKB 85 [100] is compounded by a 'retention bias' in returning participants, a phenomenon that has been reported in other biobanks (e.g. the HUNT study [101]). In addition, white Europeans are overrepresented in the UKB (94.6%) [100] compared with the UK population as a whole (81.7% in England and Wales in 2021 87 [102]). This bias towards healthy, affluent, white participants in the UKB means that caution is warranted when extrapolating the findings of this study to other populations—both in the UK and elsewhere—and highlights the need for more diverse biobank cohorts 88 [103] in which to study sex differences in muscle ageing.

HGS and MM measurement error represents another potential concern in our study. HGS readings were taken on dynamometers, many of which showed clear evidence of miscalibration/measurement drift. We took steps to account for this—both by adjusting (and in some cases excluding) HGS readings, and by controlling for dynamometer-level effects within our models. Nevertheless, we detected no compelling evidence of a systematic effect of dynamometer measurement error, and therefore it is unlikely that dynamometer effects introduced an artefactual sexual dimorphism. Most MM values used in this study are not direct measurements, but BIA estimates generated using novel predictive equations. We optimised these equations, and are confident that the MM estimates that they produce are superior to the default values provided in the UKB dataset (themselves estimates generated by predictive equations [104,105]). However, BIA estimates will inevitably be incorrect for some participants, and the possibility of some systematic bias due to prediction error (e.g. persistent under- or overestimation of MM in one sex) cannot be ruled out, although the trends apparent in the DXA data showed close accordance with our BIA estimates.

Implications for diagnosis and detection of primary sarcopenia

Our results are directly relevant to the detection and diagnosis of primary sarcopenia, a complex geriatric syndrome with no universally accepted diagnostic criteria [6,8,69]. Despite a 2018 literature review finding that nearly 70% of research studies diagnosed sarcopenia using skeletal muscle mass alone 6 [6], the pronounced sex difference in age-related MM decline calls into question the suitability of muscle mass as a sufficient diagnostic criterion for sarcopenia. For example, muscle mass more than two standard deviations below the mean for a sex-specific reference population of healthy young adults has previously been used to diagnose sarcopenia [3,48,49], but sex differences in the age-mass relationship suggest that, for the median individual of each sex, these thresholds are likely to be passed at different ages.

Furthermore, while sex-specific thresholds of absolute strength are often used in sarcopenia diagnosis reviewed in [8], it is not clear that elderly females and males in most modern societies require substantially different levels of absolute strength to complete normal daily activities. Indeed, Amara et al. (2003) 24 [24] suggested that muscle-group-specific minimum strength thresholds might instead exist, below which day-to-day living may be hindered; these they estimated at ~150 N (~15.3 kgf) for hand grip strength and ~400 N (~40.8 kgf) for the plantar flexor muscles. Interestingly, the hand grip strength threshold of ~15.3 kgf is similar to the 16 kgf threshold for females given in the most-recent EWGSOP2 guidelines, but well below the 27 kgf threshold for males [2]. In fact, of 12,823 Longitudinal Participants aged ≥ 70 yr at repeat assessment in our study, 662 of the 688 participants with a HGS reading < 150 N are female (96.2%). As females comprise only 44.5% of Longitudinal Participants ≥ 70 yr (N = 5,700), females ≥ 70 yr are almost 32 times as likely as males ≥ 70 yr to fall below the 150 N / 15.3 kgf threshold in this simple analysis (11.6% of females vs 0.37% of males). Additionally,

HGS readings of 11.2% of males ≥ 70 yr (799/7,123) are ≥ 150 N but < 27 kgf, meaning that these participants meet the sarcopenia criterion of EWGSOP2, yet remain above the functional threshold of Amara et al. It is therefore clear that the decision to use sex-specific strength thresholds must drastically impact the relative incidence of diagnosed sarcopenia in females versus males, and urgent consideration should therefore be given to the question of whether sex-independent strength thresholds represent more useful diagnostic markers of sarcopenia than sex-specific ones. Our findings also highlight the need to foreground sex in muscle ageing research, a field in which it often receives only superficial treatment, or is disregarded entirely (i.e. single-sex studies [e.g. 94,106]). Nevertheless, our findings appear to vindicate current efforts to emphasise the primary importance of absolute muscle strength and function—rather than size and/or quality—in sarcopenia definition and diagnosis [e.g. 2,69].

Online Methods

Data preparation

Data (Supp. Info. 5) for 502,461 participants were retrieved from the UKB [60] under approved UKB Project #67736. All analyses were conducted in R version 4.3.0 [107]. Age at assessment was calculated to the nearest month with the 'lubridate' R package [108] using date of assessment and year and month of birth; as day of birth is not available in UKB, day of birth was nominally designated as the 15th for all participants. Season of attendance was defined as meteorological season (i.e. Spring: March–May; Summer: June–August; Autumn: September–November; Winter: December–February). Self-reported ethnicity was used to assign participants to one of three primary ethnic group categories used in the 2021 UK Census: Asian or Asian British ('Asian'); Black, Black British, Caribbean or African ('Black'); and White ('White') [109]. For the purpose of analysis, all other participants (reporting Mixed or multiple ethnic groups, Other ethnic group, Do not know, Prefer not to answer, or NA) were grouped into a category termed 'Other'. We chose not to exclude participants based on a history of medical conditions that might potentially affect muscle size or function, as preliminary analyses indicated that doing so would substantially reduce sample size without materially affecting overall trends (e.g. removal of >36,000 participants with one-or-more self-reported cancer diagnoses; see Supp. Info. 6).

24,023 participants were excluded from analysis overall, with the remaining 478,438 participants split into those who attended an additional ('repeat') assessment in Cheadle, Newcastle, or Reading between April 2014 and March 2020 ('Longitudinal Participants', N = 45,629 participants) and those who did not ('Cross-sectional Participants', N = 432,809) (Supp. Info. 1). A 'retention bias' towards healthier participants was evident in the Longitudinal Participants, who—compared to the Cross-sectional Participants—had significantly lower baseline BMIs (26.6 ± 0.02 kg/m^2 vs 27.5 ± 0.007 kg/m^2; P < 0.001; FPR < 0.001) and significantly fewer self-reported cancer diagnoses at baseline (0.065 ± 0.001 vs 0.091 ± 0.0005; P < 0.001; FPR < 0.001). Longitudinal Participants were also significantly less socioeconomically deprived than Cross-sectional Participants, as determined using the Townsend Deprivation Score [TDS] at baseline (-1.89 ± 0.01 vs -1.27 vs 0.005 [with higher scores indicating higher deprivation]; P < 0.001; FPR < 0.001), and TDS correlates strongly with overall health [110]. These differences persisted after controlling for systematic age and sex differences between the two participant groups (p-values <0.001; FPRs <0.001—data not shown).

A 'Menopause Cohort' of 365,895 Cross-sectional Participants was also created, which included 170,101 females of known menopause status and all 195,794 males (Supp. Info. 1). Females who reported having undergone menopause outside of the typical 'menopause window' of 45–60 yr [71] were removed from the participant group. Stratification of

postmenopausal females by HRT usage history within this participant group was based on response to the question "Have you ever used hormone replacement therapy (HRT)?", with only those who answered "Yes" or "No" included in HRT analyses.

Hand grip strength

Maximum HGS readings in whole kilogram force units (kgf) were taken from seated participants using an adjustable Jamar J00105 hydraulic hand dynamometer (Lafayette Instrument Company, Lafayette, IN, USA), which measures force isometrically [111]. HGS readings were adjusted to account for observed dynamometer miscalibration and measurement drift, and a number of readings found to have been taken under extreme drift or miscalibration were excluded from the final analyses (Supp. Info. 3). Analyses were performed using the mean HGS of both hands, as left and right hand showed similar trends (Supp. Info. 3).

Muscle mass

Most muscle mass estimates presented in this study derive from Bioelectrical Impedance Analysis (BIA), a widely employed method of estimation using electrical conductance between skin-surface electrodes [112] that is considered sufficiently accurate for muscle mass measurements in large cohorts [3]. All BIA readings in the UKB dataset were taken with a Tanita Body Composition Analyser BC-418MA (Tanita Corporation, Tokyo, Japan), which uses an undisclosed, proprietary algorithm calibrated by DXA scans [104] to generate limb and whole-body estimates of muscle mass by combining observed raw impedance values (in ohms—Ω) with information on the age, sex, ethnicity, height, and body mass of the participant [105] (also see Tanita Body Composition Analyser BC-418MA Instruction Manual).

DXA scan estimates of arm muscle mass were also available for a subset of 4,253 Longitudinal Participants ('DXA-scanned participants'). All DXA scans were conducted at the Cheadle Assessment Centre on a GE-Lunar iDXA machine (GE Healthcare, Madison, Wisconsin, USA). DXA uses the differential attenuation of X-ray beams at two different energies to generate compartmental estimates of body composition [113]. While DXA technically returns compartmental estimates of lean soft tissue (LST) rather than muscle mass, a large majority of LST is muscle, with only small amounts of skin and connective tissue present in LST [114]. It is therefore considered justified in practice to treat DXA LST readings as muscle mass estimates [114,115]. A 2018 study of 4,753 DXA-scanned UKB participants found excellent accordance between leg muscle mass estimates produced by DXA and MRI ($R^2 = 0.97$), the latter of which is considered a 'gold standard' for regional body composition analysis [116]. DXA has also been shown to produce fat-free mass estimates (i.e. LST, organs, and bone [114]) that are more similar (i.e. show higher correlation and lower mean bias) to those produced by CT scans—another standard in regional body composition analysis [116]—than are those derived from BIA [117].

We used the DXA-scanned participants to compare DXA mean arm muscle mass estimates with three sets of BIA-derived muscle mass estimates:

1) direct outputs from the Tanita Body Composition Analyser BC-418MA ('Tanita'),
2) estimates generated using the equations of Powell et al. (2020) [118] ('Powell'), which predict mean arm muscle mass using raw impedance values and other anthropometric data readily available in the UKB (e.g. height and body mass; see Model C: Arms Lean Mass (g) in Supplementary Table 3 of [118]),
3) estimates generated using our own novel equations ('novel'), which were themselves based on the Powell equations (Supp. Info. 4).

The muscle mass estimates produced by our novel equations proved to be significantly more similar to the DXA-derived muscle mass estimates than did either the Powell or Tanita estimates (Supp. Info. 4). We therefore used our novel estimates of muscle mass in our primary analyses.

Muscle quality

Muscle quality can be defined either as strength divided by muscle mass [e.g. 7,36], or strength divided by muscle cross-sectional area (CSA) [e.g. 39]. Because isometric force production—as measured in HGS—is biomechanically proportional to the CSA of active muscle rather than its mass [35,119], we defined muscle quality as (mean HGS)/(mean arm muscle CSA).

Direct measurements of arm muscle CSA are not available for UKB participants, so CSA was instead modelled by dividing muscle mass by ulna length (UL). As an individual's standing height can be predicted from their UL with high accuracy [120,121], we reasoned that the reverse must also be true, and so rearranged equations from Madden et al. (2020) [121]—originally developed to predict standing height from UL—to estimate UL from participants' standing heights. These equations were highly applicable to the UKB dataset, having been generated using a multi-ethnic British cohort comprising both sexes and spanning a wide age range (21–82 yr).

Whereas standing height generally declines with age [29], an analogous decrease in UL is considered improbable [121], suggesting that the standing height:UL ratio is unlikely to remain constant throughout life. If not controlled for, this phenomenon might result in an artefactual age-related decline in arm muscle quality (as defined in this study; Supp. Info. 7). To account for this, we calculated UL using participants' predicted maximum standing heights (PMSHs), rather than their observed standing heights. PMSHs of participants were predicted using previously published equations [29] that are proven to accurately predict maximum standing height from observed standing height, sex, and age.

The six sex- and ethnicity-specific equations [121] used to predict UL from PMSH were:

Asian Females: UL = (PMSH [cm] - 77.62) / 3.26
Asian Males: UL = (PMSH [cm] - 83.58) / 3.26
Black Females: UL = (PMSH [cm] - 79.55) / 3.14
Black Males: UL = (PMSH [cm] - 85.8) / 3.14
White Females: UL = (PMSH [cm] - 95.6) / 2.77 [Footnote 1]
White Males: UL = (PMSH [cm] - 90.92) / 3.14

To avoid bias, participants in the Other ethnic group category were randomly assigned one of the three sex-specific UL calculation equations at a ratio of 1:1:1.

A substantially greater age-related decline in female muscle quality persisted whether PMSH or several alternative calculations of UL were employed (Supp. Info. 8), indicating that our novel measure of muscle quality was not compromised by the use of estimated or predicted values for UL, CSA, and maximum standing height.

Participant matching

The *matchit* function of the 'MatchIt' R package [122] was used to match female:male pairs based on their baseline similarity in age, standing height, and body mass. Nearest neighbour matching without replacement was performed, with Mahalanobis distances used as the measure of similarity. Males were arbitrarily designated as the 'Treatment' group—i.e. for each male a matching female was sought, rather than vice versa—due to there being fewer males than female Cross-sectional Participants and Longitudinal Participants (Table 1). Designating females as the Treatment group instead of males did not return meaningfully different results (data not shown). The *caliper* argument was used to prevent matching of participants whose values for a given criterion differed by more than a predefined amount (age at baseline ≤ 0.5

[Footnote 1] Originally published in Elia (2003), and recommended for use for white females by Madden et al. (2020).

yr; standing height at baseline ≤ 1 cm; body mass at baseline ≤ 1 kg). *matchit* returned 61,683 female:male pairs drawn from the Cross-sectional Participants (i.e. Matched Cross-sectional Participants, N = 123,366) and 5,435 female:male pairs drawn from the Longitudinal Participants (i.e. Matched Longitudinal Participants, N = 10,870) that proved to be exceptionally well matched based on recommended assessment criteria [122,123] (Supp. Info. 9). Ethnicity was not included as a matching criterion, as inclusion further reduced the overall diversity of the matched participant groups (data not shown).

Calculations and analyses

Comparison of means for two independent samples was performed using either Student's *t*-test (when variances were equal) or Welch's *t*-test (when variances were unequal); equality of variance was established beforehand using an *F*-test. False Positive Risk (FPR) scores were calculated using the methods of Sellke et al. (2001) [124], as discussed by Colquhoun (2019) [125], and implemented using the *pcal* function in the 'pcal' R package [126]. Total percentage changes in longitudinal data were calculated as *(repeat measure − baseline measure)/(baseline measure)*, and annual percentage changes were calculated as $(\textit{repeat measure} / \textit{baseline measure})^{(1 / \textit{Interval})} - 1$, where *Interval* is the recorded time (in years) between the two measures. α was set at 0.05 for all analyses.

Regression analysis

The relationship between each muscle trait (HGS, MM, or MQ) and age was modelled using generalised additive mixed models (GAMMs), with the muscle trait as dependent variable. The models included smooth functions of age to allow complex ageing trajectories to be inferred (Supp. Info. 2). Sexual dimorphism in ageing trajectories was assessed by comparing pairs of GAMMs [127]: the first with a single ('global') age term that could not vary by sex, and the second with sex-specific splines. GAMMs were fitted using the *bam* function of the 'mgcv' R package, with the fast restricted maximum likelihood (fREML) method utilised for smoothing parameter estimation [128].

Longitudinal models were fitted using data from two assessments per participant, which were explicitly linked within the models by including a random intercept for each participant. Random intercepts were used instead of random slopes, as the latter could not be estimated with only two datapoints per participant.

Sex, self-reported ethnicity, and season of attendance were included as fixed effect covariates in all models. UKB Assessment Centres and individual measurement devices were included in all models as random effects, reflecting the fact that they represented realised instances of many possible Assessment Centres/devices. Standing height^2 and body mass$^{2/3}$ were also included in some models, to control for the potential contribution to muscle trait differences by body size under the principle of allometric scaling [e.g. 35,70].

To determine formally whether males and females exhibit distinct ageing trajectories of muscle traits, goodness-of-fit of global and sex-specific models were compared using the Akaike Information Criterion (ΔAIC), with ΔAICs ≥8 interpreted as sufficiently strong evidence of a difference in model quality to infer a better fit to the observed data [129].

A similar approach was used to analyse the cross-sectional Menopause Participants dataset, but with 'group'—e.g. 'premenopausal females', 'postmenopausal females', and 'males'—replacing sex in the GAMMs. We fitted HGS, MM, and MQ models for each two-group combination in turn, and again interpreted low support for a global model relative to a model with group-specific trajectories as evidence of dimorphism in muscle ageing trajectories.

This approach was deemed preferable to including three groups within a single model, because inclusion of group-specific splines could potentially improve fit in such a model, even if trajectories were identical for two of the three groups.

The model-fitted values illustrated in the Figures were estimated using the *predict.bam* function of the 'mgcv' R package, and unless otherwise specified include sex- or group-specific splines, based on a White participant attending their UK Biobank assessment(s) during Spring. White and Spring were the most frequent factor levels for 'ethnicity' and 'season of attendance' within the Cross-sectional Participants, and so for consistency were used for all Figures; selecting other values for these variables would shift male and female curves by an equal amount without changing the age trajectory. The King's Computational Research, Engineering And Technology Environment (CREATE) [130] was used to generate some of the models included in this study. Full details of models and model comparisons are provided in Supp. Info. 2.

Data visualisation

All data were visualised using the *ggplot* function of the 'ggplot2' R package [131] except Extended Data 8, which was created with the *hist* base R function. Unless otherwise indicated, model-fitted values are from models including sex- or group-specific splines (e.g. HGS model $1_{\text{SEX-SPECIFIC}}$ [Supp. Info. 2] in Fig. 1a and Table 2).

Acknowledgements

We wish to thank Aleksandar Dimitrov, Neelanksha Garg, Alix Hughes, Marc Österdahl, and Richard Powell. This research has been conducted using the UK Biobank Resource under Application Number 67736. SMH is a UKRI/Medical Research Council Scientist with MRC Programme Grant MR/N021231/1 support.

References

1 Falcon, L. J. & Harris-Love, M. O. Sarcopenia and the new ICD-10-CM code: screening, staging, and diagnosis considerations. *Federal Practitioner* **34**, 24 (2017).

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

3 Janssen, I., Heymsfield, S. B. & Ross, R. Low relative skeletal muscle mass (sarcopenia) in older persons is associated with functional impairment and physical disability. *Journal of the American Geriatrics Society* **50**, 889-896 (2002).

4 Rosenberg, I. H. Sarcopenia: origins and clinical relevance. *The Journal of nutrition* **127**, 990S-991S (1997).

5 Beaudart, C. *et al.* Validation of the SarQoL®, a specific health-related quality of life questionnaire for Sarcopenia. *Journal of cachexia, sarcopenia and muscle* **8**, 238-244 (2017).

6 Han, A., Bokshan, S. L., Marcaccio, S. E., DePasse, J. M. & Daniels, A. H. Diagnostic criteria and clinical outcomes in sarcopenia research: a literature review. *Journal of clinical medicine* **7**, 70 (2018).

7 Wu, M. *et al.* Associations of muscle mass, strength, and quality with all-cause mortality in China: a population-based cohort study. *Chinese medical journal* **135**, 1358-1368 (2022).

8 Petermann-Rocha, F. *et al.* Global prevalence of sarcopenia and severe sarcopenia: a systematic review and meta-analysis. *Journal of cachexia, sarcopenia and muscle* **13**, 86-99 (2021).

9 Shafiee, G. *et al.* Prevalence of sarcopenia in the world: a systematic review and meta-analysis of general population studies. *Journal of Diabetes & Metabolic Disorders* **16**, 1-10 (2017).

10 Kvien, T. K. *et al.* The prevalence and severity of rheumatoid arthritis in Oslo. *Scandinavian journal of rheumatology* **26**, 412-418 (1997).

11 Rasch, E. K., Hirsch, R., Paulose-Ram, R. & Hochberg, M. C. Prevalence of rheumatoid arthritis in persons 60 years of age and older in the United States: effect of different methods of case classification. *Arthritis & Rheumatism: Official Journal of the American College of Rheumatology* **48**, 917-926 (2003).

12 Cigolle, C. T., Langa, K. M., Kabeto, M. U., Tian, Z. & Blaum, C. S. Geriatric conditions and disability: the Health and Retirement Study. *Annals of internal medicine* **147**, 156-164 (2007).

13 Atti, A. R. *et al.* Cognitive impairment after age 60: clinical and social correlates in the "Faenza Project". *Journal of Alzheimer's Disease* **21**, 1325-1334 (2010).

14 Bloom, D. E., Mitgang, E. & Osher, B. Demography of global ageing. (2016).

15 Tan, L.-J., Liu, S.-L., Lei, S.-F., Papasian, C. J. & Deng, H.-W. Molecular genetic studies of gene identification for sarcopenia. *Human genetics* **131**, 1-31 (2012).

16 Zempo, H. *et al.* Heritability estimates of muscle strength-related phenotypes: A systematic review and meta-analysis. *Scandinavian journal of medicine & science in sports* **27**, 1537-1546 (2017).

17 You, J.-Y. *et al.* Heritability of muscle mass in Korean parent–offspring pairs in the Fifth Korean National Health and Nutrition Examination Survey (KNHANES V). *Maturitas* **114**, 67-72 (2018).

18 Petermann-Rocha, F. *et al.* Factors associated with sarcopenia: A cross-sectional analysis using UK Biobank. *Maturitas* **133**, 60-67 (2020).

19 Choe, H. J. *et al.* Gender differences in risk factors for the 2 year development of sarcopenia in community-dwelling older adults. *Journal of Cachexia, Sarcopenia and Muscle* (2022).

20 Ni Lochlainn, M., Bowyer, R. C., Welch, A. A., Whelan, K. & Steves, C. J. Higher dietary protein intake is associated with sarcopenia in older British twins. *Age and Ageing* **52**, afad018 (2023).

21 Cheval, B. *et al.* Association of early-and adult-life socioeconomic circumstances with muscle strength in older age. *Age and Ageing* **47**, 398-407 (2018).

22 Liu, X. *et al.* Ethnic groups differences in the prevalence of sarcopenia using the AWGS criteria. *The journal of nutrition, health & aging* **24**, 665-671 (2020).

23 Sipilä, S. *et al.* Muscle and bone mass in middle-aged women: role of menopausal status and physical activity. *Journal of cachexia, sarcopenia and muscle* **11**, 698-709 (2020).

24 Amara, C. E. *et al.* Allometric scaling of strength in an independently living population age 55–86 years. *American journal of human biology* **15**, 48-60 (2003).

25 Wells, J. C. Sexual dimorphism of body composition. *Best practice & research Clinical endocrinology & metabolism* **21**, 415-430 (2007).

26 Geisler, C., Pourhassan, M., Braun, W., Schweitzer, L. & Müller, M. J. The prediction of total skeletal muscle mass in a Caucasian population–comparison of Magnetic resonance imaging (MRI) and Dual-energy X-ray absorptiometry (DXA). *Clinical Physiology and Functional Imaging* **37**, 168-172 (2017).

27 Heymsfield, S. B. *et al.* Phenotypic differences between people varying in muscularity. *Journal of cachexia, sarcopenia and muscle* **13**, 1100-1112 (2022).

28 Miall, W., Ashcroft, M., Lovell, H. & Moore, F. A longitudinal study of the decline of adult height with age in two Welsh communities. *Human Biology*, 445-454 (1967).

29 Cline, M. G., Meredith, K. E., Boyer, J. T. & Burrows, B. Decline of height with age in adults in a general population sample: estimating maximum height and distinguishing birth cohort effects from actual loss of stature with aging. *Human biology*, 415-425 (1989).

30 Kyle, U. G., Genton, L., Slosman, D. O. & Pichard, C. Fat-free and fat mass percentiles in 5225 healthy subjects aged 15 to 98 years. *Nutrition* **17**, 534-541 (2001).

31 Buffa, R., Floris, G. U., Putzu, P. F. & Marini, E. Body composition variations in ageing. *Collegium antropologicum* **35**, 259-265 (2011).

32 Kyle, U. G., Schutz, Y., Dupertuis, Y. M. & Pichard, C. Body composition interpretation: contributions of the fat-free mass index and the body fat mass index. *Nutrition* **19**, 597-604 (2003).

33 Sözen, T., Özışık, L. & Başaran, N. Ç. An overview and management of osteoporosis. *European journal of rheumatology* **4**, 46 (2017).

34 Ross, W., Marfell-Jones, M., MacDougall, J., Wenger, H. & Green, H. Physiological testing of the high performance athlete. *Kinanthropometry Champaign IL: Human Kinetics Books*, 223-308 (1991).

35 Jaric, S. Muscle strength testing. *Sports Medicine* **32**, 615-631 (2002).

36 Barbat-Artigas, S., Rolland, Y., Zamboni, M. & Aubertin-Leheudre, M. How to assess functional status: a new muscle quality index. *The journal of nutrition, health & aging* **16**, 67-77 (2012).

37 Lindle, R. *et al.* Age and gender comparisons of muscle strength in 654 women and men aged 20–93 yr. *Journal of applied physiology* **83**, 1581-1587 (1997).

38 Häkkinen, K. *et al.* Bilateral and unilateral neuromuscular function and muscle cross-sectional area in middle-aged and elderly men and women. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences* **51**, B21-B29 (1996).

39 Frontera, W. R. *et al.* Skeletal muscle fiber quality in older men and women. *American Journal of Physiology-Cell Physiology* **279**, C611-C618 (2000).

40 Doherty, T. J. The influence of aging and sex on skeletal muscle mass and strength. *Current Opinion in Clinical Nutrition & Metabolic Care* **4**, 503-508 (2001).

41 Dodds, R. M. *et al.* Grip strength across the life course: normative data from twelve British studies. *PloS one* **9**, e113637 (2014).

42 Lynch, N. *et al.* Muscle quality. I. Age-associated differences between arm and leg muscle groups. *Journal of applied physiology* **86**, 188-194 (1999).

43 Goodpaster, B. H. *et al.* The loss of skeletal muscle strength, mass, and quality in older adults: the health, aging and body composition study. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences* **61**, 1059-1064 (2006).

44 Dey, D. K., Bosaeus, I., Lissner, L. & Steen, B. Changes in body composition and its relation to muscle strength in 75-year-old men and women: a 5-year prospective follow-up study of the NORA cohort in Göteborg, Sweden. *Nutrition* **25**, 613-619 (2009).

45 Jubrias, S., Odderson, I. R., Esselman, P. C. & Conley, K. E. Decline in isokinetic force with age: muscle cross-sectional area and specific force. *Pflügers Archiv* **434**, 246-253 (1997).

46 Roberts, B. M. *et al.* Human neuromuscular aging: Sex differences revealed at the myocellular level. *Experimental gerontology* **106**, 116-124 (2018).

47 Linge, J., Heymsfield, S. B. & Dahlqvist Leinhard, O. On the definition of sarcopenia in the presence of aging and obesity—initial results from UK Biobank. *The Journals of Gerontology: Series A* **75**, 1309-1316 (2020).

48 Baumgartner, R. N. *et al.* Epidemiology of sarcopenia among the elderly in New Mexico. *American journal of epidemiology* **147**, 755-763 (1998).

49 Melton, L. J. I. *et al.* Epidemiology of sarcopenia. *Journal of the American Geriatrics Society* **48**, 625-630 (2000).

50 Gallagher, D. *et al.* Weight stability masks sarcopenia in elderly men and women. *American Journal of Physiology-Endocrinology And Metabolism* **279**, E366-E375 (2000).

51 Gallagher, D. *et al.* Appendicular skeletal muscle mass: effects of age, gender, and ethnicity. *Journal of applied physiology* **83**, 229-239 (1997).

52 Janssen, I., Heymsfield, S. B., Wang, Z. & Ross, R. Skeletal muscle mass and distribution in 468 men and women aged 18–88 yr. *Journal of applied physiology* (2000).

53 Zamboni, M. *et al.* Body composition changes in stable-weight elderly subjects: the effect of sex. *Aging clinical and experimental research* **15**, 321-327 (2003).

54 Liu, L. K. *et al.* Age-related skeletal muscle mass loss and physical performance in Taiwan: Implications to diagnostic strategy of sarcopenia in Asia. *Geriatrics & gerontology international* **13**, 964-971 (2013).

55 Shimokata, H., Ando, F., Yuki, A. & Otsuka, R. Age-related changes in skeletal muscle mass among community-dwelling Japanese: A 12-year longitudinal study. *Geriatrics & gerontology international* **14**, 85-92 (2014).

56 Lee, M. M., Jebb, S. A., Oke, J. & Piernas, C. Reference values for skeletal muscle mass and fat mass measured by bioelectrical impedance in 390 565 UK adults. *Journal of cachexia, sarcopenia and muscle* **11**, 487-496 (2020).

57 McGregor, R. A., Cameron-Smith, D. & Poppitt, S. D. It is not just muscle mass: a review of muscle quality, composition and metabolism during ageing as determinants of muscle function and mobility in later life. *Longevity & healthspan* **3**, 1-8 (2014).

58 Fabbri, E. *et al.* Early body composition, but not body mass, is associated with future accelerated decline in muscle quality. *Journal of cachexia, sarcopenia and muscle* **8**, 490-499 (2017).

59 Jones, E. J., Bishop, P. A., Woods, A. K. & Green, J. M. Cross-sectional area and muscular strength: a brief review. *Sports Medicine* **38**, 987-994 (2008).

60 Sudlow, C. *et al.* UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS medicine* **12**, e1001779 (2015).

61 Dodds, R. M., Granic, A., Robinson, S. M. & Sayer, A. A. Sarcopenia, long-term conditions, and multimorbidity: findings from UK Biobank participants. *Journal of cachexia, sarcopenia and muscle* **11**, 62-68 (2020).

62 Pei, Y.-F. *et al.* The genetic architecture of appendicular lean mass characterized by association analysis in the UK Biobank study. *Communications biology* **3**, 608 (2020).

63 Petermann-Rocha, F., Gray, S. R., Pell, J. P., Celis-Morales, C. & Ho, F. K. Biomarkers profile of people with sarcopenia: a cross-sectional analysis from UK biobank. *Journal of the American Medical Directors Association* **21**, 2017. e2011-2017. e2019 (2020).

64 Jones, G. *et al.* Genome-wide meta-analysis of muscle weakness identifies 15 susceptibility loci in older men and women. *Nature communications* **12**, 1-11 (2021).

65 Curtis, A. *et al.* Effect of dietary patterns on malnutrition, low muscle mass and sarcopenia in cancer: a cross-sectional analysis of the UK Biobank. *Proceedings of the Nutrition Society* **82**, E108 (2023).

66 Kiss, N. *et al.* Low muscle mass, malnutrition, sarcopenia, and associations with survival in adults with cancer in the UK Biobank cohort. *Journal of Cachexia, Sarcopenia and Muscle* **14**, 1775-1788 (2023).

67 Semenova, E. A., Pranckevičienė, E., Bondareva, E. A., Gabdrakhmanova, L. J. & Ahmetov, I. I. Identification and characterization of genomic predictors of sarcopenia and sarcopenic obesity using UK biobank data. *Nutrients* **15**, 758 (2023).

68 Jauffret, C. *et al.* Association between sarcopenia and risk of major adverse cardiac and cerebrovascular events-UK Biobank database. *Journal of the American Geriatrics Society* **72**, 693-706 (2024).

69 Cruz-Jentoft, A. J., Gonzalez, M. C. & Prado, C. M. Sarcopenia≠ low muscle mass. *European Geriatric Medicine* **14**, 225-228 (2023).
70 Nevill, A. M., Tomkinson, G. R., Lang, J. J., Wutz, W. & Myers, T. D. How should adult handgrip strength be normalized? Allometry reveals new insights and associated reference curves. *Medicine & Science in Sports & Exercise* (2021).
71 Chan, S., Gomes, A. & Singh, R. S. Is menopause still evolving? Evidence from a longitudinal study of multiethnic populations and its relevance to women's health. *BMC Women's Health* **20**, 1-15 (2020).
72 Yasuda, N., Glover, E. I., Phillips, S. M., Isfort, R. J. & Tarnopolsky, M. A. Sex-based differences in skeletal muscle function and morphology with short-term limb immobilization. *Journal of applied physiology* **99**, 1085-1092 (2005).
73 Loomba-Albrecht, L. A. & Styne, D. M. Effect of puberty on body composition. *Current Opinion in Endocrinology, Diabetes and Obesity* **16**, 10-15 (2009).
74 Handelsman, D. J., Hirschberg, A. L. & Bermon, S. Circulating testosterone as the hormonal basis of sex differences in athletic performance. *Endocrine reviews* **39**, 803-829 (2018).
75 Handelsman, D. J., Sikaris, K. & Ly, L. P. Estimating age-specific trends in circulating testosterone and sex hormone-binding globulin in males and females across the lifespan. *Annals of clinical biochemistry* **53**, 377-384 (2016).
76 Burger, H. G., Dudley, E. C., Cui, J., Dennerstein, L. & Hopper, J. L. A prospective longitudinal study of serum testosterone, dehydroepiandrosterone sulfate, and sex hormone-binding globulin levels through the menopause transition. *The Journal of Clinical Endocrinology & Metabolism* **85**, 2832-2838 (2000).
77 Nam, Y.-S., Lee, G., Yun, J. M. & Cho, B. Testosterone replacement, muscle strength, and physical function. *The world journal of men's health* **36**, 110-122 (2018).
78 Overend, T., Cunningham, D., Paterson, D. & Lefcoe, M. Thigh composition in young and elderly men determined by computed tomography. *Clinical Physiology* **12**, 629-640 (1992).
79 Dai, L. *et al.* Defining reference values for body composition indices by magnetic resonance imaging in UK Biobank. *Journal of Cachexia, Sarcopenia and Muscle* **14**, 992-1002 (2023).
80 Auerbach, B. M. & Sylvester, A. D. Allometry and apparent paradoxes in human limb proportions: implications for scaling factors. *American journal of physical anthropology* **144**, 382-391 (2011).
81 Monson, T. A., Brasil, M. F. & Hlusko, L. J. Allometric variation in modern humans and the relationship between body proportions and elite athletic success. *Journal of Anthropology of Sport and Physical Education* **2**, 3 (2018).
82 Maughan, R., Watson, J. & Weir, J. Muscle strength and cross-sectional area in man: a comparison of strength-trained and untrained subjects. *British journal of sports medicine* **18**, 149-157 (1984).
83 Rockenfeller, R., Günther, M., Clemente, C. J. & Dick, T. J. Rethinking the physiological cross-sectional area of skeletal muscle reveals the mechanical advantage of pennation. *Royal Society Open Science* **11**, 240037 (2024).
84 Kubo, K. *et al.* Muscle architectural characteristics in young and elderly men and women. *International journal of sports medicine* **24**, 125-130 (2003).
85 Trevino, M. *et al.* Influence of pennation angle and muscle thickness on mechanomyographic amplitude–torque relationships and sex-related differences in the vastus lateralis. *Journal of Functional Morphology and Kinesiology* **8**, 53 (2023).
86 Martin-Rodriguez, S., Gonzalez-Henriquez, J. J., Diaz-Conde, J. C., Calbet, J. A. & Sanchis-Moysi, J. The relationship between muscle thickness and pennation angle is

mediated by fascicle length in the muscles of the lower extremities. *Scientific Reports* **14**, 14847 (2024).

87 Alegre, L., Lara, A., Elvira, J. & Aguado, X. Muscle morphology and jump performance: gender and intermuscular variability. *Journal of sports medicine and physical fitness* **49**, 320 (2009).

88 Manal, K., Roberts, D. P. & Buchanan, T. S. Optimal pennation angle of the primary ankle plantar and dorsiflexors: variations with sex, contraction intensity, and limb. *Journal of applied biomechanics* **22**, 255-263 (2006).

89 Chow, R. *et al.* Sonographic studies of human soleus and gastrocnemius muscle architecture: gender variability. *European journal of applied physiology* **82**, 236-244 (2000).

90 Lieber, R. L. *Skeletal muscle structure, function, and plasticity*. (Lippincott Williams & Wilkins, 2002).

91 Evans, W. J. & Lexell, J. Human aging, muscle mass, and fiber type composition. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences* **50**, 11-16 (1995).

92 Lexell, J., Taylor, C. C. & Sjöström, M. What is the cause of the ageing atrophy?: Total number, size and proportion of different fiber types studied in whole vastus lateralis muscle from 15-to 83-year-old men. *Journal of the neurological sciences* **84**, 275-294 (1988).

93 Smith, G. I. & Mittendorfer, B. Sexual dimorphism in skeletal muscle protein turnover. *Journal of applied physiology* **120**, 674-682 (2016).

94 Raue, U., Slivka, D., Jemiolo, B., Hollon, C. & Trappe, S. Proteolytic gene expression differs at rest and after resistance exercise between young and old women. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences* **62**, 1407-1412 (2007).

95 Gim, J.-A., Lee, S., Kim, S. C., Baek, K.-W. & Yoo, J.-I. Demographic and genome wide association analyses according to muscle mass using data of the korean genome and epidemiology study. *Journal of Korean Medical Science* **37** (2022).

96 Ontan, M. S., Dokuzlar, O., Ates Bulut, E., Soysal, P. & Isik, A. T. The relationship between osteoporosis and sarcopenia, according to EWGSOP-2 criteria, in outpatient elderly. *Journal of bone and mineral metabolism* **39**, 684-692 (2021).

97 Critchlow, A. J., Hiam, D., Williams, R., Scott, D. & Lamon, S. The role of estrogen in female skeletal muscle aging: A systematic review. *Maturitas*, 107844 (2023).

98 Phillips, S., Rook, K., Siddle, N., Bruce, S. & Woledge, R. Muscle weakness in women occurs at an earlier age than in men, but strength is preserved by hormone replacement therapy. *Clinical science* **84**, 95-98 (1993).

99 Ronkainen, P. H. *et al.* Postmenopausal hormone replacement therapy modifies skeletal muscle composition and function: a study with monozygotic twin pairs. *Journal of applied physiology* **107**, 25-33 (2009).

100 Fry, A. *et al.* Comparison of sociodemographic and health-related characteristics of UK Biobank participants with those of the general population. *American journal of epidemiology* **186**, 1026-1034 (2017).

101 Åsvold, B. O. *et al.* Cohort profile update: the HUNT study, Norway. *International journal of epidemiology* **52**, e80-e91 (2023).

102 Office for National Statistics. *ONS website, statistical bulletin, Ethnic group, England and Wales: Census 2021*, <https://www.ons.gov.uk/peoplepopulationandcommunity/culturalidentity/ethnicity/bulletins/ethnicgroupenglandandwales/census2021> (2022).

103 LeBlanc, J. *et al.* Incorporating Equity, Diversity and Inclusion in a Canadian Biobank. *Biopreservation and Biobanking* (2024).

104 Rudnev, S. *et al.* Comparison of bioimpedance body composition in young adults in the Russian Children's Study. *Clinical nutrition ESPEN* **35**, 153-161 (2020).

105 UK Biobank. *UK Biobank. Body Composition Measurement Version 1.0*, <https://biobank.ndph.ox.ac.uk/ukb/ukb/docs/body_composition.pdf> (2011).

106 Soendenbroe, C. *et al.* Preserved stem cell content and innervation profile of elderly human skeletal muscle with lifelong recreational exercise. *The Journal of Physiology* **600**, 1969-1989 (2022).

107 R: A Language and Environment for Statistical Computing v. 4.2.1 (R Foundation for Statistical Computing, Vienna, Austria, 2022).

108 Grolemund, G. & Wickham, H. Dates and times made easy with lubridate. *Journal of statistical software* **40**, 1-25 (2011).

109 Office for National Statistics. *National identity, ethnic group, language and religion question development for Census 2021*, <https://www.ons.gov.uk/census/censustransformationprogramme/questiondevelopment/nationalidentityethnicgrouplanguageandreligionquestiondevelopmentforcensus2021> (2021).

110 Adams, J., Ryan, V. & White, M. How accurate are Townsend Deprivation Scores as predictors of self-reported health? A comparison with individual level data. *Journal of public health* **27**, 101-106 (2005).

111 Lafayette Instrument Company. JAMAR Hydrolic Hand Dynamometer user instructions (2004).

112 Stahn, A., Terblanche, E. & Gunga, H.-C. in *Handbook of anthropometry* 49-90 (Springer, 2012).

113 Laskey, M. A. Dual-energy X-ray absorptiometry and body composition. *Nutrition* **12**, 45-51 (1996).

114 Prado, C. M. & Heymsfield, S. B. Lean tissue imaging: a new era for nutritional assessment and intervention. *Journal of Parenteral and Enteral Nutrition* **38**, 940-953 (2014).

115 Visser, M. *et al.* One-and two-year change in body composition as measured by DXA in a population-based cohort of older men and women. *Journal of applied physiology* **94**, 2368-2374 (2003).

116 Borga, M. *et al.* Advanced body composition assessment: from body mass index to body composition profiling. *Journal of Investigative Medicine* **66**, 1-9 (2018).

117 Tewari, N., Awad, S., Macdonald, I. A. & Lobo, D. N. A comparison of three methods to assess body composition. *Nutrition* **47**, 1-5 (2018).

118 Powell, R. *et al.* Development and validation of total and regional body composition prediction equations from anthropometry and single frequency segmental bioelectrical impedance with DEXA. *medRxiv* (2020).

119 Rospars, J.-P. & Meyer-Vernet, N. Force per cross-sectional area from molecules to muscles: a general property of biological motors. *Royal Society Open Science* **3**, 160313 (2016).

120 Elia, M. The 'MUST' report. *Nutritional screening of adults: a multidisciplinary responsibility* (2003).

121 Madden, A. M. *et al.* Improved prediction equations for estimating height in adults from ethnically diverse backgrounds. *Clinical Nutrition* **39**, 1454-1463 (2020).

122 Ho, D., Imai, K., King, G. & Stuart, E. A. MatchIt: Nonparametric Preprocessing for Parametric Causal Inference. *Journal of Statistical Software* **42**, 1–28, doi:https://doi.org/10.18637/jss.v042.i08 (2011).

123 Greifer, N. *Assessing Balance*, <https://cran.r-project.org/web/packages/MatchIt/vignettes/assessing-balance.html> (2023).

124 Sellke, T., Bayarri, M. & Berger, J. O. Calibration of ρ values for testing precise null hypotheses. *The American Statistician* **55**, 62-71 (2001).
125 Colquhoun, D. The false positive risk: a proposal concerning what to do about p-values. *The American Statistician* **73**, 192-201 (2019).
126 pcal: Calibration of P-values for Point Null Hypothesis Testing (2020).
127 Lin, X. & Zhang, D. Inference in generalized additive mixed models by using smoothing splines. *Journal of the Royal Statistical Society Series B: Statistical Methodology* **61**, 381-400 (1999).
128 Wood, S. & Wood, M. S. Package 'mgcv'. *R package version* **1**, 729 (2015).
129 Burnham, K. P., Anderson, D. R. & Huyvaert, K. P. AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons. *Behavioral ecology and sociobiology* **65**, 23-35 (2011).
130 King's College London. *King's Computational Research, Engineering and Technology Environment (CREATE)*, <https://doi.org/10.18742/rnvf-m076> (2024).
131 Wickham, H. ggplot2. *Wiley interdisciplinary reviews: computational statistics* **3**, 180-185 (2011).

Tables

	Cross-sectional Participants			Matched Cross-sectional Participants			Longitudinal Participants			Matched Longitudinal Participants			Menopause Participants			
	All	**Female**	**Male**	**All**	**Female**	**Male**	**All**	**Female**	**Male**	**All**	**Female**	**Male**	**All**	**Female (Pre-)**	**Female (Post-)**	**Male**
N	432,809	237,015	195,794	123,366	61,683	61,683	45,629	23,448	22,181	10,870	5,435	5,435	365,895	53,840	116,261	195,794
Age (yr)	57.04 ± 8.09	56.93 ± 7.99	57.19 ± 8.22	56.94 ± 8.15	56.94 ± 8.15	56.94 ± 8.15	55.67 ± 7.54	54.98 ± 7.38	56.39 ± 7.64	55.54 ± 7.46	55.54 ± 7.46	55.55 ± 7.46	56.74 ± 8.21	46.47 ± 3.8	60.73 ± 5.17	57.19 ± 8.22
Interval between assessments (yr)	-	-	-	-	-	-	8.92 ± 1.76	8.92 ± 1.75	8.91 ± 1.76	8.93 ± 1.75	8.92 ± 1.76	8.94 ± 1.75	-	-	-	-
Standing height (cm)	168.3 ± 9.3	162.4 ± 6.3	175.6 ± 6.9	168.9 ± 4.8	168.8 ± 4.7	168.9 ± 4.8	169.7 ± 9.2	163.3 ± 6.2	176.5 ± 6.6	169.7 ± 4.7	169.6 ± 4.6	169.7 ± 4.7	169.5 ± 9.3	164 ± 6.4	161.9 ± 6.2	175.6 ± 6.9
Body mass (kg)	78.2 ± 16.0	71.6 ± 14.1	86.1 ± 14.4	78.0 ± 12.6	78.0 ± 12.6	78.1 ± 12.6	77.0 ± 15.0	69.7 ± 12.8	84.8 ± 13.1	75.8 ± 10.0	75.7 ± 10.0	75.8 ± 10.0	79.1 ± 16.1	71.3 ± 14.8	71.0± 13.5	86.1 ± 14.4
Ethnicity																
Asian (% of group)	10,095 (2.3)	4,993 (2.1)	5,102 (2.6)	3,630 (3)	323 (0.5)	3,307 (5.3)	625 (1.4)	262 (1.1)	363 (1.6)	238 (2.2)	19 (0.4)	219 (4)	8,851 (2.4)	1,650 (3.1)	2,099 (1.8)	5,102 (2.6)
Black (% of group)	7,215 (1.7)	4,159 (1.8)	3,056 (1.6)	2,361 (1.9)	1,085 (1.8)	1,276 (2.1)	296 (0.6)	161 (0.7)	135 (0.6)	82 (0.8)	39 (0.7)	43 (0.8)	5,936 (1.7)	1,557 (2.9)	1,323 (1.2)	3,056 (1.6)
Other inc. Mixed and Unknown (% of group)	8,262 (1.9)	4,628 (1.9)	3,634 (1.8)	2,483 (2)	833 (1.3)	1,650 (2.7)	548 (1.2)	311 (1.3)	237 (1.1)	133 (1.2)	50 (0.9)	83 (1.5)	7,022 (1.9)	1,613 (3)	1,775 (1.5)	3,634 (1.8)
White (% of group)	407,237 (94.1)	223,235 (94.2)	184,002 (94)	114,892 (93.1)	59,442 (96.4)	55,450 (89.9)	44,160 (96.8)	22,714 (96.9)	21,446 (96.7)	10,417 (95.8)	5,327 (98)	5,090 (93.7)	344,046 (94)	49,020 (91)	111,064 (95.5)	184,002 (94)

Table 1. Characteristics of the five main UK Biobank participant groups included in this study. Note that the Matched Cross-sectional Participants and Matched Longitudinal Participants are subsets of the Cross-sectional Participants and Longitudinal Participants, respectively, and that some participants are therefore included in both an unmatched and a matched group. In contrast, cross-sectional and longitudinal participant groups are completely independent, with no shared participants. Values for the longitudinal participant groups are from baseline assessment, with the exception of 'Interval between assessments'. Menopause Participants were drawn from the Cross-sectional Participants group. 'Pre-' and 'Post-' refer to self-declared pre- and postmenopausal status, respectively. Data are either simple counts ('N' and 'Ethnicity'), or mean ± the standard deviation.

	Cross-sectional Participants				Matched Cross-sectional Participants			
	Females		Males		Females		Males	
Hand grip strength	**Obs.**	**Fitt.**	**Obs.**	**Fitt.**	**Obs.**	**Fitt.**	**Obs.**	**Fitt.**
All participants [kgf] (N)	23.7 ± 0.01 (202,028)	-	39.8 ± 0.02 (166,308)	-	25.4 ± 0.03 (52,723)	-	37.1 ± 0.04 (52,097)	-
40 yr [kgf] (N)	27.4 ± 0.14 (1,969)	27.5 ± 0.24	43.7 ± 0.20 (1,934)	44.1 ± 0.25	29.1 ± 0.25 (571)	29.8 ± 0.21	40.6 ± 0.35 (594)	41.6 ± 0.26
69 yr [kgf] (N)	20.6 ± 0.07 (6,251)	20.4 ± 0.22	35.4 ± 0.10 (6,080)	35.4 ± 0.22	22.1 ± 0.13 (1,775)	21.8 ± 0.20	32.8 ± 0.18 (1,755)	33.1 ± 0.22
Difference 40–69 yr [kgf]	-6.82	-7.10	-8.24	-8.70	-7.03	-8.00	-7.76	-8.50
Decline 40–69 yr [kgf·yr^{-1}]	-0.24	-0.24	-0.28	-0.30	-0.24	-0.28	-0.27	-0.29
Difference 40–69 yr [%]	-24.9	-25.8	-18.9	-19.7	-24.1	-26.80	-19.10	-20.40
Decline 40–69 yr [%·yr^{-1}]	-0.98	-1.02	-0.72	-0.75	-0.95	-1.07	-0.73	-0.79
	Longitudinal Participants				**Matched Longitudinal Participants**			
	Females		**Males**		**Females**		**Males**	
Hand grip strength	**Obs.**	**Fitt.**	**Obs.**	**Fitt.**	**Obs.**	**Fitt.**	**Obs.**	**Fitt.**
40 yr [kgf] (N)	28.8 ± 0.39 (211)	28.3 ± 0.26	45.4 ± 0.74 (156)	45.5 ± 0.29	29.5 ± 0.98 (42)	30.4 ± 0.36	40.4 ± 1.54 (38)	42.7 ± 0.39
80 yr [kgf] (N)	19.3 ± 0.64 (70)	18.6 ± 0.29	31.6 ± 0.69 (119)	31.0 ± 0.3	18.6 ± 1.47 (16)	19.6 ± 0.4	29.2 ± 1.74 (18)	29.3 ± 0.44
Difference 40–80 yr [kgf]	-9.45	-9.7	-13.8	-14.5	-10.9	-10.8	-11.3	-13.4
Decline 40–80 yr [kgf·yr^{-1}]	-0.24	-0.24	-0.35	-0.36	-0.27	-0.27	-0.28	-0.34
Difference 40–80 yr [%]	-32.8	-34.3	-30.5	-31.9	-36.9	-35.5	-27.9	-31.4
Decline 40–80 yr [%·yr^{-1}]	-0.99	-1.04	-0.9	-0.95	-1.14	-1.09	-0.81	-0.94

Table 2. Modest sexual dimorphism in age-related strength decline. HGS summary data for each participant group in the UK Biobank cohort, showing mean for the entire cohort, at specific ages, and the differences and rates of decline in both absolute force and percentage terms. Observed ('Obs.'; mean ± SEM [N]) and GAMM-fitted ('Fitt.'; fitted value ± standard error of the fitted value) values are given. Note that fitted values assume a White participant assessed during Spring (Online Methods), and that N varies in the matched cohort single year groups due to the matching criterion for age at baseline being ± 6 months.

	Cross-sectional Participants				Matched Cross-sectional Participants			
	Females		Males		Females		Males	
Muscle Mass	**Obs.**	**Fitt.**	**Obs.**	**Fitt.**	**Obs.**	**Fitt.**	**Obs.**	**Fitt.**
All participants [kg] (N)	2.10 ± 0.001 (237,015)	-	3.53 ± 0.001 (195,794)	-	2.29 ± 0.001 (61,683)	-	3.26 ± 0.002 (61,683)	-
40 yr [kg] (N)	2.13 ± 0.01 (2,433)	2.14 ± 0.01	3.69 ± 0.01 (2,346)	3.71 ± 0.01	2.32 ± 0.01 (687)	2.33 ± 0.01	3.41 ± 0.02 (723)	3.45 ± 0.01
69 yr [kg] (N)	2.04 ± 0.003 (7,310)	2.03 ± 0.01	3.32 ± 0.01 (7,090)	3.32 ± 0.01	2.22 ± 0.01 (2,067)	2.22 ± 0.01	3.05 ± 0.01 (2,065)	3.07 ± 0.01
Difference 40–69 yr [kg]	-0.09	-0.11	-0.37	-0.39	-0.10	-0.11	-0.37	-0.38
Decline 40–69 yr [kg·yr^{-1}]	-0.003	-0.004	-0.013	-0.013	-0.003	-0.004	-0.013	-0.013
Difference 40–69 yr [%]	-4.27	-5.00	-10.00	-10.60	-4.36	-4.84	-10.75	-11.10
Decline 40–69 yr [%·yr^{-1}]	-0.15	-0.18	-0.36	-0.38	-0.15	-0.17	-0.39	-0.40
	Longitudinal Participants				**Matched Longitudinal Participants**			
	Females		**Males**		**Females**		**Males**	
Muscle Mass	**Obs.**	**Fitt.**	**Obs.**	**Fitt.**	**Obs.**	**Fitt.**	**Obs.**	**Fitt.**
40 yr [kg] (N)	2.10 ± 0.02 (230)	2.11 ± 0.01	3.67 ± 0.03 (170)	3.71 ± 0.01	2.28 ± 0.03 (47)	2.30 ± 0.01	3.37 ± 0.07 (39)	3.41 ± 0.01
80 yr [kg] (N)	1.94 ± 0.03 (84)	1.93 ± 0.01	3.03 ± 0.03 (147)	3.02 ± 0.01	2.06 ± 0.04 (20)	2.08 ± 0.01	2.64 ± 0.08 (19)	2.73 ± 0.02
Difference 40–80 yr [kg]	-0.15	-0.18	-0.64	-0.69	-0.22	-0.21	-0.73	-0.68
Decline 40–80 yr [kg·yr^{-1}]	-0.004	-0.004	-0.016	-0.017	-0.005	-0.005	-0.018	-0.017
Difference 40–80 yr [%]	-7.31	-8.50	-17.50	-18.60	-9.48	-9.23	-21.80	-20.00
Decline 40–80 yr [%·yr^{-1}]	-0.19	-0.22	-0.48	-0.51	-0.25	-0.24	-0.61	-0.56

Table 3. Age-related muscle mass decline is greater in males. MM summary data for each participant group in the UK Biobank cohort, showing mean for the entire cohort, at specific ages, and the differences and rates of decline in both absolute force and percentage terms. Observed ('Obs.'; mean ± SEM [N]) and GAMM-fitted ('Fitt.'; fitted value ± standard error of the fitted value) values are given. Note that fitted values assume a White participant assessed during Spring (Online Methods), and that N varies in the matched cohort single year groups due to the matching criterion for age at baseline being ± 6 months.

	Cross-sectional Participants				Matched Cross-sectional Participants			
	Females		Males		Females		Males	
Muscle quality	**Obs.**	**Fitt.**	**Obs.**	**Fitt.**	**Obs.**	**Fitt.**	**Obs.**	**Fitt.**
All participants [au] (N)	279 ± 0.2 (202,028)	-	310 ± 0.2 (166,308)	-	298 ± 0.3 (52,723)	-	290 ± 0.3 (52,097)	-
40 yr [au] (N)	326 ± 1.7 (1,969)	323 ± 2.9	330 ± 1.5 (1,934)	328 ± 3.0	344 ± 3.0 (571)	347 ± 2.4	308 ± 2.6 (594)	310 ± 2.6
69 yr [au] (N)	246 ± 0.9 (6,251)	242 ± 2.8	290 ± 0.9 (6,080)	287 ± 2.8	264 ± 1.7 (1,775)	260 ± 2.3	270 ± 1.5 (1,755)	270 ± 2.4
Difference 40–69 yr (au)	-80.1	-80.4	-40.2	-40.3	-80.5	-87.0	-38.8	-40.5
Decline 40–69 yr [au·yr^{-1}]	-2.76	2.77	-1.39	-1.39	-2.78	-3.00	-1.34	-1.40
Difference 40–69 yr [%]	-24.6	-24.9	-12.2	-12.3	-23.4	-25.1	-12.6	-13.1
Decline 40–69 [%·yr^{-1}]	-0.97	-0.98	-0.45	-0.45	-0.91	-0.99	-0.46	-0.48
	Longitudinal Participants				**Matched Longitudinal Participants**			
	Females		**Males**		**Females**		**Males**	
Muscle quality	**Obs.**	**Fitt.**	**Obs.**	**Fitt.**	**Obs.**	**Fitt.**	**Obs.**	**Fitt.**
40 yr [au] (N)	350 ± 4.9 (211)	344 ± 2.78	347 ± 5.6 (156)	344 ± 3.0	355 ± 11.1 (42)	365 ± 3.7	313 ± 11.3 (38)	325 ± 3.9
80 yr [au] (N)	247 ± 7.6 (68)	235 ± 3.0	286 ± 6.2 (116)	277 ± 2.9	245 ± 18.4 (16)	246 ± 4.1	283 ± 15.6 (18)	264 ± 4.4
Difference 40–80 yr [au]	-103.0	-109.0	-60.6	-67.0	-110.0	-119.0	-29.8	-60.9
Decline 40–80 yr [au·yr^{-1}]	-2.57	-2.73	-1.51	-1.68	-2.76	-2.96	-0.74	-1.52
Difference 40–80 yr [%]	-29.40	-31.70	-17.50	-19.50	-31.00	-32.5	-9.52	-18.70
Decline 40–80 yr [%·yr^{-1}]	-0.87	-0.95	-0.48	-0.54	-0.92	-0.98	-0.25	-0.52

Table 4. Age-related muscle quality decline is greater in females. MQ summary data for each participant group in the UK Biobank cohort, showing mean for the entire cohort, at specific ages, and the differences and rates of decline in both absolute force and percentage terms. Observed ('Obs.'; mean ± SEM [N]) and GAMM-fitted ('Fitt.'; fitted value ± standard error of the fitted value) values are given. Note that fitted values assume a White participant assessed during Spring (Online Methods), and that N varies in the matched cohort single year groups due to the matching criterion for age at baseline being ± 6 months.

Figures

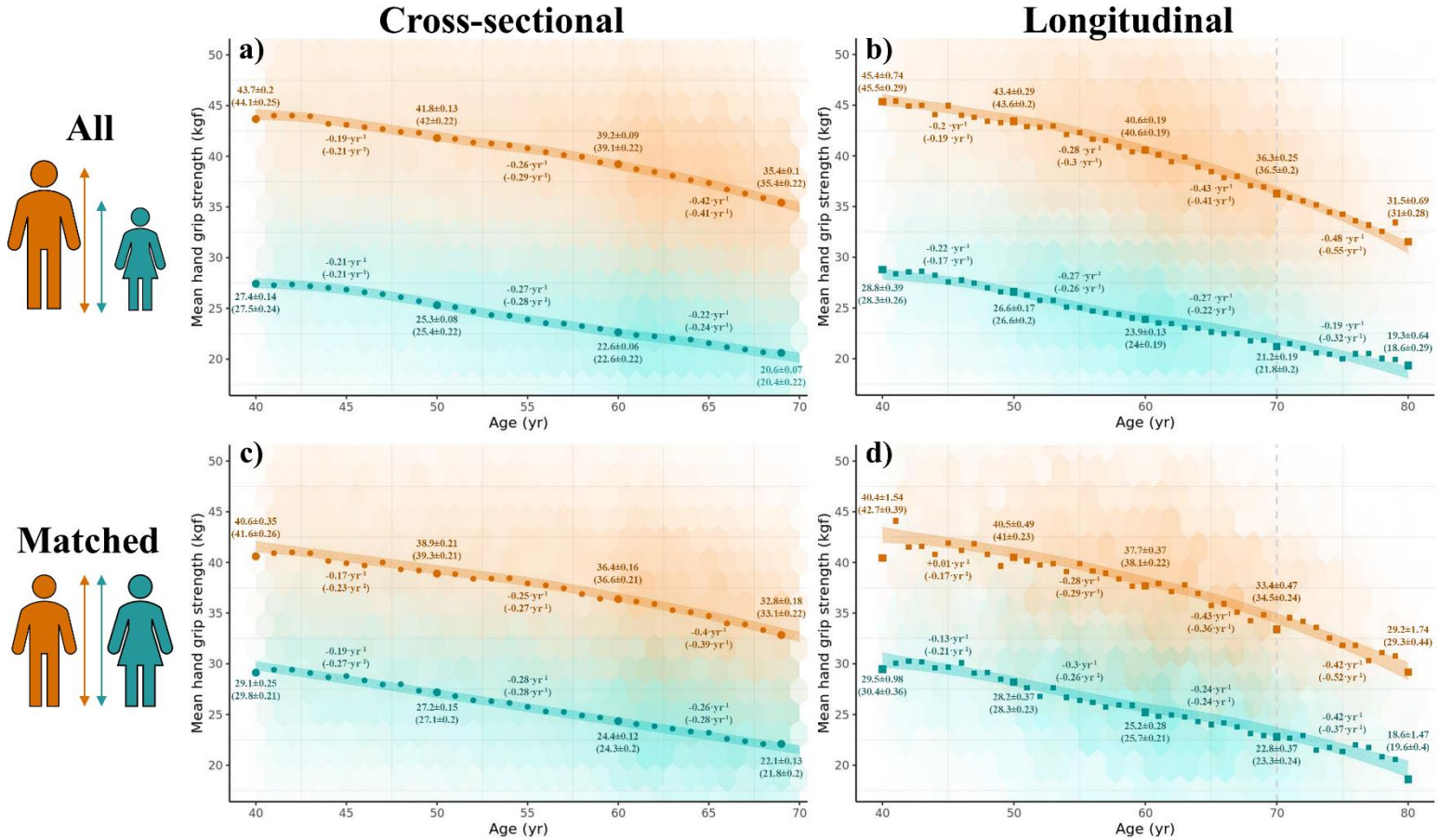


Figure 1. Modest sexual dimorphism in age-related strength decline. Sex-specific relationships between HGS and age within UK Biobank participants, as assessed both cross-sectionally (a,c) and longitudinally (b,d). Female (teal) and male (orange) observed data presented using hexagonal bin plotting, in which greater opacity indicates higher density of observations. Points represent means of all observed measures for each year age-group, and coloured bands are 95% confidence intervals for fitted relationships with fixed covariates (Online Methods). Values (± SEM) without parentheses are observed values, with equivalent fitted values given within parentheses. **a,c.** Cross-sectional observed measures of Cross-sectional Participants (a, N = 368,336) and Matched Cross-sectional Participants (c, N = 104,820) conform with fitted estimates. **b,d.** Longitudinal observed measures of Longitudinal Participants (b, N = 33,362) and Matched Longitudinal Participants (d, N = 7,958) plot both baseline and repeat HGS for each participant (unlinked). Both measures per participant were explicitly linked within the fitted models.

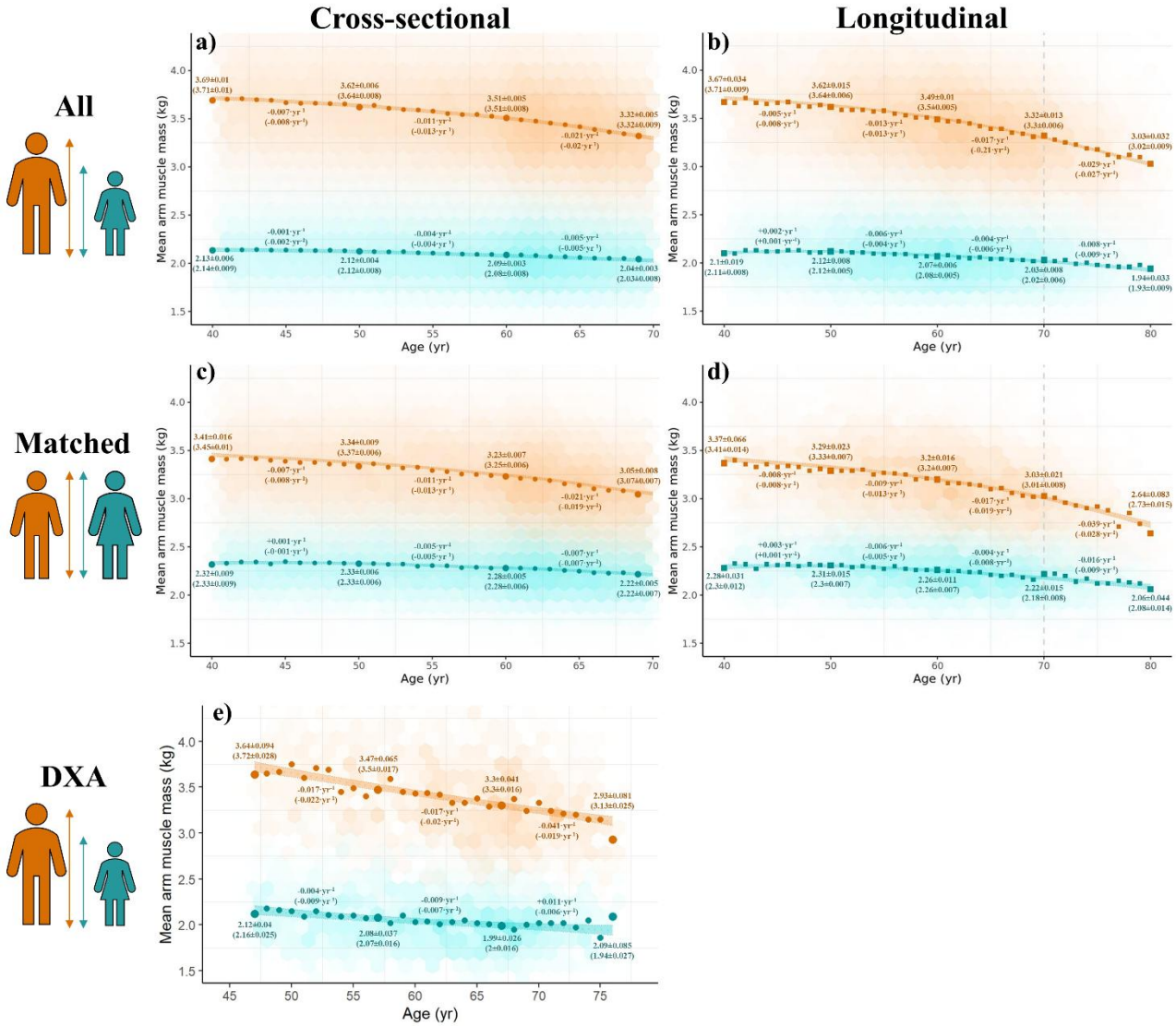


Figure 2. Age-related muscle mass decline is greater in males. Sex-specific relationships between MM and age within UK Biobank participants, as assessed both cross-sectionally (a,c,e) and longitudinally (b,d). Female (teal) and male (orange) observed data are presented using hexagonal bin plotting, in which greater opacity indicates higher density of observations. Points represent means of all observed measures for each year age-group, and coloured bands are 95% confidence intervals for fitted relationships with fixed covariates (Online Methods). Values (± SEM) without parentheses are observed values, with equivalent fitted values given within parentheses. All MM estimates are BIA-derived, except in panel e, in which data are DXA-derived. **a,c,e.** Cross-sectional observed measures of Cross-sectional Participants (a, N = 432,809), Matched Cross-sectional Participants (c, N = 123,366), and a subset of DXA-scanned Longitudinal Participants ≥44.5 and ≤78 yr analysed cross-sectionally (N = 4,253) conform with fitted estimates. **b,d.** Longitudinal observed measures of Longitudinal Participants (b, N = 44,609) and Matched Longitudinal Participants (d, N = 10,601) plot both baseline and repeat MM for each participant (unlinked). Both measures per participant were explicitly linked within the fitted models.

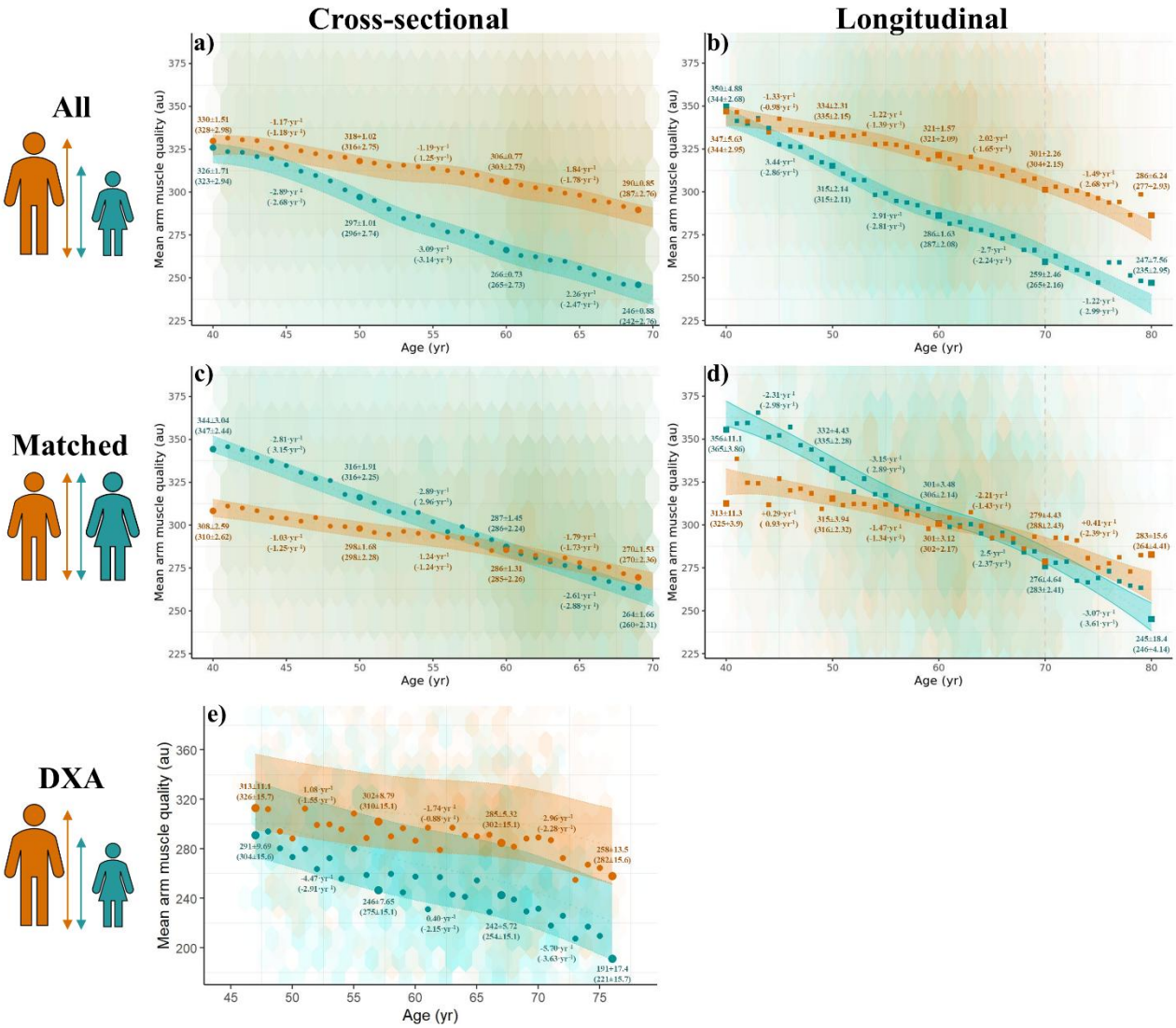


Figure 3. Age-related muscle quality decline is greater in females. Sex-specific relationships between MQ (in arbitrary units [au]) and age within UK Biobank participants, as assessed both cross-sectionally (a,c,e) and longitudinally (b,d) based on BIA (a-d) or DXA (e) measurements. Female (teal) and male (orange) observed data are presented using hexagonal bin plotting, in which greater opacity indicates higher density of observations. Points represent means of all observed measures for each year age-group, and coloured bands are 95% confidence intervals for fitted relationships with fixed covariates (Online Methods). Values (± SEM) without parentheses are observed values, with equivalent fitted values given within parentheses. **a,c,e.** Cross-sectional observed measures of Cross-sectional Participants (a, N = 368,336), Matched Cross-sectional Participants (c, N = 104,820), and a subset of DXA-scanned Longitudinal Participants ≥44.5 and ≤78 yr analysed cross-sectionally (e, N = 3,972) conform with fitted estimates. **b,d.** Longitudinal observed measures of Longitudinal Participants (b, N = 32,663) and Matched Longitudinal Participants (d, N = 7,778) plot both baseline and repeat MQ for each participant (unlinked). Both measures per participant were explicitly linked within the fitted models.

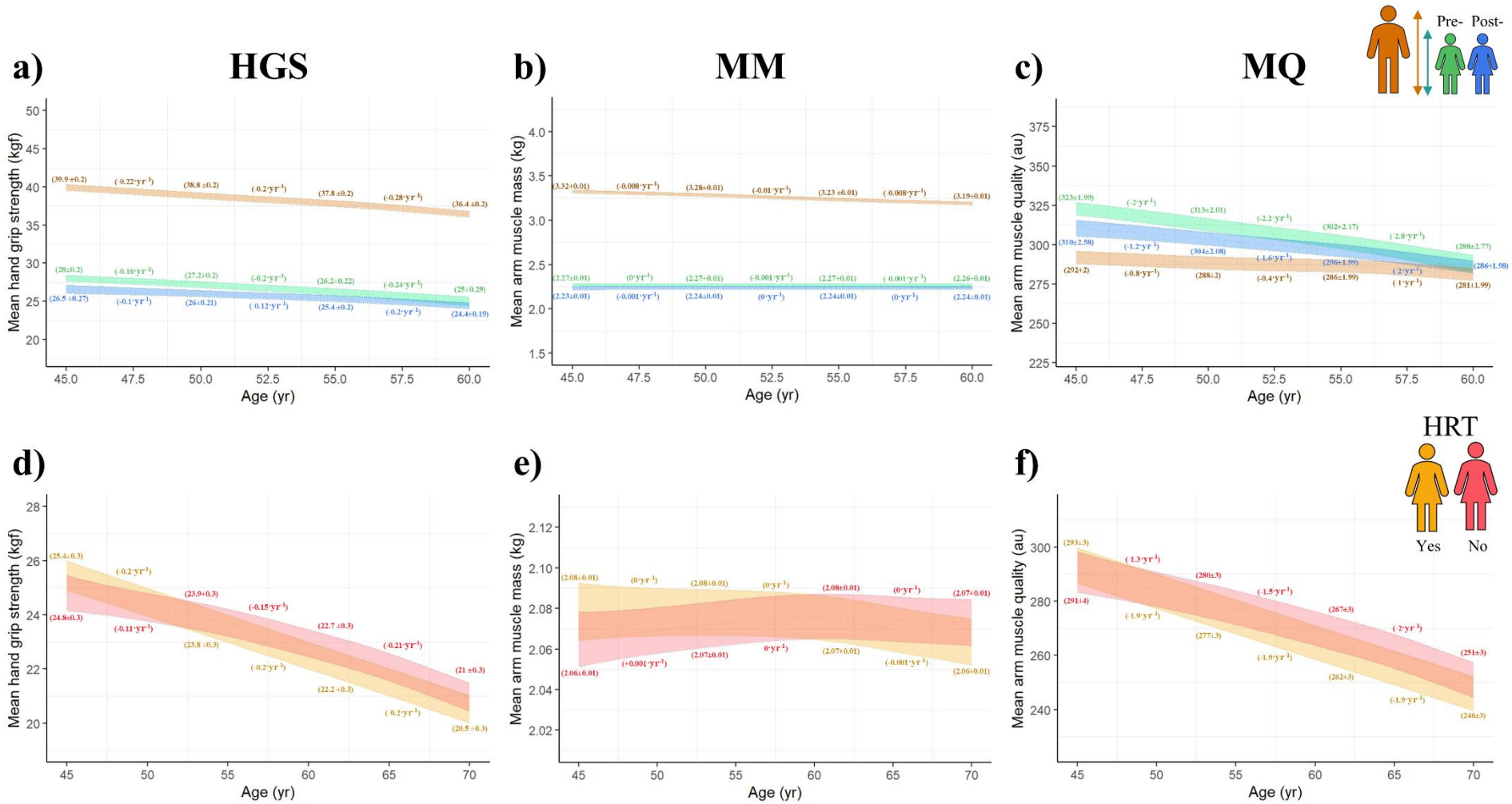
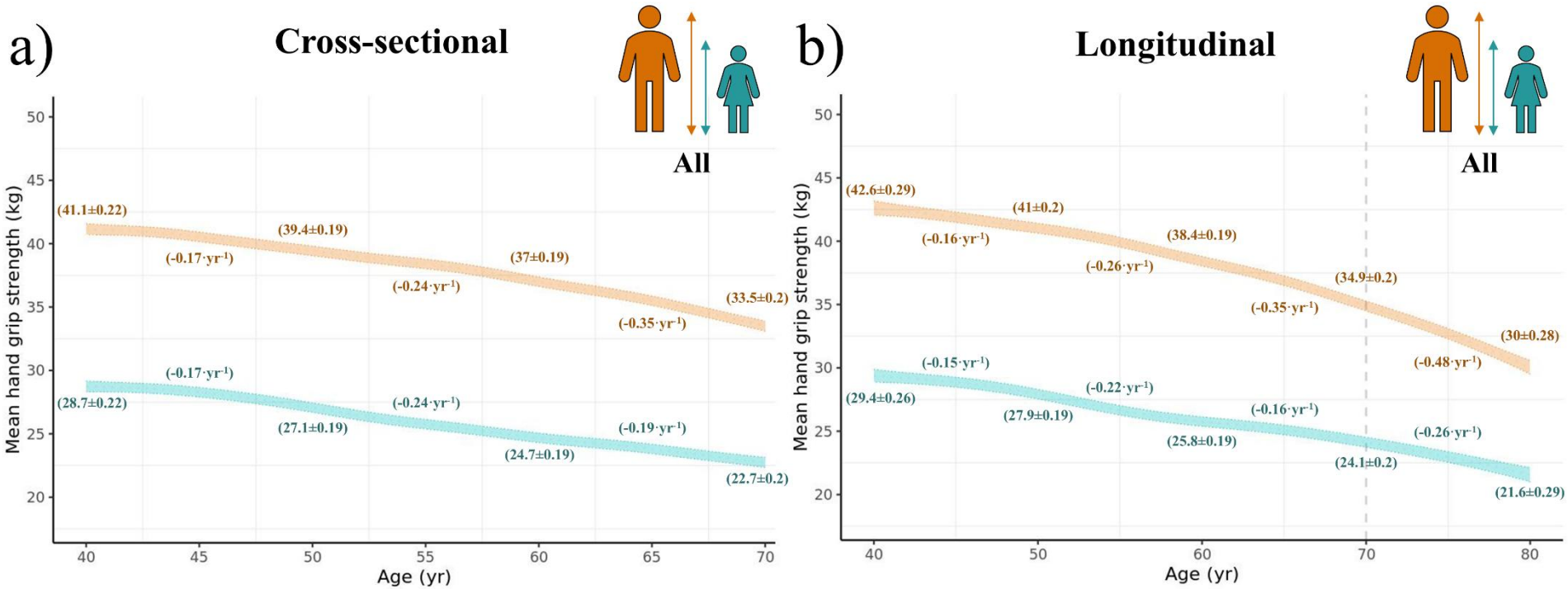


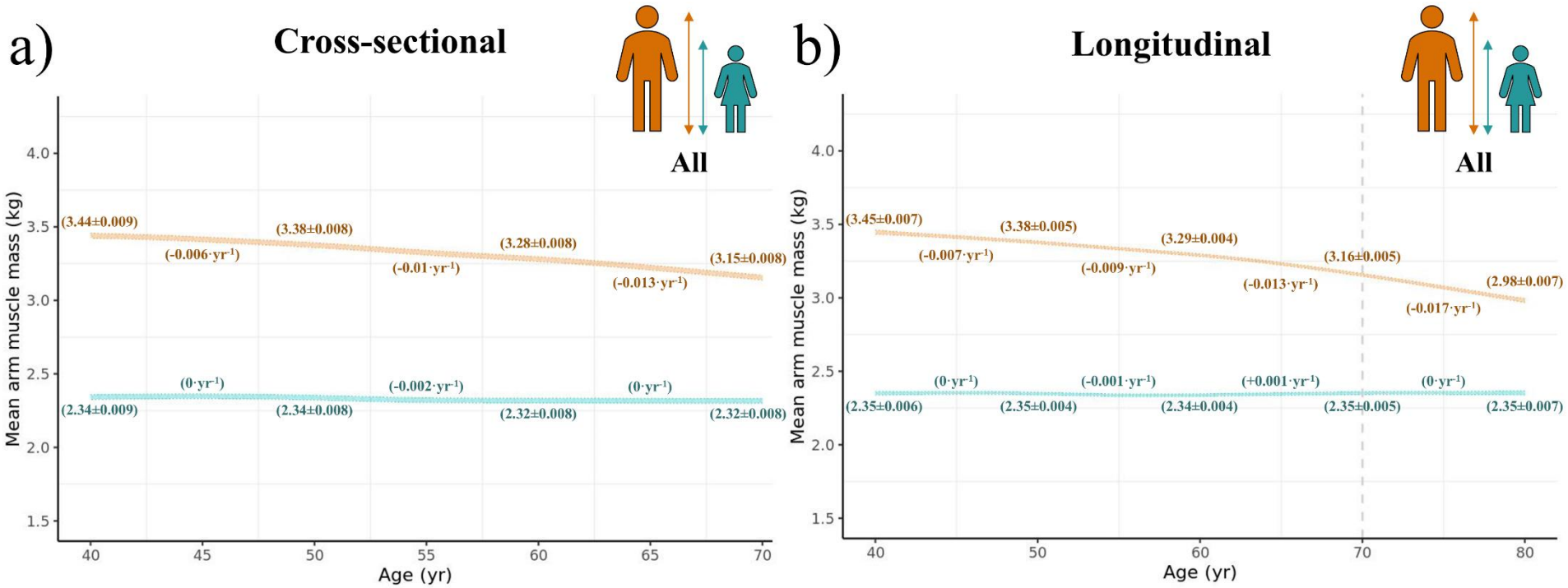
Figure 4. Females exhibit distinct muscle ageing trajectories from males irrespective of menopause status and history of HRT use. Model-fitted relationships between age and HGS (a,d), MM (b,e), and MQ (c,f) in the Menopause Participants, a subset of Cross-sectional Participants (Supp. Info. 1). Coloured bands are 95% confidence intervals (CIs) for fitted relationships with fixed covariates (Online Methods). Groups were not matched on body size, but standing height^2 (cm^2) and body mass$^{2/3}$ (kg$^{2/3}$) were included as covariates to account for the fact that premenopausal females were typically slightly larger than postmenopausal females at a given age (data not shown). Complete model specifications are provided in Supp. Info. 2. Fitted values (± SEM) are given within parentheses. **a–c.** Females are stratified by menopause status, with green and blue representing pre- and postmenopausal females, respectively; orange represents males. Standing height and body mass are fixed for plotting at 167.15 cm and 76.13

kg, respectively—calculated as *(mean premenopausal female value + mean postmenopausal female value + mean male value)/3*. One inference from these models is that some degree of accelerated decline (HGS and MQ) or decline (MM) occurs during the perimenopause, to account for the lower predicted mean values for postmenopausal females; however, our models could not capture this, as menopause status was implicitly treated as a fixed and discrete property. Note that the models plotted here were fitted to all three groups, and thus differ slightly from the pairwise models discussed in the Results. **d–f.** 115,990 postmenopausal Menopause Participants, stratified by response to the question "Have you ever used hormone replacement therapy (HRT)?". Gold and pink represent participants who answered "Yes" (N = 49,699) and "No" (N = 66,291), respectively. Standing height and body mass are fixed for plotting at 161.9 cm and 70.99 kg, respectively—calculated as *(mean "Yes" value + mean "No" value)/2*. Despite significant differences in the muscle ageing trajectories of the two groups, the widely overlapping CIs indicate that these differences are minor, at least in the case of a White female of average size (as plotted here).

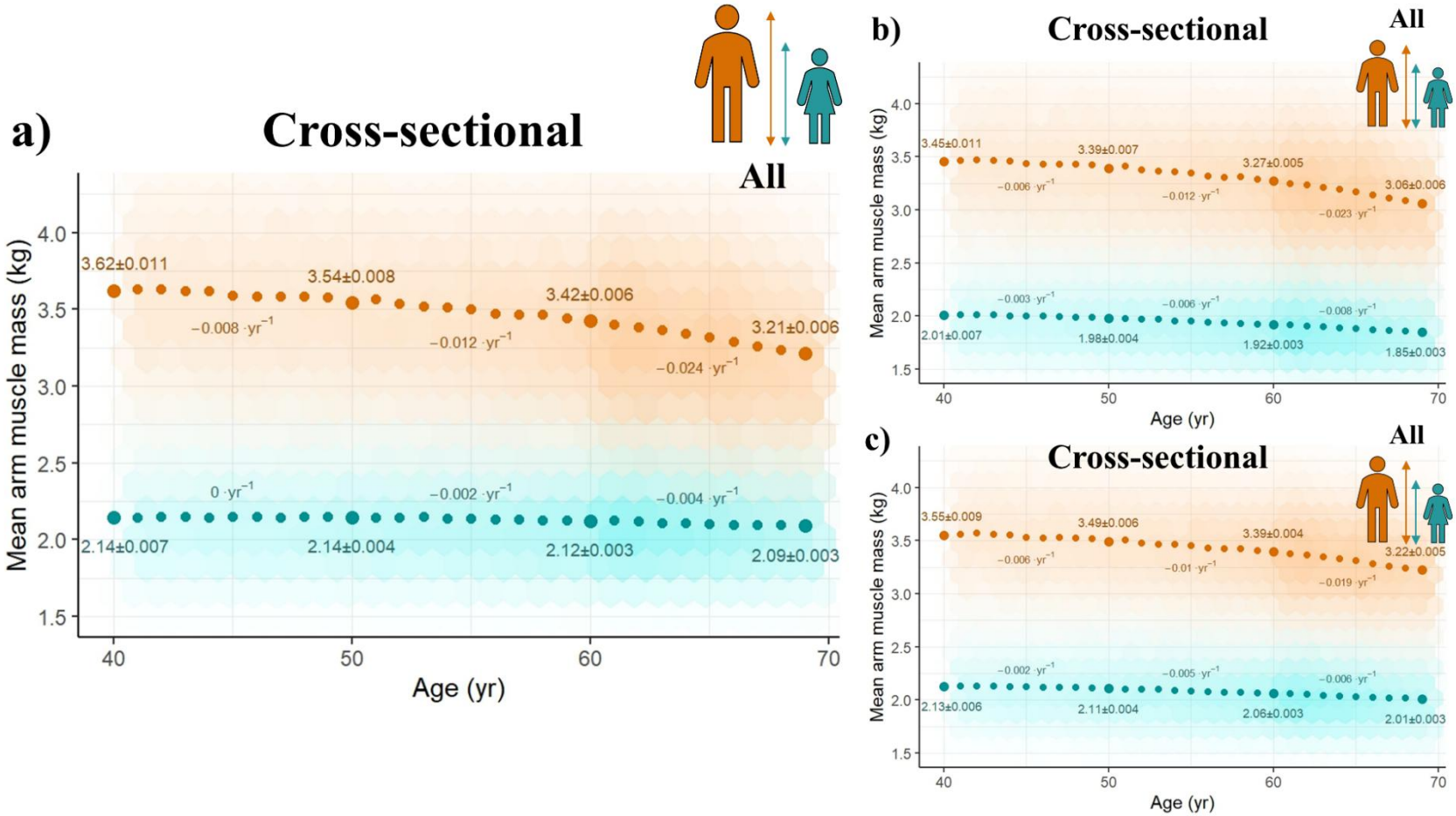
Extended Data



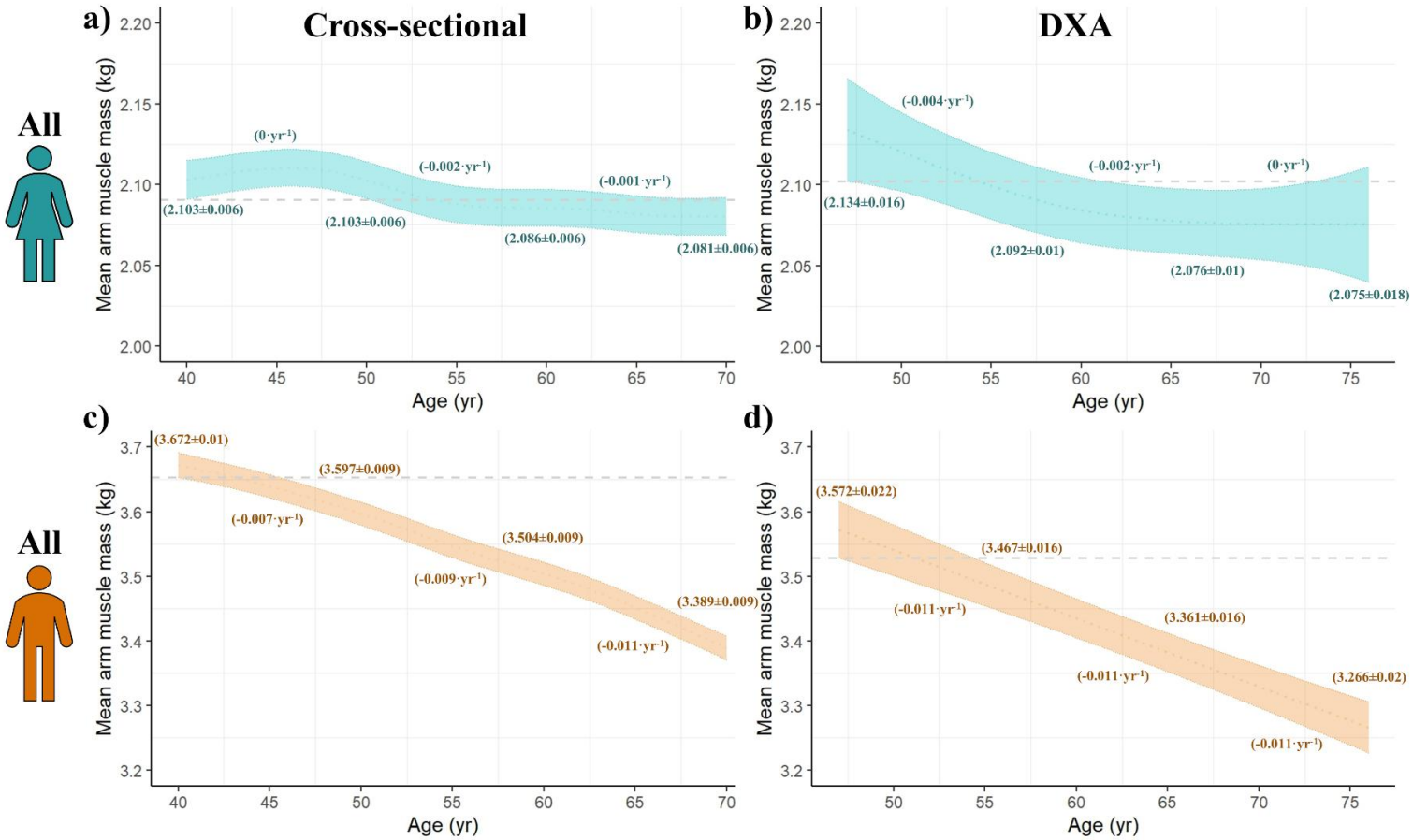
Extended Data 1. Modest sexual dimorphism in age-related strength decline persists after accounting for sex differences in body size. Model-fitted relationships between HGS and age are shown for Cross-sectional Participants (**a**) and Longitudinal Participants (**b**). Teal and orange represent females and males, respectively. Coloured bands represent the 95% confidence intervals for the fitted relationships. The models include standing height^2 (cm^2) and body mass$^{2/3}$ (kg$^{2/3}$) as covariates, to control for the potential contribution to HGS of body size differences under the principle of allometric scaling. Standing height and body mass are fixed for plotting purposes at 168.97 cm and 78.85 kg, respectively, values equal to *(mean female value + mean male value)/2* in the Cross-sectional Participants. Complete model specifications are provided in Supp. Info. 2 under headings 'HGS Model 3' (a) and 'HGS Model 4' (b). This analysis demonstrates that the general trends in age-related HGS decline presented in the main manuscript are expected to persist, even after accounting for body size differences by age and sex.



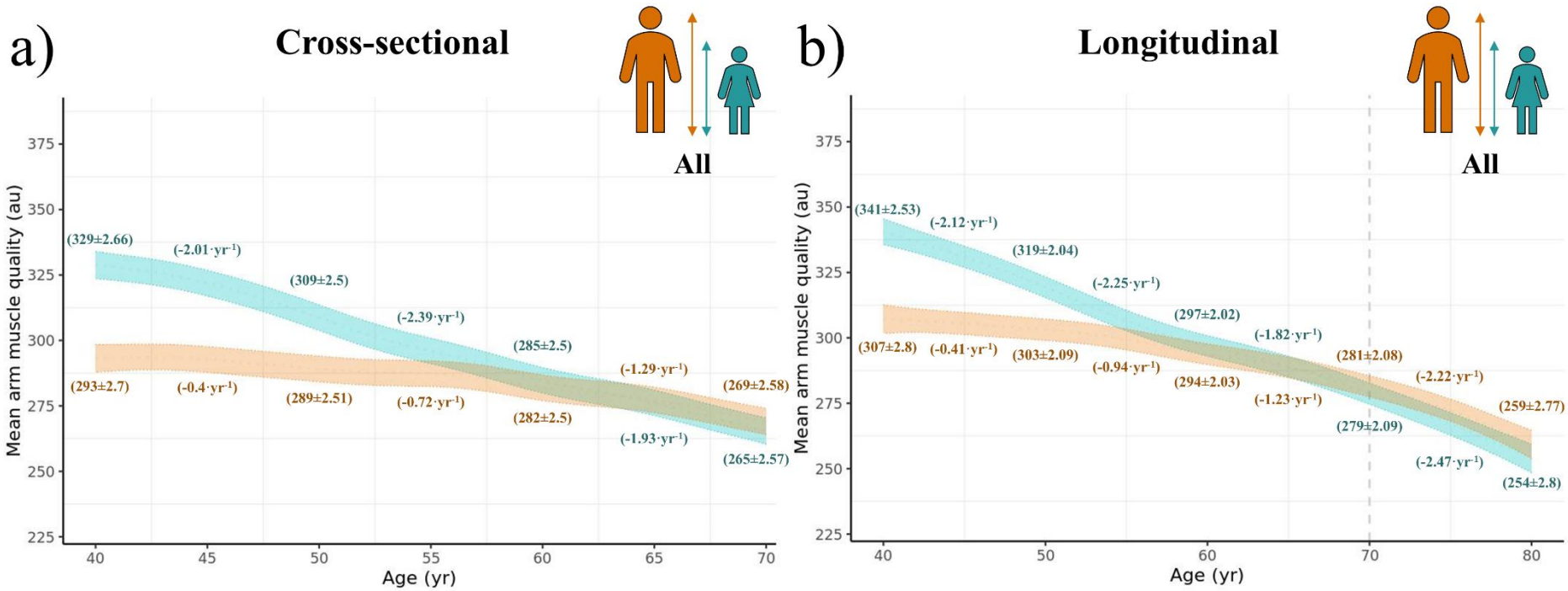
Extended Data 2. Age-related muscle decline remains substantially greater in males than females after accounting for sex differences in body size. Model-fitted relationships between MM and age are shown for Cross-sectional Participants (**a**) and Longitudinal Participants (**b**). Teal and orange represent females and males, respectively. Coloured bands represent the 95% confidence intervals for the fitted relationships. The models include standing height^2 (cm^2) and body mass$^{2/3}$ (kg$^{2/3}$) as covariates, to control for the potential contribution to MM differences by body size under the principle of allometric scaling. Standing height and body mass are fixed for plotting purposes at 168.97 cm and 78.85 kg, respectively, values equal to *(mean female value + mean male value)/2* in the Cross-sectional Participants. Complete model specifications are provided in Supp. Info. 2 under headings 'MM Model 3' (a) and 'MM Model 4' (b). Intriguingly, female MM is predicted to remain essentially constant with age, in contrast to the modest age-related MM decline in females over ~50 yr identified in our primary analyses (Fig. 2). While participant matching allows for a relationship between age and body size (e.g. approx. -0.15 cm·yr^{-1} standing height and -0.07 kg·yr^{-1} body mass in female Matched Cross-sectional Participants [data not shown]), the modelling approach utilised here does not; the discordant findings of the two approaches may suggest that age-related female MM decline reflects the inverse correlation between age and body size.



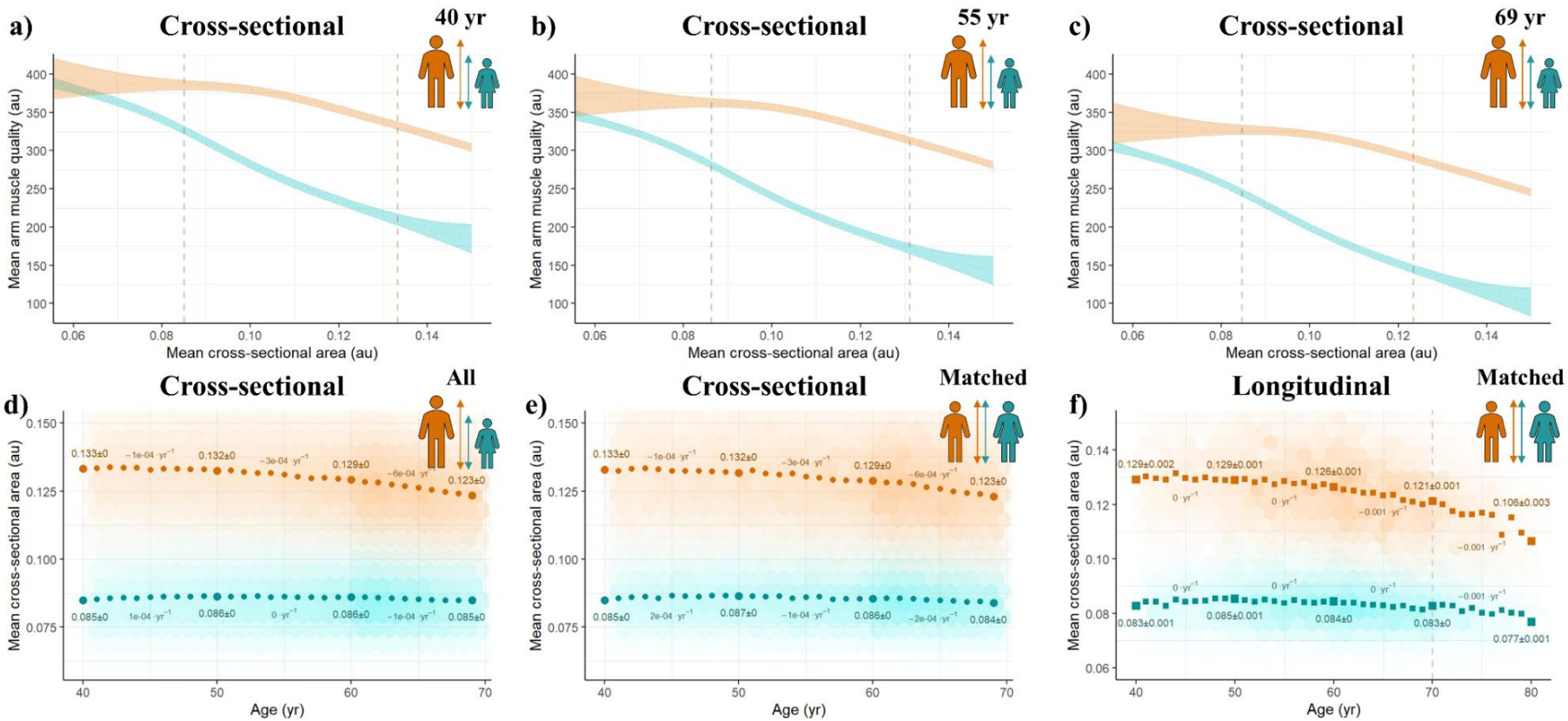
Extended Data 3. Patterns of age-related muscle mass decline are similar across three alternative sets of BIA-derived muscle mass estimates. Sex-specific relationships between muscle mass and age within the Cross-sectional Participants. Teal and orange represent females and males, respectively. The observed data are plotted using hexagonal bin plotting, with greater opacity indicating higher density of observations. Large points represent means of all observed measures for each year age-group. Values (mean ± SEM) are observed values at 40, 50, 60, and 69 yr, and inferred annual change over the corresponding 9–10 year period. *y*-axes are truncated to better visualise the general trends in the data. **a**. Direct outputs from a Tanita Body Composition Analyser BC-418MA; **b**. Estimates generated using the equations of Powell et al. (2020) (full reference in main manuscript); **c**. Estimates generated using our novel sex-specific equations, derived using unadjusted BIA impedance values (Online Methods). While the magnitude of the sexual dimorphism differs between the datasets, the primary trend of substantially greater age-related muscle mass decline in males holds across all three datasets.



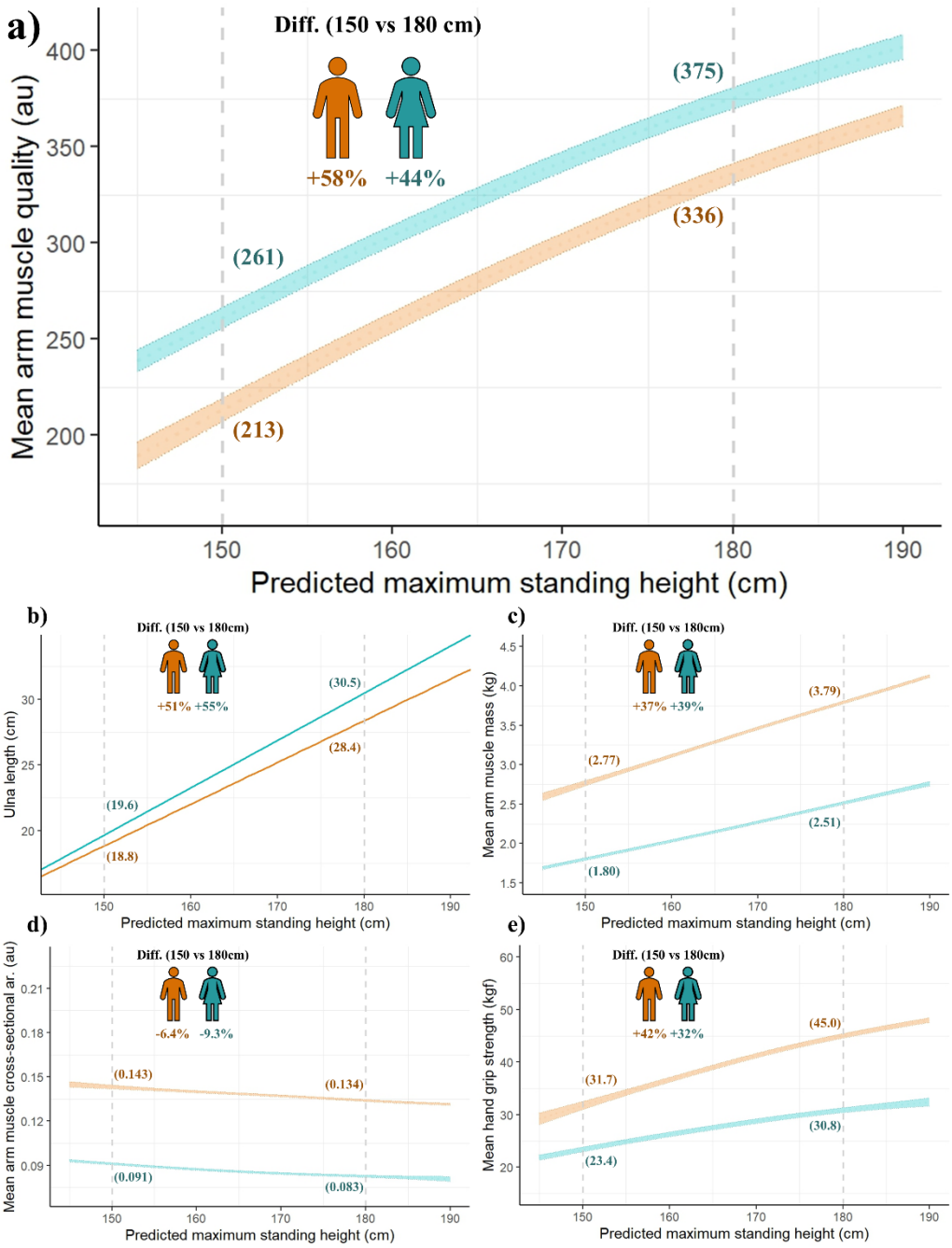
Extended Data 4. Female muscle mass exhibits minimal decline with age after accounting for age-related body-size decline. Model-fitted relationships between muscle mass and age. **a,c.** MM data from Cross-sectional Participants. **b,d.** DXA mean arm muscle mass estimates from a subset of 4,253 DXA-scanned Longitudinal Participants ≥44.5 and ≤78 yr analysed cross-sectionally. Teal and orange represent females and males, respectively. Coloured bands represent the 95% confidence intervals (CIs) for the fitted relationships. The models include standing height^2 (cm^2) and body mass$^{2/3}$ (kg$^{2/3}$) as covariates, to control for the potential contribution to muscle mass of body size differences under the principle of allometric scaling. Standing height and body mass are fixed for plotting purposes at sex-specific means for Cross-sectional Participants (females: 162.38 cm and 71.63 kg; males: 175.55 cm and 86.07 kg). In both female datasets the lower bound of the CI at the lowest age (dashed grey lines) overlaps the upper bound at the greatest age, indicating minimal difference in predicted muscle mass at 40 vs 70 yr (a) or 47 vs 76 yr (b) in the absence of age-related body-size decline. In contrast, substantial age-related muscle mass decline is predicted for males in both datasets (c,d). Complete model specifications are provided in Supp. Info. 2 under heading 'Extended Data 4'.



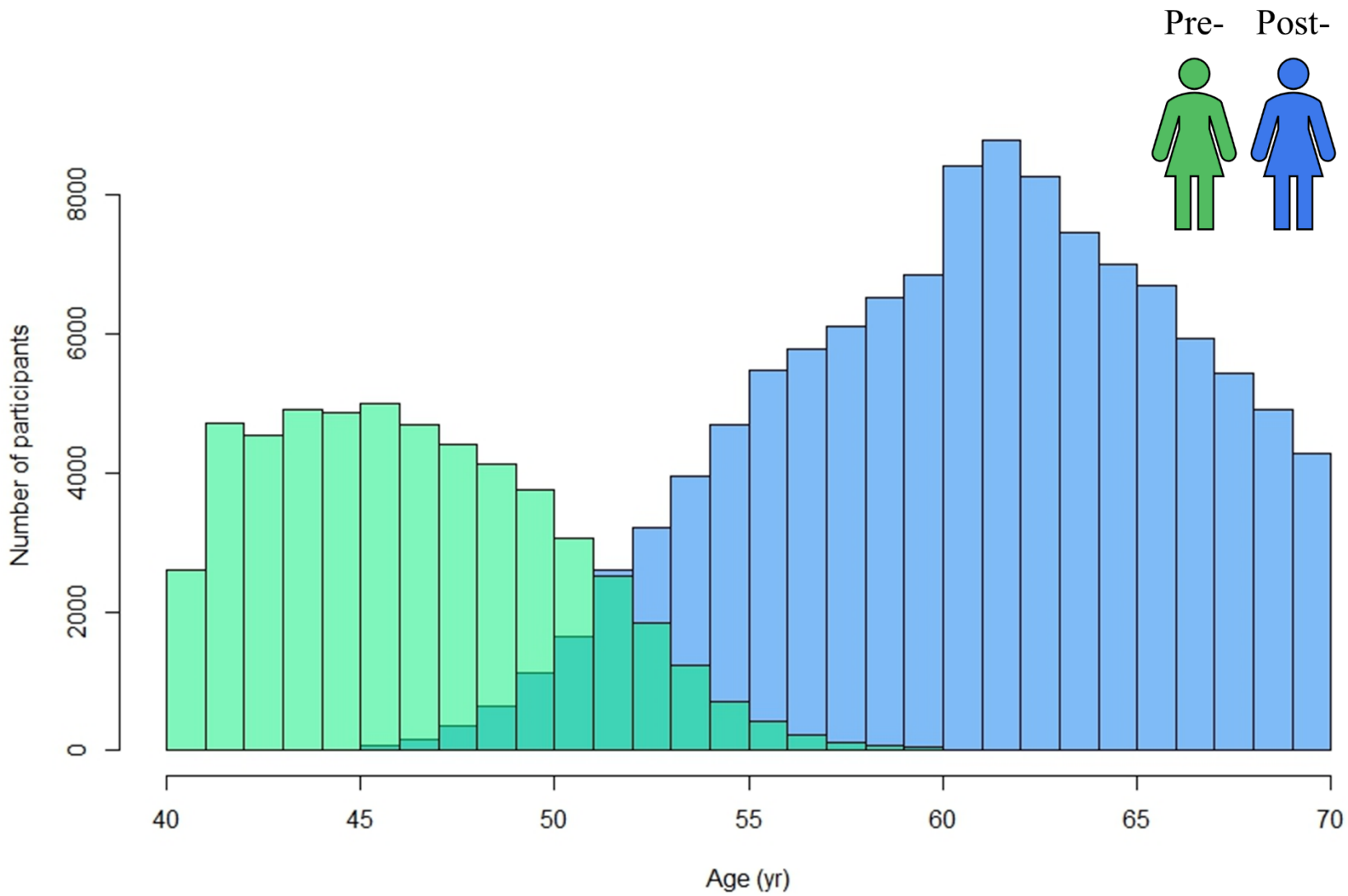
Extended Data 5. Age-related muscle quality decline remains greater in females than males after accounting for sex differences in body size. Model-fitted relationships between MQ and age are shown for Cross-sectional Participants (**a**) and Longitudinal Participants (**b**). Teal and orange represent females and males, respectively. Coloured bands represent the 95% confidence intervals for the fitted relationships. The models include standing height^2 (cm^2) and body mass$^{2/3}$ (kg$^{2/3}$) as covariates, to control for the potential contribution to MQ of body size differences under the principle of allometric scaling. Standing height and body mass are fixed for plotting at 168.97 cm and 78.85 kg, respectively, values equal to *(mean female value + mean male value)/2* in the Cross-sectional Participants. Complete model specifications are provided in Supp. Info. 2 under headings 'MQ Model 3' (a) and 'MQ Model 4' (b).



Extended Data 6. Muscle quality is inversely correlated with muscle cross-sectional area in both sexes, and females typically have smaller muscle cross-sectional areas than males. Mean arm muscle cross-sectional area (CSA) data for UKB participants. Teal and orange represent females and males, respectively. **a–c.** Model-fitted 95% confidence interval bands for relationships between MQ and CSA are shown for Cross-sectional Participants, with age fixed for plotting at 40 yr (a), 55 yr (b), and 69 yr (c). Vertical dashed lines denote observed sex-specific mean CSAs in participants of the corresponding age. Male arm muscle is predicted to produce substantially more force per unit of CSA than female muscle. For example, at 40 yr mean observed CSA of males (0.133) is ~57% greater than that of females (0.085), yet predicted male MQ at a CSA of 0.133 (333) is almost identical to predicted female MQ at 0.085 (327), despite the inverse correlation between MQ and CSA that is apparent in both sexes, but is markedly more pronounced in females. The complete model specification is provided in Supp. Info. 2 under heading 'Extended Data 6a–c'. **d–f.** Values (mean ± SEM) are observed values, and inferred annual change over the corresponding 9–10 year period. Observations are plotted using hexagonal bin plotting, with greater opacity indicating higher density of observations. The *y*-axes are truncated to visualise better the general trends in the data. Mean male CSA substantially exceeds that of females between 40–80 yr in unmatched (d) and matched (e–f) participant groups. Interestingly, age-related decline in CSA between 40–70 yr is apparent in males, but not females. .



Extended Data 7. Muscle quality is mediated by complex and sex-specific allometric relationships between standing height, ulna length, arm muscle mass, and muscle cross-sectional area. Model-fitted relationships are shown for Cross-sectional Participants. Teal and orange represent females and males, respectively. Age is arbitrarily fixed at 40 yr for plotting. Values in parentheses are fitted values at a predicted maximum standing height (PMSH—Methods) of 150 cm or 180 cm (dashed vertical lines). Coloured bands represent the 95% confidence intervals (CIs) for the fitted relationships. **a.** MQ is positively correlated with PMSH, such that MQ at a PMSH of 180 cm is predicted to be 44% greater (females) or 58% greater (males) than at 150 cm. **b-d.** Ulna length (UL, b) scales more strongly with PMSH than does MM (c), leading to an inverse relationship between muscle cross-sectional area (CSA)—defined as *MM/UL*—and PMSH (d), and implying non-isometric scaling of arm muscle. **e.** HGS correlates positively with PMSH; as *MQ = HGS/CSA*, the positive correlation between MQ and PMSH arises because greater PMSHs are generally associated with both higher HGS (numerator) and lower CSA (denominator). Complete model specifications are provided in Supp. Info. 2 under heading 'Extended Data 7'. See Methods for equations to predict UL from PMSH in white participants (as per panel b).



Extended Data 8. Age distribution of females in the Menopause Participants group, by menopause status. Green and blue represent premenopausal (N = 53,840) and postmenopausal (N = 116,261) females, respectively.

Supplementary Information 1 Participant exclusion flowcharts

Cross-sectional Participants

1. Full UK Biobank dataset (N = 502,461)

↓

2. Remove 93 participants who have withdrawn consent (N = 502,368)

↓

3. Remove 372 participants with a mismatch for genetic and self-reported sex (N = 501,996)

↓

4. Remove 16,982 participants with missing baseline data (N = 485,014):

- Unique participant ID number (‘EID’)
- Sex
- Ethnicity
- Year of birth
- Month of birth
- Date of assessment
- UK Biobank assessment centre ID number
- Standing height (cm)
- Body mass (kg)
- Waist circumference (cm)
- Hip circumference (cm)
- Hand grip dynamometer ID number
- Grip strength of left hand (kg)
- Grip strength of right hand (kg)
- BIA impedance device ID number
- BIA-derived estimate of muscle mass of left arm (kg)
- BIA-derived estimate of muscle mass of right arm (kg)

↓

5. Remove 559 participants with recorded baseline values ≤0 (N = 484,455):

- Standing height (cm)
- Body mass (kg)
- Waist circumference (cm)
- Hip circumference (cm)
- Grip strength of left hand (kg)
- Grip strength of right hand (kg)
- BIA-derived estimate of muscle mass of left arm (kg)
- BIA-derived estimate of muscle mass of right arm (kg)
- BIA-derived impedance of left arm (Ω)
- BIA-derived impedance of right arm (Ω)
- BIA-derived impedance of left leg (Ω)
- BIA-derived impedance of right leg (Ω)
- BIA-derived impedance of whole body (Ω)

↓

6. Remove 46 participants with implausible or extreme recorded baseline values (N = 484,409):

- Standing height (cm): ≤122 or ≥199 (females); ≤133 (males)
- Waist circumference (cm): ≤48 or ≥163 (females); ≤20 or ≥197 (males)
- Hip circumference (cm): ≤39 or ≥195 (males)
- Grip strength of left hand (kg): ≥63 (females); ≥86 (males)
- Grip strength of right hand (kg): ≥70 (females)
- BIA-derived estimate of muscle mass of left arm (kg): ≥5 (females)
- BIA-derived estimate of muscle mass of right arm (kg): ≥5.1 (females)
- BIA-derived impedance of left leg (Ω): ≥637 (females); ≥660 (males)
- BIA-derived impedance of right leg (Ω): ≥815 (females); ≥627 (males)
- BIA-derived impedance of whole body (Ω): ≤72 (males)

↓

7. Remove 1,960 participants whose baseline grip strength readings were taken on an errant dynamometer (ID #007336) (N = 482,449)

↓

8. Remove 2,315 participants <40 or ≥70 years of age at baseline assessment (N = 480,134)

↓

9. Remove 47,325 participants who attended a repeat assessment between 30th April 2014 and 13th March 2020 (analysed separately; see 'Longitudinal Participants' exclusion flowchart) (N = 432,809)

↓

Cross-sectional Participants (N = 432,809; 237,015 females and 195,794 males):

368,336 participants with HGS values (202,028 females and 166,308 males) (see Supp. Info. 3)
432,809 participants with MM values (237,015 females and 195,794 males)
368,336 participants with MQ values (202,028 females and 166,308 males)

Longitudinal Participants

1. Participants who attended a repeat assessment between 30th April 2014 and 13th March 2020, and who were not excluded due to withdrawal or baseline data quality (see Steps 1–9 of 'Cross-sectional Participants' exclusion flowchart) (N = 47,325)

↓

2. Remove 74 participants who attended a repeat assessment in Bristol (N = 47,251)

↓

3. Remove 1,505 participants whose repeat grip strength readings were taken on an errant dynamometer (ID #007336) (N = 45,746)

↓

4. Remove 117 participants who were not measured on one of the following BIA Impedance Devices: #001477, #001480, #001486, #003167, #003168, #003169, #003173 (N = 45,629)

↓

Longitudinal Participants (N = 45,629; 23,448 females and 22,181 males) :

40,438 participants with baseline HGS values (20,805 females and 19,633 males) (see Supp. Info. 3)
37,687 participants with repeat HGS values (19,352 females and 18,335 males) (see Supp. Info. 3)
33,362 participants with both HGS values (17,146 females and 16,216 males) (see Supp. Info. 3)
45,629 participants with baseline MM values (23,448 females and 22,181 males)
44,609 participants with repeat MM values (22,907 females and 21,702 males)
44,609 participants with both MM values (22,907 females and 21,702 males)
40,438 participants with baseline MQ values (20,805 females and 19,633 males)
36,903 participants with repeat MQ values (18,937 females and 17,966 males)
32,663 participants with both MQ values (16,780 females and 15,883 males)

Menopause Participants

1. Cross-sectional Participants (N = 432,809)

↓

2. Remove 37,712 females with a recorded response other than “Yes” or “No” to the question "Have you had your menopause (periods stopped)?" (N = 395,097)

↓

3. Remove 678 females ≥60 yr who reported being premenopausal (N = 394,419)

↓

4. Remove 9,533 postmenopausal females with a recorded response of “NA”, “prefer not to answer”, or “do not know” to the question "How old were you when your periods stopped?" (N = 384,886)

↓

5. Remove 18,991 postmenopausal females whose self-reported age at menopause was <45 yr or ≥60 yr (N = 365,895)

↓

Menopause Participants (N = 365,895; 170,101 females and 195,794 males)

310,941 participants with HGS values (144,633 females and 166,308 males) (see Supp. Info. 3)

365,895 participants with MM values (170,101 females and 195,794 males)

310,941 participants with MQ values (144,633 females and 166,308 males)

Supplementary Information 2 Regression analysis

The generalised additive mixed models (GAMMs) below are presented in the syntax of the 'mgcv' R package's *bam* function (see main manuscript for references). Underlining denotes terms included in sex- or group-specific models only, while the remaining terms are included in both global and sex- or group-specific models. 'Dynamometer instance' denotes a specific dynamometer over a pre-determined usage period: for example, four discrete usage periods ('instances') were identified for dynamometer #000194, each of which was then coded as a unique 'device' (i.e. '#000194_0'–'#000194_3') for modelling purposes (Supp. Info. 3). 'Device day of use' is a value that equals zero on the earliest assessment date of a BIA impedance device/dynamometer instance, and increments by one for each subsequent calendar day (Supp. Info. 3). *s(x, bs = "re")* indicates that covariate *x* was modelled as a random effect. Lower AIC scores indicate better model fit, and larger ΔAICs therefore indicate stronger support for the model with the lower AIC score (Online Methods). Units are as follows: age—yr; body mass$^{2/3}$—kg$^{2/3}$; CSA—au; HGS—kgf; MM—kg; MQ—au; PMSH—cm; standing height^2—cm^2. Units for DXA-derived estimates are kg for muscle mass and au for muscle quality.

HGS model 1—Cross-sectional Participants

HGS ~ sex + s(age) + s(age, by = sex) + ethnicity + season of attendance + s(UKB Assessment Centre, bs = "re") + s(dynamometer instance, bs = "re") + s(dynamometer instance, device day of use, bs = "re")
AIC$_{\text{GLOBAL}}$ 2,441,315; **AIC$_{\text{SEX-SPECIFIC}}$** 2,440,783; **ΔAIC** 532

HGS model 2—Longitudinal Participants

HGS ~ sex + s(age) + s(age, by = sex) + ethnicity + season of attendance + s(UKB Assessment Centre, bs = "re") + s(dynamometer instance, bs = "re") + s(dynamometer instance, device day of use, bs = "re") + s(participant ID, bs = "re")
AIC$_{\text{GLOBAL}}$ 439,223; **AIC$_{\text{SEX-SPECIFIC}}$** 437,287; **ΔAIC** 1,936

HGS model 3—Cross-sectional Participants (controlling for body size)

HGS ~ sex + s(age) + s(age, by = sex) + ethnicity + season of attendance + s(UKB Assessment Centre, bs = "re") + s(dynamometer instance, bs = "re") + s(dynamometer instance, device day of use, bs = "re") + standing height^2 + body mass$^{2/3}$
AIC$_{\text{GLOBAL}}$ 2,407,242; **AIC$_{\text{SEX-SPECIFIC}}$** 2,406,732; **ΔAIC** 511

HGS model 4—Longitudinal Participants (controlling for body size)

HGS ~ sex + s(age) + s(age, by = sex) + ethnicity + season of attendance + s(UKB Assessment Centre, bs = "re") + s(dynamometer instance, bs = "re") + s(dynamometer instance, device day of use, bs = "re") + s(participant ID, bs = "re") + standing height^2 + body mass$^{2/3}$
AIC$_{\text{GLOBAL}}$ 438,840; **AIC$_{\text{SEX-SPECIFIC}}$** 436,894; **ΔAIC** 1,946

HGS model 5—Matched Cross-sectional Participants

HGS ~ sex + s(age) + s(age, by = sex) + ethnicity + season of attendance + s(UKB Assessment Centre, bs = "re") + s(dynamometer instance, bs = "re") + s(dynamometer instance, device day of use, bs = "re")
AIC$_{\text{GLOBAL}}$ 690,961; **AIC$_{\text{SEX-SPECIFIC}}$** 690,900; **ΔAIC** 61

HGS model 6—Matched Longitudinal Participants

HGS ~ sex + s(age) + s(age, by = sex) + ethnicity + season of attendance + s(UKB Assessment Centre, bs = "re") + s(dynamometer instance, bs = "re") + s(dynamometer instance, device day of use, bs = "re") + s(participant ID, bs = "re")
AIC$_{\text{GLOBAL}}$ 103,641; **AIC$_{\text{SEX-SPECIFIC}}$** 103,412; **ΔAIC** 229

MM model 1—Cross-sectional Participants

MM ~ sex + s(age) + s(age, by = sex) + ethnicity + season of attendance + s(UKB Assessment Centre, bs = "re") + s(BIA impedance device, bs = "re") + s(BIA impedance device, device day of use, bs = "re")
AIC$_{\text{GLOBAL}}$ 373,361; **AIC$_{\text{SEX-SPECIFIC}}$** 368,198; **ΔAIC** 5,163

MM model 2—Longitudinal Participants

MM ~ sex + s(age) + s(age, by = sex) + ethnicity + season of attendance + s(UKB Assessment Centre, bs = "re") + s(BIA impedance device, bs = "re") + s(BIA impedance device, device day of use, bs = "re") + s(participant ID, bs = "re")
AIC$_{\text{GLOBAL}}$ -90,132; **AIC$_{\text{SEX-SPECIFIC}}$** -100,374.; **ΔAIC** 10,242

MM model 3—Cross-sectional Participants (controlling for body size)

MM ~ sex + s(age) + s(age, by = sex) + ethnicity + season of attendance + s(UKB Assessment Centre, bs = "re") + s(BIA impedance device, bs = "re") + s(BIA impedance device, device day of use, bs = "re") + standing height^2 + body mass$^{2/3}$
AIC$_{\text{GLOBAL}}$ -127,295; **AIC$_{\text{SEX-SPECIFIC}}$** -139,200; **ΔAIC** 11,905

MM model 4—Longitudinal Participants (controlling for body size)

MM ~ sex + s(age) + s(age, by = sex) + ethnicity + season of attendance + s(UKB Assessment Centre, bs = "re") + s(BIA impedance device, bs = "re") + s(BIA impedance device, device day of use, bs = "re") + s(participant ID, bs = "re") + standing height^{2} + body mass$^{2/3}$
AIC$_{\text{GLOBAL}}$ -117,911; **AIC$_{\text{SEX-SPECIFIC}}$** -131,759; **ΔAIC** 13,848

MM model 5—Matched Cross-sectional Participants

MM ~ sex + s(age) + s(age, by = sex) + ethnicity + season of attendance + s(UKB Assessment Centre, bs = "re") + s(BIA impedance device, bs = "re") + s(BIA impedance device, device day of use, bs = "re")
AIC$_{\text{GLOBAL}}$ 80,146; **AIC$_{\text{SEX-SPECIFIC}}$** 78,585; **ΔAIC** 1,561

MM model 6—Matched Longitudinal Participants

MM ~ sex + s(age) + s(age, by = sex) + ethnicity + season of attendance + s(UKB Assessment Centre, bs = "re") + s(BIA impedance device, bs = "re") + s(BIA impedance device, device day of use, bs = "re") + s(participant ID, bs = "re")
AIC$_{\text{GLOBAL}}$ -23,390; **AIC$_{\text{SEX-SPECIFIC}}$** -25,342; **ΔAIC** 1,952

MM model 7—4,253 Longitudinal Participants ≥44.5 and ≤78 yr subjected to a DXA scan (N = 2,224 [F], 2,029 [M])

(NB DXA measurement device and UKB Assessment Centre were not included as covariates in this model, as all DXA scans were performed at the Cheadle Assessment Centre, and no information was available on whether multiple DXA devices were utilised)

DXA-derived estimate of mean arm muscle mass ~ sex + s(age) + s(age, by = sex) + ethnicity + season of attendance

AIC$_{\text{GLOBAL}}$ 4,651.8; **AIC$_{\text{SEX-SPECIFIC}}$** 4,597.5; **ΔAIC** 54.3

MQ model 1—Cross-sectional Participants

MQ ~ sex + s(age) + s(age, by = sex) + ethnicity + season of attendance + s(UKB Assessment Centre, bs = "re") + s(dynamometer instance, bs = "re") + s(dynamometer instance, device day of use, bs = "re") + s(BIA impedance device, bs = "re") + s(BIA impedance device, device day of use, bs = "re")
AIC$_{\text{GLOBAL}}$ 4,145,095; **AIC$_{\text{SEX-SPECIFIC}}$** 4,141,887; **ΔAIC** 3,208

MQ model 2—Longitudinal Participants

MQ ~ sex + s(age) + s(age, by = sex) + ethnicity + season of attendance + s(UKB Assessment Centre, bs = "re") + s(dynamometer instance, bs = "re") + s(dynamometer instance, device day of use, bs = "re") + s(BIA impedance device, bs = "re") + s(BIA impedance device, device day of use, bs = "re") + s(participant ID, bs = "re")
AIC$_{\text{GLOBAL}}$ 794,360; **AIC**$_{\text{SEX-SPECIFIC}}$ 793,770; **ΔAIC** 590

MQ model 3—Cross-sectional Participants (controlling for body size)

MQ ~ sex + s(age) + s(age, by = sex) + ethnicity + season of attendance + s(UKB Assessment Centre, bs = "re") + s(dynamometer instance, bs = "re") + s(dynamometer instance, device day of use, bs = "re") + s(BIA impedance device, bs = "re") + s(BIA impedance device, device day of use, bs = "re") + standing height^2 *+ body mass*$^{2/3}$
AIC$_{\text{GLOBAL}}$ 4,038,448; **AIC**$_{\text{SEX-SPECIFIC}}$ 4,034,707; **ΔAIC** 3,741

MQ model 4—Longitudinal Participants (controlling for body size)

MQ ~ sex + s(age) + s(age, by = sex) + ethnicity + season of attendance + s(UKB Assessment Centre, bs = "re") + s(dynamometer instance, bs = "re") + s(dynamometer instance, device day of use, bs = "re") + s(participant ID, bs = "re") + s(BIA impedance device, bs = "re") + s(BIA impedance device, device day of use, bs = "re") + standing height^2 *+ body mass*$^{2/3}$
AIC$_{\text{GLOBAL}}$ 793,321; **AIC**$_{\text{SEX-SPECIFIC}}$ 792,881; **ΔAIC** 440

MQ model 5—Matched Cross-sectional Participants

MQ ~ sex + s(age) + s(age, by = sex) + ethnicity + season of attendance + s(UKB Assessment Centre, bs = "re") + s(dynamometer instance, bs = "re") + s(dynamometer instance, device day of use, bs = "re") + s(BIA impedance device, bs = "re") + s(BIA impedance device, device day of use, bs = "re")
AIC$_{\text{GLOBAL}}$ 1,174,546; **AIC**$_{\text{SEX-SPECIFIC}}$ 1,173,498; **ΔAIC** 1,048

MQ model 6—Matched Longitudinal Participants

MQ ~ sex + s(age) + s(age, by = sex) + ethnicity + season of attendance + s(UKB Assessment Centre, bs = "re") + s(dynamometer instance, bs = "re") + s(dynamometer instance, device day of use, bs = "re") + s(BIA impedance device, bs = "re") + s(BIA impedance device, device day of use, bs = "re") + s(participant ID, bs = "re")
AIC$_{\text{GLOBAL}}$ 188,917; **AIC**$_{\text{SEX-SPECIFIC}}$ 188,701; **ΔAIC** 216

MQ model 7—3,972 Longitudinal Participants ≥45.4 and ≤78 yr subjected to a DXA scan (N = 2,061 [F], 1,911 [M])

(NB DXA measurement device and UKB Assessment Centre were not included as covariates in this model, as all DXA scans were performed at the Cheadle Assessment Centre, and no information was available on whether multiple DXA devices were utilised)

DXA-derived estimate of mean arm muscle quality ~ sex + s(age) + s(age, by = sex) + ethnicity + season of attendance + s(dynamometer instance, bs = "re") + s(dynamometer instance, device day of use, bs = "re")
AIC$_{GLOBAL}$ 44,188; **AIC$_{SEX\text{-}SPECIFIC}$** 44,165; **ΔAIC** 23

HGS menopause model—Menopause Participants (controlling for body size)

HGS ~ group + s(age) + s(age, by = group) + ethnicity + season of attendance + s(UKB Assessment Centre, bs = "re") + s(dynamometer instance, bs = "re") + s(dynamometer instance, device day of use, bs = "re") + standing height^2 + body mass$^{2/3}$*
*1. Premenopausal females vs Postmenopausal females: **AIC$_{GLOBAL}$** 876,019.1; **AIC$_{GROUP\text{-}SPECIFIC}$** 876,007.7; **ΔAIC** 11.4
2. Postmenopausal females vs Males: **AIC$_{GLOBAL}$** 1,763,732; **AIC$_{GROUP\text{-}SPECIFIC}$** 1,763,260; **ΔAIC** 473
3. Premenopausal females vs Males: **AIC$_{GLOBAL}$** 1,433,089; **AIC$_{GROUP\text{-}SPECIFIC}$** 1,433,072; **ΔAIC** 17.4
4. Postmenopausal females (HRT) vs Postmenopausal females (no HRT): **AIC$_{GLOBAL}$** 594,725.5; **AIC$_{GROUP\text{-}SPECIFIC}$** 594,709.6; **ΔAIC** 15.9
5. Postmenopausal females (HRT) vs Males: **AIC$_{GLOBAL}$** 1,411,573; **AIC$_{GROUP\text{-}SPECIFIC}$** 1,411,387; **ΔAIC** 186
6. Postmenopausal females (no HRT) vs Males: **AIC$_{GLOBAL}$** 1,498,187; **AIC$_{GROUP\text{-}SPECIFIC}$** 1,497,791; **ΔAIC** 396

MM menopause model—Menopause Participants (controlling for body size)

MM ~ group + s(age) + s(age, by = group) + ethnicity + season of attendance + s(UKB Assessment Centre, bs = "re") + s(BIA impedance device, bs = "re") + s(BIA impedance device, device day of use, bs = "re") + standing height^2 + body mass$^{2/3}$*
*1. Premenopausal females vs Postmenopausal females: **AIC$_{GLOBAL}$** -195,675.839; **AIC$_{GROUP\text{-}SPECIFIC}$** -195,675.915; **ΔAIC** 0.076
2. Postmenopausal females vs Males: **AIC$_{GLOBAL}$** -44,516; **AIC$_{GROUP\text{-}SPECIFIC}$** -48,716; **ΔAIC** 4,200
3. Premenopausal females vs Males: **AIC$_{GLOBAL}$** -7,879; **AIC$_{GROUP\text{-}SPECIFIC}$** -8,429; **ΔAIC** 550
4. Postmenopausal females (HRT) vs Postmenopausal females (no HRT): **AIC$_{GLOBAL}$** -133,347.8; **AIC$_{GROUP\text{-}SPECIFIC}$** -133,374.7; **ΔAIC** -26.8
5. Postmenopausal females (HRT) vs Males: **AIC$_{GLOBAL}$** -6,301; **AIC$_{GROUP\text{-}SPECIFIC}$** -7,850; **ΔAIC** 1,549
6. Postmenopausal females (no HRT) vs Males: **AIC$_{GLOBAL}$** -13,455; **AIC$_{GROUP\text{-}SPECIFIC}$** -16,181; **ΔAIC** 2,726

MQ menopause model—Menopause Participants (controlling for body size)

MQ ~ group + s(age) + s(age, by = group) + ethnicity + season of attendance + s(UKB Assessment Centre, bs = "re") + s(dynamometer instance, bs = "re") + s(dynamometer instance, device day of use, bs = "re") + s(BIA impedance device, bs = "re") + s(BIA impedance device, device day of use, bs = "re") + standing height^2 + body mass$^{2/3}$*

*1. Premenopausal females vs Postmenopausal females: **AIC$_{GLOBAL}$** 1,582,850.9; **AIC$_{GROUP\text{-}SPECIFIC}$** 1,582,839.7; **ΔAIC** 11.2
2. Postmenopausal females vs Males: **AIC$_{GLOBAL}$** 2,898,594; **AIC$_{GROUP\text{-}SPECIFIC}$** 2,898,195; **ΔAIC** 399
3. Premenopausal females vs Males: **AIC$_{GLOBAL}$** 2,316,656; **AIC$_{GROUP\text{-}SPECIFIC}$** 2,316,406; **ΔAIC** 250
4. Postmenopausal females (HRT) vs Postmenopausal females (no HRT): **AIC$_{GLOBAL}$** 1,079,356.6; **AIC$_{GROUP\text{-}SPECIFIC}$** 1,079,343; **ΔAIC** 13.6
5. Postmenopausal females (HRT) vs Males: **AIC$_{GLOBAL}$** 2,279,735; **AIC$_{GROUP\text{-}SPECIFIC}$** 2,279,567; **ΔAIC** 168
6. Postmenopausal females (no HRT) vs Males: **AIC$_{GLOBAL}$** 2,430,711; **AIC$_{GROUP\text{-}SPECIFIC}$** 2,430,522; **ΔAIC** 189

Fig. 4a—Menopause Participants (HGS, controlling for body size)

HGS ~ group + s(age) + s(age, by = group) + ethnicity + season of attendance + s(UKB Assessment Centre, bs = "re") + s(dynamometer instance, bs = "re") + s(dynamometer instance, device day of use, bs = "re") + standing height^2 + body mass$^{2/3}$*

*'Premenopausal females', 'Postmenopausal females', and 'Males'

Fig. 4b—Menopause Participants (MM, controlling for body size)

MM ~ group + s(age) + s(age, by = group) + ethnicity + season of attendance + s(UKB Assessment Centre, bs = "re") + s(BIA impedance device, bs = "re") + s(BIA impedance device, device day of use, bs = "re") + standing height^2 + body mass$^{2/3}$*

*'Premenopausal females', 'Postmenopausal females', and 'Males'

Fig. 4c—Menopause Participants (MQ, controlling for body size)

MQ ~ group + s(age) + s(age, by = group) + ethnicity + season of attendance + s(UKB Assessment Centre, bs = "re") + s(dynamometer instance, bs = "re") + s(dynamometer instance, device day of use, bs = "re") + s(BIA impedance device, bs = "re") + s(BIA impedance device, device day of use, bs = "re") + standing height^2 + body mass$^{2/3}$*

*'Premenopausal females', 'Postmenopausal females', and 'Males'

Extended Data 4 a,c—a) Cross-sectional Participants (females only); c) Cross-sectional Participants (males only)

MM ~ s(age) + ethnicity + season of attendance + s(UKB Assessment Centre, bs = "re") + s(BIA impedance device, bs = "re") + s(BIA impedance device, device day of use, bs = "re") + standing height^2 + body mass$^{2/3}$

Extended Data 4 b,d—b) 2,224 female Longitudinal Participants ≥44.5 and ≤78 yr subjected to a DXA scan; d) 2,029 male Longitudinal Participants ≥44.5 and ≤78 yr subjected to a DXA scan

(NB DXA measurement device and UKB Assessment Centre were not included as covariates in this model, as all DXA scans were performed at the Cheadle Assessment Centre, and no information was available on whether multiple DXA devices were utilised)

DXA-derived estimate of mean arm muscle mass ~ s(age) + ethnicity + season of attendance + standing height^{2} + body mass$^{2/3}$

Extended Data 6a–c—Cross-sectional Participants

MQ ~ s(CSA) + sex + s(age) + s(CSA, by = sex) + s(age, by = sex) + ethnicity + season of attendance + s(UKB Assessment Centre, bs = "re") + s(dynamometer instance, bs = "re") + s(dynamometer instance, device day of use, bs = "re") + s(BIA impedance device, bs = "re") + s(BIA impedance device, device day of use, bs = "re")

Extended Data 7a—Cross-sectional Participants

MQ ~ s(PMSH) + sex + s(age) + s(PMSH, by = sex) + s(age, by = sex) + ethnicity + season of attendance + s(UKB Assessment Centre, bs = "re") + s(dynamometer instance, bs = "re") + s(dynamometer instance, device day of use, bs = "re") + s(BIA impedance device, bs = "re") + s(BIA impedance device, device day of use, bs = "re")

Extended Data 7c—Cross-sectional Participants

MM ~ s(PMSH) + sex + s(age) + s(PMSH, by = sex) + s(age, by = sex) + ethnicity + season of attendance + s(UKB Assessment Centre, bs = "re") + s(BIA impedance device, bs = "re") + s(BIA impedance device, device day of use, bs = "re")

Extended Data 7d—Cross-sectional Participants

CSA ~ s(PMSH) + sex + s(age) + s(PMSH, by = sex) + s(age, by = sex) + ethnicity + season of attendance + s(UKB Assessment Centre, bs = "re") + s(BIA impedance device, bs = "re") + s(BIA impedance device, device day of use, bs = "re")

Extended Data 7e—Cross-sectional Participants

HGS ~ s(PMSH) + sex + s(age) + s(PMSH, by = sex) + s(age, by = sex) + ethnicity + season of attendance + s(UKB Assessment Centre, bs = "re") + s(dynamometer instance, bs = "re") + s(dynamometer instance, device day of use, bs = "re")

Supplementary Information 3 Adjustment of hand grip strength readings

When plotting average HGS values against date of assessment for individual dynamometers, we observed evidence of drift in HGS readings for many devices (Fig. SI3.I). We therefore undertook to identify and remove readings taken under extreme drift, and to implement a HGS correction to minimise the impact of device drift on the remaining readings. Here, and throughout this Supplementary Information section, 'average HGS' refers to the average (mean) of left and right hand readings taken at a single assessment.

UK Biobank have provided dates at which dynamometers were recalibrated (Resources 2431–2433). For the purposes of quality control (QC) and adjustment, we treated measurements taken prior to a device's earliest recalibration date, between consecutive recalibration dates, and after its latest recalibration date as separate device 'instances'.

We first performed QC and adjustment of baseline HGS readings. We excluded readings taken on devices used for <500 participants (6,840 [1.4%] of 496,905 participants with baseline readings), and participants without HGS measures for both hands. We manually inspected plots for each dynamometer instance and identified clear temporal outliers and errant readings for exclusion (covering 10,495 participants in total) (e.g. Fig. SI3.I). Next, we excluded readings taken on dynamometer instances used for <250 participants (total 5,556 participants).

We quantified the drift exhibited by each dynamometer instance by fitting a regression model for average HGS that included linear, quadratic, and cubic terms for 'device day of use' (a value that equals zero on the earliest assessment date of a dynamometer instance, and increments by one for each subsequent calendar day). To control for participant-level differences that may influence HGS independently of systematic dynamometer effects we included sex, height (cm), weight (kg), age (yr), age^2 (yr^2), and self-reported ethnic background [b] as covariates. We could not include Assessment Centre or season of attendance as covariates, since both are correlated with device day of use (dynamometers were periodically moved between Assessment Centres) and lead to volatile cubic fits. A small number of dynamometer instances ('000193', '000194_1' and '000196_0') were highly unbalanced in terms of sex; for these dynamometer instances the cubic model was fitted to male data only.

Average HGS values were adjusted in a two-stage process that will be described in more detail below: (i) by removing readings taken under the most extreme drift, and (ii) by re-fitting the cubic models for drift, and correcting the remaining readings (Figure SI3.II).

(i) A preliminary adjustment was made to average HGS values using the estimated model parameters. Namely, the fitted parameter values for device day of use, device day of use^2, and device day of use^3 were used to calculate the contribution of device drift to each individual observation; this contribution was then subtracted from the observed value to give a preliminary adjusted value. Note that this process makes an explicit assumption that dynamometers are correctly calibrated at the beginning of each dynamometer instance. Subsequently, we identified and removed the 5% of all remaining 471,611 participants subject to the largest preliminary adjustments (i.e. absolute value of adjustment). We recorded the threshold adjustment value for exclusion (5.51 kgf). The preliminary adjustment values were then discarded.

[b] 'Asian', 'black', 'Chinese', 'mixed', 'other', 'unknown', or 'white', as determined from the 'Ethnic background' data field (Data-Field 21000).

(ii) We excluded any dynamometer instances with <250 observations remaining after step (i), and re-fitted the cubic regression model (identical in form to that described above) to the remaining readings for each dynamometer instance. This final model was used to generate refined adjustments for each observed average HGS reading. Finally, we performed post-adjustment filtering. First, we calculated the mean HGS across participants for each dynamometer instance, both pre- and post-adjustment. To identify and mitigate the most problematic dynamometers, we excluded dynamometer instances with the largest absolute change in mean value until 5% of participants had been excluded (10 of 204 dynamometer instances; implied threshold for dynamometer instance mean adjustment of 2.71 kgf). Second, we removed individual participants whose average HGS reading was adjusted by at least 5.51 kgf (i.e. the previously identified threshold adjustment value for exclusion).

This process of QC and adjustment resulted in adjusted baseline readings for 422,364 participants. The distribution of adjustments is shown in Figure SI3.III. The average adjustment was +0.33 kgf. In total, 168,862 participants (40.0%) had their HGS reading adjusted to a lower value and 249,847 (59.2%) had their reading adjusted to a higher value. 383,721 participants (90.9%) had their reading adjusted by <2.5 kgf.

HGS readings for the repeat assessment were adjusted using an identical process. However, due to the larger sample size at baseline, we applied the participant-specific and device instance-specific adjustment thresholds calculated using the baseline values (5.51 kgf and 2.71 kgf, respectively).

Note that the process of adjusting and excluding HGS readings described in this Supplementary Information was performed prior to the creation of the participant groups for our primary analyses (Supp. Info. 1). The substantial number of participants with missing adjusted HGS readings (e.g. ~15% of Cross-sectional Participants) were retained for the MM analyses where MM estimates were available (Supp. Info. 1).

While adjustment of HGS readings inevitably altered individual HGS-age trajectories (i.e. those of most Longitudinal Participants), Figs SI3.IV–V show that cohort-level HGS-age trajectories are very consistent between the unadjusted and adjusted HGS datasets, when assessed in the Cross-sectional Participants. We conclude that the general cohort-level trends in HGS and MQ reported in this study are not biased by our HGS adjustments.

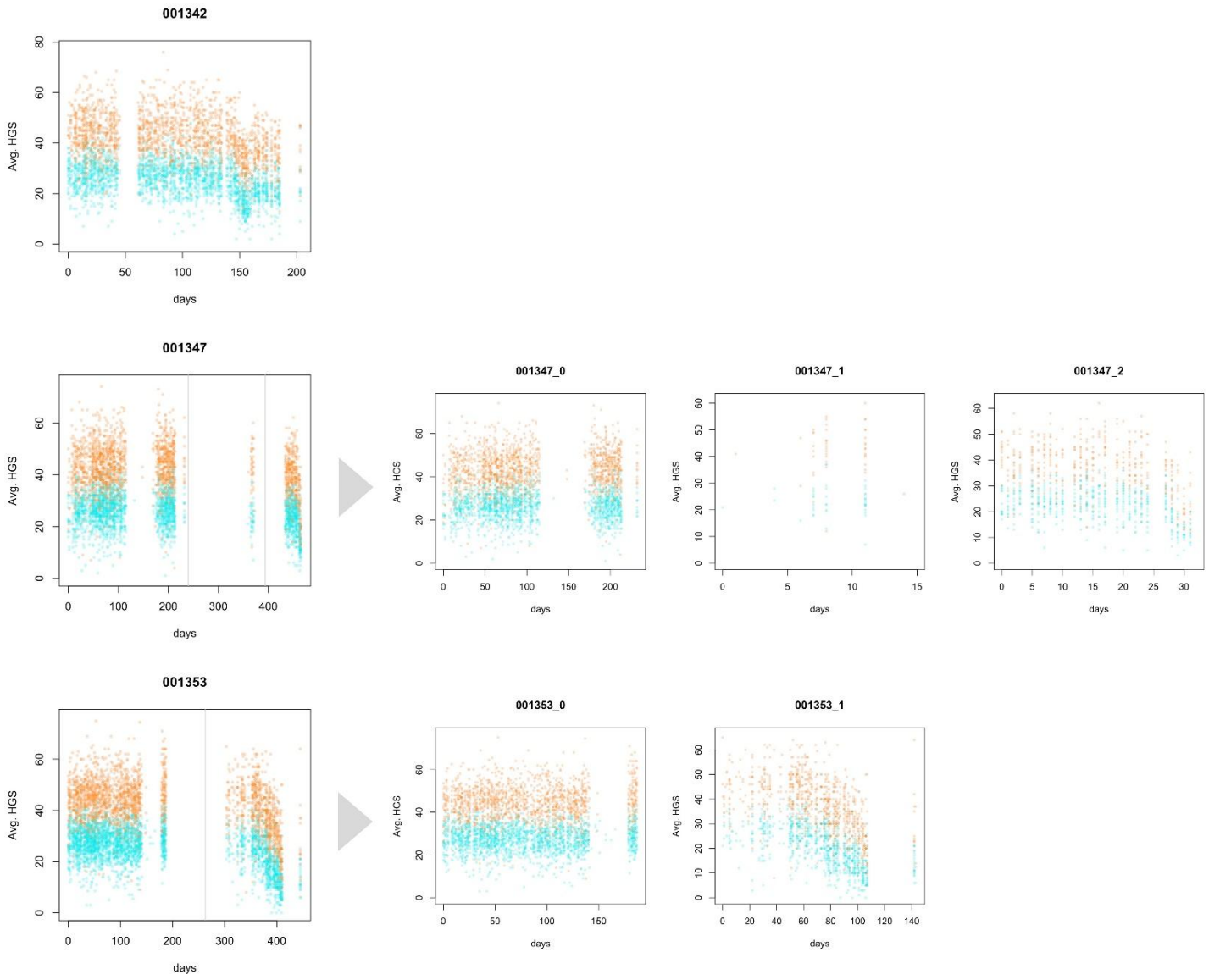

001342
Avg. HGS
days
001347
001347_0
001347_1
001347_2
001353
001353_0
001353_1

Figure SI3.I. Example of dynamometer drift. Plots of unadjusted average HGS values for three illustrative dynamometer devices (#001342, #001347, and #001353) are shown in the left column. Each point represents a participant; *y*-axis, average HGS in kgf; *x*-axis, 'device day of use'; point colour, sex (male = blue, female = pink). Vertical grey lines indicate dates at which UK Biobank recorded a calibration check (and if necessary, recalibration) of the device. Where calibration checks occurred, we split the data into 'dynamometer instances' with independent 'device days of use'. This is illustrated by the grey arrows and the plots on their right, which display the same data, but as separate dynamometer instances. Note that for device #001342 there were no recorded calibration checks, and all data is therefore treated as a single dynamometer instance. Note that dynamometer instance #001347_1 contains readings for <250 participants, and is therefore excluded from subsequent analysis (as described in detail in the Supplementary Information text).

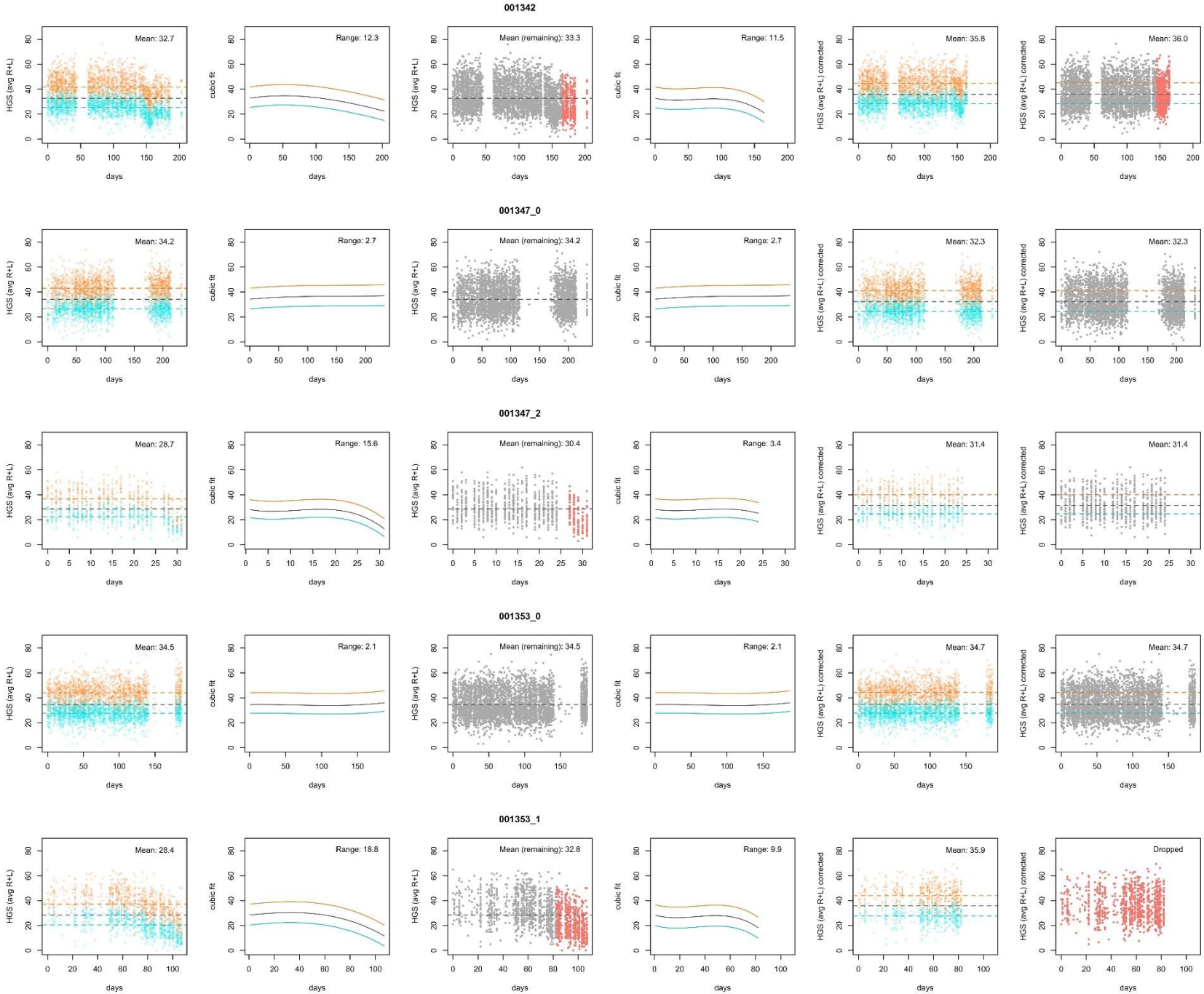

001342
Mean: 32.7
Range: 12.3
Mean (remaining): 33.3
Range: 11.5
Mean: 35.8
Mean: 36.0
001347_0
Mean: 34.2
Range: 2.7
Mean (remaining): 34.2
Range: 2.7
Mean: 32.3
Mean: 32.3
001347_2
Mean: 28.7
Range: 15.6
Mean (remaining): 30.4
Range: 3.4
Mean: 31.4
Mean: 31.4
001353_0
Mean: 34.5
Range: 2.1
Mean (remaining): 34.5
Range: 2.1
Mean: 34.7
Mean: 34.7
001353_1
Mean: 28.4
Range: 18.8
Mean (remaining): 32.8
Range: 9.9
Mean: 35.9
Dropped
HGS (avg R+L)
HGS (avg R+L) corrected
cubic fit
days

Figure SI3.II. Illustration of HGS QC and adjustment. These plots illustrate the subsequent steps of the QC and adjustment process for each of the dynamometer instances shown in Fig. SI3.I (excluding dynamometer instance #001347_1, which was excluded due to low numbers). **First column**: unadjusted data, plotted as described for Fig. SI3.I. Dashed lines indicate mean observed average HGS values (blue = male, red = female, grey = all). **Second column**: preliminary cubic fit to the observed values. Indicative lines are shown for males (blue), females (red), and all (grey), but since a single cubic model is fitted per dynamometer instance, these represent the same line shifted in position on the *y*-axis. **Third column**: unadjusted average HGS values are plotted again, coloured according to whether they are taken forward to the final cubic model; red points are excluded due to being among the 5% of participants (across all dynamometer instances) with the largest absolute preliminary adjustment value. **Fourth column**: refined cubic fit to the observed values, based on remaining participants only. **Fifth column**: average HGS values plotted after adjustment, but before the final filtering step. **Sixth column**: adjusted average HGS values are plotted again, coloured according to whether they are retained in the final dataset; red points are excluded due to being measured in a dynamometer instance among the top 5% according to overall mean adjustment (e.g. #001353_1 here) or due to exhibiting individual adjustments larger than the threshold identified from the preliminary fit (e.g. #001342 here). A more detailed description is provided in the Supplementary Information text.

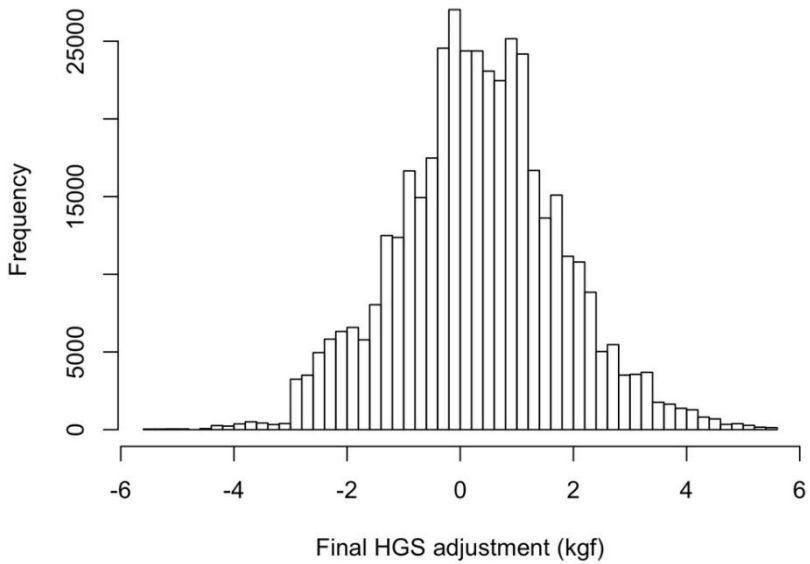
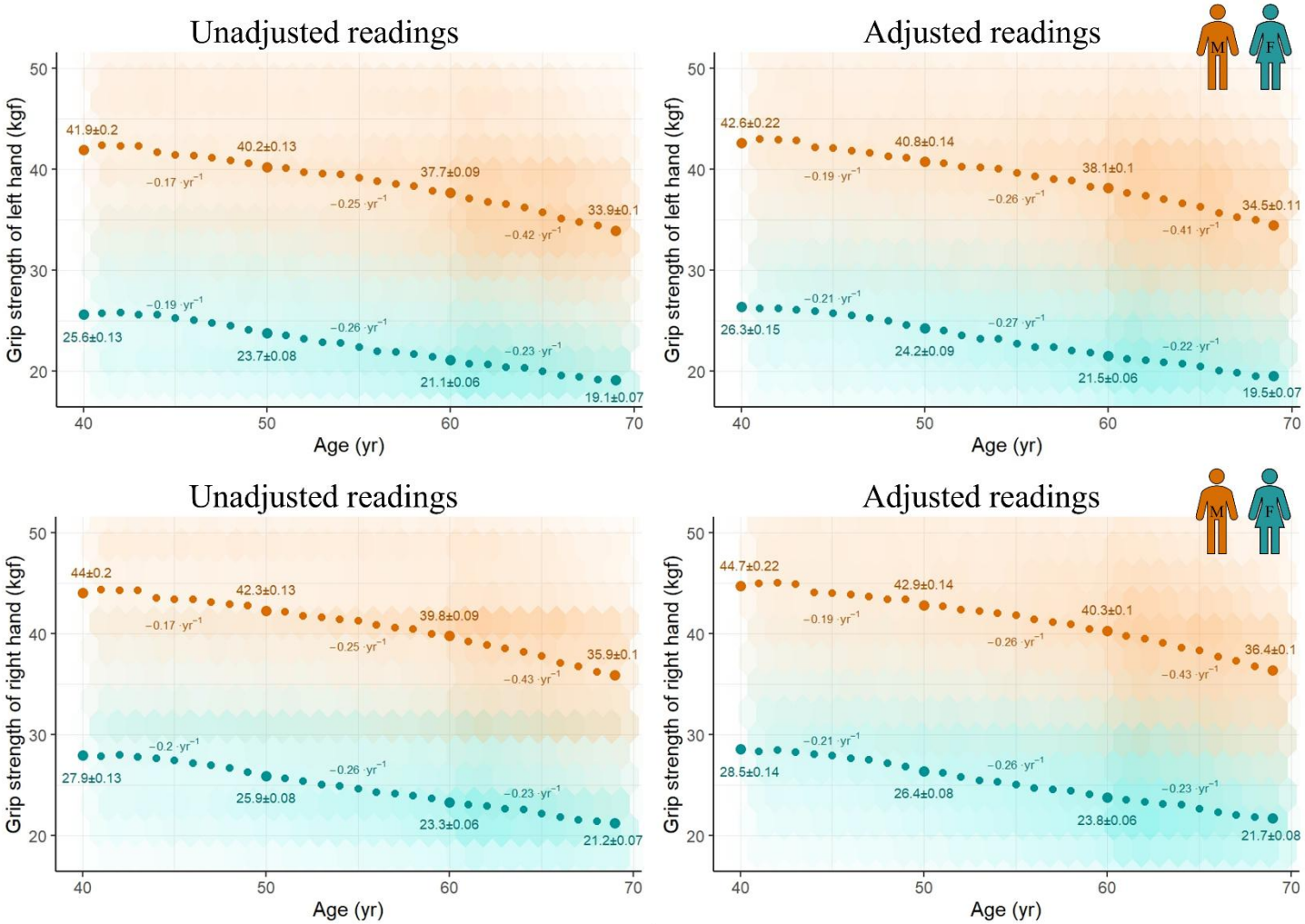
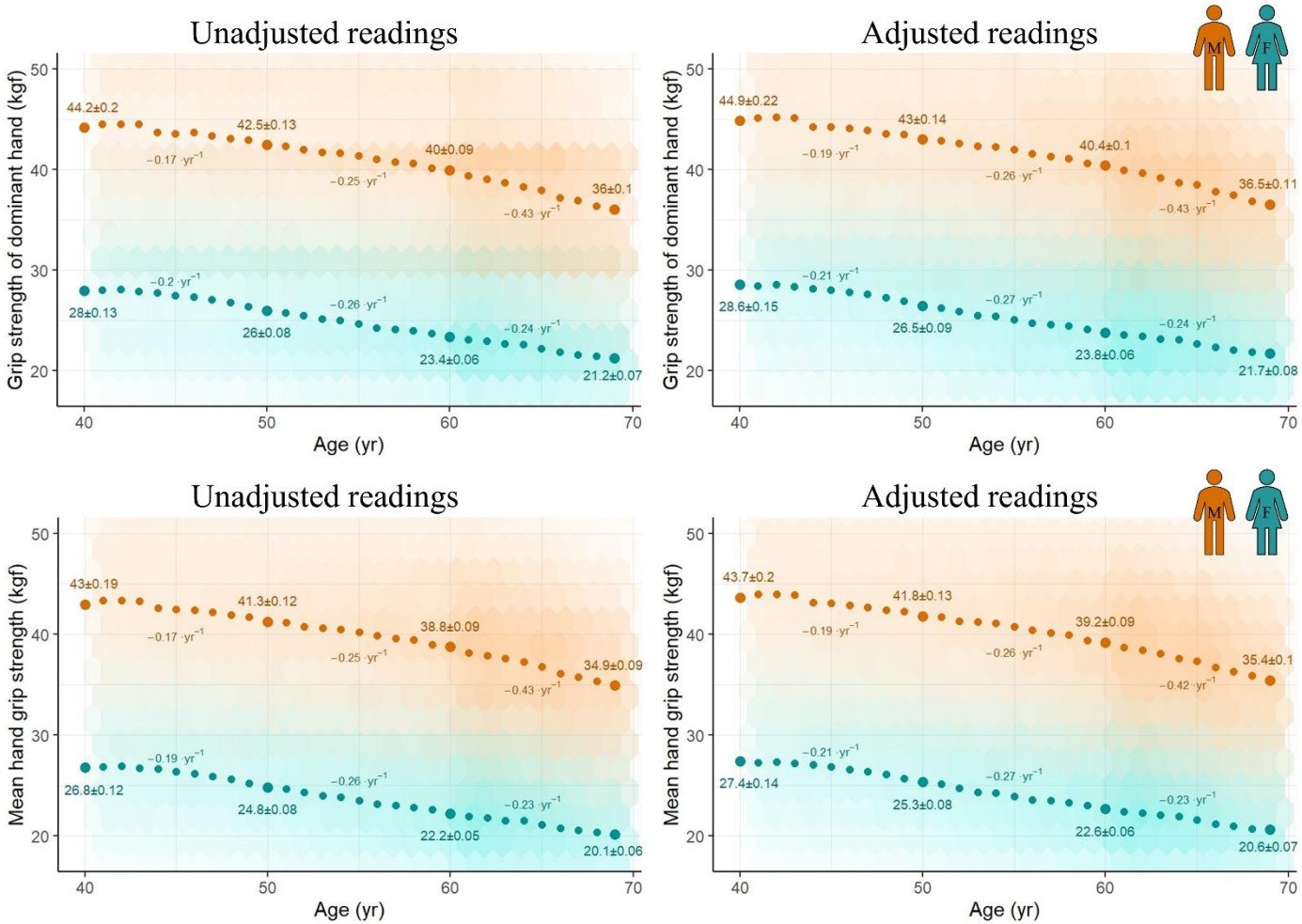


Figure SI3.III. Distribution of HGS adjustments. Histogram of adjustment values for the baseline HGS measures of 422,364 UKB participants.



SI3.IV. HGS-age trajectories are similar for unadjusted and adjusted HGS readings (left and right hand measures). Sex-specific relationships between grip strength of the left (top row) and right (bottom row) hand and age in the Cross-sectional Participants. Teal and orange represent females and males, respectively. Values (mean ± SEM) are observed values at 40, 50, 60, and 69 yr, and inferred annual change over the corresponding period. Observations are plotted using hexagonal bin plotting, with greater opacity indicating higher density of observations. *y*-axes are truncated to visualise better the general trends in the data.



SI3.V. HGS-age trajectories are similar for unadjusted and adjusted HGS readings (dominant hand and mean measures). Sex-specific relationships between grip strength of the dominant hand (top row) and mean HGS (bottom row) and age in the Cross-sectional Participants. Teal and orange represent females and males, respectively. Values (mean ± SEM) are observed values at 40, 50, 60, and 69 yr, and inferred annual change over the corresponding period. Observations are plotted using hexagonal bin plotting, with greater opacity indicating higher density of observations. *y*-axes are truncated to visualise better the general trends in the data. Dominant hand was determined by participants' responses to the question "Are you right or left handed?" (Data-Field 1707); participants who responded "Use both right and left hands equally" or "Prefer not to answer" were excluded from the dominant hand plots.

Supplementary Information 4 Muscle mass adjustment and estimation

Adjustment of BIA impedance values

A preliminary analysis of 45,686 participants who attended a repeat assessment revealed an apparent MM increase in an unexpectedly large proportion (~37.3%) of the 44,702 participants for whom two Tanita MM estimates were available. This phenomenon was particularly pronounced in females (~45.2%), especially those who were younger at baseline assessment (Fig. SI4.I). Because a widespread trend of age-related MM increase seemed biologically implausible—and also contradicted our cross-sectional finding of age-related MM decline—we decided to investigate further.

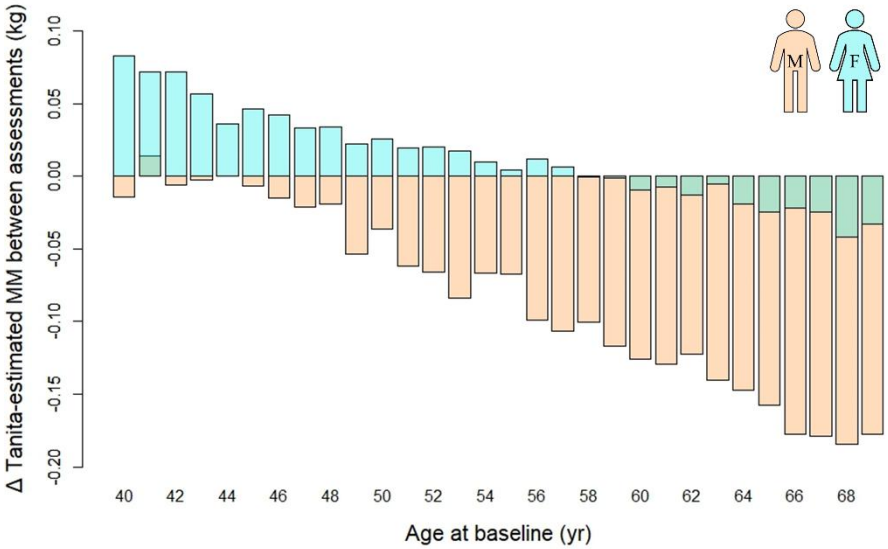


Figure SI4.I. Apparent age-related MM increase in over 1/3rd of participants. N = 44,702. Bars indicate average change in MM between baseline and repeat assessment for participants group by whole year of age at baseline. Cyan and orange represent females (N = 22,967) and males (21,735), respectively, with green indicating areas of overlap. Average interval between assessments was 8.92 ± 1.76 (SD) yr. MM estimates were generated by a Tanita Body Composition Analyser BC-418MA.

We first stratified the 44,702 participants by the UKB Assessment Centre at which their repeat assessment was performed. Age-related MM decline was observed on average for males at all three Assessment Centres. Female age-related MM decline was also observed on average in females assessed at the Reading (N = 6,288) and Newcastle (N = 11,292) centres, whereas females who attended the Cheadle (N = 27,122) centre tended to exhibit age-related MM increase on average (Fig. SI4.IIa). Two of the seven BIA impedance devices used at repeat assessments were based at Reading (#001480 and #003169), two at Newcastle (#003167 and #003168), and three at Cheadle (#001477, #001486, and #003173). None of these devices was transferred between Assessment Centres during the assessment period. Females assessed at Cheadle were reported to have—on average—greater arm MM at repeat than at baseline assessment, regardless of the BIA impedance device used at repeat assessment; in contrast, all four of the devices based at Reading and Newcastle were associated with age-related female MM decline (Fig. SI4.IIb). Male age-related MM decline held on average across all seven BIA impedance devices, but was markedly less pronounced in participants assessed at Cheadle (Fig. SI4.IIb). The MM estimates accorded with the lower raw arm impedance values (Ω) generally observed in the Cheadle data (Fig. SI4.IIc).

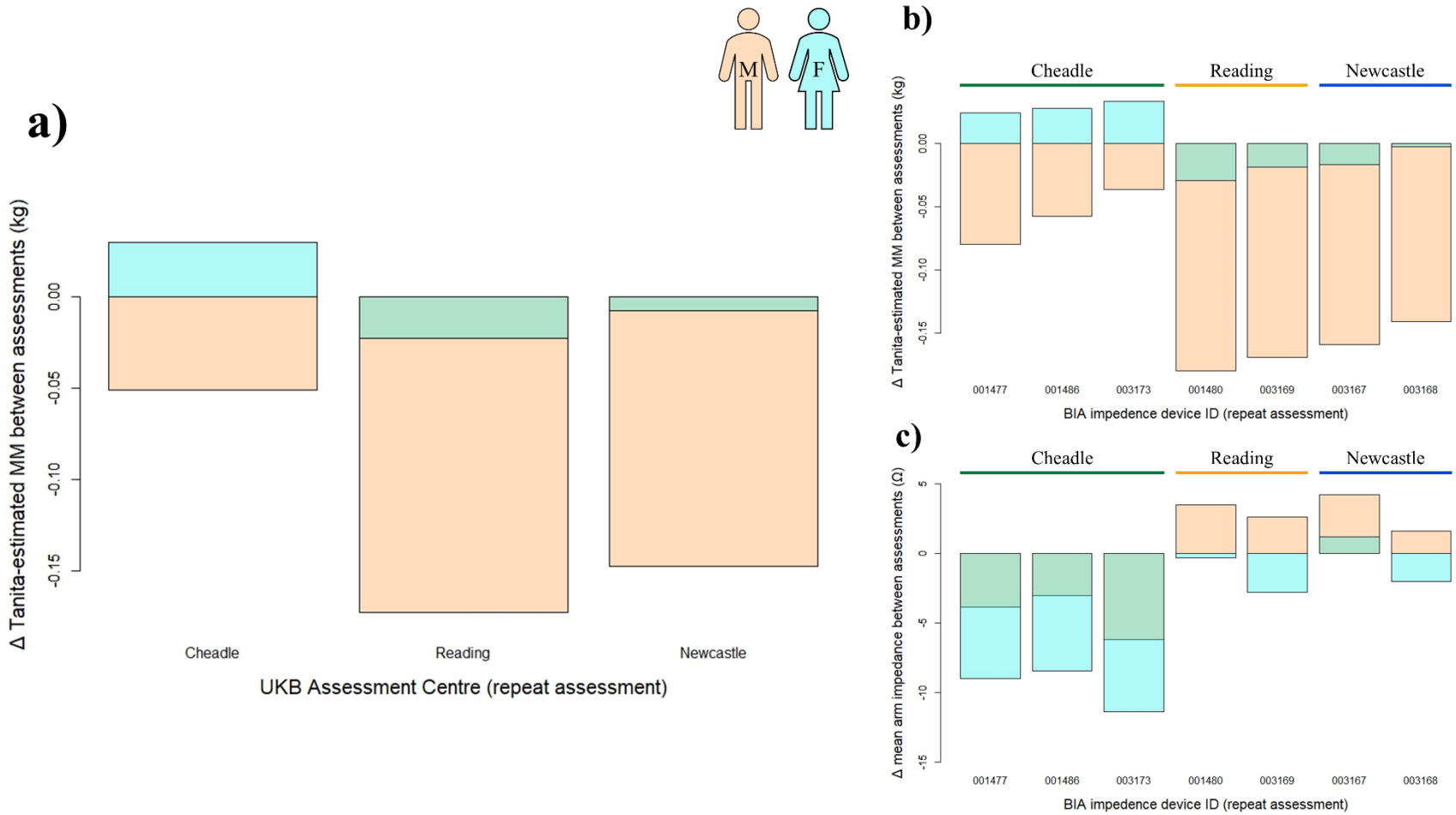


Figure SI4.II. Tanita BIA devices based in Cheadle give aberrant readings. 44,702 participants assessed longitudinally, with an average of 8.92 ± 1.76 (SD) yr between baseline and repeat assessment. Cyan and orange represent females (N = 22,967) and males (21,735), respectively, with green indicating areas of overlap. **a,b.** Females whose repeat assessment was conducted in Cheadle were reported to have—on average—greater arm MM at repeat than at baseline assessment, regardless of the BIA impedance device used at repeat assessment, whereas average female arm MM was lower at repeat assessment across all four BIA impedance devices base in Reading and Newcastle. Reported loss of MM at repeat assessment in males was also markedly less pronounced in Cheadle. **c**. Raw arm impedance data accord with the MM estimates, as lower impedance values are associated with higher estimated MM, due to the relatively low impedance of muscle tissue *versus* fat and bone.

We next checked whether differences between Cheadle and the other two Assessment Centres remained after controlling for potential demographic differences between them. The *matchit* function of the R 'MatchIt' R package (referenced in main manuscript) was used to match Cheadle participants with participants of the same sex who were assessed at either Reading *or* Newcastle ('Other'), based on similarity across predefined criteria, namely age at baseline; interval between baseline and repeat assessment ('Interval'); standing height at baseline; and body mass at baseline. Nearest neighbour matching without replacement was performed, with Mahalanobis distances used as the measure of similarity. The *caliper* argument was used to prevent matching of participants whose values for a given criterion differed by more than a predefined amount (age at baseline: ≤0.5 yr; Interval: ≤0.45 yr; standing height at baseline: ≤1 cm; body mass at baseline: ≤1 kg). This matching procedure returned 4,722 same-sex Cheadle:Other participant pairs (N = 9,444) that were well-matched across all matching criteria (all SMDs ≥-0.0017 and ≤0.0267; all VRs ≥0.9852 and ≤1.0008; see Griefer [2023]—referenced in main manuscript). Nevertheless, a clear difference was still apparent between Cheadle and the other two Assessment Centres (Fig. SI4.III, overleaf), confirming that the discrepancy was not the result of demographic differences in the matching criteria between Assessment Centres. We thus concluded that BIA measurement error in Cheadle was likely the cause of the discrepancy, and that adjustment of the Cheadle BIA impedance values was therefore warranted.

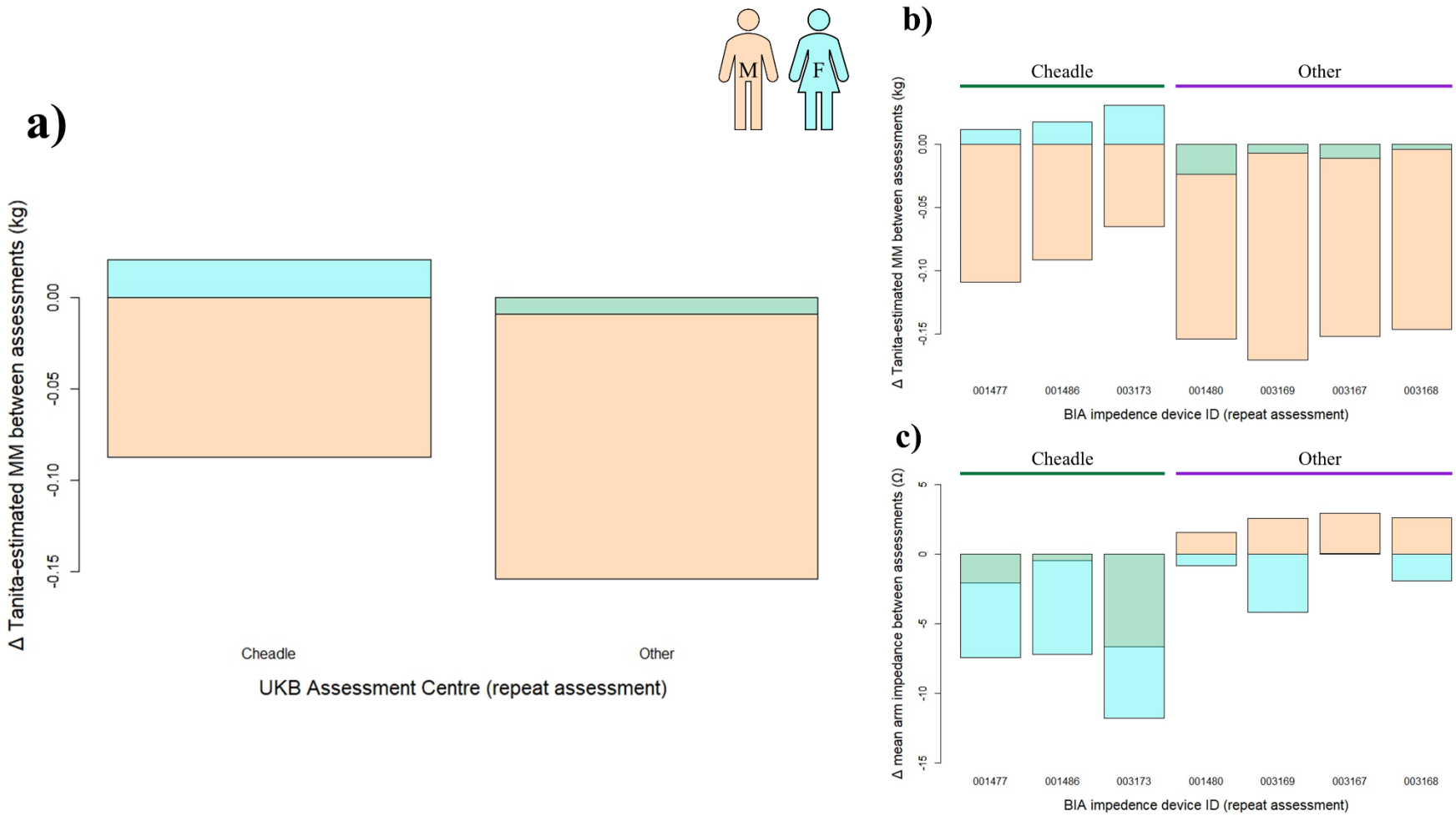


Figure SI4.III. Discrepancies between UKB Assessment Centres persist after controlling for demographic differences. N = 9,444, comprising 4,722 same-sex Cheadle:Other participant pairs matched on similarity in age and body size at baseline, as well as interval (yr) between baseline and repeat assessment. 'Other' denotes participants whose repeat assessment was conducted at either Reading *or* Newcastle. Cyan and orange represent females (N = 5,110; 2,555 pairs) and males (N = 4,334; 2,167 pairs), respectively, with green indicating areas of overlap. The figure panels are analogous to those in Fig. SI4.IIa–c, and so demonstrate that the aberrant Cheadle data are not explicable by demographic differences in the matching criteria between Cheadle and the other two Assessment Centres.

As no information was available on BIA impedance device performance, accuracy, or calibration history, it was necessary for us to develop our own procedure for adjusting impedance values from the BIA impedance devices based at Cheadle.

We first formally assessed whether the data supported the necessity of adjustments at the level of BIA impedance device and/or Assessment Centre. To do so, we used the *lm* base R function to fit multiple linear regression models to various subsets of the repeat assessment data, with mean arm impedance (Ω) as the dependent variable, and age and sex as covariates. We then tested whether including BIA impedance device *or* Assessment Centre as an additional covariate in these models improved their performance relative to their respective, simpler (nested) models which did not include them. Likelihood-ratio tests—as performed using the *lrtest* function of the 'lmtest' R package (version 0.9-40: Hothorn et al. 2015)—were used for comparative assessment of model performance. Inclusion of Assessment Centre as a covariate significantly improved model performance in the full participant dataset (P <0.001; FPR <0.001), but not in a model fitted to Reading and Newcastle data only (P = 0.58; FPR = 0.5). Furthermore, inclusion of BIA impedance device as a covariate did not improve performance in this model (P = 0.02; FPR = 0.18). We therefore concluded that the four BIA impedance devices based at Reading and Newcastle behaved in an essentially similar manner. We found that including BIA impedance device as a covariate significantly improved performance in a model fitted to Cheadle data only (P <0.001; FPR <0.001), but not when readings from BIA impedance device #003173 were omitted (i.e. with only readings from devices #001477 and #001486 included; P = 0.01; FPR = 0.11). We therefore deemed it appropriate to make two separate adjustments to the Cheadle impedance values: one to impedances values from device #003173, and a second to impedance values from devices #001477 and #001486.

Calculation of MM (see 'Generating novel muscle mass estimates' below), required several impedance measures, namely: whole-body impedance, 'arms impedance index' (i.e. height^2 [in metres^2] divided by the sum of the left and right arm impedance values), and 'legs impedance index' (i.e. height^2 [in metres^2] divided by the sum of the left and right leg impedance values). It was therefore necessary to calculate separate adjusted values of summed arm impedance, summed leg impedance, and whole-body impedance. For each of these impedance measures, we sought to calculate two simple scalar adjustment values for the correction of values ascertained at Cheadle (i.e. independently of age, sex, ethnicity, or other covariates): one for participants assessed using impedance device #003173, and one for participants assessed using either #001477 or #001486.

To calculate adjustment values for the summed arm impedance data, we again performed participant matching on the initial pool of 45,686 participants, in the manner described above. 'Cheadle' and 'Other' participants of the same sex were matched by age at baseline (≤ 0.5 yr), Interval (≤ 0.45 yr), summed arm impedance at baseline (≤ 2 Ω), and Townsend Deprivation Score (a measure of material deprivation: min. -6.26; max. 10.1; mean -1.89) at recruitment (≤ 1). The matching procedure returned 3,746 Cheadle:Other participant pairs. Our approach implicitly assumed that any within-pair discrepancy in summed arm impedance at repeat assessment was the result of measurement error in Cheadle. Therefore, the summed arm impedance values of each pair member at repeat assessment were essentially treated as 'observed' (Cheadle participant) and

‘expected’ (Other participant) values. The average difference between these observed and expected values was then calculated across participant pairs, with the two mean differences—i.e. one mean for the Cheadle participants assessed on impedance device #003173, and a second for those assessed on either #001477 or #001486—subsequently used as the final adjustment values. A comparable procedure was performed to calculate adjustment values for the summed leg impedance data (2,385 Cheadle:Other pairs), but with summed leg impedance (≤ 2 Ω) replacing summed arm impedance, and with number of days per week of vigorous physical activity (10+ minutes) (‘vigorous physical activity’) at baseline (≤ 1 day) included as an additional matching parameter. This procedure was then performed for a third and final time to calculate adjustment values for whole-body impedance data (2,883 Cheadle:Other pairs), with whole-body impedance at baseline (≤ 1 Ω) replacing summed arm/leg impedance as the impedance-related matching parameter, and without the inclusion of vigorous physical activity at baseline as a matching parameter. Note that a new matched subset was generated in each case (e.g. the ‘leg’ subset was not generated from the ‘arm’ subset), and that these subsets were used only to calculate adjustment values (i.e. none of these three matched subsets was included in any subsequent analyses).

All three subsets were well-matched across their respective matching criteria (all SMDs ≥-0.0058 and ≤0.0353; all VRs ≥0.9514 and ≤1.0045). Significance tests (Cheadle vs Other)—on age at baseline, Interval, number of days per week of moderate physical activity (10+ minutes) (‘moderate physical activity’) at baseline, vigorous physical activity at baseline, Townsend Deprivation Score at recruitment, and the relevant impedance value at baseline—were all unequivocally non-significant in the three matched subsets (all *P*-values and FPRs >0.09), with the exception of moderate physical activity at baseline in the ‘legs’ subset (*P* = 0.01; FPR = 0.12). Underlying demographic differences between the Cheadle and Other participant groups were therefore considered sufficiently minor to justify treating the impedance values of Other participants as the ‘expected’ impedance values of their Cheadle counterparts. The final adjustment values—as subsequently applied to the respective raw impedance values—are given in Table SI4.I.

Cheadle BIA impedance device #	Summed arm impedance	Summed leg impedance	Whole-body impedance
003173	+19.81	+4.29	+7.74
001477 / 001486	+15	+3.19	+6.88

Table SI4.I. Adjustment values (Ω) for raw impedance values from the three BIA impedance devices based in Cheadle.

Once adjusted, the Cheadle impedance data proved to be broadly similar to the Reading and Newcastle impedance data (Fig. SI4.IV, overleaf), which we interpreted as evidence of a successful and appropriate adjustment. These adjusted impedance data were then used in the generation of novel estimates of mean arm muscle mass.

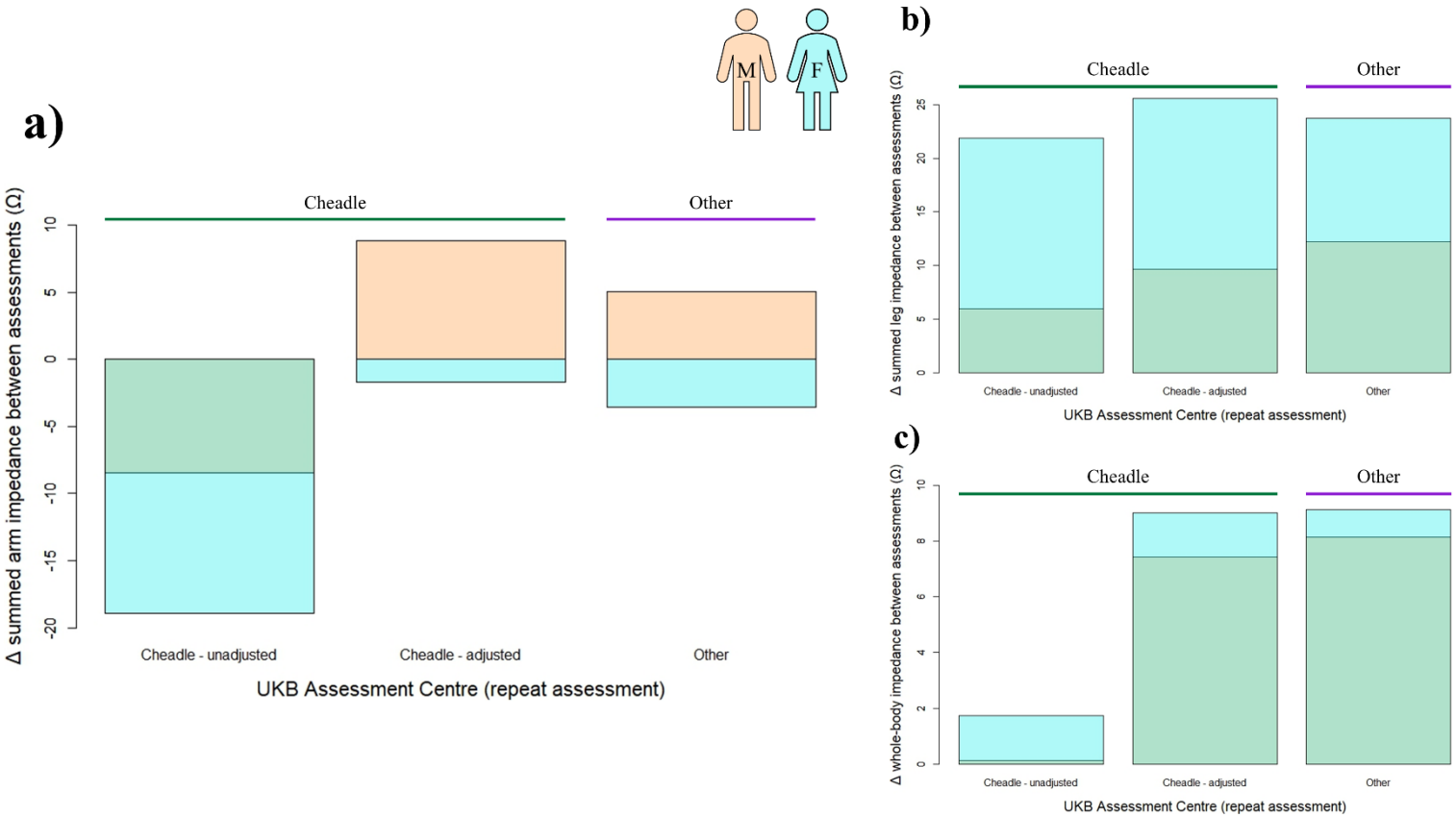


Figure SI4.IV. Adjustment of BIA impedance values minimises differences between UKB Assessment Centres. N = 9,444. These participants are the same as those shown in Fig. III, i.e. 4,722 same-sex Cheadle:Other participant pairs matched on similarity in age and body size at baseline, as well as interval (yr) between baseline and repeat assessment. 'Other' denotes participants whose repeat assessment was conducted at either Reading *or* Newcastle. Cyan and orange represent females (N = 5,110; 2,555 pairs) and males (N = 4,334; 2,167 pairs), respectively, with green indicating areas of overlap. The adjusted Cheadle BIA impedance values accord more closely with the Other values than do the unadjusted Cheadle BIA impedance values, particularly in the case of summed arm (**a**) and whole-body (**c**) impedance. The difference in summed leg impedance values between Cheadle and the other two Assessment Centres was much more modest (**b**).

Generating novel muscle mass estimates

We generated novel MM estimates using sex-specific equations based on those of Powell et al. (2020) ('Powell' equations; see main manuscript). Sex-specific equations produced substantially better MM estimates in preliminary analysis than did a single equation with sex included as a covariate (data not shown). Derivation and validation of equations was performed in a subset of 4,182 participants who were subjected to both a BIA reading and a DXA scan during their repeat assessment, and for whom all other necessary impedance and anthropometric measures were available. DXA-derived MM estimates were used as reference values against which to develop the predictive equations.

To develop a female-specific equation, we first split the female participants (N = 2,179) 1:1 into Derivation and Validation groups. The following linear model was then fitted to the Derivation group using the *lm* base R function:

DXA-derived total arm MM estimate (g) ~ age (yr) + standing height (cm) + body mass (kg) + waist circumference (cm) + hip circumference (cm) + arms impedance index (au) + legs impedance index (au) + whole-body impedance (Ω)

where 'arms impedance index' = (height^2 [in metres^2] / summed arm impedance), and 'legs impedance index' = (height^2 [in metres^2] / summed leg impedance). The *ols_step_best_subset* function of the 'olsrr' R package (version 0.5.3: Hebbali 2020) was then used to determine which set of independent variables to retain in the final model, with AIC values as the objective function for selection of model terms. The *ols_step_best_subset* function works by fitting all possible models given the independent variables supplied, and returning the optimal model (in this case, the model with the lowest AIC value) for each number of independent variables (i.e. for models fitted using 1, 2 … k independent variables, where k = no. independent variables). This final model was then used to estimate total arm MM (g) in the Validation group via the *predict* function in the R 'stats' base package. Finally, both DXA-derived total arm MM (g) estimates and our own total arm MM (g) estimates were divided by 2,000, to convert values to mean arm MM (kg).

This modelling procedure was performed twice in the females: once using unadjusted Ω values, and once using adjusted Ω values (see previous section 'Adjustment of BIA impedance values'). Analogous analyses were also performed for males (N = 2,003), resulting in a total of four models, the coefficients of which are given in the Table SI4.II.

	Females (unadjusted Ω values)	Females (adjusted Ω values)	Reference model: females (Powell)	Males (unadjusted Ω values)	Males (adjusted Ω values)	Reference model: males (Powell)
Intercept	614.1	355.7	2,479	2,012	2,491	6,738
Age (yr)	-4.425	-2.374	-7.9	-14.19	-17.95	-15.47
Standing height (cm)	15.31	15.21	0	21.26	0	0
Body mass (kg)	22.51	22.28	34.5	20.97	18.56	62.5
Waist circumference (cm)	4.353	4.25	2.008	-12.63	-9.963	-18.93
Hip circumference (cm)	-13.6	-13.28	-19.87	10.04	15.15	-31.58
Arms impedance index (au)	478,800	525,600	724,356	747,600	969,500	929,870
Legs impedance index (au)	0	0	-118,528	-222,000	-135,300	-388,873
Whole-body impedance (Ω)	-1.573	-1.538	-1.196	-4.332	-1.876	-3.742

Table SI4.II. Coefficients in final models. Covariates that were dropped from final models are indicated by a coefficient of zero. The coefficients of the Powell equations are also provided for reference.

Comparative analysis of muscle mass estimates

To determine which set of BIA-derived MM estimates was most similar to the DXA MM estimates, we calculated mean prediction error, maximum prediction error, and root mean square error (RMSE—calculated using the *rsme* function of the 'Metrics' R package [version 0.1.4: Hamner et al. 2022]) for the Tanita estimates, the Powell estimates (with unadjusted and adjusted Ω values), and our novel estimates (with unadjusted and adjusted Ω values in the corresponding derived models). We also performed pairwise model comparisons of predictive accuracy with Diebold-Mariano tests using the *dm.test* function of the 'forecast' package (version 8.21: Hyndman et al. 2020). Our novel MM estimates proved to be significantly more similar to the DXA MM estimates than did the Powell and Tanita MM estimates, regardless of whether unadjusted or adjusted impedance values were used for model derivation; however, no difference in predictive accuracy was observed between these two models (Tables SI4.III–IV). Nevertheless, we opted to use the MM estimates generated using adjusted impedance values for our subsequent analyses, due to the

superior mean prediction error, maximum prediction error, and RMSE values, along with the strong prior evidence that adjustment of the Cheadle BIA impedance data was warranted (see 'Adjustment of BIA impedance values').

	Mean prediction error (kg)	**Max. prediction error (+ve or -ve) (kg)**	**Root Mean Square Error (RMSE) (kg)**
Tanita estimates	-0.031	1.569	0.236
Powell estimates (unadjusted Ω values)	0.118	1.598	0.259
Powell estimates (adjusted Ω values)	0.18	1.435	0.291
Novel estimates (unadjusted Ω values)	-0.007	1.35	0.244
Novel estimates (adjusted Ω values)	0.003	1.2	0.211

Table SI4.III. Analysis of similarity of DXA-derived MM estimates to five other sets of MM estimates. 'Error' was defined as a participant's DXA MM estimate—implicitly treated as the 'true' MM value—minus their corresponding BIA-derived MM estimate.

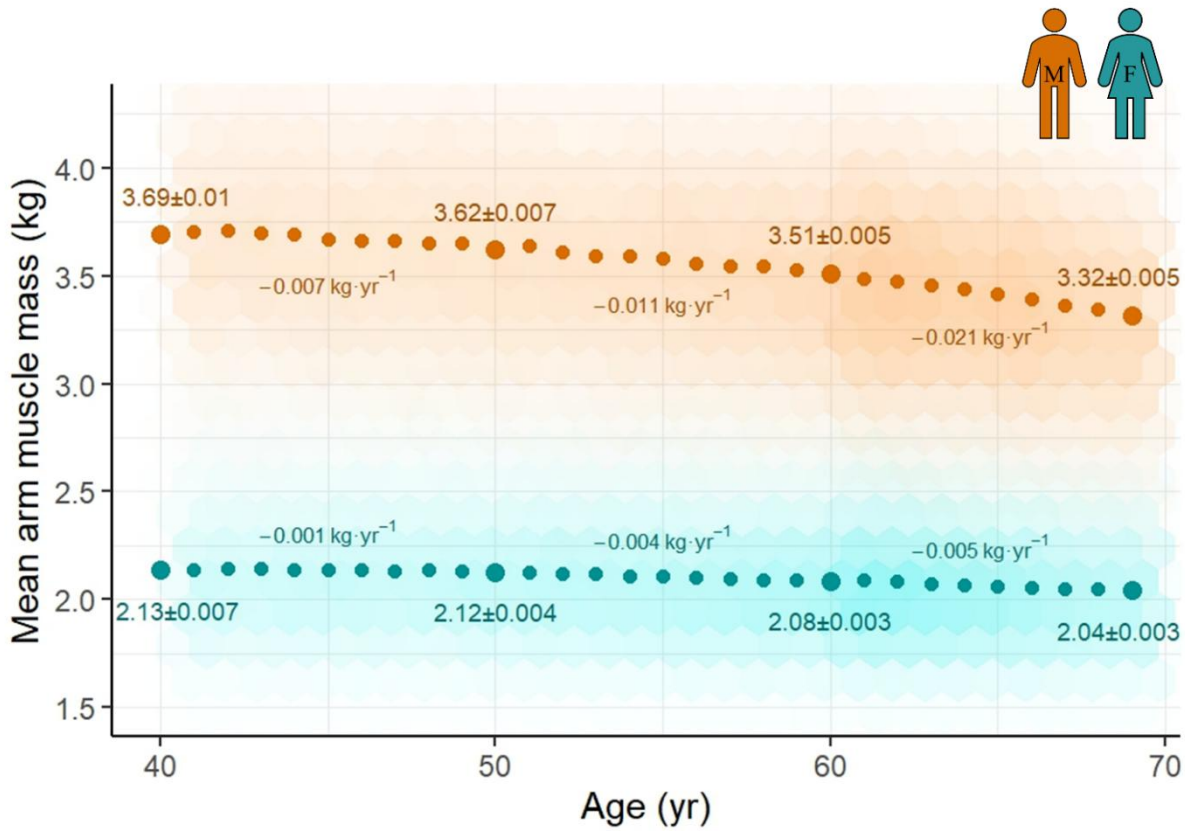
	Tanita estimates	**Powell estimates (unadjusted Ω values)**	**Powell estimates (adjusted Ω values)**	**Novel estimates (unadjusted Ω values)**	**Novel estimates (adjusted Ω values)**
Tanita estimates	-	-	-	-	-
Powell estimates (unadjusted Ω values)	< 0.001 (< 0.001)	-	-	-	-
Powell estimates (adjusted Ω values)	< 0.001 (< 0.001)	< 0.001 (< 0.001)	-	-	-
Novel estimates (unadjusted Ω values)	< 0.001 (< 0.001)	< 0.001 (< 0.001)	< 0.001 (< 0.001)	-	-
Novel estimates (adjusted Ω values)	< 0.001 (< 0.001)	< 0.001 (< 0.001)	< 0.001 (< 0.001)	0.73 (0.5)	-

Table SI4.IV. Results of Diebold-Mariano tests of predictive accuracy. Test results are reported as '*P*-value (FPR score)'.

Supplementary Information 5 UK Biobank Data-Fields included in this study

eid; sex_f31; year_of_birth_f34; hand_grip_dynamometer_device_id_f38; impedance_device_id_f43; hand_grip_strength_left_f46; hand_grip_strength_right_f47; waist_circumference_f48; hip_circumference_f49; standing_height_f50; month_of_birth_f52; date_of_attending_assessment_centre_f53; uk_biobank_assessment_centre_f54; number_of_selfreported_cancers_f134; townsend_deprivation_index_at_recruitment_f189; number_of_daysweek_of_moderate_physical_activity_10_minutes_f884; number_of_daysweek_of_vigorous_physical_activity_10_minutes_f904; handedness_chiralitylaterality_f1707; had_menopause_f2724; ever_used_hormonereplacement_therapy_hrt_f2814; age_at_menopause_last_menstrual_period_f3581; ethnic_background_f21000; body_mass_index_bmi_f21001; weight_f21002; genetic_sex_f22001; impedance_of_whole_body_f23106; impedance_of_leg_right_f23107; impedance_of_leg_left_f23108; impedance_of_arm_right_f23109; impedance_of_arm_left_f23110; arm_predicted_mass_right_f23122; arm_fat_mass_left_f23124; arm_lean_mass_left_f23250; arm_lean_mass_right_f23254

Supplementary Information 6 Muscle mass of participants with no self-reported history of cancer



SI6 Excluding participants with self-reported history of cancer has little effect on population-level MM trends. Scatterplot of the relationship between MM and age in 396,177 Cross-sectional Participants with no self-reported history of cancer (i.e. those with a value of '0' for 'Number of self-reported cancers'—Data-Field 134). Teal and orange represent females (N = 213,212) and males (N = 182,965), respectively. Values (mean ± SEM) are observed values at 40, 50, 60, and 69 yr, and inferred annual change over the corresponding 9–10 year period. Observations are plotted using hexagonal bin plotting, with greater opacity indicating higher density of observations. *y*-axes are truncated to visualise better the general trends in the data. Note that the results are essentially identical to those of the Cross-sectional Participants as a whole (main manuscript, Fig. 2a).

Supplementary Information 7 Ulna length and muscle quality

The example below illustrates how estimates of MQ are affected by the use of observed standing height *versus* predicted maximum standing height (PMSH) in MQ calculations.

In this example, MQ is estimated at both 40 and 69 yr—i.e. at both ends of the Cross-sectional Participants age range—using the mean female Cross-sectional Participant values for HGS (23.65 kgf) and MM (2.1 kg). These values are then used to estimate MQ using i) observed standing height and ii) PMSH. Even in the absence of true MQ decline (i.e. no change of HGS, UL, and MM), general age-related decline in standing height will affect MQ—as estimated in this study—due to the lower *estimated* UL of older, shorter participants (Online Methods). In this example, a 69-year-old female of mean observed standing height for her age group (160.43 cm) is predicted to have a MQ value 6.7% lower (-18.98 au) than that of a 40-year-old female of mean observed standing height for her age group (164.5 cm) (283.23 vs 264.25 au), with this difference wholly due to the shorter estimated ulna length (UL) of the older female (25.15 vs 23.46 cm—1.69 cm). If the equivalent analysis is performed using PMSH rather than observed standing height, the difference in predicted UL is attenuated (25.15 vs 24.16 cm—0.99 cm), and the difference in MQ reduces to -3.91% (-11.08 au). While a birth cohort effect (i.e. a trend towards younger individuals attaining greater maximum standing heights) probably contributes to the lower average observed standing height of older UKB participants, this example nevertheless demonstrates how an artefactual age-related decline in MQ may arise simply from participants experiencing a normal age-related reduction in standing height. Calculating UL using PMSH rather than observed standing height in this study was thus warranted to minimise the impact of this phenomenon.

Mean values for female Cross-sectional Participants

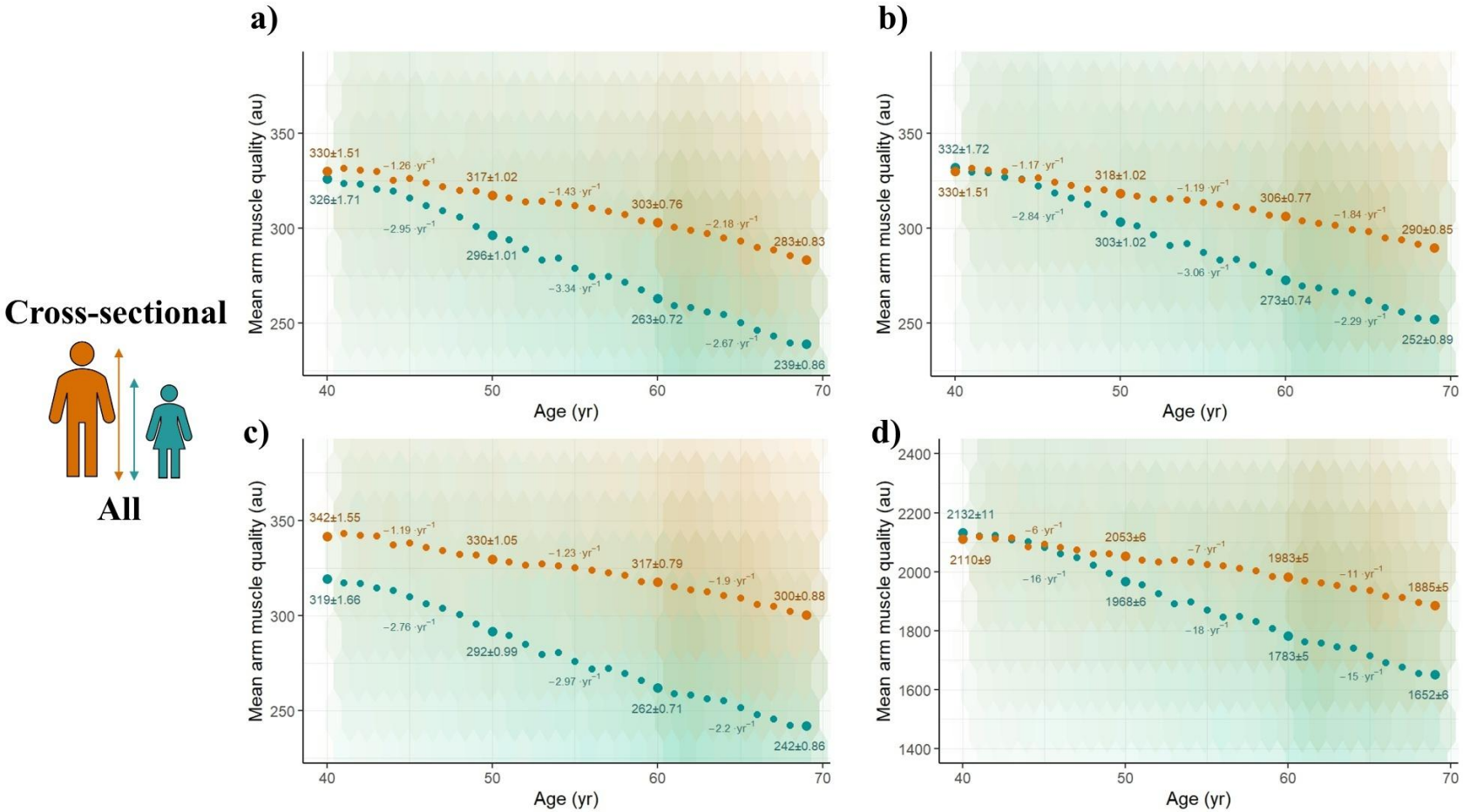
HGS (kgf) = 23.65
MM (kg) = 2.1
Observed standing height (cm) at 40 yr = 164.5
Observed standing height (cm) at 69 yr = 160.43
PMSH (cm) at 40 yr = 164.53
PMSH (cm) at 69 yr = 162.37
UL (cm) at 40 yr, calculated using observed standing height = 25.15
UL (cm) at 69 yr, calculated using observed standing height = 23.46
UL (cm) at 40 yr, calculated using PMSH = 25.15
UL (cm) at 69 yr, calculated using PMSH = 24.16

MQ calculations
CSA = MM / UL
MQ = HGS / CSA

MQ (observed standing height)
At 40 yr
CSA = 2.1 / 25.15 = 0.0835
MQ = 23.65 / 0.0835 = 283.23 au
At 69 yr
CSA = 2.1 / 23.46 = 0.0895
MQ = 23.65 / 0.0895 = 264.25 au
Difference
-18.98 au (-6.7%)

MQ (PMSH)
At 40 yr
CSA = 2.1 / 25.15 = 0.0835
MQ = 23.65 / 0.0835 = 283.23 au
At 69 yr
CSA = 2.1 / 24.16 = 0.0869
MQ = 23.65 / 0.0869 = 272.15 au
Difference
-11.08 au (-3.91%)

Supplementary Information 8 Muscle quality calculated with alternative estimates of ulna length



SI8 Muscle quality decline in greater in females irrespective of how ulna length is estimated. Sex-specific relationships between muscle quality and age are shown for the Cross-sectional Participants. Teal and orange represent females and males, respectively. Values (mean ± SEM) are observed values at 40, 50, 60, and 69 yr, and inferred

annual change over the corresponding 9–10 year period. Large points represent means of all observed measures for each year age-group. The observed data are plotted using hexagonal bin plotting, with greater opacity indicating higher density of observations. *y*-axes are truncated to visualise better the general trends in the data. **a.** Muscle quality is calculated using observed standing height (cm) rather than predicted maximum standing height (PMSH—Methods; Supp. Info. 7). **b.** Female ulna length (UL) is estimated using a re-arranged version of the Madden et al. (2020) (full reference in main manuscript) equation given in Table 5, Equation F (Female)—i.e. *UL = (PMSH-84.67)/3.14*—as opposed to the Elia (2003) (full reference in main manuscript) equation used in the primary analyses (Online Methods). Male muscle quality estimates are unchanged. **c.** The Madden et al. equations for Asian, black, and white individuals have ethnicity-specific slopes (Asian = 3.26; black/white = 3.14), and sex-and-ethnicity-specific intercepts (Asian females = 77.62; Asian males = 83.58; black females = 79.55; black males = 85.8; white females = 84.67; white males = 90.92). UL is calculated using the appropriate ethnicity-specific slope, and an intermediate intercept of *(female intercept+male intercept)/2)*—i.e. UL *(Asian) = (PMSH-((77.62+ 83.58)/2))/3.26*; *UL (black) = (PMSH-((79.55+85.8)/2))/3.14*; and *UL (white) = (PMSH-((84.67+90.92)/2))/3.14*—to eliminate any possible artefactual sexual dimorphism arising from the use of sex-specific equations. **d.** UL is not estimated, and muscle quality is instead calculated as *HGS/(MM/PMSH)*, assuming simple isometric scaling of UL with PMSH. The primary finding of substantially greater age-related muscle quality decline in females persists across all scenarios, confirming that it is not simply an artefact of our UL estimates.

Supplementary Information 9 Participant matching

```
Call:
MatchIt::matchit(formula = treat ~ age_months_0_0 + standing_height_f50_0_0 +
    weight_f21002_0_0, data = temp, method = "nearest", distance = "mahalanobis",
    replace = F, caliper = c(age_months_0_0 = 6, standing_height_f50_0_0 = 1,
        weight_f21002_0_0 = 1), std.caliper = F)

Summary of Balance for All Data:
                        Means Treated Means Control Std. Mean Diff. Var. Ratio eCDF Mean eCDF Max
age_months_0_0               686.2493      683.1100          0.0318     1.0572    0.0115   0.0287
standing_height_f50_0_0      175.5497      162.3820          1.9217     1.1835    0.2416   0.6881
weight_f21002_0_0             86.0683       71.6298          0.9998     1.0430    0.1068   0.4493

Summary of Balance for Matched Data:
                        Means Treated Means Control Std. Mean Diff. Var. Ratio eCDF Mean eCDF Max Std. Pair Dist.
age_months_0_0               683.3229      683.3396         -0.0002     1.0008    0.0003   0.0012          0.0194
standing_height_f50_0_0      168.9428      168.8152          0.0186     1.0269    0.0024   0.0172          0.0265
weight_f21002_0_0             78.0576       78.0028          0.0038     0.9993    0.0004   0.0045          0.0211

Sample Sizes:
           Control Treated
All         237015  195794
Matched      61683   61683
Unmatched   175332  134111
Discarded        0       0
```

```
Call:
MatchIt::matchit(formula = treat ~ age_months_0_0 + standing_height_f50_0_0 +
    weight_f21002_0_0, data = temp, method = "nearest", distance = "mahalanobis",
    replace = F, caliper = c(age_months_0_0 = 6, standing_height_f50_0_0 = 1,
        weight_f21002_0_0 = 1), std.caliper = F)

Summary of Balance for All Data:
                        Means Treated Means Control Std. Mean Diff. Var. Ratio eCDF Mean eCDF Max
age_months_0_0               676.6890      659.7453          0.1849     1.0702    0.0472   0.0898
standing_height_f50_0_0      176.5289      163.2949          2.0185     1.1287    0.3156   0.7077
weight_f21002_0_0             84.8177       69.6595          1.1586     1.0438    0.1538   0.5075

Summary of Balance for Matched Data:
                        Means Treated Means Control Std. Mean Diff. Var. Ratio eCDF Mean eCDF Max Std. Pair Dist.
age_months_0_0               666.5715      666.4591          0.0012     1.0012    0.0011   0.0037          0.0313
standing_height_f50_0_0      169.7187      169.5867          0.0201     1.0166    0.0033   0.0208          0.0551
weight_f21002_0_0             75.7672       75.7386          0.0022     0.9959    0.0006   0.0083          0.0349

Sample Sizes:
           Control Treated
All          23448   22181
Matched       5435    5435
Unmatched    18013   16746
Discarded        0       0
```

SI9 Successful matching of female:male pairs on baseline age, height, and body mass. Summaries of the matching procedures used to generate the Matched Cross-sectional Participants (left) and the Matched Longitudinal Participants (right) participant groups (Online Methods). The fields ‘age_months_0_0’, ‘standing_height_f50_0_0’, and ‘weight_f21002_0_0’ denote baseline values for age (months), standing height (cm), and body mass (kg), respectively. Standard Mean Differences between -0.0002 and 0.0201 and Variance Ratios between 0.9959 and 1.0269 indicate excellent balance (i.e. successful matching), as per the recommendations of Greifer (2023) (full reference in main manuscript).