Development of a diverse osteoporosis screening tool for older US adults from the health and retirement study

Margaret Gough Courtney, Josephine Roberts, K. Godde [*]

Department of Sociology and Anthropology, University of La Verne, 1950 Third St., La Verne, CA, USA

A B S T R A C T

Existing osteoporosis screening tools have limitations, including using race as a predictor, and development on homogeneous samples. This biases risk assessment of osteoporosis in diverse populations and increases health inequities. We develop a tool that relies on variables easily learned during point-of-care, known by individuals, and with negligible racial bias. Data from the 2012–2016 waves of the population-based cohort Health and Retirement Study (HRS) were used to build a predictive model of osteoporosis diagnosis on a 75 % training sample of adults ages 50–90. The model was validated on a 25 % holdout sample and a cross-sectional sample of American individuals ages 50–80 from the National Health and Nutrition Examination Survey (NHANES). Sensitivity and specificity were compared across sex and race/ethnicity. The model has high sensitivity in the HRS holdout sample (89.9 %), which holds for those identifying as female and across racial/ethnic groups. Specificity is 57.9 %, and area under the curve (AUC) is approximately 0.81. Validation in the NHANES sample using empirically measured osteoporosis produced relatively good values of sensitivity, specificity, and consistency across groups. The model was used to create a publicly-available, open-source tool called the Osteoporosis Health Equality (& Equity) Evaluation (OsteoHEE). The model provided high sensitivity for osteoporosis diagnosis, with consistently high results for those identifying as female, and across racial/ethnic groups. Use of this tool is expected to improve equity in screening and increase access to bone density scans for those at risk of osteoporosis. Validation on alternative samples is encouraged.

1. Introduction

There are well-known inequities in disease diagnosis across race, ethnicity, and socioeconomic status (SES). One degenerative disease receiving increased attention, where there are disparate diagnosis rates, is osteoporosis, which is characterized by low bone mineral density and an increased risk of fracture [1]. For many individuals a fragility fracture is the first indication of osteoporosis, yet, even after a fragility fracture many individuals go undiagnosed in the United States [2]. Furthermore, post-fracture diagnosis (like pre-fracture diagnosis) is not evenly distributed across the population [3,4]. Rather, receiving a post-fragility fracture osteoporosis diagnosis is more common among women identifying as White/European American compared to women identifying as Black/African American [5,6].

According to fundamental cause theory both racism and socioeconomic status (SES) are fundamental causes of health disparities in the United States [7,8]. Fundamental causes involve flexible resources that can be used to reduce the likelihood of developing disease

and/or improve outcomes after the disease occurs. They can be economic in nature (e.g., income or wealth) but also social (e.g., human and social capital). The implications of the theory and its extensions are that although SES is a fundamental cause of health disparities, it is also tied up with racism in U.S. society, such that racism contributes to differences in SES, but racism also contributes to differences in health independent of SES [7]. Thus, in attempting to understand the disparities seen for osteoporosis diagnosis, it is necessary to consider the role of racism, along with the role of SES. Indeed, recent studies have provided strong evidence of inequities in osteoporosis diagnosis by race that appear to stem from systemic racism to a greater extent than economic resources [9,10].

SES, as conceived as social class, also plays a role for health in other ways, for example generating a tension between health consumerism among those of higher social class and a degree of social suffering for those of lower social class [11]. Health consumers treat health care like a service and health as an investment. They focus on optimizing health and preventing disease. While enacted by individuals, health consumerism is supported by the health care system. Research suggests it is more accessible to those with greater socioeconomic resources [11]. By contrast, those with fewer socioeconomic resources have been shown to embody social suffering as a result of limited power in the economy and other institutional settings. Knowledge and empowerment can help with health improvements, but within a system of unequal power and access to resources, challenges continue to exist [11]. In such a context, many researchers have sought to provide screening tools that could be used by individuals or their doctors to help assess risk of diseases [e.g. Ref. [12]]; interventions targeted at self-advocacy in the health care setting [13,14]; and provide information to individuals with fewer resources and lower amounts of power in an effort to empower them in the health care system [15,16]. We argue this is a valuable (though limited) approach, but that existing screening tools for assessing risk of osteoporosis suffer from numerous flaws. In this study we describe some of the limitations of prior screening tools, develop an improved screening tool, and highlight the importance of the tool as one of several strategies that can be deployed to reduce inequities in osteoporosis diagnosis. A key question we answer in this study is whether our newly developed osteoporosis screening tool performs similarly, or better, compared to existing osteoporosis screening tools, and whether it can be validated in another dataset. In addition, we examine the extent to which our new tool can be leveraged to improve health equity across racial/ethnic groups.

Recent work calls into question the utility and validity of race as a predictor variable in medical screening tools for the United States. As race is a social construct, it does not reflect underlying population histories important to human genetics. Instead, race, and particularly systemic/structural racism, reflects how people are treated on the basis of their visual characteristics and the biases within a particular society, which can translate to differences in health outcomes [17]. Much is known about the prevalence of systemic/structural racism in biomedicine and how that affects bone density, patient care, preventative treatment, use of risk factors, and fracture outcomes [[5],e.g. Ref. [18]], underscoring the need to address these issues.

T-scores from bone mineral density (BMD) screening also have limitations; they were developed on people identifying as Non-Hispanic White and are thus biased as they do not capture the full scope of human variation and may not be appropriate for other identities [19]. Moreover, the role that vitamin D supplementation plays for bone health has been redefined [20–23] as the underlying premise for its use was predicated on misunderstandings of how polymorphisms and human variation manifest [23], making a reevaluation of how bone health is treated a necessity.

Existing osteoporosis screening tools suffer from designs that limit applicability. Two of the most popular screening tools for

Table 1
Comparison of components and performance across osteoporosis-related screening assessment tools.

Tool	Predicts Osteoporosis or Fracture?	Components/Variables	Sensitivity	Specificity	AUC	Originally Developed for Both Sexes?
GARVAN (no BMD)	Fracture	Age, sex, body weight, history of prior fracture after age 50, history of falls in the past year, femoral neck BMD (optional)	75·5 % (women)[a] 29·4 % (men)[a] (Holloway-Kew et al., 2019)	45·6 % (women)[a] 82·7 % (men)[a] (Holloway-Kew et al., 2019)	Median 0·76 (Nguyen & Eisman 2017)	Yes
ORAI	Osteoporosis	Age, weight, noncurrent estrogen user	93·3 % (Chavda et al., 2022)	46·4 % (Chavda et al., 2022)	0·79 (Cadarette et al., 2000)	No
OSIRIS	Osteoporosis	Age, weight, current estrogen user, prior low-impact fracture	78·5 % (Chavda et al., 2022)	51·4 % (Chavda et al., 2022)	0·71 (Sedrine et al., 2002)	No
SCORE	Osteoporosis	Age, weight, non-Black race, rheumatoid arthritis, previous rib/wrist/hip fracture, never used estrogen	89·0 % (Chavda et al., 2022)	50·0 % (Chavda et al., 2022)	0·81 (Lydick et al., 1998)	No
FRAX	Fracture	Age, sex, race/ethnicity, weight, height, history of prior fractures, parental hip fracture, current smoking, glucocorticoid use, rheumatoid arthritis, secondary osteoporosis, alcohol use, femoral neck BMD (optional)	Average of 10·25 % from Jiang et al. (2017) meta-analysis	Average of 97·02 % from Jiang et al. (2017) meta-analysis	Median 0·69 (Nguyen & Eisman 2017)	Yes
OsteoHEE	Osteoporosis	Sex, age, weight, psychological problems, arthritis, thyroid disease, mobility index, allostatic load, having seen a dentist in the prior two years	89·9 % (HRS)[b] 88.5 % (NHANES)	57·9 % (HRS)[b] 49.2 % (NHANES)	0·81 (HRS)[b] 0.69 (NHANES)	Yes

[a] Fragility fractures.
[b] HRS holdout sample.

identifying risk of osteoporosis or osteoporosis-related fracture in the United States (SCORE and FRAX) include race as a predictor, or have separate calculators by racial identity [24], which introduces bias into the applicability/accuracy of the results as it captures societal treatment along with health biomarkers. Although alternative assessments exist for identifying risk of osteoporosis or osteoporosis-related fracture, the three next most popular measures (GARVAN, ORAI, OSIRIS) all suffer from significant limitations, such as a reliance on very few predictors or homogeneous development samples (see Table 1) [24].

The FRAX is one of the most well-known screening tools in the realm of osteoporosis, although it predicts future osteoporosis-related fracture risk, not osteoporosis per se. Yet, its performance is poor; according to a recent meta-analysis, the tool's sensitivity (meaning the ability of the tool to detect the health problem it is intended to detect—in this case fracture) averages only 10·25 % [25]. GARVAN, like FRAX, is used for predicting fractures; a recent study showed relatively modest sensitivity for women (75·5 %) and poor sensitivity for men (29·4 %) [26]. Unlike FRAX and GARVAN, SCORE is designed as an assessment for osteoporosis risk, not future fracture. Its sensitivity is very good at 89 %, but it includes race as a predictor and was developed on a rather homogeneous sample of approximately 1400 postmenopausal women [27], Like SCORE, ORAI and OSIRIS are osteoporosis screening tools, rather than fracture risk screening tools. Both include limited predictor variables, as noted in Table 1, and were developed and/or tested on fairly homogeneous samples of postmenopausal women [27]. Thus, existing tools are not well designed to account for the diverse experiences of individuals in the United States, or across the world, and many existing tools have been developed without regard to the population identifying as male.

Collectively, the known biases affecting patients with low bone density, and the limitations of existing tools, call for a new low bone density screening tool that provides health equity in knowing one's osteoporosis risk. Further, in the spirit of preventative medicine, increased risk should be detected early, rather than after significant bone loss when fracture becomes inevitable. Thus, this paper introduces a newly developed tool for assessing bone health, Osteoporosis Health Equality (& Equity) Evaluation (OsteoHEE), that uses updated biological/medical knowledge, a preventative medicine approach, and addresses the biases that arise from the systemic/structural racism that exist in medicine, in order to provide health equity without sacrificing predictive ability. Our approach to this tool was to include variables that have negligible racial bias, are easily learned during point-of-care, and that are known by the lay public, so that the tool can be used in a doctor's office, as well as in the privacy of one's home. We built OsteoHEE using an age range spanning middle-to older-age, in order to capture any effects of accelerated aging. We developed this tool using the longitudinal Health and Retirement Study (HRS) and validated the tool with a holdout sample from the HRS, and a sample from the National Health and Nutrition Examination Survey (NHANES). The latter validation purposely used comparable, but not identical, measures in order to show flexibility of application of the OsteoHEE tool. We further evaluate the results from the tool to quantify and examine any racial and sex biases to ensure health equity. One of the primary purposes of the tool is to empower patients to advocate for their health by requesting bone scans, which have been shown to reduce health inequities over use of risk factors alone when seeking an osteoporosis diagnosis in older adults [28].

2. Materials and methods

2.1. Health and retirement study

To build the prediction model we used data from the HRS, collected by the University of Michigan's Institute for Social Research [for more information on response rates and study design see 29]. We followed best practices for making predictive models, by splitting the data into training (75 %) and test (25 %) samples. The data were obtained from 1) the RAND Longitudinal file, a cleaned and harmonized version of the core and exit interviews of the HRS [30]; 2) the RAND fat files [31], which provide the data for osteoporosis diagnosis; 3) the cross-wave race/ethnicity study [32]; 4) the Biomarker Study [33], which includes components of the allostatic load
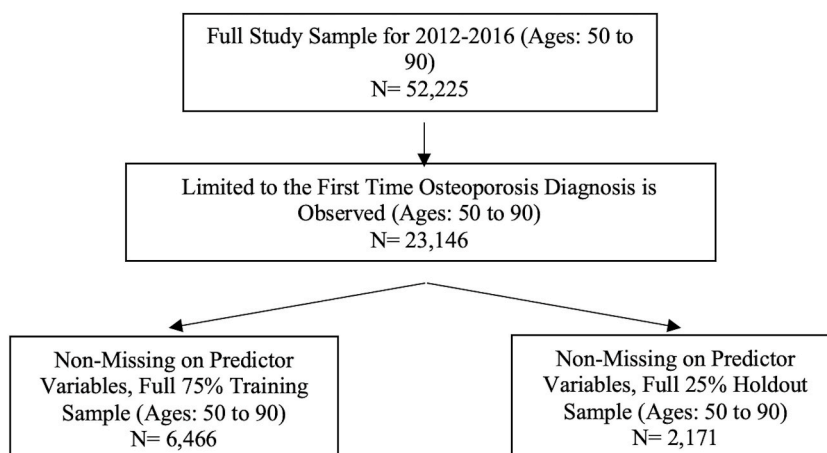

Fig. 1. Sample size flow chart.

variable; and 5) the Life History Mail Survey for the thyroid disease variable [34]. Core data were collected biennially and span the years 2012 through 2016, coinciding with when the osteoporosis data were collected. The training sample includes individuals ages 50–90 (N = 6466) with information on the predictor variables and osteoporosis outcome, while the holdout sample includes 2171 individuals ages 50–90 (see Fig. 1). This age range was selected to ensure that osteoporosis was developing or had already developed as a result of accelerated or normal aging. While accelerated aging as a result of chronic stress was found to be negligible in this sample [35], we still control for any effects undetected in prior work as this area of research on osteoporosis is in its infancy. About 54 % of the sample identifies as female. The racial/ethnic composition of the training sample is fairly consistent with the national distribution for older adults; approximately 80 % of the sample identifies as White/European American, 11 % as Black/African American, and 9 % as another race/ethnicity. We are unable to report the racial/ethnic composition of the sample at a finer level due to sample size and privacy restrictions. Informed consent was obtained from participants by the researchers consistent with their ethics board approval (for HRS, see: https://hrs.isr.umich.edu/sites/default/files/biblio/HRS_IRB_Information-10-2017.pdf; for NHANES, described below, see: https://wwwn.cdc.gov/nchs/data/nhanes/2013-2014/documents/mec_consent_form.pdf). As the data set for HRS is secondary, the research was deemed exempt by the University of La Verne Institutional Review Board (protocol # 2019-13-CAS). For NHANES, that portion of the study qualified as Not Human Subjects Research through the University of La Verne Institutional Review Board, due to it being deidentified secondary data, and did not require review. This paper follows the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) reporting guidelines.

To capture a broad set of potential risk factors, variables were selected based on epidemiological theory and published literature. The initial set of variables was screened by a filter method (Spearman's correlation) to remove redundant variables, as well as those with negligible statistical impact on the outcome variable of osteoporosis diagnosis [for more information on this methodology and variables see Ref. [35]]. We then estimated a model with the remaining predictors and dropped all non-significant predictors in an effort to balance predictive capabilities and parsimony. Some known risk factors for osteoporosis did not make it into final models as they were found to have negligible Spearman's correlations with osteoporosis diagnosis and/or had no impact on the key predictor variable in this sample. Thus, for example, variables relating to exercise, a history of smoking, and alcohol consumption were not statistically important in this sample [for more information on this, see [28]].

The outcome variable is a measure of osteoporosis diagnosis coded in the following manner: = 1 if the respondent reports an osteoporosis diagnosis and reports having had a bone density scan; = 0 if the respondent reports an osteoporosis diagnosis but no bone density scan or if the respondent reports no osteoporosis diagnosis. The final model includes the following predictors from the 2012–2016 waves: sex (1 = male or 0 = female; these are the only identities reported in the HRS, which is a limitation), age in years, weight in kilograms (self-reported, continuous), whether the respondent reported having ever had/received a diagnosis of emotional, nervous, or psychiatric problems from a doctor (0 = no or 1 = yes), whether the respondent reported ever having arthritis or receiving such a diagnosis from a doctor (0 = no or 1 = yes), whether the respondent reported that they ever had thyroid disease (5 = no or 1 = yes; from the 2017 Life History Mail Survey), the mobility index (continuous, range of 0–5), allostatic load (continuous, range of 0–4), and whether the respondent had seen a dentist in the previous 2 years (0 = no or 1 = yes). The mobility index is constructed from responses to questions about difficulty walking one block, walking several blocks, walking across the room, climbing one flight of stairs, and climbing several flights of stairs. Allostatic load is an index wherein values above the 75th percentile on the following measures are assigned a value of 1, and the index is created by summing across the items: HDL cholesterol (mg/dL), total cholesterol (mg/dL), blood pressure (=1 if respondent had either 75th percentile or higher systolic blood pressure (mmHg) or diastolic blood pressure (mmHg)) and waist circumference ≥35 inches for females and ≥40 inches for males (metric modified from McCrory et al. [36]).

2.2. National Health and Nutrition Examination Survey

The NHANES validation sample is derived from the 2013–14 and 2017–18 rounds of the cross-sectional NHANES (the 2015–2016 wave lacks certain key variables and could not be included). The sample is unweighted because the prediction model from HRS is being applied to the data for validation. As such, the sample includes individuals ages 50–80 (N = 4157), about 49 % of whom identify as female. The racial/ethnic composition of the NHANES data set includes 42 % identifying as White/European American, 22 % as Black/African American, and 36 % as another race/ethnicity. We create variables in NHANES that correspond closely to the HRS to improve validation, with the exception of using measured osteoporosis versus self-reported diagnosis: osteoporosis as measured by a bone mineral density scan (measured as part of the examination survey; 0 = no or 1 = yes; the outcome), sex (1 = male or 0 = female), age in years (continuous), weight in kilograms (continuous; measured as part of the examination survey), Patient Health Questionnaire (PHQ) depression scale score ≥5 (0 = no or 1 = yes; described below), whether a health professional ever told the respondent they have arthritis (0 = no or 1 = yes), whether the respondent has ever been told by a health professional they have thyroid disease (0 = no or 1 = yes), a mobility index (continuous, ranging from 0 to 3; sums affirmative answers to difficulty walking ¼ mile, walking up 10 stairs, and walking across a room), allostatic load (continuous, ranging from 0 to 5), and whether the respondent has seen a dentist in the last 2 years (0 = no or 1 = yes; constructed from the variable asking when the respondent last saw a dentist). The depression score is derived from the PHQ measure that screens for depressive symptoms and for which each question refers to the prior 2 weeks. Like in HRS, allostatic load is an index. In NHANES it includes the following five components: systolic blood pressure (mmHg), HDL cholesterol (mg/dL), total cholesterol (mg/dL), high-sensitivity *C*-reactive protein (mg/L), and waist circumference ≥35 inches for females and ≥40 inches for males.

Data used in the study are available from the original source. For HRS data, the RAND Longitudinal data set is publicly available through HRS (https://hrsdata.isr.umich.edu/data-products/rand?_ga=2.8281138.1210917905.1669763625-224554619.

1609868080), as are the RAND Fat Files (https://hrsdata.isr.umich.edu/data-products/rand?_ga=2.8281138.1210917905. 1669763625-224554619.1609868080) and the Life History Mail Survey (https://hrsdata.isr.umich.edu/data-products/2017-fall-life-history-mail-survey). The cross-wave race/ethnicity data and the Biomarker Study are available through a restricted-use data agreement (https://hrs.isr.umich.edu/data-products/restricted-data/available-products). NHANES data are publicly available for download (https://wwwn.cdc.gov/nchs/nhanes/default.aspx). The statistical code used in the analyses is available upon request from the authors, pending review by the HRS Restricted Data Agreement Disclosure personnel.

2.3. Analytic method

A multivariable logistic regression was applied to the HRS training sample. Model fit and discrimination were evaluated using a McFadden's adjusted Pseudo R^2 and a C statistic. Multicollinearity was assessed using the variance inflation factor (VIF). Listwise deletion addressed missing data. Fig. 1 depicts how we arrived at the final sample size; Appendix Table A demonstrates the similarity of the analytic sample and the sample before listwise deletion. To address the complex sampling design of the HRS and reduce response bias, we estimated models using sample stratum and sample PSU (from the RAND longitudinal data), and weights (from the Biomarker Study) in Stata 16·0, using *svy* commands, which also produce robust standard errors that help account for repeat observations. There are limited repeat observations in the data set because we restrict the data to individuals who either never report osteoporosis across the three waves or are reporting it for the first time; most respondents who reported an osteoporosis diagnosis reported it in 2012, the first year it was collected. We plotted the residuals against time to assess whether repeat observations were a concern for estimation, and the pattern was random, suggesting they were not.

The prediction equation derived from the HRS training sample was first validated via the HRS holdout sample and subsequently the NHANES sample. The results were examined for accuracy, sensitivity, specificity, and the area under the curve (AUC). Within each validation sample, we tested whether sensitivity and specificity differed across sex and categories of race/ethnicity using the same prediction commands but limited to the relevant subgroups.

There are different approaches to choosing cutoff values for classifying a case in the screening tool literature. Tools like SCORE and

Table 2
Descriptive statistics for main model for full training sample in Health and Retirement Study (N = 6466).

Variable	Mean (SE) or Proportion
Osteoporosis	
No	0·8602162
Yes	0·1397838
Age	63·42164 (0·2726646)
Weight in Kilograms	82·62309 (0·3476965)
Sex	
Male	0·4576681
Female	0·5423319
Race/Ethnicity	
White/European American	0·7982251
Black/African American	0·1093924
Another race/ethnicity	0·0923825
Reports emotional, nervous, or psychological problems	
No	0·7815341
Yes	0·2184659
Reports arthritis	
No	0·4557241
Yes	0·5442759
Reports thyroid disease	
No	0·9735408
Yes	0·0264592
Mobility index	0·9876317 (0·0256055)
Allostatic load	1·747098 (0·0155549)
Seen dentist in previous 2 years	
No	0·3374844
Yes	0·6625156
Osteoporosis Diagnosis in 2012	
No	0·8806030
Yes	0·1193970
Osteoporosis Diagnosis in 2014[a]	
No	0·4627329
Yes	0·5372671
Osteoporosis Diagnosis in 2016	
No	0·8494440
Yes	0·1505560

[a] Proportion reporting osteoporosis appears higher as these are only new respondents to the osteoporosis question.

ORAI use cutoff values that provide a sensitivity of approximately 90 % [37,38]. Other tools calibrate the cutoff to the predicted probability in the population (e.g., using a median risk score that reflects population prevalence) [39]. We use the median predicted probability (0·091) as a cutoff value for osteoporosis classification. In our study the median predicted probability achieves both features—it provides us with a sensitivity of approximately 90 %, and it is close to the prevalence of osteoporosis among the older population, which is approximately 10 % [40]. We tested higher and lower thresholds, but higher thresholds resulted in unacceptably low sensitivity, and lower thresholds resulted in unacceptably low specificity.

3. Results

Descriptive statistics for the HRS training sample are shown in Table 2. Approximately 14 % of the sample reported an osteoporosis diagnosis. Average age was about 63 years, with about 54 % of the sample identifying as female. Health conditions were fairly common with 54 % reporting an arthritis diagnosis, 22 % reporting emotional, nervous, or psychological problems, 3 % reporting thyroid disease, and an average allostatic load of 1·75 (range: 0–4). The coefficients from the prediction model that are used for the OsteoHEE screening tool are shown in Table 3. Identifying as male, higher weight in kilograms, and higher allostatic load [an inverse relationship likely due to waist circumference and which did not have an issue with multicollinearity in this study; see discussion in [9] for more information about this relationship with osteoporosis diagnosis] are all associated with lower odds of osteoporosis diagnosis, while older age, psychological problems, arthritis, thyroid disease, greater mobility problems, and having seen a dentist in the prior two years were all linked to higher odds of an osteoporosis diagnosis. The pseudo R^2 is 0·2289, which is excellent [41], and the *C*-statistic (AUC) is 0·8218, which is very good. The VIF was less than 4.

In the HRS holdout sample we find that 62.2 % of observations are correctly classified, with a sensitivity of 89·9 %, and a specificity of 57·9 % (Table 4). The false positive rate (FPR) is 42·1 %; we argue this is a reasonable tradeoff for capturing a high percentage of true cases, given that a positive screening result should lead to a DXA scan, which is a non-invasive procedure that is often covered by Medicare (see Discussion for elaboration on this point). The AUC is good at 0·8108. Sensitivity is similar or better for those identifying as female (96·0 %) and across racial/ethnic groups (90·7 % for those identifying as White/European American, 84·6 % for those identifying as Black/African American, and approximately 85·0 % for those identifying as another race/ethnicity). Sensitivity is a bit lower for those identifying as Black/African American or another race/ethnicity, likely due to sample size, but classification is still good. Sensitivity for those identifying as male was unfortunately very low (13·6 %), leading us to focus the screening tool on those identifying as female.

In the validation with NHANES data we find that sensitivity is 88·5 %, specificity is lower at 49·2 %, and the false positive rate is a bit higher at 50·8 % (Table 4). The AUC is 0·6898. Sensitivity is generally similar or better for those identifying as female (100·0 %) and across racial/ethnic groups (88·0 % for those identifying as White/European American, 76·5 % for those identifying as Black/African American, and approximately 94·9 % for those identifying as another race/ethnicity), although the sensitivity for those identifying as Black/African American is lower than desired, but higher than other tools [42]. Sensitivity for those identifying as male was low as expected, though not as low as in the HRS holdout sample (39·3 %). For comparison, in Appendix Table B we show validation results in NHANES using self-reported osteoporosis diagnosis as the outcome (0 = no, 1 = yes), demonstrating slightly improved sensitivity and specificity, and a higher AUC. Although the NHANES validation results using measured osteoporosis are more variable than those from the HRS validation (and the AUCs lower), the overall consistency is still a strength, given that in NHANES the correlation between measured osteoporosis and self-reported osteoporosis is only 0.19 in this sample.

Table 3
Logistic regression model to predict osteoporosis using Health and Retirement Study training sample (N = 6466).

	B (SE)	*p*	95 % CI
Sex			
Male	−1·920447 (0·1782225)	<0·001	−2·275121, −1·565773
Age	0·0358944 (0·0045038)	<0·001	0·0269316, 0·0448573
Weight in kilograms	−0·0264736 (0·003505)	<0·001	−0·0334487, −0·0194984
Reports emotional, nervous, or psychological problems			
Yes	0·4268404 (0·1067511)	<0·001	0·214399, 0·6392819
Reports arthritis			
Yes	0·678786 (0·1251579)	<0·001	0·4297137, 0·9278582
Reports thyroid disease			
No	−0·6974424 (0·1739718)	<0·001	−1·043657, −0·3512276
Mobility index	0·1698371 (0·0309006)	<0·001	0·108343, 0·2313313
Allostatic load	−0·1180246 (0·0586959)	0·048	−0·2348331, −0·001216
Seen dentist in previous 2 years			
Yes	0·5460753 (0·10292)	<0·001	0·341258, 0·7508926
Constant	−1·897108 (0.4808921)	<0·001	−2·854114, −0·9401021

Pseudo R^2 = 0·2289.
AUC = 0·8218.

Table 4
Sensitivity and specificity across samples, with 95 % confidence intervals.

	HRS Holdout Sample		NHANES Validation Sample (with Measured Bone Mineral Density)	
	Sensitivity (95 % CI)	Specificity (95 % CI)	Sensitivity (95 % CI)	Specificity (95 % CI)
Full Sample	89·9 % (85.8–93.1 %)	57·9 % (55.6–60.2 %)	88.5 % (82.2–93.2 %)	49.2 % (47.7–50.8 %)
Male Identity	13·6 % (2.9–34.9 %)	96·0 % (94.5–97.2 %)	39.3 % (21.5–59.4 %)	88·2 % (86.7–89.5 %)
Female Identity	96·0 % (92.9–98.0 %)	22·9 % (20.3–25.7 %)	100.0 % (97.0–100.0 %)	6.6 % (5.5–7.8 %)
White/European American Identity	90·7 % (86.3–94.1 %)	56·3 % (53.5–59.0 %)	88.0 % (79.6–93.9 %)	47.3 % (44.9–49.7 %)
Black/African American Identity	84·6 % (69.5–94.1 %)	63·0 % (57.9–67.9 %)	76.5 % (50.1–93.2 %)	55·6 % (52.5–58.9 %)
Another Racial/Ethnic Identity	85·0 % (66.9–98.7 %)	58·6 % (51.5–65.5 %)	94.9 % (82.7–99.4 %)	47.7 % (45.1–50.3 %)

4. Discussion

The goal of this study was to develop an osteoporosis screening tool that improved on the performance of prior tools, including a range of variables that would be known to individual patients and knowable to health care providers, while also not including race as a predictor, as some prior screening tools have done. The model underlying the OsteoHEE tool performed well on the full sample, demonstrating sensitivity of roughly 90 % in the validation samples, while keeping specificity between 49 and 58 %, and AUC at 0·81 in the HRS. Perhaps most importantly, a high level of sensitivity was seen across racial and ethnic groups, demonstrating the utility of the tool for a diverse population. Although sensitivity was much lower for those identifying as male, and thus those identifying as male have not been incorporated into the screening interface at this time, sensitivity was extremely high for those identifying as female (96 % or higher). The development and validation of OsteoHEE on the HRS and NHANES make this tool generalizable to the greater United States population of older adults. OsteoHEE can be accessed by clinicians, the lay public, and researchers at the following link, which includes the equations and calculations published in this paper: https://osteohee.shinyapps.io/OsteoHee/.

OsteoHEE provides demonstrable improvements compared to other osteoporosis-related screening and assessment tools used in current practice. For example, sensitivity is higher than all of the tools in Table 1 except ORAI [25–27], and the difference in sensitivity between OsteoHEE and ORAI is small (3·4 %). Additionally, ORAI is limited by being built on a sample of postmenopausal women and using only three variables for prediction: age, weight, and noncurrent estrogen user. Specificity is better in OsteoHEE than it is in all other tools except FRAX [25–27], but FRAX has extremely low sensitivity [25] and a low AUC [43]. Furthermore, the AUC is higher in OsteoHEE than in the other tools except SCORE for which it is identical (0·81) [37,38,43]. The FPR for OsteoHEE ranges from 42 to 45 %, which we argue is an acceptable trade-off for capturing most cases of osteoporosis. Furthermore, this rate is actually lower than most other screening tools. For example, the FPRs for GARVAN, ORAI, OSIRIS, and SCORE are all roughly 50 % or higher (see Table 1; FPR = 1-Specificity).

The increased performance of this tool comes from examining people identifying as male and female in a single model, which allows for coefficients to be produced that juxtapose those identities. While the male identity performance is poor, and thus not ready to be released for use, we have included the predictions for people identifying as female into the OsteoHEE tool as the performance for that identity is increased over other tools. Furthermore, the improved performance also comes from the inclusion of important predictors of osteoporosis, such as thyroid disease and allostatic load, while not including inappropriate variables such as race and body mass index.

It is well established that medicine in the United States is greatly affected by structural/systemic racism [44], and thus, screening tools across different diagnoses need to account for this bias. Ignoring the existence/effect of structural/systemic racism exacerbates current inequities and stalls efforts to correct them. The deployment of a user-friendly screening tool, such as OsteoHEE, that was specifically developed to minimize racial bias is one step toward improving racial and ethnic equity in terms of osteoporosis diagnosis. Racism, like socioeconomic status, is a fundamental cause of health disparities [7], but tools that are conscious of this can give providers more objective information about risk factors for osteoporosis diagnosis and improve equity. Furthermore, as a free tool OsteoHEE is accessible to individuals with fewer resources, such as those in lower social classes. The knowledge gained from using the tool can help empower these individuals and provide them with greater opportunity to engage in the health care system as a health consumer, like their more advantaged counterparts [11]. It is our hope that this tool opens the door to more inclusive screening along with additional validation of tools focusing on their performance across racial identities and making adjustments to ensure similar performance across racial groups. We encourage the testing of this tool on additional samples and by other researchers so that the focus on equity we provide here is extended to other instruments and studies.

The limitations of this tool include continuous coefficients that are not necessarily linearly related to the log odds. While this is a violation of the assumptions of logistic regression, the impact of this is a lower estimate of the relationship between the predictor and outcome variables that may lead to a greater risk of a Type II error [45], which means our model reflects the minimum relationship between the dependent and independent variables. However, we provided more robust standard errors for construction of confidence intervals, if researchers/practitioners desire to use these. Further, we validated these coefficients and showed good performance across multiple samples, which yielded better performances than other tools, and indicates minimal impact from this phenomenon. Another issue is a lack of inclusion of genetic information regarding osteoporosis, which would certainly have increased the accuracy of this tool and hopefully decrease the rate of false positives, which could also be addressed by variables capturing a more comprehensive life history. Moreover, the diagnoses reported by respondents may use biased T-scores, which is an issue we hope researchers will address in the near future through re-examination of T-score cutoffs using a diverse sample. Additionally, the sample sizes of the subgroups are

a bit small for achieving high sensitivity, although the reported sensitivity is quite good for these groups. What is more, the arthritis variables included all types, which introduces heterogeneity. We further used a binary variable for sex that simplifies the underlying biological continuum due to the information available in the HRS. Finally, we measured the performance of our results with an oversimplified race/ethnicity variable due to sample size that negated our ability to examine performance at a more granular level.

5. Conclusions

In sum, this publicly available, user-friendly screening tool does an excellent job of predicting risk of osteoporosis for middle-to-older-aged women, across racial identities, and can be leveraged to positively change individuals' interactions with the health care system, empowering patients to seek out screening and diagnosis through a bone density scan, regardless of whether they hold a privileged identity or greater access to socioeconomic resources. We anticipate this tool can be particularly beneficial for individuals who have historically been marginalized in the health care system, including those with low socioeconomic status, and individuals from marginalized racial and ethnic groups. Together with changes to the health care system, such as greater diversity, equity, and inclusion training for clinicians, this tool can play a role in reducing inequities in osteoporosis diagnosis. Earlier, and more equitable, diagnosis will improve prospects for treatment and prevention of disease progression. However, the tool has limitations, in that the continuous coefficients do not necessarily have a linear relationship with the log odds, a lack of genetic information is included, T-score cutoffs used may be biased, sample sizes could be improved with a larger and diverse sample, the arthritis variable was too general, the sex variable was binary, and the race/ethnicity variable was oversimplified. These are limitations we sought to mitigate and/or were restricted by the nature of using secondary data. Since this is a new screening tool, we hope other researchers will continue to validate the tool on other samples to demonstrate the strength of this approach.

Ethical approval and consent to participate.

Informed consent was obtained from participants by the HRS and NHANES researchers consistent with their ethics board approvals (for HRS, see: https://hrs.isr.umich.edu/sites/default/files/biblio/HRS_IRB_Information-10-2017.pdf; for NHANES, described below, see: https://wwwn.cdc.gov/nchs/data/nhanes/2013-2014/documents/mec_consent_form.pdf). As the data set for HRS used in this paper is secondary, the research was deemed exempt by the University of La Verne Institutional Review Board (protocol # 2019-13-CAS). For NHANES, that portion of the study qualified as Not Human Subjects Research through the University of La Verne Institutional Review Board, due to it being deidentified secondary data, and did not require review.

Data availability statement

The data should be requested from the Health and Retirement Study and the National Health and Nutrition Examination Survey.

CRediT authorship contribution statement

Margaret Gough Courtney: Writing - review & editing, Writing - original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Josephine Roberts:** Writing - review & editing, Writing - original draft. **K. Godde:** Writing - review & editing, Writing - original draft, Visualization, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:Kanya Godde Chrisco and Margaret Gough Courtney reports financial support was provided by National Institute on Aging. Kanya Godde Chrisco and Margaret Gough Courtney reports financial support and administrative support were provided by University of La Verne.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2023.e23806.

References

[1] W.A. Stini, Osteoporosis in biocultural perspective, Annu. Rev. Anthropol. 24 (1995) 397–421, https://doi.org/10.1146/annurev.an.24.100195.002145.

[2] C.W. Gillespie, P.E. Morin, Osteoporosis-related health services utilization following first hip fracture among a cohort of privately-insured women in the United States, 2008-2014: an observational study, J. Bone Miner. Res. 32 (2017) 1052–1061, https://doi.org/10.1002/jbmr.3079.

[3] E.M. Lewiecki, S.F. Erb, Racial disparities and inequalities in the management of patients with osteoporosis, Orthop. Nurs. 41 (2022) 125–134, https://doi.org/10.1097/NOR.0000000000000832.

[4] K.N. Ruiz-Esteves, J. Teysir, D. Schatoff, E.W. Yu, S.-A.M. Burnett-Bowie, Disparities in osteoporosis care among postmenopausal women in the United States, Maturitas 156 (2022) 25–29, https://doi.org/10.1016/j.maturitas.2021.10.010.

[5] R.G. Miller, B.H. Ashar, J. Cohen, M. Camp, C. Coombs, E. Johnson, C.R. Schneyer, Disparities in osteoporosis screening between at-risk African-American and White women, J. Gen. Intern. Med. 20 (2005) 847–851, https://doi.org/10.1111/j.1525-1497.2005.0157.x.

[6] A.S. Mudano, L. Casebeer, F. Patino, J.J. Allison, N.W. Weissman, C.I. Kiefe, S. Person, D. Gilbert, K.G. Saag, Racial disparities in osteoporosis prevention in a managed care population, South. Med. J. 96 (2003) 445–451, https://doi.org/10.1097/01.SMJ.0000053918.93363.B0.

[7] J.C. Phelan, B.G. Link, Is racism a fundamental cause of inequalities in health? Annu. Rev. Sociol. 41 (2015) 311–330, https://doi.org/10.1146/annurev-soc-073014-112305.

[8] J.C. Phelan, B.G. Link, P. Tehranifar, Social conditions as fundamental causes of health inequalities: theory, evidence, and policy complications, J. Health Soc. Behav. 51 (2010) S28–S40, https://doi.org/10.1177/0022146510383498.

[9] M. Gough Courtney, J. Roberts, K. Godde, Structural inequity and socioeconomic status link to osteoporosis diagnosis in a population-based cohort of middle-older-age Americans, Inq. J. Health Care Organ. Provis. Financ. 60 (2023), 004695802311155719, https://doi.org/10.1177/00469580231155719.

[10] M. Gough Courtney, J. Roberts, Y. Quintero, K. Godde, Childhood family environment and osteoporosis in a population-based cohort study of middle to older age Americans, J. Bone Miner. Res. Plus. 7 (2023), e10735, https://doi.org/10.1002/jbm4.10735.

[11] C.H. Merrild, M.B. Risør, P. Vedsted, R.S. Andersen, Class, social suffering, and health consumerism, Med. Anthropol. 35 (6) (2015) 517–528. https://www.tandfonline.com/doi/full/10.1080/01459740.2015.1102248 (Accessed 20 March 2023).

[12] E. Ngo, M.B.-T. Truong, H. Nordeng, Use of decision support tools to empower pregnant women: systematic review, J. Med. Internet Res. 22 (2020), e19436, https://doi.org/10.2196/19436.

[13] E.K. Schmidt, J. Faieta, K. Tanner, Scoping review of self-advocacy education interventions to improve care, OTJR Occup. Participation Health 40 (2020) 50–56, https://doi.org/10.1177/1539449219860583.

[14] T.L. Hagan, E. Medberry, Patient education vs. Patient experiences of self-advocacy: changing the discourse to support cancer survivors, J. Cancer Educ. 31 (2016) 375–381, https://doi.org/10.1007/s13187-015-0828-x.

[15] D.H. Solomon, J.M. Polinski, M. Stedman, C. Truppo, L. Breiner, C. Egan, S. Jan, M. Patel, T.W. Weiss, Y. Chen, M.A. Brookhart, Improving care of patients at-risk for osteoporosis: a randomized controlled trial, J. Gen. Intern. Med. 22 (2007) 362–367, https://doi.org/10.1007/s11606-006-0099-7.

[16] M.A. Lopez-Olivo, J.K.A. des Bordes, H. Lin, T. Rizvi, R.J. Volk, M.E. Suarez-Almazor, Comparison of multimedia and printed patient education tools for patients with osteoporosis: a 6-month randomized controlled trial, Osteoporos. Int. 31 (2020) 857–866, https://doi.org/10.1007/s00198-019-05210-4.

[17] P.A. Braveman, E. Arkin, D. Proctor, T. Kauh, N. Holm, Systemic and structural racism: definitions, examples, health damages, and approaches to dismantling: study examines definitions, examples, health damages, and dismantling systemic and structural racism, Health Aff. 41 (2022) 171–178, https://doi.org/10.1377/hlthaff.2021.01394.

[18] D.J. Burgess, S.S. Fu, M. van Ryn, Why do providers contribute to disparities and what can be done about it? J. Gen. Intern. Med. 19 (2004) 1154–1159, https://doi.org/10.1111/j.1525-1497.2004.30227.x.

[19] S.E. Noel, M.P. Santos, N.C. Wright, Racial and ethnic disparities in bone health and outcomes in the United States, J. Bone Miner. Res. 36 (2021) 1881–1905, https://doi.org/10.1002/jbmr.4417.

[20] C.E. Powe, M.K. Evans, J. Wenger, A.B. Zonderman, A.H. Berg, M. Nalls, H. Tamez, D. Zhang, I. Bhan, S.A. Karumanchi, N.R. Powe, R. Thadhani, Vitamin D–binding protein and vitamin D status of Black Americans and White Americans, N. Engl. J. Med. 369 (2013) 1991–2000, https://doi.org/10.1056/NEJMoa1306357.

[21] L.A. Burt, E.O. Billington, M.S. Rose, D.A. Raymond, D.A. Hanley, S.K. Boyd, Effect of high-dose vitamin D supplementation on volumetric bone density and bone strength: a randomized clinical trial, JAMA 322 (2019) 736–745, https://doi.org/10.1001/jama.2019.11889.

[22] M.S. LeBoff, S.H. Chou, E.M. Murata, C.M. Donlon, N.R. Cook, S. Mora, I.-M. Lee, G. Kotler, V. Bubes, J.E. Buring, J.E. Manson, Effects of supplemental vitamin D on bone health outcomes in women and men in the VITamin D and OmegA-3 TriaL (VITAL), J. Bone Miner. Res. 35 (2020) 883–893, https://doi.org/10.1002/jbmr.3958.

[23] L.L. Brown, B. Cohen, D. Tabor, G. Zappalà, P. Maruvada, P.M. Coates, The vitamin D paradox in Black Americans: a systems-based approach to investigating clinical practice, research, and public health - expert panel meeting report, BMC Proc. 12 (2018) 6, https://doi.org/10.1186/s12919-018-0102-4.

[24] S.-J. Chen, Y.-J. Chen, C.-H. Cheng, H.-F. Hwang, C.-Y. Chen, M.-R. Lin, Comparisons of different screening tools for identifying fracture/osteoporosis risk among community-dwelling older people, Medicine (Baltim.) 95 (2016), e3415, https://doi.org/10.1097/MD.0000000000003415.

[25] X. Jiang, M. Gruner, F. Trémollieres, W. Pluskiewicz, E. Sornay-Rendu, P. Adamczyk, P.F. Schnatz, Diagnostic accuracy of FRAX in predicting the 10-year risk of osteoporotic fractures using the USA treatment thresholds: a systematic review and meta-analysis, Bone 99 (2017) 20–25, https://doi.org/10.1016/j.bone.2017.02.008.

[26] K.L. Holloway-Kew, Y. Zhang, A.G. Betson, K.B. Anderson, D. Hans, N.K. Hyde, G.C. Nicholson, N.A. Pocock, M.A. Kotowicz, J.A. Pasco, How well do the FRAX (Australia) and Garvan calculators predict incident fractures? Data from the Geelong Osteoporosis Study, Osteoporos. Int. 30 (2019) 2129–2139, https://doi.org/10.1007/s00198-019-05088-2.

[27] S. Chavda, B. Chavda, R. Dube, Osteoporosis screening and fracture risk assessment tool: its scope and role in general clinical practice, Cureus 14 (2022) 1–11, https://doi.org/10.7759/cureus.26518.

[28] K. Godde, M. Gough Courtney, J. Roberts, Health insurance coverage as a social determinant of osteoporosis diagnosis in a population-based cohort study of older American adults, J. Appl. Gerontol. Off. J. South. Gerontol. Soc. 42 (2022) 302–312, https://doi.org/10.1177/07334648221132792.

[29] A. Sonnega, J.D. Faul, M.B. Ofstedal, K.M. Langa, J.W. Phillips, D.R. Weir, Cohort profile: the health and retirement study (HRS), Int. J. Epidemiol. 43 (2014) 576–585, https://doi.org/10.1093/ije/dyu067.

[30] Health and Retirement Study, RAND HRS Longitudinal File 2016 (V2) public use dataset, Prod. Distrib. Univ. Mich. Funding Natl. Inst. Aging Grant Number NIA U01AG009740 Ann Arbor MI (2020).

[31] Health and Retirement Study, RAND HRS fat files public use dataset, Prod. Distrib. Univ. Mich. Funding Natl. Inst. Aging Grant Number NIA U01AG009740 Ann Arbor MI (2021). https://www.rand.org/well-being/social-and-behavioral-policy/centers/aging/dataprod/enhanced-fat.html. (Accessed 4 February 2021).

[32] Health and Retirement Study, Cross-Wave Race and Ethnicity File, Prod. Distrib. Univ. Mich. Funding Natl. Inst. Aging Grant Number NIA U01AG009740 Ann Arbor MI, 2014. https://hrsdata.isr.umich.edu/data-products/2008-telomere-data. (Accessed 11 October 2021).

[33] Health and Retirement Study, Biomarker Data 2006-2016, Prod, Distrib. Univ. Mich. Funding Natl. Inst. Aging Grant Number NIA U01AG009740, Ann Arbor MI, 2020.

[34] Health and Retirement Study, 2017 Fall Life History Mail Survey Occupation and Industry Data | Public Use Dataset, Prod, Distrib. Univ. Mich. Funding Natl. Inst. Aging Grant Number NIA U01AG009740 Ann Arbor MI, 2017. https://hrs.isr.umich.edu/data-products/restricted-data/available-products/11273. (Accessed 4 February 2021).

[35] M. Gough Courtney, Y. Quintero, K. Godde, Assessing the roles of demographic, social, economic, environmental, health-related, and political factors on risk of osteoporosis diagnosis among older adults, Arch. Osteoporosis 16 (2021) 1–12, https://doi.org/10.1007/s11657-021-01042-0.

[36] C. McCrory, G. Fiorito, C. Ni Cheallaigh, S. Polidoro, P. Karisola, H. Alenius, R. Layte, T. Seeman, P. Vineis, R.A. Kenny, How does socio-economic position (SEP) get biologically embedded? A comparison of allostatic load and the epigenetic clock(s), Psychoneuroendocrinology 104 (2019) 64–73, https://doi.org/10.1016/j.psyneuen.2019.02.018.

[37] E. Lydick, K. Cook, J. Turpin, M. Melton, R. Stine, C. Byrnes, Development and validation of a simple questionnaire to facilitate identification of women likely to have low bone density, Am. J. Manag. Care 4 (1998) 37–48.

[38] S.M. Cadarette, S.B. Jaglal, N. Kreiger, W.J. McIsaac, G.A. Darlington, J.V. Tu, Development and validation of the Osteoporosis Risk Assessment Instrument to facilitate selection of women for bone densitometry, CMAJ Can. Med. Assoc. J. J. Assoc. Medicale Can. 162 (2000) 1289–1294.

[39] D.M. Hafeman, J. Merranko, T.R. Goldstein, D. Axelson, B.I. Goldstein, K. Monk, M.B. Hickey, D. Sakolsky, R. Diler, S. Iyengar, D.A. Brent, D.J. Kupfer, M. W. Kattan, B. Birmaher, Assessment of a person-level risk calculator to predict new-onset bipolar spectrum disorder in youth at familial risk, JAMA Psychiatr. 74 (2017) 841–847, https://doi.org/10.1001/jamapsychiatry.2017.1763.

[40] N.C. Wright, A.C. Looker, K.G. Saag, J.R. Curtis, E.S. Delzell, S. Randall, B. Dawson-Hughes, The recent prevalence of osteoporosis and low bone mass in the United States based on bone mineral density at the femoral neck or lumbar spine, J. Bone Miner. Res. 29 (2014) 2520–2526, https://doi.org/10.1002/jbmr.2269.

[41] D. McFadden, Quantitative methods for analysing travel behaviour of individuals, in: Cowles Foundation Discussion Paper No. 474, New Haven: Yale University, New Haven, CT, 1977.

[42] A.R. Cass, A.J. Shepherd, C.A. Carlson, Osteoporosis risk assessment and ethnicity: validation and comparison of 2 clinical risk stratification instruments, J. Gen. Intern. Med. 21 (2006) 630–635, https://doi.org/10.1111/j.1525-1497.2006.00459.x.

[43] T.V. Nguyen, J.A. Eisman, Fracture risk assessment: from population to individual, J. Clin. Densitom. 20 (2017) 368–378, https://doi.org/10.1016/j.jocd.2017.06.023.

[44] R. Yearby, Racial disparities in health status and access to healthcare: the continuation of inequality in the United States due to structural racism, Am. J. Econ. Sociol. 77 (2018) 1113–1152, https://doi.org/10.1111/ajes.12230.

[45] R.G. Long, The crux of the method: assumptions in ordinary least squares and logistic regression, Psychol. Rep. 103 (2008) 431–434, https://doi.org/10.2466/pr0.103.2.431-434.