

Grip strength, walking pace, and hospital-record-defined incident degenerative cervical myelopathy: a UK Biobank cohort study

Simin Yang

Southern Medical University

Le Kang

Southern Medical University

Yuqian Yang

Southern Medical University

Jiahao Du

Southern Medical University

Zhenhui Wang

Southern Medical University

Shengxuan Ding

Southern Medical University

Xiangwei He

Southern Medical University

Shuizhong Cen

`censzh3@alumni.sysu.edu.cn`

Southern Medical University

Research Article

Keywords: degenerative cervical myelopathy, grip strength, walking pace, hospital records, risk markers, UK Biobank, cohort study

Posted Date: July 7th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9706454/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at BMC Neurology on September 12th, 2026.
See the published version at <https://doi.org/10.1186/s12883-026-05305-6>.

Grip strength, walking pace, and hospital-record-defined incident degenerative cervical
myelopathy: a UK Biobank cohort study

Article type: Research article

Running title: Grip strength, walking pace, and coded DCM risk

Authors

Simin Yang^{1*}, Le Kang², Yuqian Yang¹, Jiahao Du³, Zhenhui Wang³, Shengxuan Ding¹,
Xiangwei He¹, Shuizhong Cen^{3*}

Affiliations

1. Neurosurgery Center, National Key Clinical Specialty, Engineering Research Center of
Diagnostic and Therapeutic Technology and Devices for Cerebrovascular Diseases in Ministry
of Education, Guangdong Provincial Key Laboratory on Brain Function Repair and
Regeneration, Zhujiang Hospital Institute of Neurosurgery, Zhujiang Hospital, Southern Medical
University, Guangzhou, China.

2. Department of Anesthesiology, Zhujiang Hospital, Southern Medical University, Guangzhou,
China.

3. Department of Spinal Surgery, Orthopedic Medical Center, Zhujiang Hospital, Southern
Medical University, Guangzhou, China.

Correspondence: Simin Yang, Neurosurgery Center, Zhujiang Hospital, Southern Medical
University, Guangzhou 510282, China. Email: ysmlorain@126.com; Shuizhong Cen,

Department of Spinal Surgery, Orthopedic Medical Center, Zhujiang Hospital, Southern Medical University, Guangzhou 510282, China. Email: censshzh3@alumni.sysu.edu.cn

Abstract

Background: Degenerative cervical myelopathy (DCM) is often recognized late. Grip strength and usual walking pace are simple motor-functional measures, but their associations with hospital-record-defined DCM endpoint layers and their incremental risk-enrichment value remain uncertain.

Methods: We studied 501,936 UK Biobank participants with linked hospital records. The primary endpoint was hospital-record-defined incident DCM; primary-position incident DCM and diagnosis-procedure-paired surgically treated DCM were supportive layers. Cox models assessed grip strength and walking pace. Post hoc sensitivity analyses excluded the first 2 and 5 years of follow-up and adjusted for available smoking status. Prediction analyses compared a base clinical model with models augmented by grip strength, walking pace, or both markers.

Results: Among 501,936 participants, 978 incident DCM, 639 primary-position DCM, and 552 surgically treated DCM events occurred. In 494,527 complete cases, grip strength per 5 kg and walking pace per category were inversely associated with incident DCM (HR 0.845, 95% CI 0.810-0.881; and HR 0.824, 95% CI 0.785-0.866). Associations remained directionally consistent after excluding the first 2 and 5 years; after the 5-year lag, HRs were 0.870 and 0.830. The combined model improved validation Harrell's C-index from 0.6280 to 0.6560 and 10-year AUC from 0.6272 to 0.6548, but calibration indicated underprediction and decision-curve benefit was limited to thresholds 0.001-0.003.

Conclusions: Grip strength and usual walking pace were associated with hospital-record-defined incident DCM in UK Biobank. These markers support modest risk enrichment and earlier clinical suspicion under low-threshold evaluation contexts, but not stand-alone screening, triage, or replacement of neurological examination and imaging.

Keywords: degenerative cervical myelopathy; grip strength; walking pace; hospital records; risk markers; UK Biobank; cohort study

Introduction

Degenerative cervical myelopathy (DCM) is the most common cause of chronic spinal cord dysfunction in adults and encompasses age-related degenerative narrowing of the cervical canal with secondary spinal cord compromise. [1-13] Across primary care, neurology, rehabilitation, and spine practice, patients frequently present with a mixture of hand clumsiness, gait disturbance, balance change, sensory symptoms, and long-tract signs rather than with an immediately recognized label of myelopathy. [3-7, 9-14] This diagnostic ambiguity matters because delayed recognition may allow additional disability to accumulate before spine-directed evaluation is undertaken. [3, 5-7, 12, 13]

That clinical problem motivates interest in simple measures of motor reserve. Grip strength and usual walking pace are attractive because they are inexpensive, widely available, easy to interpret, and closely aligned with the functional domains most commonly impaired in DCM. [15-24] Grip strength has broad prognostic value in aging populations, [17, 23, 24] and gait speed or pace is similarly linked to function and survival. [15, 22] In DCM specifically, hand dexterity and gait disturbance are not peripheral features; they are central manifestations of cord dysfunction and are repeatedly used to characterize severity, recovery, and outcome. [9-11, 14-16, 18-21, 25]

Prior population-based work has linked grip strength and walking pace to incident DCM. [26] In contrast to a single-endpoint association analysis, the present study evaluated whether the same markers remain informative in a primary-position endpoint, a surgically treated secondary endpoint, a contextual diagnosis-to-surgery capture layer, and a utility-oriented framework beyond a parsimonious clinical baseline model.

We therefore hypothesized that grip strength and usual walking pace would retain independent associations across these clinically ordered endpoint layers and would contribute modest incremental value beyond a simple baseline clinical model.

Methods

Study design and data source

We conducted a population-based prospective cohort study within UK Biobank, a large volunteer cohort with linked hospital follow-up data. [27] The manuscript follows STROBE guidance for observational studies and uses the TRIPOD framework for its prediction component. [28, 29] All analyses were performed on participant-level analytic files derived within the UK Biobank Research Analysis Platform.

Baseline demographic and clinical variables were derived from the participant-level dataset. Endpoint reconstruction was based on linked hospital diagnosis and procedure records. These endpoints should therefore be interpreted as hospital-record-defined endpoints rather than clinically adjudicated DCM diagnoses. The analytic workflow separated baseline covariate extraction from record-level endpoint reconstruction and then harmonized the resulting participant-level files into a unified time-to-event framework.

Endpoint reconstruction and prespecified hierarchy

The primary endpoint was hospital-record-defined incident DCM using a diagnosis in any hospital coding position after baseline. The primary-position sensitivity endpoint was hospital-record-defined incident DCM restricted to a primary-position diagnosis. The key secondary endpoint was diagnosis-procedure-paired surgically treated DCM, defined by DCM diagnosis and qualifying cervical procedure codes within a prespecified operative window. The diagnosis-to-surgery capture layer among participants with incident DCM was retained as a contextual analysis rather than as evidence of true clinical progression.

Detailed ICD-10 and OPCS-4 code lists, record-level reconstruction rules, operative windows, and endpoint counts are reported in Supplementary Table S1 (Additional file 1). The main diagnosis-procedure pairing window was -30 to +180 days, and the strict sensitivity window was 0 to +90 days. Because MRI findings, modified Japanese Orthopaedic Association scores, neurological examination findings, and clinical adjudication were unavailable, the coding hierarchy was used to improve specificity but not to claim adjudicated clinical diagnosis. Early feasibility work showed that a surgery-only primary endpoint would materially reduce the event base, so hospital-record-defined incident DCM was retained as the disease-occurrence endpoint and surgically treated DCM was positioned as a clinically consequential secondary layer. [21, 26, 30-33]

Predictors and covariates

The two focal predictors were grip strength and usual walking pace. Grip strength was modeled per 5-kg increment using the available maximal grip value. Walking pace was modeled per category using the harmonized walking-pace variable available in the analytic dataset. The final baseline clinical model included age, sex, body mass index (BMI), waist circumference, alcohol intake, diabetes, and hypertension.

These covariates were retained because they were clinically plausible, highly complete, and stable across the common analytic cohort. Broader candidate variables, completeness profiles, and reasons for retention or exclusion are summarized in Supplementary Table S2 (Additional file 1). Frailty index was not included in the final model. We additionally fitted an available-smoking sensitivity model using baseline smoking status coded as never, previous, or current. Physical activity and socioeconomic deprivation were not available in the audited analytic variable set used for this analysis; rather than adding an incompletely audited post hoc covariate layer, we treated these factors as residual-confounding limitations. Additional endpoint and covariate-completeness summaries are shown in Supplementary Tables S1-S2 (Additional file 1) and Supplementary Figures S1-S2 (Additional file 2). The accelerometer and body-composition enhancement-boundary summary is provided in Supplementary Table S3 (Additional file 1), with corresponding visual summaries in Supplementary Figures S3-S4 (Additional file 2).

Association analyses

We fitted Cox proportional hazards models for the primary endpoint, the primary-position sensitivity endpoint, the surgically treated DCM endpoint, and the contextual diagnosis-to-surgery capture analysis. Model 1 included age and sex. Model 2 additionally included BMI, waist circumference, alcohol intake, diabetes, and hypertension. Model 3 was the fully adjusted model and additionally included grip strength and usual walking pace.

For each endpoint, the corresponding association model used the endpoint-specific complete-case sample for the variables included in that model. No multiple imputation was performed. This approach was considered acceptable because missingness was minimal for the retained variables, and the fully adjusted primary-endpoint model retained 494,527 participants with 944

events. Proportional hazards assumptions were evaluated using scaled Schoenfeld residuals and standard diagnostic plots. [34]

To address potential reverse causality from undiagnosed or subclinical DCM at baseline, we added lagged sensitivity analyses for the primary endpoint that excluded participants with endpoint occurrence or loss from observation within the first 2 and 5 years after baseline and re-anchored follow-up at the corresponding landmark among participants still under observation. The same fully adjusted covariate set was used. We also added an available-smoking sensitivity model for the primary endpoint, with smoking entered as an ordinal baseline variable among participants coded as never, previous, or current smokers. These sensitivity analyses are presented as supportive analyses rather than as new primary analyses.

Prediction analyses

Prediction analyses were anchored to incident DCM under the any-position definition. To avoid denominator drift across candidate models, we prespecified a single common complete-case prediction cohort formed by requiring nonmissing values for all variables used across the compared models. This yielded 494,527 participants with 944 primary endpoint events. The cohort was randomly split 70:30 into derivation (346,168 participants; 660 events) and validation (148,359 participants; 284 events).

We compared four prespecified models: (1) base clinical model; (2) base plus grip strength; (3) base plus usual walking pace; and (4) base plus both markers. Supportive utility checks were performed for the combined model using the primary-position DCM endpoint and the surgically treated DCM endpoint.

Discrimination, calibration, and decision analysis

Discrimination was summarized with Harrell's C-index and validation 10-year AUC. [35, 36] Before calculating validation discrimination, the risk-score direction was checked so that higher model scores corresponded to higher predicted risk. The 10-year AUC was calculated on the validation set among participants who either experienced the endpoint by 10 years or remained event-free beyond 10 years; participants censored before 10 years were excluded from that calculation.

Calibration of the combined model was summarized in the validation cohort using decile-based calibration, calibration intercept and slope, a weighted observed-to-predicted ratio, and the maximum absolute decile gap. [37, 38] Observed risk in each decile was estimated from Kaplan–Meier risk at the target horizon. Decision-curve analysis was then used to quantify net benefit across very low clinically plausible decision thresholds, with the main display focused on the 0.001 to 0.005 range and the effective threshold window reported when model net benefit exceeded both treat-all and treat-none strategies. [39, 40]

Ethics

UK Biobank was approved by the North West Multi-Centre Research Ethics Committee (11/NW/0382), and all participants provided written informed consent. The present study used de-identified data within the approved UK Biobank Research Analysis Platform under UK Biobank Application Number 1162548. The study was conducted in accordance with the Declaration of Helsinki and relevant UK Biobank governance requirements.

Use of generative AI and AI-assisted technologies

During preparation of the manuscript, ChatGPT (OpenAI) was used only for language editing to improve readability and fluency. It was not used for data extraction, statistical analysis, result

generation, or scientific interpretation. The authors reviewed and edited all content and take full responsibility for the submitted manuscript.

Results

Cohort construction and endpoint hierarchy

The endpoint-reconstruction cohort comprised 501,936 participants with an available baseline date. Within this cohort there were 978 hospital-record-defined incident DCM events under the any-position primary endpoint, 639 incident primary-position DCM events, 552 incident surgically treated DCM events under the main pairing definition, and 545 under the strict pairing definition. The participant-level master table therefore preserved disease-occurrence, specificity, surgery-defined severity, and contextual diagnosis-to-surgery capture layers in a single hierarchy.

The cohort assembly and hospital-record-defined endpoint hierarchy are summarized in Figure 1.

The common complete-case prediction cohort included 494,527 participants and 944 primary endpoint events (Table 1). The endpoint-transition representation yielded 974 transitions from no hospital-record-defined DCM to incident DCM, 548 diagnosis-to-surgery capture transitions after incident DCM, and only 4 direct no-DCM-to-surgery captures. This pattern supports interpreting incident DCM as the primary disease-occurrence endpoint and the surgical layer as a downstream diagnosis-procedure-paired capture layer.

Primary endpoint: hospital-record-defined incident DCM (any-position)

In the fully adjusted model, both grip strength and usual walking pace retained strong independent associations with hospital-record-defined incident DCM. Grip strength was associated with a lower hazard of incident DCM per 5-kg increment (HR 0.845, 95% confidence

interval [CI] 0.810-0.881; $P = 2.52e-15$). Usual walking pace likewise showed an inverse association per category increase (HR 0.824, 95% CI 0.785-0.866; $P = 7.35e-15$) (Table 2; Figure 2).

These were not fragile estimates from a severely eroded analytic subset. The fully adjusted primary-endpoint model retained 494,527 participants and 944 events, indicating that the central association analysis rested on a large and stable denominator.

Lagged sensitivity analyses reduced concern that the signal was driven only by undiagnosed DCM already present near baseline. After excluding the first 2 years, grip strength (HR 0.860, 95% CI 0.823-0.899; $P = 1.63e-11$) and usual walking pace (HR 0.828, 95% CI 0.786-0.872; $P = 1.34e-12$) remained inversely associated with the primary endpoint. After excluding the first 5 years, the associations remained directionally consistent for grip strength (HR 0.870, 95% CI 0.829-0.913; $P = 2.03e-08$) and usual walking pace (HR 0.830, 95% CI 0.783-0.880; $P = 3.17e-10$) (Supplementary Table S4 (Additional file 1)).

In the available-smoking sensitivity model, which additionally adjusted for baseline smoking status among participants coded as never, previous, or current smokers, the primary-endpoint associations remained similar for grip strength (HR 0.846, 95% CI 0.812-0.883; $P = 6.86e-15$) and usual walking pace (HR 0.830, 95% CI 0.790-0.873; $P = 3.38e-13$) (Supplementary Table S4 (Additional file 1)).

High-specificity sensitivity endpoint: incident DCM (primary-position)

The more specific primary-position endpoint produced the same directional pattern. Grip strength remained independently protective (HR 0.842, 95% CI 0.800-0.887; $P = 6.36e-11$), as did usual walking pace (HR 0.832, 95% CI 0.782-0.884; $P = 3.27e-09$) (Table 2; Figure 3).

This agreement is important because it argues against the primary signal being driven solely by the broader any-position coding rule. Instead, it supports the interpretation that the two focal markers remain informative when the endpoint is stress-tested against a narrower diagnosis definition.

Secondary endpoint: surgically treated DCM

The surgically treated DCM endpoint also retained stable associations for the two focal markers. Grip strength was associated with a lower hazard of surgically treated DCM (HR 0.858, 95% CI 0.812-0.907; $P = 5.71e-08$), and usual walking pace again showed a protective association (HR 0.846, 95% CI 0.789-0.907; $P = 2.71e-06$) (Table 2; Figure 3).

Conceptually, this matters because the same two markers were not restricted to broad disease occurrence alone. The inverse association extended into a surgically treated endpoint that is closer to severity and operative disease burden.

Contextual diagnosis-to-surgery capture analysis

The contextual diagnosis-to-surgery capture analysis did not retain stable independent associations after hospital-record-defined incident DCM had already been captured. Grip strength yielded an HR of 1.026 (95% CI 0.973-1.083; $P = 0.342$), and usual walking pace yielded an HR of 1.058 (95% CI 0.957-1.170; $P = 0.268$) (Table 2; Figure 3).

We therefore interpret this layer as peri-diagnostic diagnosis-to-surgery capture rather than as true clinical progression. This interpretation is supported by the median interval of 0 days from incident DCM coding to surgery coding in the 1-to-2 transition, indicating that many records likely reflected diagnostic and operative coding around the same care episode.

Prediction performance and incremental utility

In the validation cohort, the base clinical model yielded a Harrell's C-index of 0.6280 and a 10-year AUC of 0.6272. Adding grip strength improved these metrics to 0.6445 and 0.6457, respectively. Adding usual walking pace improved them to 0.6469 and 0.6430. The combined model performed best, with a Harrell's C-index of 0.6560 and a 10-year AUC of 0.6548 (Table 3; Figures 4 and 5).

Relative to the base model, the combined model improved the validation C-index by 0.0280 and the 10-year AUC by 0.0275. Grip alone and walking pace alone each provided smaller but directionally concordant gains. We interpret these increments as modest risk-enrichment gains, not as evidence for an actionable stand-alone prediction tool.

Supportive endpoint checks were directionally consistent. For the combined model, the validation 10-year AUC was 0.6564 for the primary-position incident DCM endpoint and 0.6380 for the surgically treated DCM endpoint (Table 4). These supportive checks reinforce the interpretation that the utility signal was not restricted to a single disease definition.

Calibration and decision analysis

Calibration of the combined model in validation showed a 10-year calibration intercept of 0.0001 and a calibration slope of 1.3804. The weighted observed-to-predicted ratio was 1.4878, indicating underprediction on average, although the maximum absolute decile gap remained small at 0.0009 (Table 4; Figure 6).

Decision-curve analysis showed that the combined model had the broadest positive net-benefit window, but the gain was concentrated in a narrow very-low-threshold range of approximately 0.001 to 0.003 (Figure 7). Thus, the prediction layer is best interpreted as limited utility enrichment under very low decision thresholds rather than as a stand-alone screening or triage platform.

Endpoint-transition and diagnosis-to-surgery capture structure

The endpoint-transition backbone was dominated by transition from no hospital-record-defined DCM to incident DCM and then into the diagnosis-procedure-paired surgical layer. There were 974 0-to-1 transitions and 548 1-to-2 captures, but only 4 direct 0-to-2 captures (Figure 8). Because the median 1-to-2 interval was 0 days, this structure should be read as a coding and care-episode capture framework rather than as proof of the biological tempo of clinical progression.

Discussion

The present findings have a clinically practical but deliberately restrained interpretation. Lower grip strength and slower usual walking pace identified individuals at higher risk of hospital-record-defined incident DCM in a large population cohort, and the same signal remained visible when the endpoint was made more specific and when surgically treated DCM was examined as a secondary clinical layer. The key contribution is that the same two functional markers remained informative across a hierarchical hospital-record endpoint structure and a risk-enrichment framework.

Prior population-based work has linked grip strength and walking pace to incident DCM. [26] In contrast to a single-endpoint association analysis, the present study evaluated whether the same markers remain informative in a primary-position endpoint, a surgically treated secondary endpoint, a contextual diagnosis-to-surgery capture layer, and a utility-oriented framework beyond a parsimonious clinical baseline model.

That endpoint hierarchy is clinically relevant, but it should be interpreted within the limits of hospital coding. DCM is heterogeneous, and surgery-related outcomes should not automatically be treated as interchangeable with broad disease-occurrence endpoints. [1-7, 9-14, 25] Our data

support a disciplined structure: hospital-record-defined incident DCM is the primary disease-occurrence endpoint; primary-position DCM is a specificity layer; surgically treated DCM is a clinically consequential secondary endpoint; and the diagnosis-to-surgery capture analysis is contextual rather than a true progression endpoint.

The biological interpretation is plausible, but residual confounding and reverse causality remain possible. Grip strength and usual walking pace may partly reflect overall health status, sarcopenia, or frailty-related processes. Lagged sensitivity analyses showed that the association persisted after excluding early events, which reduces but does not eliminate the possibility that low grip strength or slow walking pace reflected undiagnosed baseline DCM. The available-smoking sensitivity analysis also preserved the main associations, although physical activity and socioeconomic deprivation were not available in the audited analytic variable set used for this analysis and were therefore retained as residual-confounding limitations.

The risk-enrichment analyses also deserve a measured reading. The combined model did not approach the performance of a high-accuracy screening tool, and we do not present this paper as a machine-learning, artificial-intelligence, or actionable clinical decision model. The absolute gains in discrimination were modest, calibration showed average underprediction, and decision-curve benefit was concentrated in a narrow very-low-threshold range. [35-40] The prediction results therefore support only modest risk enrichment.

This point is important for clinical translation. Grip strength and usual walking pace should not be used as independent triage tools replacing neurological examination, imaging, or specialist assessment. Their more realistic value is to enrich clinical suspicion in settings where the threshold for further evaluation is intentionally low—for example, when patients already present

with subtle hand dysfunction, gait change, neck symptoms, or other nonspecific neurologic complaints. [3, 5-7, 9-13]

The null result in the diagnosis-to-surgery capture layer is equally informative. Once patients have already entered the hospital-record-defined incident DCM state, surgery coding is influenced by symptom burden, imaging severity, clinician judgment, access to care, patient preference, operative decision making, and the timing of administrative coding. [14, 18-21, 25, 31-33] Given that the median interval from incident DCM coding to surgery coding was 0 days, this layer is better interpreted as peri-diagnostic surgical capture than as clinical progression.

This study has several strengths. It was based on a large prospective cohort with linked hospital follow-up, a record-level endpoint reconstruction process, a participant-level time-to-event structure, and a common complete-case prediction cohort that prevented denominator drift across candidate models. The analysis also remained focused on two clinically interpretable markers tested across a clinically ordered endpoint structure.

It also has limitations. First, hospital-coded DCM endpoints remain imperfect proxies for clinically adjudicated disease; MRI findings, mJOA scores, neurological examination, and adjudicated clinical diagnosis were unavailable. Second, low grip strength and slow walking pace could reflect undiagnosed or subclinical DCM at baseline; lagged analyses reduce this concern but cannot remove it completely. Third, walking pace was self-reported rather than instrumented. Fourth, physical activity and socioeconomic deprivation were not available in the audited analytic variable set used for this analysis; rather than adding an incompletely audited post hoc covariate layer, we treated these factors as residual-confounding limitations despite the available-smoking sensitivity model. Fifth, the prediction layer relied on internal derivation/validation, showed average underprediction, and should not be interpreted as a

deployable clinical tool. Sixth, the surgically treated endpoint captures severity only indirectly, and the diagnosis-to-surgery capture layer should not be overinterpreted as true progression.

Overall, these findings support a restrained but clinically useful interpretation. Grip strength and usual walking pace form a concise functional profile that remains informative across a prespecified hospital-record endpoint hierarchy and provides modest incremental value for risk enrichment beyond a simple clinical baseline model. The principal contribution of the study lies in integrating these markers into a clinically ordered endpoint and utility framework while explicitly acknowledging the limits of coded endpoints.

Conclusions

Grip strength and usual walking pace were independently associated with hospital-record-defined incident DCM in UK Biobank. The association persisted under a more specific disease definition, within a surgically treated secondary endpoint, and in lagged sensitivity analyses excluding early events.

Together, these two simple functional markers provided modest incremental value for risk enrichment beyond a parsimonious clinical baseline model. Their main value lies in supporting earlier clinical suspicion under low-threshold evaluation contexts, not in replacing specialist evaluation, imaging, or neurological examination and not in establishing a high-performance stand-alone screening model.

List of abbreviations

AUC, area under the curve; BMI, body mass index; CI, confidence interval; DCM, degenerative cervical myelopathy; ICD-10, International Classification of Diseases, Tenth Revision; mJOA,

354 modified Japanese Orthopaedic Association; OPCS-4, Office of Population Censuses and
355 Surveys Classification of Interventions and Procedures, version 4; RAP, Research Analysis
356 Platform; STROBE, Strengthening the Reporting of Observational Studies in Epidemiology;
357 TRIPOD, Transparent Reporting of a multivariable prediction model for Individual Prognosis Or
358 Diagnosis; UK Biobank, United Kingdom Biobank.

359 **Declarations**

360 **Ethics approval and consent to participate**

361 Human Ethics and Consent to Participate declarations: UK Biobank was approved by the North
362 West Multi-Centre Research Ethics Committee (11/NW/0382), and all participants provided
363 written informed consent. The present study used de-identified data within the approved UK
364 Biobank Research Analysis Platform under UK Biobank Application Number 1162548. The
365 study was conducted in accordance with the Declaration of Helsinki and relevant UK Biobank
366 governance requirements.

367 **Clinical trial number**

368 Clinical trial number: not applicable.

369 **Consent for publication**

370 Not applicable. This article does not contain any individual person's identifiable data.

371 **Availability of data and materials**

372 The data that support the findings of this study are available from UK Biobank, but restrictions
373 apply to the availability of these data, which were used under licence for the current study and

are therefore not publicly available. Researchers can apply for access directly through UK Biobank. Custom analytic code used in this study is available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

This study was supported by the Guangdong Basic and Applied Basic Research Foundation (2022A1515012118) and the National Science Fund for Distinguished Young Scholars (82202657). The funders had no role in study design, data collection, data analysis, data interpretation, manuscript preparation, or the decision to submit the article for publication.

Authors' contributions

SY conceptualized the study, designed the methodology, developed software and analytic code, performed formal analysis, curated and validated data, prepared visualizations, and drafted the manuscript. LK contributed to visualization, drafting, and critical revision. YY contributed to software, validation, and critical revision. JD contributed to investigation and critical revision. ZW contributed to investigation, software, and critical revision. SD contributed to investigation, software, and critical revision. XH contributed to investigation and critical revision. SC supervised the study, administered the project, acquired funding, and critically revised the manuscript. All authors read and approved the final manuscript.

Acknowledgements

This research has been conducted using the UK Biobank Resource under Application Number 1162548. The authors thank the UK Biobank participants and staff.

Authors' information

Not applicable.

Additional files

Additional file 1. DOCX. Supplementary tables and figure legends. This file contains Supplementary Tables S1-S4 and legends for Supplementary Figures S1-S4.

Additional file 2. PDF. Supplementary figures. This file contains Supplementary Figures S1-S4, including endpoint code spectrum, covariate completeness, accelerometer subset boundary, and body-composition subset boundary.

References

1. Fehlings MG, Tetreault LA, Wilson JR, Skelly AC. Cervical spondylotic myelopathy: current state of the art and future directions. *Spine (Phila Pa 1976)*. 2013;38(22 Suppl 1):S1-S8. <https://doi.org/10.1097/BRS.0b013e3182a7e9e0>.
2. Nouri A, Tetreault L, Singh A, Karadimas SK, Fehlings MG. Degenerative cervical myelopathy: epidemiology, genetics, and pathogenesis. *Spine (Phila Pa 1976)*. 2015;40(12):E675-E693. <https://doi.org/10.1097/BRS.0000000000000913>.
3. Davies BM, Mowforth OD, Smith EK, Kotter MRN. Degenerative cervical myelopathy. *BMJ*. 2018;360:k186. <https://doi.org/10.1136/bmj.k186>.

- 413 4. Wilson JR, Badhiwala JH, Moghaddamjou A, Martin AR, Fehlings MG. Degenerative
414 cervical myelopathy: a review of the latest advances and future directions in management.
415 *Neurospine*. 2019;16(3):494-505. <https://doi.org/10.14245/ns.1938240.120>.
- 416 5. Milligan J, Ryan K, Fehlings M, Bauman C. Degenerative cervical myelopathy: diagnosis and
417 management in primary care. *Can Fam Physician*. 2019;65(9):619-624.
- 418 6. Kane SF, Abadie KV, Willson A. Degenerative cervical myelopathy: recognition and
419 management. *Am Fam Physician*. 2020;102(12):740-750.
- 420 7. Grodzinski B, Stubbs DJ, Davies BM. Most degenerative cervical myelopathy remains
421 undiagnosed, particularly amongst the elderly: modelling the prevalence of degenerative cervical
422 myelopathy in the United Kingdom. *J Neurol*. 2023;270(1):311-319.
423 <https://doi.org/10.1007/s00415-022-11349-8>.
- 424 8. Theodore N. Degenerative cervical spondylosis. *N Engl J Med*. 2020;383(2):159-168.
425 <https://doi.org/10.1056/NEJMr2003558>.
- 426 9. Fehlings MG, Tetreault LA, Riew KD, Middleton JW, Wang JC. A Clinical Practice
427 Guideline for the Management of Degenerative Cervical Myelopathy: Introduction, Rationale,
428 and Scope. *Global Spine J*. 2017;7(3 Suppl):21S-27S.
429 <https://doi.org/10.1177/2192568217703088>.
- 430 10. Fehlings MG, Tetreault LA, Riew KD, Middleton JW, Aarabi B, Arnold PM, et al. A
431 Clinical Practice Guideline for the Management of Patients With Degenerative Cervical
432 Myelopathy: Recommendations for Patients With Mild, Moderate, and Severe Disease and
433 Nonmyelopathic Patients With Evidence of Cord Compression. *Global Spine J*. 2017;7(3
434 Suppl):70S-83S. <https://doi.org/10.1177/2192568217701914>.

435 11. Williams J, D'Amore P, Redlich N, Darlow M, Suwak P, Sarkovich S, et al. Degenerative
436 Cervical Myelopathy: Evaluation and Management. *Orthop Clin North Am.* 2022;53(4):509-521.
437 <https://doi.org/10.1016/j.ocl.2022.05.007>.

438 12. Pope DH, Mowforth OD, Davies BM, Kotter MRN. Diagnostic Delays Lead to Greater
439 Disability in Degenerative Cervical Myelopathy and Represent a Health Inequality. *Spine (Phila*
440 *Pa 1976).* 2020;45(6):368-377. <https://doi.org/10.1097/BRS.0000000000003305>.

441 13. Gallagher DO, Taghlabi KM, Bondar K, Saifi C. Degenerative Cervical Myelopathy: A
442 Concept Review and Clinical Approach. *Clin Spine Surg.* 2024;37(1):1-8.
443 <https://doi.org/10.1097/BSD.0000000000001463>.

444 14. Johansen TO, et al. Long-term results after surgery for degenerative cervical myelopathy.
445 *Neurosurgery.* 2024;94(3):454-460. <https://doi.org/10.1227/neu.0000000000002712>.

446 15. Choy WJ, Chen L, de Oliveira CQ, Verhagen AP, Damodaran O, Anderson DB. Gait
447 assessment tools for degenerative cervical myelopathy: a systematic review. *J Spine Surg.*
448 2022;8(1):149-162. <https://doi.org/10.21037/jss-21-109>.

449 16. Jimenez KA, Kwon JW, Yoon J, Lee HM, Moon SH, Suk KS, et al. Handgrip strength
450 correlated with falling risk in patients with degenerative cervical myelopathy. *J Clin Med.*
451 2021;10(9):1980. <https://doi.org/10.3390/jcm10091980>.

452 17. Bohannon RW. Grip strength: an indispensable biomarker for older adults. *Clin Interv*
453 *Aging.* 2019;14:1681-1691. <https://doi.org/10.2147/CIA.S194543>.

454 18. Liang G, Liao T, Ye Y, Cai Y, Chen J, Chang Y. Hand Dexterity Recovery Capacity for
455 Degenerative Cervical Myelopathy With Varying Levels of Impairment: A Prospective 1-Year
456 Follow-up Study. *Neurospine.* 2025;22(1):202-210. <https://doi.org/10.14245/ns.2448682.341>.

457 19. Cole TS, Almefty KK, Godzik J, Muma AH, Hlubek RJ, Martinez-Del-Campo E, et al.
458 Functional improvement in hand strength and dexterity after surgical treatment of cervical
459 spondylotic myelopathy: a prospective quantitative study. *J Neurosurg Spine*. 2020;32(6):907-
460 913. <https://doi.org/10.3171/2019.10.SPINE19685>.

461 20. Davies BM, McHugh M, Elgheriani A, Kolas AG, Tetreault LA, Hutchinson PJA, et al.
462 Reported Outcome Measures in Degenerative Cervical Myelopathy: A Systematic Review. *PLoS*
463 *One*. 2016;11(8):e0157263. <https://doi.org/10.1371/journal.pone.0157263>.

464 21. Wu JC, Ko CC, Yen YS, Huang WC, Chen YC, Liu L, et al. Epidemiology of cervical
465 spondylotic myelopathy and its risk of causing spinal cord injury: a national cohort study.
466 *Neurosurg Focus*. 2013;35(1):E10. <https://doi.org/10.3171/2013.4.FOCUS13122>.

467 22. Studenski S, Perera S, Patel K, Rosano C, Faulkner K, Inzitari M, et al. Gait speed and
468 survival in older adults. *JAMA*. 2011;305(1):50-58. <https://doi.org/10.1001/jama.2010.1923>.

469 23. Bohannon RW. Muscle strength: clinical and prognostic value of hand-grip dynamometry.
470 *Curr Opin Clin Nutr Metab Care*. 2015;18(5):465-470.
471 <https://doi.org/10.1097/MCO.0000000000000202>.

472 24. Sayer AA, Kirkwood TBL. Grip strength and mortality: a biomarker of ageing? *Lancet*.
473 2015;386(9990):226-227. [https://doi.org/10.1016/S0140-6736\(14\)62349-7](https://doi.org/10.1016/S0140-6736(14)62349-7).

474 25. Tetreault L, Palubiski LM, Kryshtalskyj M, Idler RK, Martin AR, Ganau M, et al. Significant
475 predictors of outcome following surgery for the treatment of degenerative cervical myelopathy: a
476 systematic review of the literature. *Neurosurg Clin N Am*. 2018;29(1):115-127.e35.
477 <https://doi.org/10.1016/j.nec.2017.09.020>.

26. Lin C, Li J, Zhu Z, Zhou J, Wang F, Wei F, et al. Association between grip strength and walking pace with incidence of degenerative cervical myelopathy: a UK Biobank observational study. *Eur Spine J*. Published online July 9, 2024. <https://doi.org/10.1007/s00586-024-08374-8>.
27. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, 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 Med*. 2015;12(3):e1001779. <https://doi.org/10.1371/journal.pmed.1001779>.
28. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP; STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *PLoS Med*. 2007;4(10):e296. <https://doi.org/10.1371/journal.pmed.0040296>.
29. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): the TRIPOD statement. *BMJ*. 2015;350:g7594. <https://doi.org/10.1136/bmj.g7594>.
30. Grodzinski B, Stubbs DJ, Davies BM. Previous surgical exposure and the onset of degenerative cervical myelopathy: a propensity-matched case-control analysis nested within the UK Biobank cohort. *Global Spine J*. 2025;15(5):2594-2603. <https://doi.org/10.1177/21925682241304335>.
31. Singh A, Tetreault L, Fehlings MG, Fischer DJ, Skelly AC. Risk factors for development of cervical spondylotic myelopathy: results of a systematic review. *Evid Based Spine Care J*. 2012;3(3):35-42. <https://doi.org/10.1055/s-0032-1327808>.
32. Al-Tamimi YZ, Guilfoyle M, Seeley H, Laing RJ. Measurement of long-term outcome in patients with cervical spondylotic myelopathy treated surgically. *Eur Spine J*. 2013;22(11):2552-2557. <https://doi.org/10.1007/s00586-013-2965-4>.

501 33. Veeravagu A, Connolly ID, Lamsam L, Li A, Swinney C, Azad TD, et al. Surgical outcomes
502 of cervical spondylotic myelopathy: an analysis of a national, administrative, longitudinal
503 database. *Neurosurg Focus*. 2016;40(6):E11. <https://doi.org/10.3171/2016.3.FOCUS1669>.

504 34. Grambsch PM, Therneau TM. Proportional hazards tests and diagnostics based on weighted
505 residuals. *Biometrika*. 1994;81(3):515-526. <https://doi.org/10.1093/biomet/81.3.515>.

506 35. Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing
507 models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med*.
508 1996;15(4):361-387. [https://doi.org/10.1002/\(SICI\)1097-0258\(19960229\)15:4<361::AID-](https://doi.org/10.1002/(SICI)1097-0258(19960229)15:4<361::AID-SIM168>3.0.CO;2-4)
509 [SIM168>3.0.CO;2-4](https://doi.org/10.1002/(SICI)1097-0258(19960229)15:4<361::AID-SIM168>3.0.CO;2-4).

510 36. Uno H, Cai T, Tian L, Wei LJ. Evaluating prediction rules for t-year survivors with censored
511 regression models. *J Am Stat Assoc*. 2007;102(478):527-537.
512 <https://doi.org/10.1198/016214507000000149>.

513 37. Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW. Calibration: the
514 Achilles heel of predictive analytics. *BMC Med*. 2019;17:230. [https://doi.org/10.1186/s12916-](https://doi.org/10.1186/s12916-019-1466-7)
515 [019-1466-7](https://doi.org/10.1186/s12916-019-1466-7).

516 38. Van Calster B, Nieboer D, Vergouwe Y, De Cock B, Pencina MJ, Steyerberg EW. A
517 calibration hierarchy for risk models was defined: from utopia to empirical data. *J Clin*
518 *Epidemiol*. 2016;74:167-176. <https://doi.org/10.1016/j.jclinepi.2015.12.005>.

519 39. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction
520 models. *Med Decis Making*. 2006;26(6):565-574. <https://doi.org/10.1177/0272989X06295361>.

521 40. Vickers AJ, van Calster B, Steyerberg EW. A simple, step-by-step guide to interpreting
522 decision curve analysis. *Diagn Progn Res*. 2019;3:18. [https://doi.org/10.1186/s41512-019-0064-](https://doi.org/10.1186/s41512-019-0064-7)
523 [7](https://doi.org/10.1186/s41512-019-0064-7).

525 Tables

526 Editable tables are provided below. Figures are provided separately as individual files.

527 **Table 1 Baseline characteristics of the analytic cohort**

Characteristic	Overall	No incident DCM	Incident DCM
Participants	494,527	493,583	944
Age, years	56.5 (8.1)	56.5 (8.1)	58.6 (7.6)
Male sex	225,184 (45.5%)	224,664 (45.5%)	520 (55.1%)
BMI, kg/m ²	27.4 (4.8)	27.4 (4.8)	28.6 (5.3)
Waist circumference, cm	90.2 (13.4)	90.2 (13.4)	93.9 (14.0)
Diabetes	25,679 (5.2%)	25,578 (5.2%)	101 (10.7%)
Hypertension	119,035 (24.1%)	118,706 (24.0%)	329 (34.9%)
Grip strength, kg	32.7 (11.3)	32.7 (11.3)	31.4 (11.4)
Usual walking pace			
Slow	39,692 (8.0%)	39,492 (8.0%)	200 (21.2%)
Steady/average	260,489 (52.7%)	259,994 (52.7%)	495 (52.4%)
Brisk	192,838 (39.0%)	192,596 (39.0%)	242 (25.6%)
Alcohol intake frequency			
Daily/almost daily	100,735 (20.4%)	100,544 (20.4%)	191 (20.2%)
3-4/week	114,400 (23.1%)	114,212 (23.1%)	188 (19.9%)
1-2/week	127,806 (25.8%)	127,588 (25.8%)	218 (23.1%)
1-3/month	55,165 (11.2%)	55,065 (11.2%)	100 (10.6%)
Special occasions only	56,792 (11.5%)	56,659 (11.5%)	133 (14.1%)
Never	39,199 (7.9%)	39,087 (7.9%)	112 (11.9%)

528 Common complete-case prediction cohort. Values are shown as mean (SD) or n (%). Analytic
529 cohort n = 494,527; primary endpoint events = 944. Rare special-code categories were omitted
530 from the main baseline table for readability.

531

532 Table 2 Primary and supportive association results for grip strength and usual walking pace

Endpoint	Variable	Analytic n	Events	HR (95% CI)	P value
Hospital-record-defined incident DCM (any-position)	Grip strength (per 5 kg)	494,527	944	0.845 (0.810-0.881)	2.52e-15
Hospital-record-defined incident DCM (any-position)	Usual walking pace (per category)	494,527	944	0.824 (0.785-0.866)	7.35e-15
Primary-position hospital-record-defined incident DCM	Grip strength (per 5 kg)	494,527	616	0.842 (0.800-0.887)	6.36e-11
Primary-position hospital-record-defined incident DCM	Usual walking pace (per category)	494,527	616	0.832 (0.782-0.884)	3.27e-09
Surgically treated DCM (main)	Grip strength (per 5 kg)	494,527	538	0.858 (0.812-0.907)	5.71e-08
Surgically treated DCM (main)	Usual walking pace (per category)	494,527	538	0.846 (0.789-0.907)	2.71e-06
Diagnosis-to-surgery capture among hospital-record-defined incident DCM	Grip strength (per 5 kg)	944	534	1.026 (0.973-1.083)	0.342
Diagnosis-to-surgery capture among hospital-record-defined incident DCM	Usual walking pace (per category)	944	534	1.058 (0.957-1.170)	0.268

533 Fully adjusted estimates for the primary and supportive hospital-record-defined endpoint layers.

534 The diagnosis-to-surgery capture layer is shown for context and should not be interpreted as true

535 clinical progression.

536 Table 3 Prediction-model discrimination and incremental gain

Model	Validation Harrell's C-index	Validation 10- year AUC	ΔC-index vs base	ΔAUC vs base
Base clinical model	0.6280	0.6272	0.0000	0.0000
Base + grip strength	0.6445	0.6457	0.0165	0.0185
Base + usual walking pace	0.6469	0.6430	0.0189	0.0158
Base + grip strength + usual walking pace	0.6560	0.6548	0.0280	0.0275

Validation discrimination metrics for the base clinical model and the two motor-functional reserve augmentation strategies. Harrell's C-index values are reported after checking that higher model scores corresponded to higher predicted risk.

Table 4 Calibration and supportive utility summary

Section	Metric	Value
Combined model	Validation Harrell's C-index	0.6560
Combined model	Validation 10-year AUC	0.6548
Combined model	Decile-based calibration intercept	0.0001
Combined model	Decile-based calibration slope	1.3804
Combined model	Weighted observed/predicted ratio	1.4878
Combined model	Maximum absolute decile gap	0.0009
Combined model	Effective decision-threshold window	0.001-0.003
Supportive endpoint	Incident DCM (primary-only) validation 10-year AUC	0.6564
Supportive endpoint	Surgically treated DCM (main) validation 10-year AUC	0.6380

Calibration and supportive utility results for the combined model (base + grip strength + usual walking pace). Supportive endpoint checks are reported as validation 10-year AUC values for the high-specificity endpoint and the main surgically treated DCM support layer. These metrics indicate modest risk enrichment, not an actionable stand-alone clinical tool.

Figure legends

546 Figures are supplied as separate files. Additional files contain supplementary tables and
547 supplementary figures.

548 Figure 1. Study design and endpoint hierarchy. Overview of cohort assembly, record-level
549 endpoint reconstruction, clinically ordered endpoint layers, and the common complete-case
550 prediction cohort.

551 Figure 2. Adjusted associations with hospital-record-defined incident degenerative cervical
552 myelopathy. Fully adjusted estimates for the primary endpoint, hospital-record-defined incident
553 DCM (any-position), adjusted for age, sex, body mass index, waist circumference, alcohol intake
554 frequency, diabetes, hypertension, grip strength, and usual walking pace. Analytic cohort n =
555 494,527; primary endpoint events = 944.

556 Figure 3. Associations of grip strength and usual walking pace across supporting hospital-record-
557 defined endpoint layers. Grouped estimates for primary-position incident DCM, surgically
558 treated DCM, and the contextual diagnosis-to-surgery capture layer. The two focal markers
559 retained a concordant protective direction in the primary-position and surgically treated layers
560 but not in the contextual capture analysis.

561 Figure 4. Validation Harrell's C-index across the four prediction models. Best performance was
562 achieved by the combined model (base + grip strength + usual walking pace), followed by the
563 walking-pace and grip-only augmented models. Validation cohort n = 148,359; primary endpoint
564 events = 284.

565 Figure 5. Validation 10-year discrimination and incremental gain. Panel A shows validation 10-
566 year AUC across the four models. Panel B shows incremental gain over the base clinical model
567 in Harrell's C-index and 10-year AUC. The combined model delivered the largest gain in both
568 discrimination metrics.

569 Figure 6. Calibration of the combined model. Decile-based calibration of the model combining
570 the base clinical variables with grip strength and usual walking pace. Observed risk remained
571 above predicted risk across deciles, indicating average underprediction despite preserved
572 discrimination.

573 Figure 7. Decision-curve analysis. Validation-set net benefit across the clinically relevant very-
574 low-threshold range. The combined model had the broadest positive net-benefit window, but
575 positive net benefit was concentrated at thresholds approximately 0.001 to 0.003.

576 Figure 8. Endpoint-transition and diagnosis-to-surgery capture summary. The dominant coded
577 endpoint structure was transition from no hospital-record-defined DCM to incident DCM and
578 then into the diagnosis-procedure-paired surgical layer; direct 0-to-2 capture was rare. The 1-to-2
579 median interval of 0 days supports interpreting this layer as peri-diagnostic capture rather than
580 true clinical progression.

Figure 1. Study design and endpoint hierarchy

UK Biobank DCM analysis: endpoint reconstruction, time-to-event structures, prediction cohort, and hospital-record endpoint hierarchy.

A. Cohort and endpoint reconstruction

Baseline cohort
Participants with baseline date
n = 501,936

Endpoint reconstruction
Raw DCM diagnosis participants: 1,241
Incident DCM (any-position): 978
Surgery-main: 552
Surgery-strict: 545

B. Time-to-event structures

Transition	Events
0 -> 1 transitions	974
0 -> 2 direct	4
1 -> 2 transitions	548

C. Prediction cohort

Complete-case prediction cohort
Complete-case cohort: 494,527
Primary endpoint events: 944
Surgery-main events: 538

D. Endpoint hierarchy

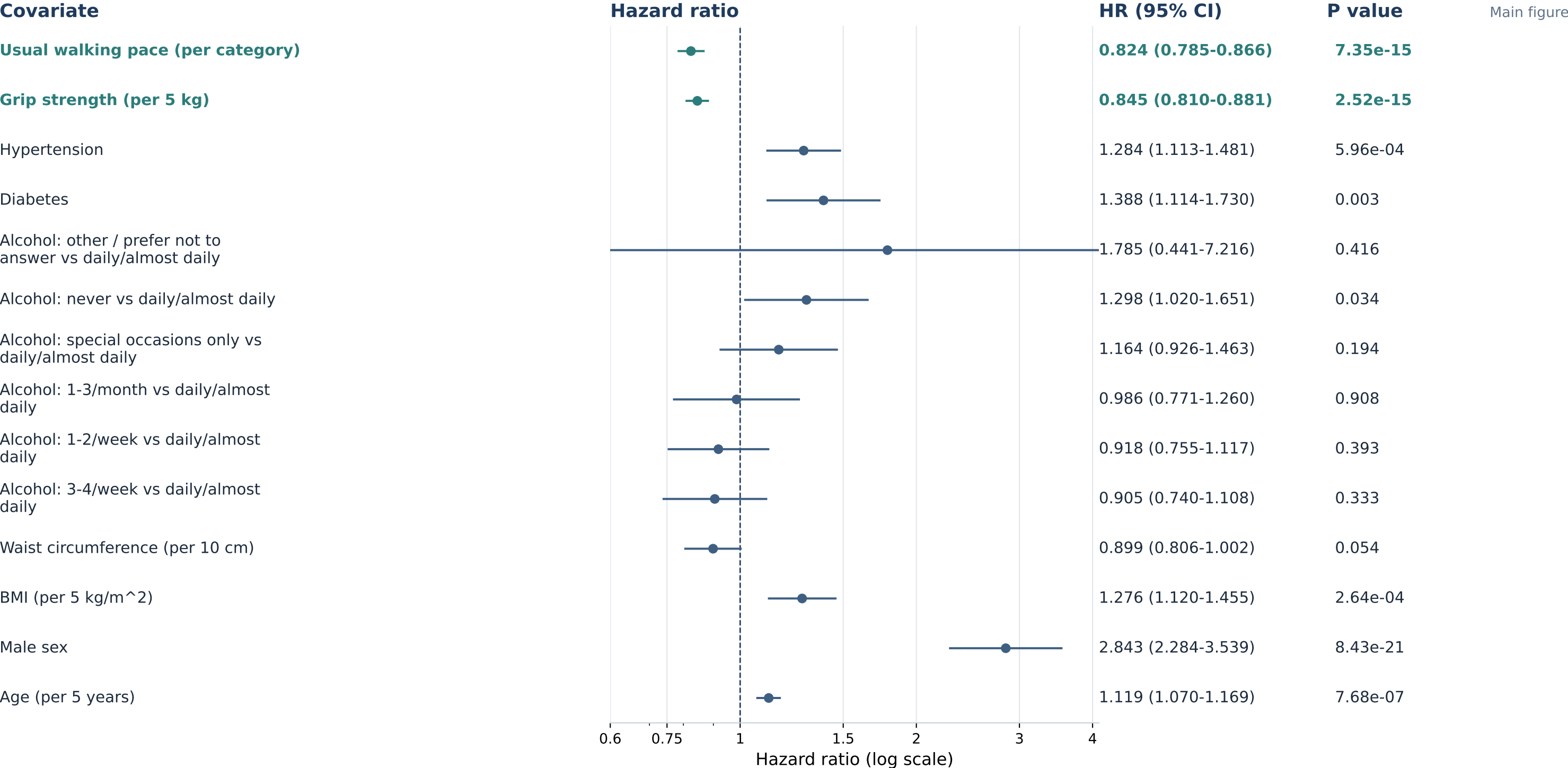
Role	Endpoint	n
Primary	DCM (any-position)	978
Sensitivity	DCM (primary-only)	639
Key secondary	Surgery-main	552
Strict	Surgery-strict	545

Manuscript analysis backbone

- Hospital-record-defined incident DCM (any-position) was retained as the primary endpoint.
- Grip strength and usual walking pace formed the two-pillar motor-functional reserve axis.
- Diagnosis-to-surgery capture and enhancement layers were retained as contextual/supportive analyses.

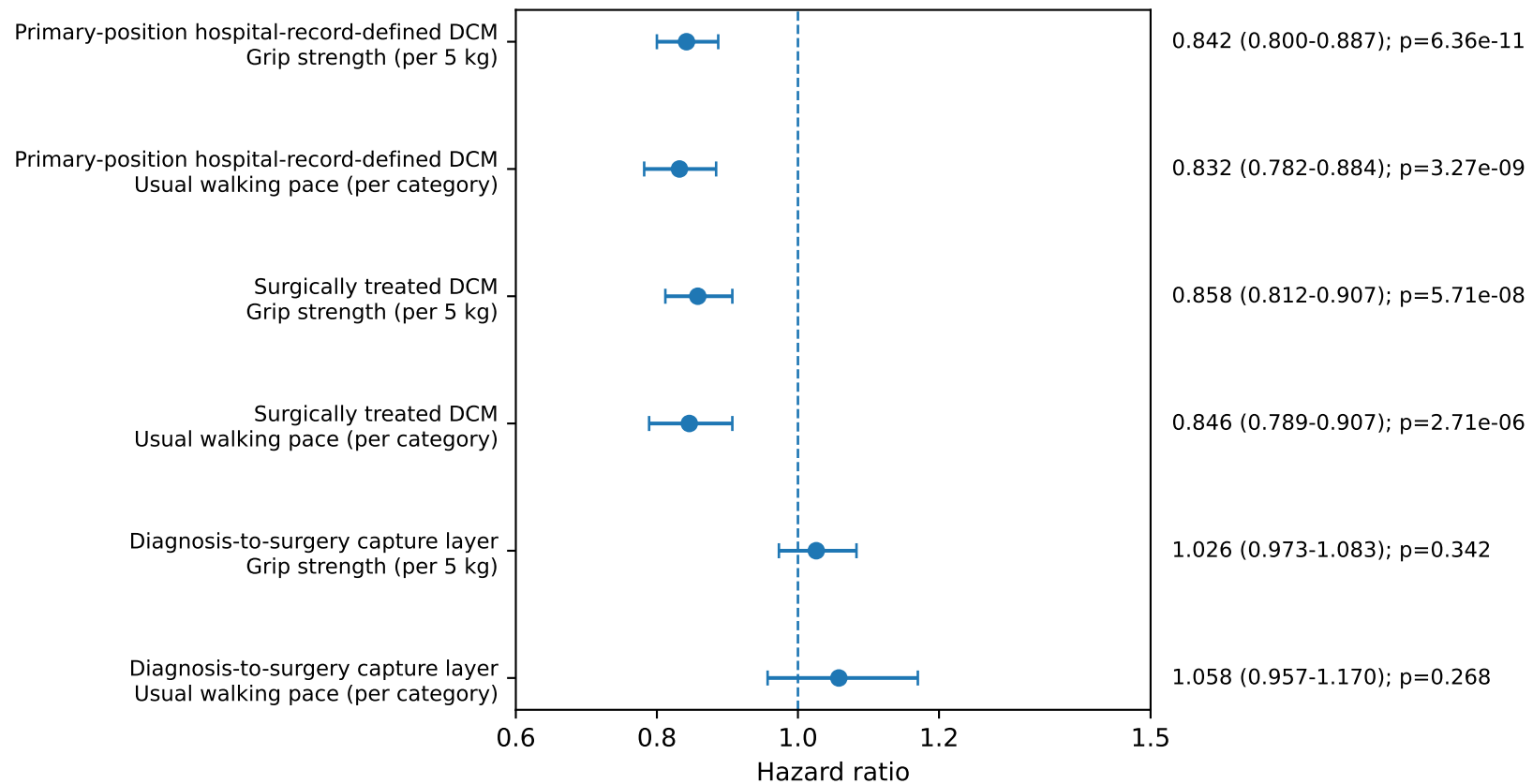
Figure 2. Adjusted associations with hospital-record-defined incident DCM

Fully adjusted estimates for hospital-record-defined incident DCM (any-position).



Adjusted for age, sex, body mass index, waist circumference, alcohol intake frequency, diabetes, hypertension, grip strength, and usual walking pace. Analytic cohort n = 494,527; primary endpoint events = 944.

Supporting endpoint layers for grip strength and usual walking pace



Validation discrimination across prediction models

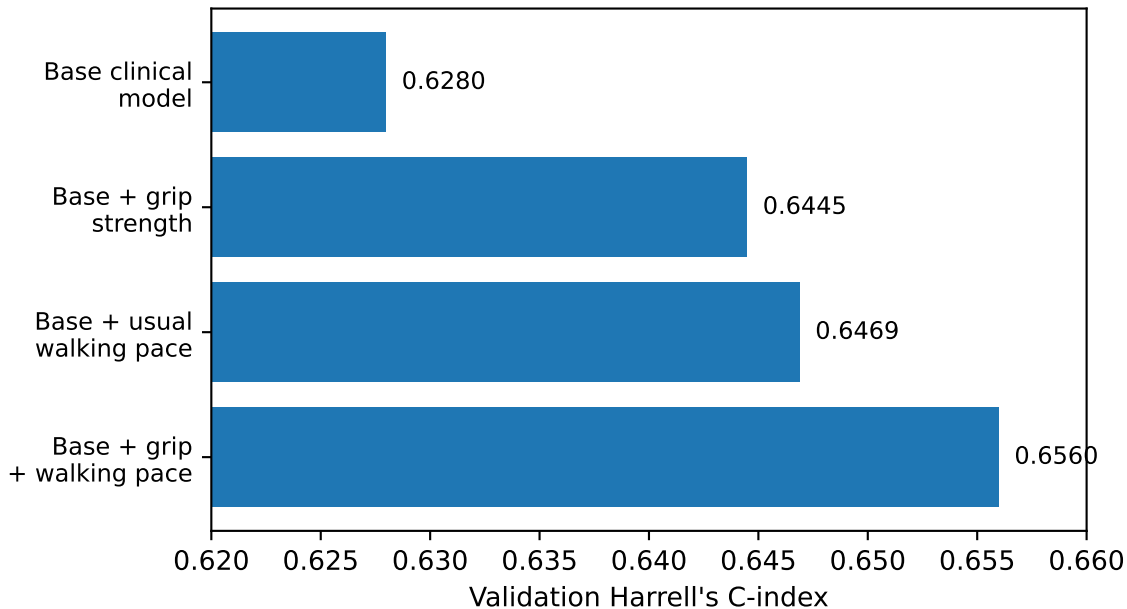
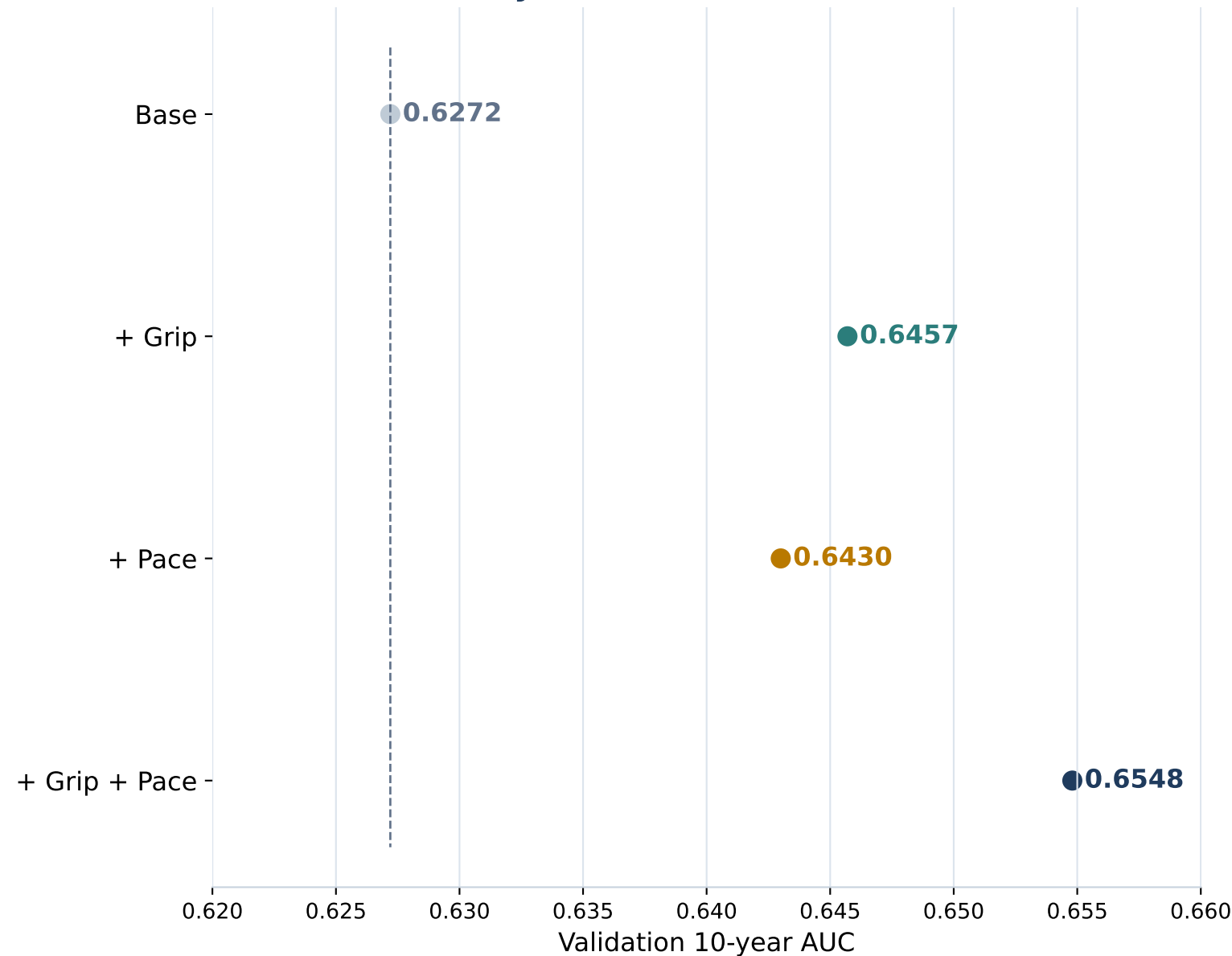


Figure 5. Validation 10-year discrimination and incremental gain

Main figure

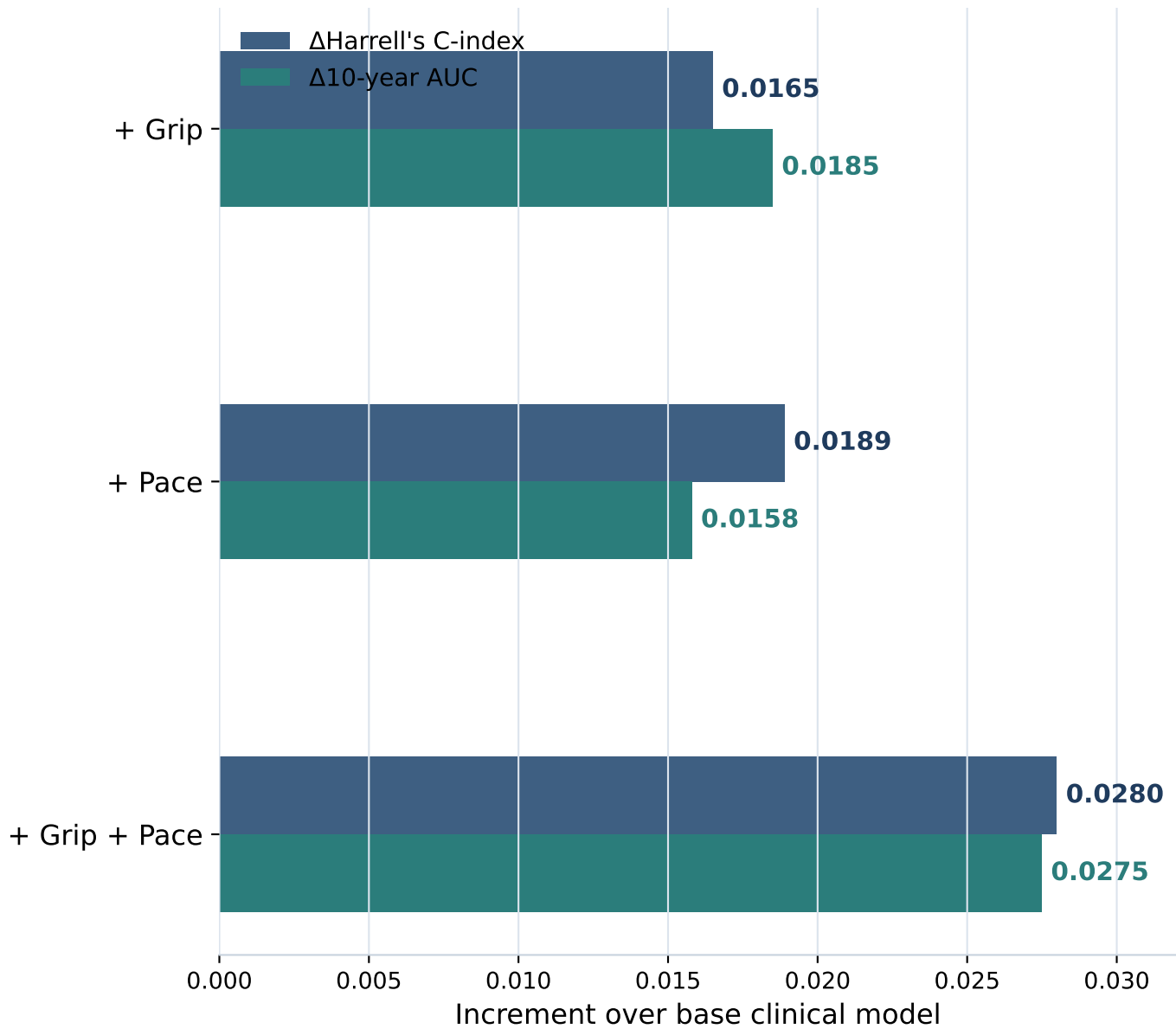
Validation 10-year AUC and incremental discrimination associated with grip strength and usual walking pace.

A. Validation 10-year AUC



Validation cohort: n = 148,359; primary endpoint events = 284.

B. Increment over the base clinical model



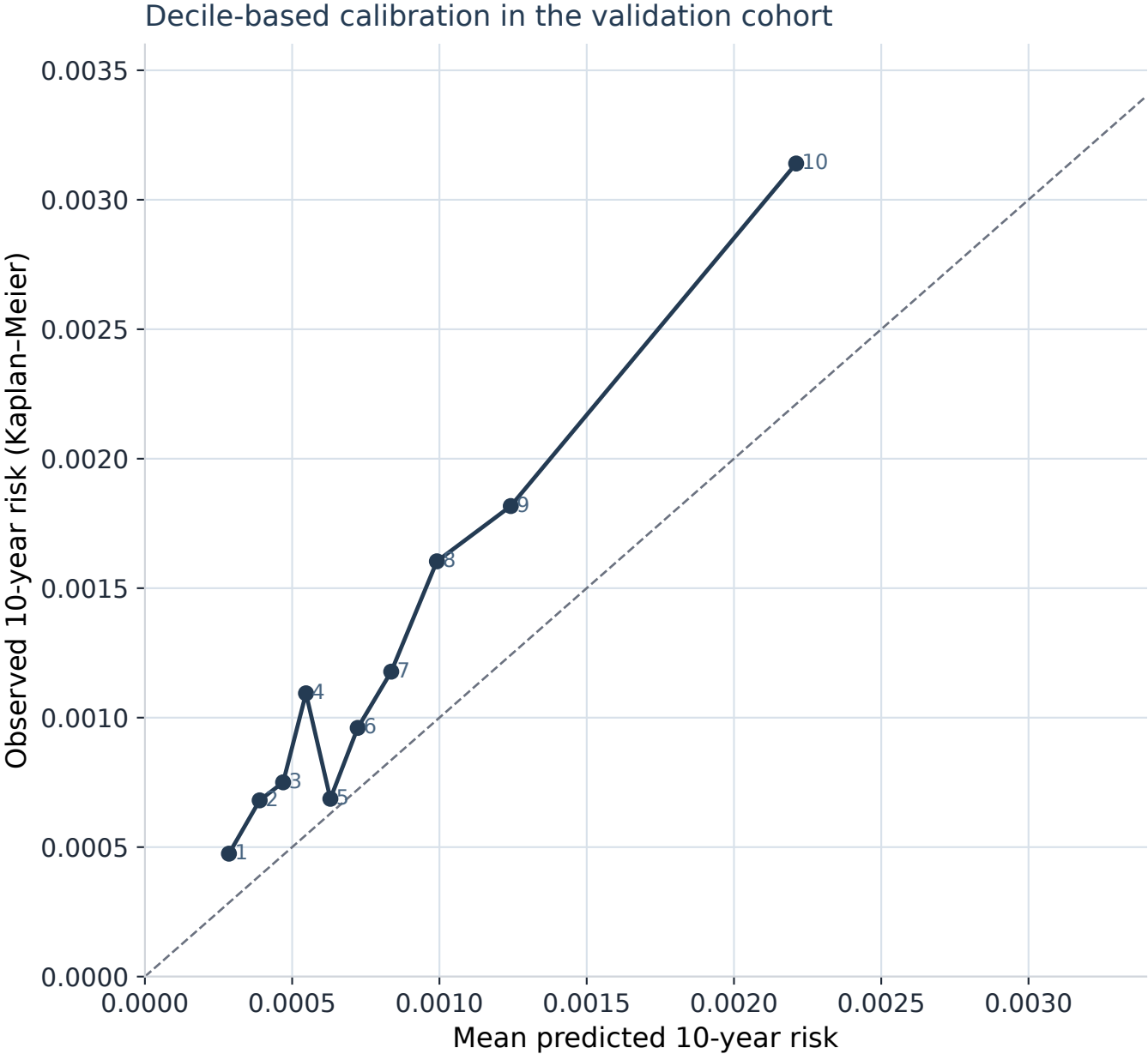
The combined model delivered the largest gain in both validation 10-year AUC and Harrell's C-index.

Figure 6. Calibration of the combined model

Decile-based calibration of the model combining the base clinical variables with grip strength and usual walking pace.

Calibration summary

Main figure



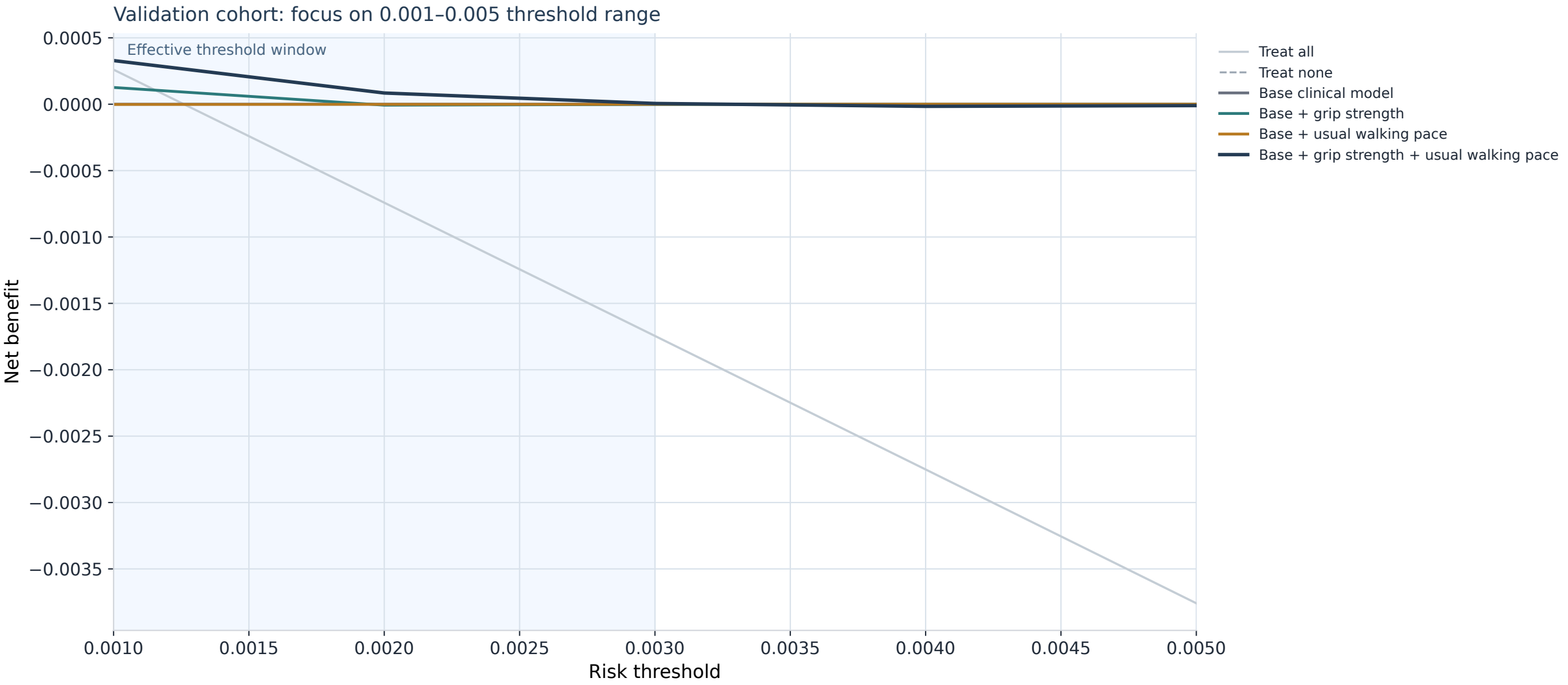
Validation sample	148,359
Primary events	284
Intercept	0.0001
Slope	1.3804
Observed / predicted ratio	1.4878
Maximum absolute decile gap	0.0009

Observed risk remained consistently above predicted risk, indicating systematic underprediction despite preserved discrimination.

Figure 7. Decision-curve analysis

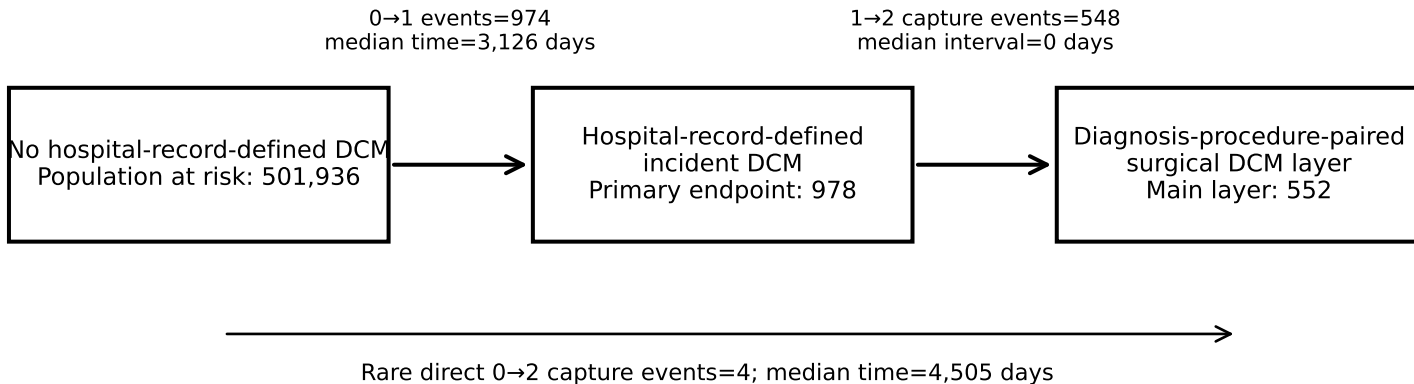
Main figure

Validation-set net benefit across the clinically relevant very-low-threshold range. The combined model had the broadest positive net-benefit window.



The combined model yielded positive net benefit across thresholds 0.001–0.003, whereas gains outside this very-low-threshold range were limited.

Endpoint-transition and diagnosis-to-surgery capture summary



The surgical layer is interpreted as diagnosis-to-surgery capture under the coding/pairing framework, not as proof of true clinical progression from first symptoms to operation.

Supplementary Files

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

- [Additionalfile1supplementarytablesandfigurelegends.docx](#)
- [Additionalfile2supplementaryfigures.pdf](#)