

Association of clinical outcome assessments of mobility capacity and incident disability in community-dwelling older adults - a systematic review and meta-analysis

Tobias Braun, PhD; Christian Thiel, PhD; Raphael Simon Peter, PhD; Carolin Bahns, MSc; Gisela Büchele, PhD; Kilian Rapp, MD, MPH, PhD; Clemens Becker, MD, PhD; Christian Grüneberg, PhD

Author Affiliations:

Department of Applied Health Sciences, Hochschule für Gesundheit (University of Applied Sciences), Bochum, Germany (Braun, Thiel, Grüneberg);

Department of Clinical Gerontology, Robert-Bosch-Hospital, Stuttgart, Germany (Braun, Rapp, Becker);

Institute of Epidemiology and Medical Biometry, Ulm University, Ulm, Germany (Peter, Büchele);

Digital Geriatric Medicine, Medical Clinic, Heidelberg University, Germany (Becker);

Department of Therapy Science I, Brandenburg Technical University Cottbus - Senftenberg, Senftenberg, Germany (Bahns)

HSD Hochschule Döpfer (University of Applied Sciences), Department of Health, Cologne, Germany (Braun);

Faculty of Sports Science, Ruhr-University Bochum, Bochum, Germany (Thiel).

Corresponding Author: Tobias Braun, PhD, Department of Applied Health Sciences, Hochschule für Gesundheit (University of Applied Sciences), Bochum, Gesundheitscampus 6-8, 44801 Bochum, Germany.

Tel. business: +49 (0) 234 77727 629

Fax: +49 (0) 234 77727 829

E-mail: tobias.braun@hs-gesundheit.de

Manuscript word count (not including title, abstract, acknowledgement, references, tables, and figure legends): 3.473 words

Key Points

QUESTION What are the associations between clinical outcome assessments of mobility capacity and incident disability in community-dwelling older adults?

FINDINGS In this systematic review and meta-analysis that included 40 reports and data of 85,515 older adults, the risk ratios of incident disability on activities of daily living were 1.23, 1.30, 1.15, and 1.07 for usual gait speed, Short Physical Performance Battery, Timed Up and Go test, and Chair Rise Test, respectively, per one conventional unit.

MEANING Common assessments of mobility capacity may help identify older people at risk of incident disability and should be routinely established in regular health examinations.

Abstract

IMPORTANCE The predictive value of common performance-based outcome assessments of mobility capacity on incident disability in activities of daily living in community-dwelling older adults remains uncertain.

OBJECTIVE To synthesize all available research on the association between mobility capacity and incident disability in non-disabled older adults.

DATA SOURCES MEDLINE, EMBASE and CINAHL databases were searched without any limits or restrictions.

STUDY SELECTION Published reports of longitudinal cohort studies that estimated a direct association between baseline mobility capacity, assessed with a standardized outcome assessment, and subsequent development of disability, including initially non-disabled older adults.

DATA EXTRACTION AND SYNTHESIS Data extraction was completed by independent pairs of reviewers. The risk of bias was assessed using the Quality in Prognosis Studies (QUIPS) tool. Random-effect models were used to explore the objective. The certainty of evidence was assessed using GRADE.

MAIN OUTCOME AND MEASURES The main outcome measures were the pooled relative risks (RR) per one conventional unit per mobility assessment for incident disability in activities of daily living.

RESULTS A total of 40 reports were included, evaluating 85,515 and 78,379 participants at baseline and follow-up, respectively (median mean age: 74.6 years). The median disability rate at follow-up was 12.0% (IQR: 5.4%–23.3%). The overall risk of bias was judged as low, moderate and high in 6 (15%), 6 (15%), and 28 (70%) reports, respectively.

For usual and fast gait speed, the RR per -0.1 m/s was 1.23 (95% CI: 1.18–1.28; 26,638 participants) and 1.28 (95% CI: 1.19–1.38; 8,161 participants), respectively. Each point decrease in Short Physical Performance Battery score increased the risk of incident disability by 30% (RR = 1.30, 95% CI: 1.23–

1.38; 9,183 participants). The RR of incident disability by each second increase in Timed Up and Go test and Chair Rise Test performance was 1.15 (95% CI: 1.09–1.21; 30,426 participants) and 1.07 (95% CI: 1.04–1.10; 9,450 participants), respectively.

CONCLUSIONS AND RELEVANCE Among community-dwelling non-disabled older adults, a poor mobility capacity is a potent modifiable risk factor for incident disability. Mobility impairment should be mandated as a quality indicator of health for older people.

Background

Activities of daily living (ADL), such as bathing, dressing, and toileting, are essential for older individuals to remain independent.¹ Disability can be defined on the basis of having difficulty with one or more ADLs or needing (or receiving) help with a functional task, such as transferring, eating, and walking.² ADL disability is associated with increased mortality, institutionalization, greater use of care, and a lower quality of life.^{1,3} Disability is a complex, multidimensional, and dynamic process that can be prevented, delayed, or improved through evidence-based interventions, which are most effective if they are based on personalized risk stratification.^{4,5}

Mobility impairment, defined as reduced mobility capacity, is one of the strongest modifiable predictors for incident ADL disability in older adults.⁶⁻⁹ Mobility capacity can be used for stratification to start interventions.¹⁰ There are established methods to quantify mobility capacity such as single-component or multi-component clinical outcome assessments (COA) of walking, balance, transferring, or lower-extremity functioning, including comfortable or fast gait speed, chair rise performance, the Timed Up and Go test (TUG), or the Short Physical Performance Battery (SPPB).^{10,11} These performance-based measures of supervised capacity are relatively simple to perform and accepted by older people.^{8,12}

There is solid evidence for a longitudinal association between physical functioning and disability onset and/or progression.^{6,7,9,13,14} In a recent meta-analysis, Wang et al.⁷ reported pooled longitudinal associations between physical performance measures (low/slow vs. high/fast) and ADL dependency for assessments of mobility, including the SPPB (odds ratio (OR): 3.5, 95% CI: 2.5–4.9), TUG (3.4, 1.9–6.3), gait speed (2.3, 1.6–3.4), and Chair Rise Test (CRT; 1.9, 1.6–2.2). However, these authors did not include all relevant studies into the meta-analyses and they calculated pooled OR, which do not approximate risk ratios (RR) when the outcome is frequent and are sometimes misinterpreted.¹⁵ Thus, the predictive value of common COAs of mobility capacity on incident ADL disability in community-dwelling older adults remains uncertain. Therefore, the objective of this study was to

systematically review the literature on the association between mobility capacity and incident ADL disability in community-dwelling older adults.

Methods

The systematic review protocol was prospectively registered with PROSPERO (CRD42020160490). Analysis and reporting were informed by recent recommendations¹⁶ and performed according to the PRISMA 2020 statement.¹⁷

Eligibility criteria

Reports were eligible if they met the following criteria: (1) Design: Published reports of longitudinal cohort studies that estimated a direct association between baseline mobility capacity and subsequent development of disability over a follow-up period of 1 to 6 years. (2) Population: Community-dwelling older adults, ≥ 60 years, non-disabled at baseline. (3) Prognostic factor: At least one COA of mobility capacity. (4) Outcome: At least one measure of ADL disability at follow-up, including nursing home admission, use of home care services or care dependency.

We excluded reports (1) not written in English or German, (2) solely focusing on individuals with clinical disorders such as dementia, and (3) from institutionalized settings such as hospitals or nursing homes. A full description of eligibility criteria is given in the eText 1 (Supplement).

Information sources and search strategy

We searched MEDLINE (OVID), EMBASE (OVID), and CINAHL (EBSCOhost) from inception to February 18, 2021. We used a combination of search terms related to “community-dwelling older adults”, “mobility capacity”, “disability” and “prediction”, informed by published search filters.^{16,18-20} The search strategy was kept broad, reviewed and revised by a senior information specialist, adjusted for each database, and is reported in the Supplement (eTables 1, 2, and 3). We imposed no limits or restrictions on any of the searches.

We also screened reference lists of a convenience sample of 16 related reviews (listed in eTable 4; 1,048 references) from the authors' private archives and manually screened reference lists of included reports for additional reports.

Selection process and data extraction

References were stored in EndNote X7.7. After duplicate removal by EndNote, pairs of reviewers (TB, CBa, and a third person) independently screened the titles and abstracts of identified records to remove obviously irrelevant reports (then using an over-inclusive approach at this stage). The level of inter-rater agreement between two reviewers was calculated by means of a kappa statistic.

A full-text review of potentially relevant reports was conducted by one reviewer (TB) and validated by a second reviewer (CBa). If necessary, missing information was requested from the authors. The reasons for full-text exclusion were recorded. At all steps, disagreement between reviewers was resolved by discussion. The methods of data extraction are reported in the eText 2 (Supplement).

Quality assessment

The risk of bias (RoB) of individual reports was assessed, as recommended,¹⁶ with the Quality in Prognosis Studies (QUIPS; six domains).²¹ The overall RoB was judged as proposed by Grooten et al.²² If all domains were classified as having low RoB, or up to one moderate RoB, the overall RoB was low. If one or more domains were classified as having high RoB, or ≥ 3 moderate RoB, the overall RoB was high. All reports in between were classified as having moderate RoB. The QUIPS tool was completed by one reviewer (TB).

The methodological quality of included reports was assessed using the Newcastle-Ottawa scale (NOS) for cohort studies.²³ A detailed description of how we applied the NOS in this review is given in the eText 3 in the Supplement. The NOS was assessed by one reviewer (TB).

The Grading of Recommendations, Assessment, Development and Evaluations (GRADE) approach for prognosis studies was used to assess the overall certainty of evidence of the meta-analyses as high, moderate, low, or very low.²⁴

Data synthesis and analysis

The primary outcome was any effect measure (eg, hazard ratio (HR), OR, RR) of baseline mobility capacity and incident ADL disability at follow-up for each pair predictor variable and outcome. In the meta-analysis, all results were re-expressed as RRs per unit of the respective mobility outcome assessment (eg, 0.1 m/s for gait speed).

The decision which reports/analyses were eligible for grouped synthesis was based on the following criteria: (a) A minimum of 5 reports for a given standardized mobility assessment. Analyses based on single items of a comprehensive assessment, such as single items of the POMA, were not eligible. (b) In case of different follow-up periods of the same sample, we used the longer period. (c) If results were reported for the total sample and clustered subgroups (eg, by gender), we only used the total sample. (d) If unadjusted and adjusted associations were reported, we used the most fully adjusted estimations.

Associations for different units of measurement (eg, per 1 m/s vs. per 0.1 m/s, walking time over 4 m vs. 6 m vs. 20 feet) were converted to one conventional unit per mobility assessment method (eg, 0.1 m/s for gait speed). Associations reported for ordinal categorical exposures (eg, for thirds, quarters, or by established cut-points) were transferred into continuous measures of association using methods of dose-response meta-analysis.²⁵ ORs were converted into RRs using the marginal risk for the outcome reported in the respective study. Finally, the standard error of estimates (if not reported) was calculated from reported confidence intervals or p-values. Combined estimates from random effects meta-analyses were calculated using the inverse variance method and the DerSimonian-Laird estimator for τ^2 and visualized as forest plots for each mobility assessment method. Statistical analyses were conducted using R (R Foundation for Statistical Computing, Vienna,

Austria) version 3.6.3 with the packages ‘dosresmeta’ version 2.0.1 and ‘meta’ version 4.18-2. SPSS version 23.0 (IBM corp., Armonk, NY, USA) was used for descriptive data analysis. We conducted sensitivity meta-analyses restricted to trials with an overall low RoB.

To assess the RoB across studies (small-study effects), we generated funnel plots for meta-analyses including data of at least 10 reports of varying size and used Egger’s test for funnel plot asymmetry.²⁶ In case of asymmetry, we reviewed the characteristics of the reports to check for possible reasons, such as publication bias.²⁶

Results

Study selection

The database search identified 12,038 records (flow chart in eFigure 1 in the Supplement). After removal of duplicates, two pairs of reviewers screened the titles and abstracts of 9,187 records, with a kappa agreement of 0.56 (95% CI: 0.50–0.62, n = 8,777, TB and CBa) and 0.52 (95% CI: 0.37–0.67, n = 410, TB and LH), respectively. Full-text assessment of 318 articles led to the inclusion of 32 reports. The screening of the reference lists of 16 relevant reviews and the included reports led to the full-text assessment of 51 additional reports, of which 8 were included in this review. All 329 reports excluded after full-text assessment are listed in eTables 5 and 6 (Supplement) with reasons for exclusion.

Study characteristics

We included 40 reports based on 33 different longitudinal cohort studies (key and additional characteristics in eTable 7, eTable 8, and eTable 9 in the Supplement, respectively).²⁷⁻⁶⁶ Reports came from the US (15, 38%), Japan (10, 25%), Italy (4, 10%), or other single countries (11, 28%). Twenty-four (60%) reports were published between 2011 and 2021. The median follow-up was 3 years (IQR: 2–5; range: 1–6).

The total number of included participants at baseline and at follow-up were 85,515 (median of studies: 921; IQR: 465–1,929; range: 102–25,031) and 78,379 (median: 799; IQR: 453–1,655; range: 75–25,031), respectively. The median completion rate was 93% (IQR: 81%–100%; range: 56%–100%). Two reports included only women.^{38,55} All other reports included women and men, with a median proportion of women of 60% (IQR: 53%–70%; range 44%–100%). The minimal inclusion age of the single studies ranged from 60 years to 85 years, and the median mean age of participants was 74.6 years (IQR: 72.4–79.7; 37 reports).

Seventeen reports (43%) excluded older people with cognitive impairment, as indicated by a low MMSE score or the diagnosis of dementia. The median mean MMSE score at baseline was 26.9 points (IQR: 26.4–27.8; 20 reports).

The most frequent mobility assessment was usual gait speed (23 reports), followed by SPPB (12 reports), CRT (8 reports), fast gait speed (7 reports), and the TUG (7 reports). Balance tests (eg, one-leg standing, functional reach test) were each used in less than 5 reports. As expected, there was high heterogeneity in gait speed assessment conduction, i.e., with respect to distance, counting, or setting (eTable 10 in the Supplement).

Incident ADL disability according to (modified) Katz criteria was used as an outcome in 23 reports (58%), with a median of 15.7% of participants disabled at follow-up (IQR: 8.5%–27.5%). In those reports, ADL disability mainly was assessed by the self-reported inability or difficulty to perform ≥ 1 basic ADL, such as bathing, dressing, eating, toileting, transferring, or walking (overview of disability tasks in the eTable 11 in the Supplement). Incident long term care insurance (LTCI) certification was used in 9 reports (23%; all studies from Japan or Korea), with a median LTCI certification rate at follow-up of 9.6% (IQR: 3.8%–14.3%). Three reports (8%) recorded nursing home admissions/institutionalizations (cumulative event rates: 3.0%, 3.8%, and 16.6%). The median disability rate at follow-up, regardless of the disability outcome, was 12.0% (IQR: 5.4%–23.3%).

Methodological quality and RoB in studies

The mean NOS score was 7.5 ± 1.6 (range: 3–9) points, with 33 reports (83%) rated with ≥ 7 points (rating per report in eTable 12 in the Supplement).

RoB was relatively low concerning the measurement of the prognostic factor and the statistical analysis and reporting. For the other four categories, RoB was rather moderate or high, especially concerning study population and study attrition. The overall RoB was judged as low, moderate, and high in 6 (15%), 6 (15%), and 28 (70%) reports, respectively ('RoB summary figure' eFigure 2 and 'RoB graph' in eFigure 3). The certainty of evidence for each prognostic factor is described the eTable 13.

Results of single studies

In total, the included reports presented 161 single effect estimations between baseline mobility capacity and ADL disability at follow-up (eTable 14 in the Supplement). Unless otherwise specified, all data were derived from the primary reference of each included report.

Results of syntheses

Gait speed (usual)

Slower baseline gait speed at usual pace was significantly associated with disability at follow-up (RR per -0.1 m/s = 1.23, 95% CI: 1.18–1.28, $I^2 = 76\%$, 19 reports, 26,638 participants, Figure 1 panel A, moderate certainty of evidence).^{27, 28, 33, 38, 39, 45, 46, 48, 50, 52, 53, 55-57, 59-61, 64, 65} Egger's test ($t = 2.76$, $P = 0.01$) indicated small-study effects for this analysis, which we attributed to true heterogeneity between reports after visual inspection of the funnel plot and a review of the study characteristics (eFigure 4 in the Supplement).

The association was even higher when only reports with an overall low RoB were included in the meta-analysis (RR per -0.1 m/s = 1.34, 95% CI: 1.19–1.51, $I^2 = 0\%$, 2 reports, 4,442 participants, Figure 1 panel B).^{50, 53}

Gait speed (fast)

For fast gait speed, the RR for incident disability at follow-up per -0,1 m/s was 1.28 (95% CI: 1.19–1.38, $I^2 = 81\%$, 7 reports, 8,161 participants, Figure 2 panel A, low certainty of evidence).^{29,35,41–43,47,59}

Based on reports with an overall low RoB, the RR did not change considerably (1.25, 95% CI: 1.18–1.32, $I^2 = 0\%$, 2 reports, 1,567 participants, Figure 2 panel B).^{43,47}

Short Physical Performance Battery

Each point decrease in SPPB score increased the risk of incident disability by 30% (RR per 1-point decrease = 1.30, 95% CI: 1.23–1.38, $I^2 = 61\%$, 11 reports, 9,183 participants, Figure 3, moderate certainty of evidence).^{30,32,44,45,48,52,56,60,61,63,64}. All reports were judged to have an overall high RoB.

There was no evidence of small-study effects (funnel plot in eFigure 5 in the Supplement; Egger test: $t = 1.84$, $P = 0.09$).

Chair Rise Test

There was a statistically significant association between baseline Chair Rise Test performance and disability at follow-up (RR per 1-second increase = 1.07, 95% CI: 1.04–1.10, $I^2 = 61\%$, 7 reports, 9,450 participants, Figure 4, moderate certainty of evidence).^{28,36,41,42,51,55,56}

Timed Up and Go test

The RR per 1-second increase in TUG performance (poorer performance) was 1.15 (95% CI: 1.09–1.21, $I^2 = 32\%$, 5 reports, 30,426 participants, Figure 5 panel A, high certainty of evidence).^{47-49,51,58}

This association remained stable when only reports with an overall low RoB were included in the analysis (RR = 1.17, 95% CI: 1.08–1.25, $I^2 = 55\%$, 3 reports, 30,179 participants, Figure 5 panel B).^{47,49,51}

Discussion

In this systematic review and meta-analysis of 40 reports, including more than 85,000 community-dwelling older adults, we found that mobility capacity was significantly associated with incident ADL disability. The pooled relative risk per one conventional unit was calculated for four common, feasible, standardized, and performance-based outcome assessments, including gait speed at usual and fast pace, SPPB, TUG, and CRT. The results are based on a heterogeneous sample of studies with an overall relatively high risk of bias.

Each 0.1 m/s decrease in usual and fast gait speed increased the risk of subsequent disability by 23% and 28%, respectively. Our results are in agreement with the findings of a large meta-analysis of individual participant data from 7 independent cohort studies,¹³ including 27,220 community-dwelling older adults. The age-adjusted relative risk reduction per 0.1 m/s greater speed at usual pace for bathing or dressing dependence over 3 years was 32% in men and 26% in women. In another meta-analysis, Wang et al.⁷ reported a pooled OR of 1.64 per 0.1 m/s lower speed for ADL disability (2 reports). The pooled OR for older people with “low vs. high” gait speed was 2.33 (6 reports).⁷ These associations differ from our findings since they are expressed in ORs, data on usual and fast speed was pooled, and the associations for different units of measurement were not converted to one conventional unit, as in the present analysis.⁷

For the SPPB, our meta-analysis of 11 reports suggests that each 1-point decrease increased the risk of incident disability by 30%. Our analysis extends the findings of two previous reviews reporting on the predictive ability of the SPPB for ADL disability.^{7,67} In the most recent review, the OR per 1-point decrease for incident disability was 1.12 (7 reports).⁷ In a prospective cohort study, the SPPB was associated with different disability subtypes, with adjusted HRs ranging from 1.10 for transient disability to 1.35 for long-term disability.⁶⁸

Based on 7 reports, our findings suggest that each additional second needed to complete the CRT increases the risk of ADL disability by 7%. A previous review, which was based on 3 reports and dichotomized CRT results, also found a strong association (slow vs. fast, OR = 1.90).⁷

For TUG we found a pooled RR of 1.15 per 1-second increase in test completion that remained stable in the sensitivity analysis excluding those reports with a high risk of bias. A systematic review and a recent cohort study of 1,084,875 older adults found that slower TUG speed is associated with incident ADL disability in older adults.^{7,69}

In essence, our analyses show that all included assessments of mobility capacity have a predictive value for ADL disability. Strong associations have also been reported between mobility and mortality by others.⁶⁹⁻⁷²

Assessments of mobility capacity should be implemented in regular health and/or physical evaluations of older adults to identify people at risk of disability. The general practitioner and other qualified health professionals may perform such checks in various clinical or community settings.^{51,71}

Although we included only assessments considered feasible in such circumstances, feasibility varies between instruments. For example, the TUG cannot be completed by a significant number of community-dwelling older adults due to physical and/or cognitive abilities, and insufficient space can limit the testing in a patient's home environment.⁷³ Approximately 20% of community-dwelling older adults are unable to complete 5 chair rises.⁷⁴ Gait speed assessment is quick and feasible, but not trivial. The testing protocol has a relevant impact on the recorded gait speed, especially the starting modalities (standing vs. walking), walking pace (usual vs. fast), distance and setting.^{75,76} For example, compared with real-world walking, patients consistently walk faster during 'normal' walking in (supervised) laboratory settings.⁷⁷ In agreement with other reviews,^{7,70} we observed large heterogeneity in test protocols. In clinical care and research, a standardized test protocol is essential for obtaining a reliable and valid assessment of gait speed or other mobility parameters.⁷⁵

Walking-related digital mobility outcomes or unsupervised assessment of mobility with mobile health technologies, such as real-world walking speed recorded with wearable sensors or chair rise activities recorded with domestic-integrated devices, may overcome the limitations of conventional (supervised) COAs, such as differences between gait speed test protocols.^{78,79} Moreover, digital (unsupervised) assessments of mobility appear to be more accurate and sensitive to change, and thus may emerge as complementary assessments to support risk stratification and monitoring of physical capacity of older people.⁷⁸ Methods for the non-supervised assessment of gait speed and other digital gait parameters in an older individual's usual environment are under development.^{80,81} The SPPB's use for non-supervised assessment is doubtful.⁸²

For clinical implementation, evidence of predictive validity for disability of each assessment should be considered. Some of the included studies performed direct comparisons between COA of mobility capacity. For example, Minneci et al.⁵² compared the prognostic value of usual gait speed and the SPPB, and reported non-significant differences between AUC values for incident disability (0.73 versus 0.71). The direct comparison by Makizako et al.⁵¹ showed that the predictive value of the TUG (HR = 2.24, 95% CI: 1.42–3.53) is better than of the CRT (HR = 1.88, 95% CI: 1.11–3.20).

COAs of mobility capacity have also been used in composite prediction models, which might be more accurate than single parameters alone. For example, Jonkman et al.⁸³ included gait speed and CRT in a clinical prediction model for older people of 65–75 years that showed an AUC of 0.72 for the onset of functional decline. Another clinical prediction model that included CRT and balance performance in midlife (at 53 years) showed good performance (AUC = 0.74) for late-life disability, indicating that mobility capacity may improve risk stratification already at younger ages.⁸⁴

Limitations

This study has some limitations that are inherent to meta-analyses, including heterogeneity in populations and variable endpoint definitions across the included reports. For example, gait speed was assessed differently across the studies and ADL disability was measured in more than 10 ways.

The generalizability of our findings is limited since people with cognitive impairment and people who deceased were excluded from some included studies. We could analyze data only from authors who replied to our request and the results might be influenced by reporting bias since trial registers were not searched and conference abstracts were not included. We did, however, perform an extensive screening of reference lists and related reviews.

Conclusions

Our findings suggest that the SPPB, gait speed at usual or fast pace, the TUG, and the CRT identify people at risk of incident disability. Mobility impairment is a major and modifiable risk factor for the loss of autonomy, impaired quality of life, and mortality. It should be mandated as a quality indicator of health for older people.

Author Contributions:

Dr Braun and Dr Peter had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Concept and design: Braun, Thiel, Becker, Grüneberg.

Acquisition of data: Braun, Bahns.

Analysis of data: Braun, Peter, Büchele.

Interpretation of data: Braun, Thiel, Peter, Bahns, Büchele, Rapp, Becker, Grüneberg.

Drafting of the manuscript: Braun.

Critical revision of the manuscript for important intellectual content: Braun, Thiel, Peter, Bahns, Büchele, Rapp, Becker, Grüneberg.

Conflict of Interest Disclosures: None declared.

Funding/Support: This work was supported by the 'Wilhelm-Stiftung für Rehabilitationsforschung', Essen, Germany.

Role of the Funder/Sponsor: The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Additional Contributions: Lisa Happe, Katja Ehrenbrusthoff, Elena Cramer, Kathrin Stiller, and Jérôme Camerlynck (HS Gesundheit Bochum) contributed to the data extraction. Gabriele Meyer (HS Gesundheit Bochum) contributed to the development of the search strategy.

Figure legends:

Figure 1. Association of usual gait speed with incident disability (panel A: all studies; panel B: only studies with an overall low risk of bias)

Figure 2: Association of fast gait speed with incident disability (panel A: all studies; panel B: only studies with an overall low risk of bias)

Figure 3: Association of Short Physical Performance Battery with incident disability

Figure 4: Association of Chair Rise Test with incident disability

Figure 5: Association of Timed Up and Go test with incident disability (panel A: all studies; panel B: only studies with an overall low risk of bias)

References

1. Guralnik JM, Fried LP, Salive ME. Disability as a public health outcome in the aging population. *Annu Rev Public Health*. 1996;17:25-46. doi:10.1146/annurev.pu.17.050196.000325.
2. Gill TM. Assessment of function and disability in longitudinal studies. *J Am Geriatr Soc*. 2010;58 Suppl 2:12. doi:10.1111/j.1532-5415.2010.02914.x.
3. Fried LP, Guralnik JM. Disability in older adults: evidence regarding significance, etiology, and risk. *J Am Geriatr Soc*. 1997;45(1):92-100. doi:10.1111/j.1532-5415.1997.tb00986.x.
4. Olsen PØ, Termannsen A-D, Bramming M, Tully MA, Caserotti P. Effects of resistance training on self-reported disability in older adults with functional limitations or disability – a systematic review and meta-analysis. *Eur Rev Aging Phys Act*. 2019;16(1):1-25. doi:10.1186/s11556-019-0230-5.
5. Hardy SE, Dubin JA, Holford TR, Gill TM. Transitions between states of disability and independence among older persons. *Am J Epidemiol*. 2005;161(6):575-584. doi:10.1093/aje/kwi083.
6. den Ouden MEM, Schuurmans MJ, Arts IEMA, van der Schouw YT. Physical performance characteristics related to disability in older persons: a systematic review. *Maturitas*. 2011;69(3):208-219. doi:10.1016/j.maturitas.2011.04.008.
7. Wang DXM, Yao J, Zirek Y, Reijnierse EM, Maier AB. Muscle mass, strength, and physical performance predicting activities of daily living: a meta-analysis. *J Cachexia Sarcopenia Muscle*. 2020;11(1):3-25. doi:10.1002/jcsm.12502.
8. Vermeulen J, Neyens JCL, van Rossum E, Spreuwenberg MD, Witte LP de. Predicting ADL disability in community-dwelling elderly people using physical frailty indicators: a systematic review. *BMC Geriatr*. 2011;11(1):33. doi:10.1186/1471-2318-11-33.
9. Cavanaugh EJ, Richardson J, McCallum CA, Wilhelm M. The Predictive Validity of Physical Performance Measures in Determining Markers of Preclinical Disability in Community-Dwelling Middle-Aged and Older Adults: A Systematic Review. *Phys Ther*. 2018. doi:10.1093/ptj/pzy109.

10. Freiburger E, Sieber CC, Kob R. Mobility in Older Community-Dwelling Persons: A Narrative Review. *Front Physiol.* 2020;11. doi:10.3389/fphys.2020.00881.
11. Chung J, Demiris G, Thompson HJ. Instruments to assess mobility limitation in community-dwelling older adults: a systematic review. *J Aging Phys Act.* 2015;23(2):298-313. doi:10.1123/japa.2013-0181.
12. Walton MK, Powers JH, Hobart J, et al. Clinical Outcome Assessments: Conceptual Foundation- Report of the ISPOR Clinical Outcomes Assessment - Emerging Good Practices for Outcomes Research Task Force. *Value Health.* 2015;18(6):741-752. doi:10.1016/j.jval.2015.08.006.
13. Perera S, Patel KV, Rosano C, et al. Gait Speed Predicts Incident Disability: A Pooled Analysis. *J Gerontol A Biol Sci Med Sci.* 2016;71(1):63-71. doi:10.1093/gerona/glv126.
14. Savino E, Volpato S, Zuliani G, Guralnik JM. Assessment of mobility status and risk of mobility disability in older persons. *Curr Pharm Des.* 2014;20(19):3099-3113.
15. Holmberg MJ, Andersen LW. Estimating Risk Ratios and Risk Differences: Alternatives to Odds Ratios. *JAMA.* 2020;324(11):1098-1099. doi:10.1001/jama.2020.12698.
16. Riley RD, Moons KGM, Snell KIE, et al. A guide to systematic review and meta-analysis of prognostic factor studies. *BMJ.* 2019;364:k4597. doi:10.1136/bmj.k4597.
17. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ.* 2021;372:n71. doi:10.1136/bmj.n71.
18. van de Glind EMM, van Munster BC, Spijker R, Scholten RJPM, Hooft L. Search filters to identify geriatric medicine in Medline. *J Am Med Inform Assoc.* 2012;19(3):468-472. doi:10.1136/amiainl-2011-000319.
19. Ingui BJ, Rogers MA. Searching for clinical prediction rules in MEDLINE. *J Am Med Inform Assoc.* 2001;8(4):391-397. doi:10.1136/jamia.2001.0080391.
20. Geersing G-J, Bouwmeester W, Zuithoff P, et al. Search filters for finding prognostic and diagnostic prediction studies in Medline to enhance systematic reviews. *PLoS ONE.* 2012;7(2):e32844. doi:10.1371/journal.pone.0032844.

21. Hayden JA, van der Windt DA, Cartwright JL, Côté P, Bombardier C. Assessing bias in studies of prognostic factors. *Ann Intern Med*. 2013;158(4):280-286. doi:10.7326/0003-4819-158-4-201302190-00009.
22. Grooten WJA, Tseli E, Ång BO, et al. Elaborating on the assessment of the risk of bias in prognostic studies in pain rehabilitation using QUIPS-aspects of interrater agreement. *Diagn Progn Res*. 2019;3:5. doi:10.1186/s41512-019-0050-0.
23. Wells GA, Tugwell P, O'Connell D, et al. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomized studies in meta-analyses. 2015. http://www.ohri.ca/programs/clinical_epidemiology/. Accessed March 2, 2020.
24. Iorio A, Spencer FA, Falavigna M, et al. Use of GRADE for assessment of evidence about prognosis: rating confidence in estimates of event rates in broad categories of patients. *BMJ*. 2015;350:h870. doi:10.1136/bmj.h870.
25. Greenland S, Longnecker MP. Methods for trend estimation from summarized dose-response data, with applications to meta-analysis. *Am J Epidemiol*. 1992;135(11):1301-1309. doi:10.1093/oxfordjournals.aje.a116237.
26. Sterne JAC, Sutton AJ, Ioannidis JPA, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ*. 2011;343:d4002. doi:10.1136/bmj.d4002.
27. Abe T, Kitamura A, Taniguchi Y, et al. Pathway from gait speed to incidence of disability and mortality in older adults: A mediating role of physical activity. *Maturitas*. 2019;123:32-36. doi:10.1016/j.maturitas.2019.02.002.
28. Akune T, Muraki S, Oka H, et al. Incidence of certified need of care in the long-term care insurance system and its risk factors in the elderly of Japanese population-based cohorts: The ROAD study. *Geriatr Gerontol Int*. 2014;14(3):695-701. doi:10.1111/ggi.12155.

29. Artaud F, Singh-Manoux A, Dugravot A, Tzourio C, Elbaz A. Decline in Fast Gait Speed as a Predictor of Disability in Older Adults. *J Am Geriatr Soc*. 2015;63(6):1129-1136. doi:10.1111/jgs.13442.
30. Balzi D, Lauretani F, Barchielli A, et al. Risk factors for disability in older persons over 3-year follow-up. *Age Ageing*. 2010;39(1):92-98. doi:10.1093/ageing/afp209.
31. Beauchamp MK, Jette AM, Ward RE, et al. Predictive Validity and Responsiveness of Patient-Reported and Performance-Based Measures of Function in the Boston RISE Study. *J Gerontol A Biol Sci Med Sci*. 2015;70(5):616-622. doi:10.1093/gerona/glu227.
32. Björkman M, Jyväkorpi SK, Strandberg TE, Pitkälä KH, Tilvis RS. Sarcopenia Indicators as Predictors of Functional Decline and Need for Care among Older People. *J Nutr Health Aging*. 2019;23(10):916-922. doi:10.1007/s12603-019-1280-0.
33. Buchman AS, Leurgans SE, Wang T, et al. Motor function is the primary driver of the associations of sarcopenia and physical frailty with adverse health outcomes in community-dwelling older adults. *PLoS ONE*. 2021;16(2):e0245680. doi:10.1371/journal.pone.0245680.
34. Chaudhry SI, McAvay G, Ning Y, Allore HG, Newman AB, Gill TM. Geriatric impairments and disability: The cardiovascular health study. *J Am Geriatr Soc*. 2010;58(9):1686-1692. doi:10.1111/j.1532-5415.2010.03022.x.
35. Chu L-W, Chiu AYY, Chi I. Impact of falls on the balance, gait, and activities of daily living functioning in community-dwelling Chinese older adults. *J Gerontol A Biol Sci Med Sci*. 2006;61(4):399-404. doi:10.1093/gerona/61.4.399.
36. da Silva Alexandre T, Corona LP, Nunes DP, Santos JLF, Duarte YAdO, Lebrão ML. Gender differences in incidence and determinants of disability in activities of daily living among elderly individuals: SABE study. *Arch Gerontol Geriatr*. 2012;55(2):431-437. doi:10.1016/j.archger.2012.04.001.

37. DeVore PA. Ability of a computerized geriatric assessment to predict need for change in living status among elderly living at home. *South Med J*. 1994;87(7):743-748. doi:10.1097/00007611-199407000-00014.
38. Diem SJ, Lui L-Y, Langsetmo L, et al. Effects of Mobility and Cognition on Maintenance of Independence and Survival Among Women in Late Life. *J Gerontol A Biol Sci Med Sci*. 2018;73(9):1251-1257. doi:10.1093/gerona/glx209.
39. Doi T, Nakakubo S, Tsutsumimoto K, et al. Spatio-temporal gait variables predicted incident disability. *Journal of neuroengineering and rehabilitation*. 2020;17(1):11. doi:10.1186/s12984-020-0643-4.
40. Donoghue OA, Savva GM, Cronin H, Kenny RA, Horgan NF. Using timed up and go and usual gait speed to predict incident disability in daily activities among community-dwelling adults aged 65 and older. *Archives of Physical Medicine and Rehabilitation*. 2014;95(10):1954-1961. doi:10.1016/j.apmr.2014.06.008.
41. Gill TM, Richardson ED, Tinetti ME. Evaluating the risk of dependence in activities of daily living among community-living older adults with mild to moderate cognitive impairment. *J Gerontol A Biol Sci Med Sci*. 1995;50(5):M235-41. doi:10.1093/gerona/50a.5.m235.
42. Gill TM, Williams CS, Tinetti ME. Assessing risk for the onset of functional dependence among older adults: The role of physical performance. *J Am Geriatr Soc*. 1995;43(6):603-609. doi:10.1111/j.1532-5415.1995.tb07192.x.
43. Gill TM, Allore HG, Holford TR, Guo Z. Hospitalization, restricted activity, and the development of disability among older persons. *JAMA*. 2004;292(17):2115-2124. doi:10.1001/jama.292.17.2115.
44. Guralnik JM, Ferrucci L, Simonsick EM, Salive ME, Wallace RB. Lower-extremity function in persons over the age of 70 years as a predictor of subsequent disability. *N Engl J Med*. 1995;332(9):556-561. doi:10.1056/NEJM199503023320902.

45. Guralnik JM, Ferrucci L, Pieper CF, et al. Lower extremity function and subsequent disability: Consistency across studies, predictive models, and value of gait speed alone compared with the short physical performance battery. *J Gerontol A Biol Sci Med Sci*. 2000;55(4):M221-31.
46. Heiland EG, Welmer A-K, Wang R, et al. Association of mobility limitations with incident disability among older adults: A population-based study. *Age Ageing*. 2016;45(6):812-819. doi:10.1093/ageing/afw076.
47. Hoshi M, Hozawa A, Kuriyama S, et al. The predictive power of physical function assessed by questionnaire and physical performance measures for subsequent disability. *Aging Clin Exp Res*. 2012;24(4):345-353. doi:10.3275/8104.
48. Huang W-N, Perera S, VanSwearingen J, Studenski S. Performance measures predict onset of activity of daily living difficulty in community-dwelling older adults. *J Am Geriatr Soc*. 2010;58(5):844-852. doi:10.1111/j.1532-5415.2010.02820.x.
49. Lee JE, Chun H, Kim YS, et al. Association between Timed Up and Go Test and Subsequent Functional Dependency. *J Korean Med Sci*. 2020;35(3):e25. doi:10.3346/jkms.2020.35.e25.
50. Makizako H, Shimada H, Doi T, Tsutsumimoto K, Suzuki T. Impact of physical frailty on disability in community-dwelling older adults: A prospective cohort study. *BMJ Open*. 2015;5(9):e008462. doi:10.1136/bmjopen-2015-008462.
51. Makizako H, Shimada H, Doi T, et al. Predictive Cutoff Values of the Five-Times Sit-to-Stand Test and the Timed "Up & Go" Test for Disability Incidence in Older People Dwelling in the Community. *Phys Ther*. 2017;97(4):417-424. doi:10.2522/ptj.20150665.
52. Minneci C, Mello AM, Mossello E, et al. Comparative study of four physical performance measures as predictors of death, incident disability, and falls in unselected older persons: The insufficienza Cardiaca negli Anziani Residenti a Dicomano Study. *J Am Geriatr Soc*. 2015;63(1):136-141. doi:10.1111/jgs.13195.

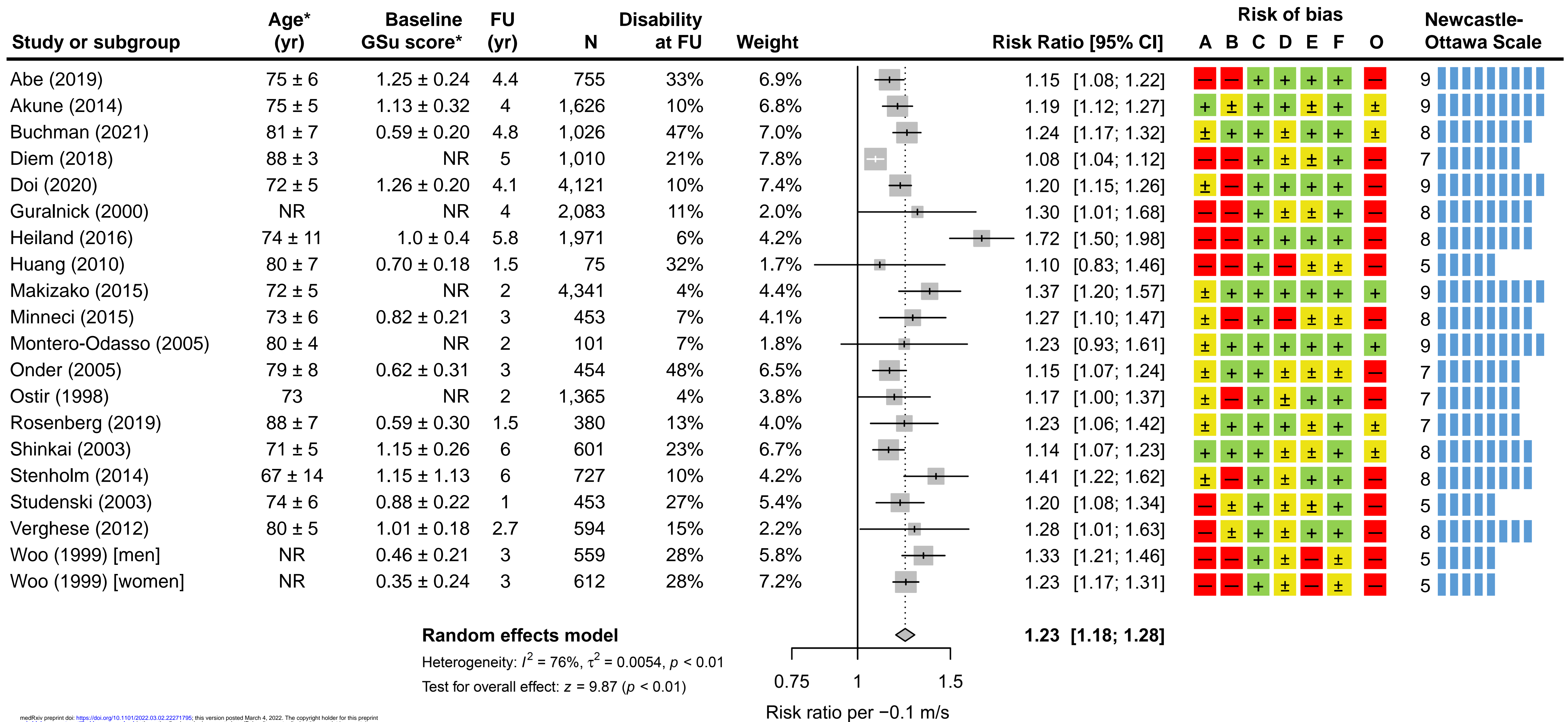
53. Montero-Odasso M, Schapira M, Soriano ER, et al. Gait velocity as a single predictor of adverse events in healthy seniors aged 75 years and older. *J Gerontol A Biol Sci Med Sci*. 2005;60(10):1304-1309. doi:10.1093/gerona/60.10.1304.
54. Moriya S, Murata A, Kimura S, Inoue N, Miura H. Predictors of eligibility for long-term care funding for older people in Japan. *Australas J Ageing*. 2013;32(2):79-85. doi:10.1111/j.1741-6612.2012.00601.x.
55. Onder G, Penninx BWJH, Ferrucci L, Fried LP, Guralnik JM, Pahor M. Measures of physical performance and risk for progressive and catastrophic disability: Results from the Women's Health and Aging Study. *J Gerontol A Biol Sci Med Sci*. 2005;60(1):74-79.
56. Ostir GV, Markides KS, Black SA, Goodwin JS. Lower body functioning as a predictor of subsequent disability among older Mexican Americans. *J Gerontol A Biol Sci Med Sci*. 1998;53(6):M491-5. doi:10.1093/gerona/53a.6.m491.
57. Rosenberg T, Montgomery P, Hay V, Lattimer R. Using frailty and quality of life measures in clinical care of the elderly in Canada to predict death, nursing home transfer and hospitalisation - the frailty and ageing cohort study. *BMJ Open*. 2019;9(11):e032712. doi:10.1136/bmjopen-2019-032712.
58. Sakamoto R, Okumiya K, Ishine M, et al. Predictors of difficulty in carrying out basic activities of daily living among the old-old: A 2-year community-based cohort study. *Geriatr Gerontol Int*. 2016;16(2):214-222. doi:10.1111/ggi.12462.
59. Shinkai S, Kumagai S, Fujiwara Y, et al. Predictors for the onset of functional decline among initially non-disabled older people living in a community during a 6-year follow-up. *Geriatr Gerontol Int*. 2003;3(s1):S31-S39. doi:10.1111/j.1444-0594.2003.00094.x.
60. Stenholm S, Guralnik JM, Bandinelli S, Ferrucci L. The Prognostic Value of Repeated Measures of Lower Extremity Performance: Should We Measure More Than Once? *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*. 2014;69(7):894-899. doi:10.1093/gerona/glt175.

61. Studenski S, Perera S, Wallace D, et al. Physical performance measures in the clinical setting. *J Am Geriatr Soc.* 2003;51(3):314-322. doi:10.1046/j.1532-5415.2003.51104.x.
62. Tsutsumimoto K, Doi T, Shimada H, et al. Combined Effect of Slow Gait Speed and Depressive Symptoms on Incident Disability in Older Adults. *J Am Med Dir Assoc.* 2016;17(2):123-127. doi:10.1016/j.jamda.2015.08.012.
63. Vaarst J, Boyle E, Vestergaard S, et al. Does physical performance and muscle strength predict future personal and nursing care services in community-dwelling older adults aged 75+? *Scand J Public Health.* 2021:1403494820979094. doi:10.1177/1403494820979094.
64. Verghese J, Holtzer R, Lipton RB, Wang C. Mobility stress test approach to predicting frailty, disability, and mortality in high-functioning older adults. *J Am Geriatr Soc.* 2012;60(10):1901-1905. doi:10.1111/j.1532-5415.2012.04145.x.
65. Woo J, Ho SC, Yu AL. Walking speed and stride length predicts 36 months dependency, mortality, and institutionalization in Chinese aged 70 and older. *J Am Geriatr Soc.* 1999;47(10):1257-1260. doi:10.1111/j.1532-5415.1999.tb05209.x.
66. Zhang F, Ferrucci L, Culham E, Metter EJ, Guralnik J, Deshpande N. Performance on five times sit-to-stand task as a predictor of subsequent falls and disability in older persons. *Journal of Aging and Health.* 2013;25(3):478-492. doi:10.1177/0898264313475813.
67. Gawel J, Vengrow D, Collins J, Brown S, Buchanan A, Cook C. The short physical performance battery as a predictor for long term disability or institutionalization in the community dwelling population aged 65 years old or older. *Phys Ther Rev.* 2012;17(1):37-44. doi:10.1179/1743288X11Y.0000000050.
68. Gill TM, Murphy TE, Barry LC, Allore HG. Risk factors for disability subtypes in older persons. *J Am Geriatr Soc.* 2009;57(10):1850-1855. doi:10.1111/j.1532-5415.2009.02443.x.
69. Chun S, Shin DW, Han K, et al. The Timed Up and Go test and the ageing heart: Findings from a national health screening of 1,084,875 community-dwelling older adults. *European Journal of Preventive Cardiology.* 2021;28(2):213-219. doi:10.1177/2047487319882118.

70. Veronese N, Stubbs B, Volpato S, et al. Association Between Gait Speed With Mortality, Cardiovascular Disease and Cancer: A Systematic Review and Meta-analysis of Prospective Cohort Studies. *J Am Med Dir Assoc*. 2018;19(11):981-988.e7. doi:10.1016/j.jamda.2018.06.007.
71. Pavasini R, Guralnik J, Brown JC, et al. Short Physical Performance Battery and all-cause mortality: systematic review and meta-analysis. *BMC Med*. 2016;14(1):215. doi:10.1186/s12916-016-0763-7.
72. Cooper R, Kuh D, Hardy R. Objectively measured physical capability levels and mortality: systematic review and meta-analysis. *BMJ*. 2010;341:c4467. doi:10.1136/bmj.c4467.
73. Rockwood K, Awalt E, Carver D, MacKnight C. Feasibility and measurement properties of the functional reach and the timed up and go tests in the Canadian study of health and aging. *J Gerontol A Biol Sci Med Sci*. 2000;55(2):70-73.
74. Guralnik JM, Simonsick EM, Ferrucci L, et al. A short physical performance battery assessing lower extremity function: association with self-reported disability and prediction of mortality and nursing home admission. *J Gerontol*. 1994;49(2):85-94.
75. Stuck AK, Bachmann M, Füllemann P, Josephson KR, Stuck AE. Effect of testing procedures on gait speed measurement: A systematic review. *PLoS One*. 2020;15(6):e0234200. doi:10.1371/journal.pone.0234200.
76. Krumpoch S, Lindemann U, Rappl A, Becker C, Sieber CC, Freiburger E. The effect of different test protocols and walking distances on gait speed in older persons. *Aging Clin Exp Res*. 2020. doi:10.1007/s40520-020-01703-z.
77. Rojer AGM, Coni A, Mellone S, et al. Robustness of In-Laboratory and Daily-Life Gait Speed Measures over One Year in High Functioning 61- to 70-Year-Old Adults. *Gerontology*. 2021:1-10. doi:10.1159/000514150.
78. Warmerdam E, Hausdorff JM, Atrsaei A, et al. Long-term unsupervised mobility assessment in movement disorders. *The Lancet Neurology*. 2020;19(5):462-470. doi:10.1016/S1474-4422(19)30397-7.

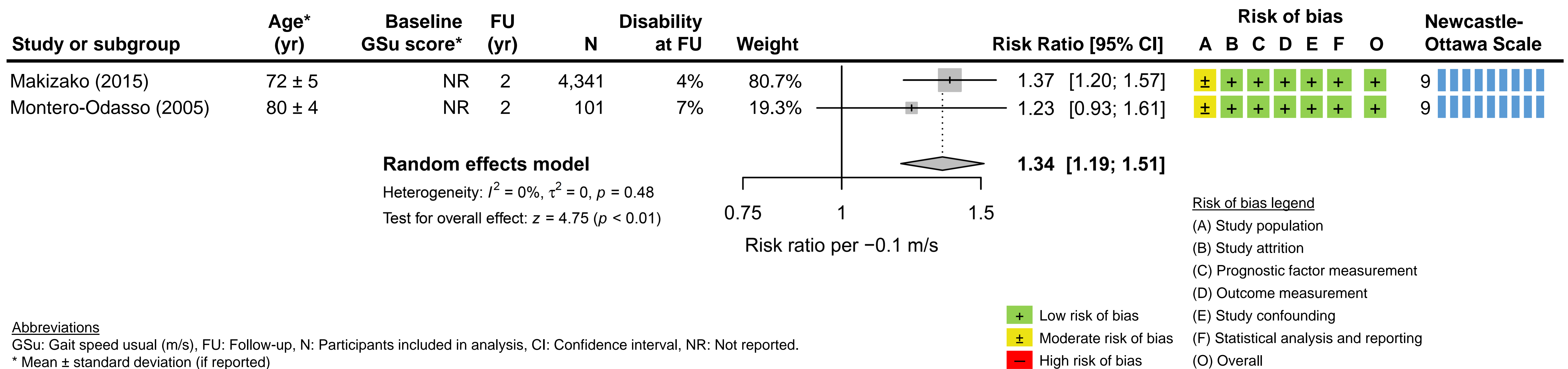
79. Polhemus A, Ortiz LD, Brittain G, et al. Walking on common ground: a cross-disciplinary scoping review on the clinical utility of digital mobility outcomes. *npj Digit Med*. 2021;4(1):1-14. doi:10.1038/s41746-021-00513-5.
80. Rochester L, Mazzà C, Mueller A, et al. A Roadmap to Inform Development, Validation and Approval of Digital Mobility Outcomes: The Mobilise-D Approach. *Digit Biomark*. 2020;4(Suppl 1):13-27. doi:10.1159/000512513.
81. Maetzler W, Pilotto A. Digital assessment at home - mPower against Parkinson disease. *Nat Rev Neurol*. 2021;17(11):661-662. doi:10.1038/s41582-021-00567-9.
82. Jung H-W, Roh H, Cho Y, et al. Validation of a Multi-Sensor-Based Kiosk for Short Physical Performance Battery. *J Am Geriatr Soc*. 2019. doi:10.1111/jgs.16135.
83. Jonkman NH, Colpo M, Klenk J, et al. Development of a clinical prediction model for the onset of functional decline in people aged 65-75 years: pooled analysis of four European cohort studies. *BMC Geriatr*. 2019;19(1):179. doi:10.1186/s12877-019-1192-1.
84. Dodds RM, Kuh D, Sayer AA, Cooper R. Can measures of physical performance in mid-life improve the clinical prediction of disability in early old age? Findings from a British birth cohort study. *Exp Gerontol*. 2018;110:118-124. doi:10.1016/j.exger.2018.06.001.

Panel A



medRxiv preprint doi: <https://doi.org/10.1101/2022.03.02.22271795>; this version posted March 4, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Panel B



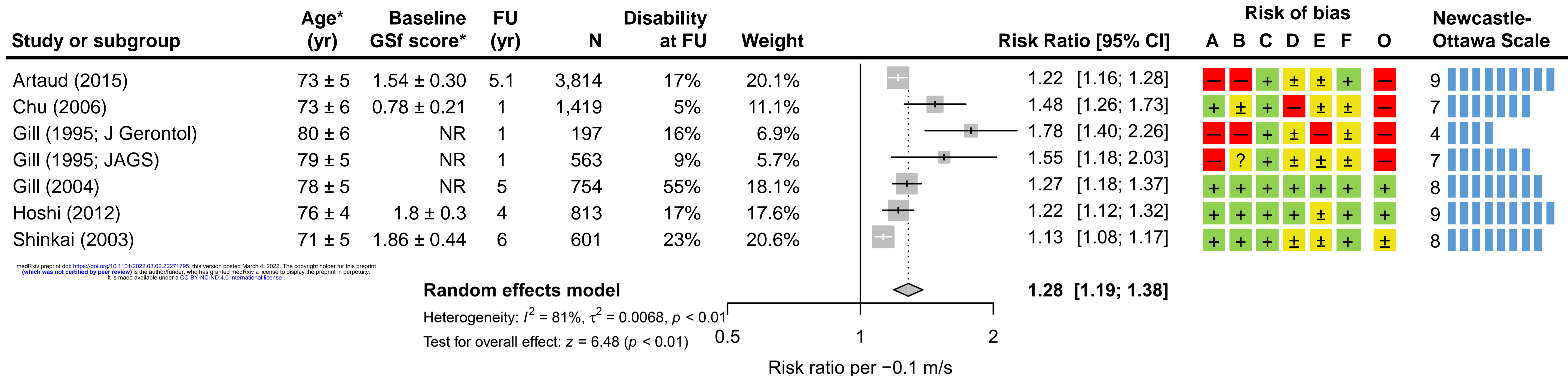
Risk of bias legend
 (A) Study population
 (B) Study attrition
 (C) Prognostic factor measurement
 (D) Outcome measurement
 (E) Study confounding
 (F) Statistical analysis and reporting
 (O) Overall

Abbreviations

GSu: Gait speed usual (m/s), FU: Follow-up, N: Participants included in analysis, CI: Confidence interval, NR: Not reported.
 * Mean ± standard deviation (if reported)

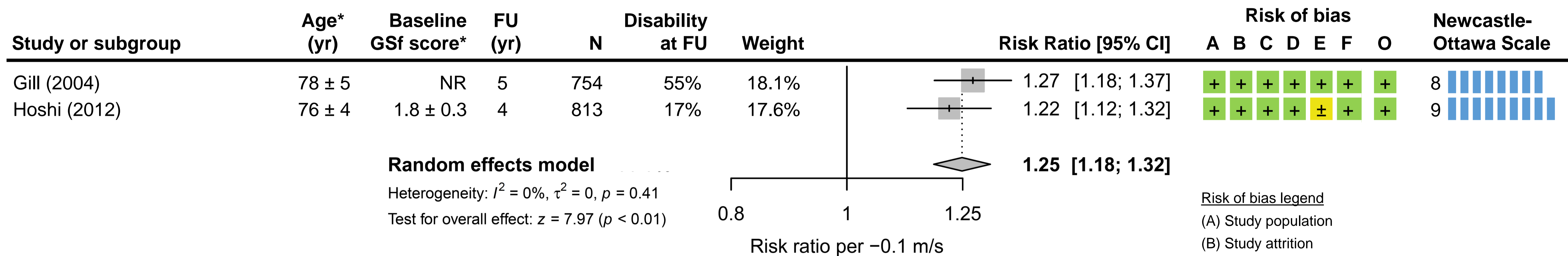
+ Low risk of bias
 ± Moderate risk of bias
 — High risk of bias

Panel A



medRxiv preprint doi: <https://doi.org/10.1101/2022.03.02.22271795>; this version posted March 4, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a [CC-BY-NC-ND 4.0 International license](#).

Panel B



Risk of bias legend

- (A) Study population
- (B) Study attrition
- (C) Prognostic factor measurement
- (D) Outcome measurement
- (E) Study confounding
- (F) Statistical analysis and reporting
- (O) Overall

- + Low risk of bias
- ± Moderate risk of bias
- High risk of bias

Abbreviations:

GSf: Gait speed fast (m/s), FU: Follow-up, N: Participants included in analysis, CI: Confidence interval, NR: Not reported.

* Mean ± standard deviation (if reported)

Study or subgroup	Age* (yr)	Baseline SPPB score*	FU (yr)	N	Disability at FU	Weight	Risk Ratio [95% CI]	Risk of bias							Newcastle- Ottawa Scale
								A	B	C	D	E	F	O	
Balzi (2010)	74 ± 7	NR	3	848	9%	6.8%	1.53 [1.32; 1.78]	±	±	+	±	±	+	—	8
Bjorkman (2019)	83 ± 4	9.0 ± 2.6	4	262	26%	10.8%	1.15 [1.07; 1.24]	—	—	+	—	±	—	—	7
Guralnik (2000) [Iowa]	NR	NR	4	1,121	11%	8.8%	1.35 [1.21; 1.51]	—	—	+	±	±	+	—	8
Guralnik (2000) [N. Carolina]	NR	NR	4	962	10%	8.7%	1.35 [1.21; 1.52]	—	—	+	±	±	+	—	8
Guralnik (2000) [New Haven]	NR	NR	6	455	17%	7.4%	1.25 [1.09; 1.44]	—	—	+	±	±	+	—	8
Guralnik (1995)	77	NR	4	1,122	10%	8.2%	1.33 [1.18; 1.50]	—	—	+	±	±	+	—	8
Huang (2010)	80 ± 7	7.5 ± 1.9	1.5	75	32%	3.8%	1.18 [0.92; 1.50]	—	—	+	—	±	±	—	5
Minnecci (2015)	73 ± 6	9.3 ± 2.1	3	453	7%	5.7%	1.33 [1.12; 1.59]	±	—	+	—	±	±	—	8
Ostir (1998)	73	NR	2	1,365	4%	6.3%	1.38 [1.17; 1.62]	±	—	+	±	+	+	—	7
Stenholm (2014)	67 ± 14	10.7 ± 1.7	6	727	10%	8.3%	1.54 [1.36; 1.73]	±	—	+	±	+	+	—	8
Studenski (2003)	74 ± 6	8.4 ± 2.3	1	453	27%	10.1%	1.16 [1.07; 1.27]	—	±	+	±	±	+	—	5
Vaarst (2021)	82 ± 4	9.7 ± 2.4	4.1	323	38%	7.4%	1.26 [1.09; 1.45]	—	+	+	+	±	+	—	8
Verghese (2012)	80 ± 5	9.9 ± 1.7	2.7	594	15%	7.8%	1.23 [1.08; 1.40]	—	±	+	±	+	+	—	8

Random effects model

Heterogeneity: $I^2 = 61\%$, $\tau^2 = 0.0062$, $p < 0.01$

Test for overall effect: $z = 9.17$ ($p < 0.01$)

0.75 1 1.5
Risk ratio per 1 point decrease

1.30 [1.23; 1.38]

Risk of bias legend

- (A) Study population
- (B) Study attrition
- (C) Prognostic factor measurement
- (D) Outcome measurement
- (E) Study confounding
- (F) Statistical analysis and reporting
- (O) Overall

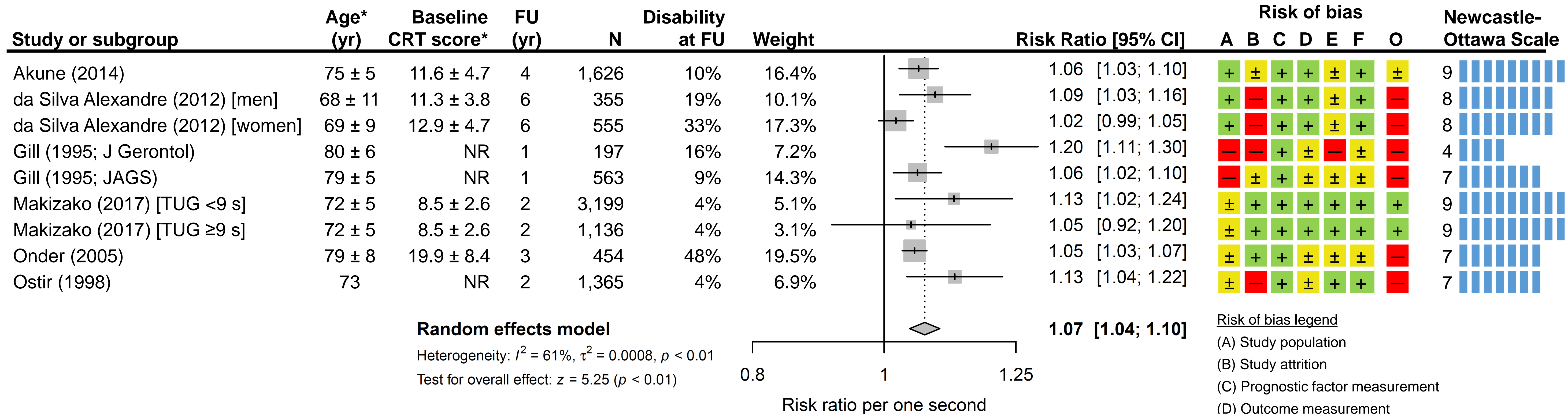
- + Low risk of bias
- ± Moderate risk of bias
- High risk of bias

Abbreviations:

SPPB: Short physical performance battery (points); FU: Follow-up; N: Participants included in analysis; CI: Confidence interval; NR: Not reported

* Mean ± standard deviation (if reported)

medRxiv preprint doi: <https://doi.org/10.1101/2022.03.02.22271795>; this version posted March 4, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

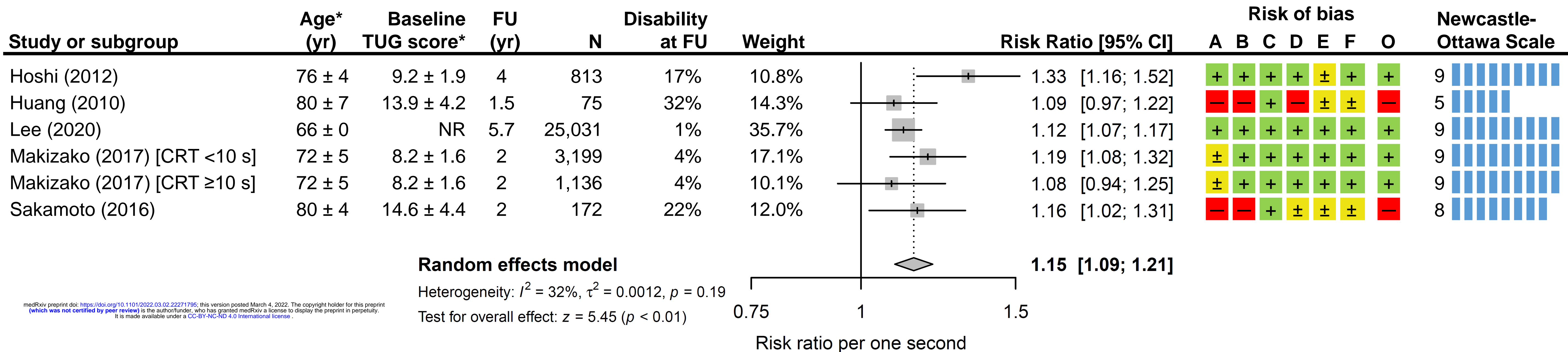


Risk of bias legend
 (A) Study population
 (B) Study attrition
 (C) Prognostic factor measurement
 (D) Outcome measurement
 (E) Study confounding
 (F) Statistical analysis and reporting
 (O) Overall

+ Low risk of bias
± Moderate risk of bias
- High risk of bias

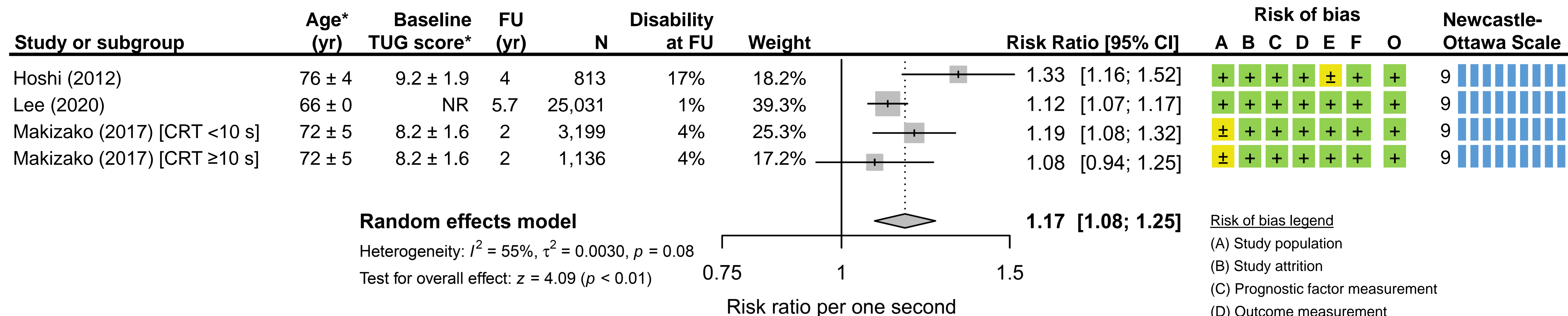
Abbreviations
 TUG: Timed up and go test (sec), CRT: Chair rise test, FU: Follow-up, N: Participants included in analysis, CI: Confidence interval, NR: Not reported.
 * Mean ± standard deviation (if reported)

Panel A



medRxiv preprint doi: <https://doi.org/10.1101/2022.03.02.22271795>; this version posted March 4, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a [CC-BY-NC-ND 4.0 International license](#).

Panel B



Risk of bias legend

- (A) Study population
- (B) Study attrition
- (C) Prognostic factor measurement
- (D) Outcome measurement
- (E) Study confounding
- (F) Statistical analysis and reporting
- (O) Overall

Abbreviations

TUG: Timed up and go test (sec), CRT: Chair rise test, FU: Follow-up, N: Participants included in analysis, CI: Confidence interval, NR: Not reported.

* Mean ± standard deviation (if reported)

- + Low risk of bias
- ± Moderate risk of bias
- High risk of bias