Is Looking Older than One's Actual Age a Sign of Poor Health?

Stephen W. Hwang, MD, MPH[1,2], Mina Atia[1], Rosane Nisenbaum, PhD[1], Dwayne E. Pare, MA[3], and Steve Joordens, PhD[3]

[1]Centre for Research on Inner City Health, The Keenan Research Centre in the Li Ka Shing Knowledge Institute of St. Michael's Hospital, Toronto, Ontario, Canada; [2]Division of General Internal Medicine, Department of Medicine, University of Toronto, Toronto, Canada; [3]Centre for Computational Cognitive Neuroscience, University of Toronto Scarborough, Toronto, Canada.

BACKGROUND: Physicians often begin the physical examination with an assessment of whether a patient looks older than his or her actual age. This practice suggests an implicit assumption that patients who appear older than their actual age are more likely to be in poor health.

OBJECTIVE: To determine the sensitivity and specificity of apparent age for the detection of poor health status.

DESIGN: Cross-sectional.

PATIENTS: A total of 126 outpatients (ages 30–70) from four primary care clinics and one general internal medicine clinic at an academic medical institution.

MEASUREMENTS: With the patient's actual age provided, physicians (n=58 internal medicine residents and general internal medicine faculty) viewed patient photographs and assessed how old each patient looked. For each physician, we examined the sensitivity and specificity of the difference between how old the patient looked and the patient's actual age for the detection of poor health, defined using SF-12 physical health and mental health scores.

RESULTS: Using the threshold of looking ≥5 years older than actual age and with poor health defined as an SF-12 score ≥2.0 SD below age group norms, median sensitivity was 29% (IQR, 19% to 35%), median specificity 82% (IQR, 77% to 88%), median positive likelihood ratio 1.7 (IQR, 1.3 to 2.2), and median negative likelihood ratio 0.9 (IQR, 0.8 to 0.9). Using the threshold of looking ≥10 years older than actual age, median sensitivity was 5% (IQR, 2% to 9%) and median specificity was 99% (IQR, 96% to 100%).

CONCLUSIONS: The diagnostic value of apparent age depends on how many years older than his or her actual age a patient looks. A physician's assessment that a patient looks ≥10 years older than his or her actual age has very high specificity for the detection of poor health.

INTRODUCTION

Physicians are trained to begin the physical examination with a general inspection of the patient, which often includes an assessment of whether the patient "appears his/her stated age" or "appears older than his/her stated age"[1,2]. Implicit in this practice are the assumptions that patients who appear older than their actual age are more likely to be in poor health and patients who appear their actual age are less likely to be in poor health. However, we are unaware of any research that has directly examined the validity of these assumptions. We conducted this study to determine whether a physician's assessment that a patient looks older than his or her actual age is a sensitive or specific predictor of poor health status, as determined using the SF-12.

METHODS

Patient Participants

Patients aged 30 to 70 years were recruited in February to April 2009 at four primary care clinics and one general internal medicine clinic, all of which were affiliated with an academic teaching hospital in a large urban center. Patients were approached in clinic waiting areas during times that an investigator was available and asked if they were interested in participating in the study. Patients were excluded if they could not speak or read English. After informed consent was obtained, participants completed a survey that obtained information on demographic characteristics, smoking status, chronic health conditions, and health status. Age in years (rounded to the nearest whole number) was calculated based on date of birth and date of survey completion. A digital photograph was taken of each participant, showing a frontal view of the individual's face with a neutral facial expression.

The health status of each patient participant was determined using the self-administered version of the 12-Item Short Form Health Survey (SF-12). This instrument generates a physical composite score and mental composite score, which are well-validated measures of general physical health and general mental health, respectively[3]. Higher physical and mental composite scores indicate better health and are normalized to a mean of 50 and a standard deviation (SD) of 10 in the United States adult general population[3]. We defined poor health status as the presence of a low physical health score and/or a low mental health score. Physical and mental health scores were classified as low based on four different criteria: ≥0.5 SD (≥5 points), ≥1.0 SD (≥10 points), ≥1.5 SD (≥15 points), and ≥2.0 SD (≥20 points) below the patient's age group norm. Validation studies of the SF-12 have shown that

scores in this range are found in individuals with serious physical and mental health impairments due to conditions such as severe congestive heart failure, severe diabetes, or clinical depression[3].

Physician Participants

Physician participants were randomly selected internal medicine residents and general internal medicine faculty physicians at the University of Toronto. Residents were in the first, second, or third year of postgraduate training in internal medicine. Sampling was stratified to obtain approximately equal numbers of resident and faculty participants. Physicians were sent an e-mail inviting them to participate in the study through a web-based computer program. The program presented physicians with one patient photograph at a time. To reduce any systematic bias due to rater fatigue, the computer program randomized the order in which the photographs were presented each time the study was administered, so that each physician viewed the photographs in a different random order. Physicians were asked to rate all 126 patient photographs and were blinded to the specific objectives of the study and to the patients' health status. The physician received the following instructions with each image: "This patient is [actual age] years of age. How old do you think this patient *looks*? Note: You are NOT guessing this patient's actual age, since that is already given." Examples of actual images presented to physicians are shown in the Appendix. Physicians were able to enter any whole number between 15 and 100, inclusive. There was no time limit to view each photo, but physicians were encouraged to complete the study in a single sitting. After viewing the photographs, physicians were asked to provide their personal demographic information.

All patient and physician participants gave written informed consent and received a small honorarium for participating in the study. This study was approved by the Research Ethics Board of St. Michael's Hospital, Toronto, Canada.

Statistical Analyses

"Apparent age difference" was calculated by subtracting the patient's actual age from each physician's assessment of the patient's apparent age. The degree to which physicians agreed in their assessments of apparent age difference was determined using the intraclass correlation coefficient estimated from a two-way random effects model.

We explored the association between certain patient characteristics and the patient's mean apparent age difference (calculated on the basis of his or her ratings by all physicians). Patient characteristics that were examined were age, race, education, household income, employment status, smoking status, count of chronic medical conditions, and appearance factors. Appearance factors included sex, grey or white hair, facial hair, male pattern baldness, and glasses, as determined by two investigators (SWH and MA) who viewed each photograph independently and conferred to resolve any discrepancies. Male pattern baldness was defined as an appearance consistent with type II or higher on the Norwood Classification scale[4]. The associations between patient characteristics and the mean apparent age difference were examined using

Pearson correlation (for age), t-tests (for sex, smoking status, grey or white hair, facial hair, and male pattern baldness), and ANOVA (for race, education, household income, employment status, and count of chronic medical conditions).

For each physician, we calculated the sensitivity, specificity, positive likelihood ratio, and negative likelihood ratio of rating a patient as looking ≥1 year, ≥5 years, and ≥10 years older than actual age for the detection of poor health status. Analyses were conducted using four different criteria for poor health status: SF-12 scores ≥0.5 SD, ≥1.0 SD, ≥1.5 SD, and ≥2.0 SD below the age group norm. For the entire group of physicians, the median and interquartile range (IQR) for sensitivity, specificity, positive likelihood ratio, and negative likelihood ratio were calculated. The area under the receiver-operator curve, an overall measure of how well a diagnostic test discriminates between ill and healthy patients, was determined for each physician by calculating the c statistic using logistic regression. Analyses were done comparing the performance of resident and faculty physicians. Additional analyses were conducted for patients <50 years old compared to those ≥50 years old, and for the outcomes of poor physical health status alone and poor mental health status alone; these results are not presented because they were very similar to our overall results. SAS version 9.2 (SAS Institute Inc., Cary, NC) and SPSS version 16 (SPSS, Inc., Chicago, IL) were used to perform statistical analyses.

RESULTS

Characteristics of the 126 patient participants are shown in Table 1. Mean age was 46.2 years (SD, 9.0). The patients' SF-12 physical composite scores were on average 5.3 points (SD, 12.0) below age group norms, and their SF-12 mental composite scores were on average 6.3 points (SD, 12.4) below age group norms. The proportion of patients who met each of the four different criteria for poor health status is shown in Table 1. A total of 43 patients (34%) had a physical and/or mental health score that was ≥2.0 SD below the age group norm.

Fifty-eight physicians participated in the study (response rate, 73%). Fifty-one physicians rated all 126 photographs, and 7 physicians rated some but not all photographs. Demographic information was provided by 24 internal medicine residents and 29 general internal medicine faculty physicians, of whom 62% were male and 66% were white. Age ranged from 24 to 70 years, with a mean age of 28 years (SD, 2) among resident physicians and 45 years (SD, 10) among faculty physicians.

The mean apparent age difference for each patient, based on his or her ratings by all physicians, ranged from 8.9 years younger to 10.3 years older than actual age, with a mean value across all patients of 0.5 years older than actual age. Analyses of the association between patient characteristics and mean apparent age difference revealed that younger age (Pearson correlation coefficient -0.20, p=0.03), low household income (F=3.5, p=0.02), and a high count of chronic conditions (F=3.5, p=0.05) were associated with being perceived as appearing older than actual age. Sex, race, education, employment status, smoking status, grey or white hair, facial hair, male pattern baldness, and wearing glasses were not significantly

Table 1. Characteristics of Patient Participants (n=126)

Characteristic	Number (%)
Sex*	
Male	63 (50)
Race	
White	80 (65)
Black	12 (10)
Hispanic	8 (6)
Other	26 (19)
Education	
Not a high school graduate	20 (16)
High school graduate	17 (14)
Some university/college	33 (26)
University/college degree	56 (44)
Annual household income	
≤$20,000	45 (36)
$20,001 to $60,000	35 (28)
$60,001 to $100,000	21 (17)
≥$100,000	25 (20)
Employment status	
Working full time or part time	63 (50)
On disability	43 (34)
Homemaker	5 (4)
Student	3 (2)
Unemployed	5 (4)
Retired	7 (6)
Smoking status	
Daily or occasional	52 (41)
Non-smoker	63 (50)
No response	11 (9)
Grey or white hair*	58 (46)
Facial hair*	29 (23)
Male pattern baldness*	40 (32)
Glasses*	29 (23)
Count of chronic conditions†	
0	64 (51)
1	37 (29)
2	12 (10)
≥3	13 (10)
Physical health score‡	
≥0.5 SD (5 points) below norm	51 (41)
≥1.0 SD (10 points) below norm	42 (33)
≥1.5 SD (15 points) below norm	28 (22)
≥2.0 SD (20 points) below norm	22 (18)
Mental health score‡	
≥0.5 SD (5 points) below norm	58 (46)
≥1.0 SD (10 points) below norm	50 (40)
≥1.5 SD (15 points) below norm	34 (27)
≥2.0 SD (20 points) below norm	26 (21)
Physical and/or mental health score‡	
≥0.5 SD (5 points) below norm	79 (63)
≥1.0 SD (10 points) below norm	70 (56)
≥1.5 SD (15 points) below norm	53 (42)
≥2.0 SD (20 points) below norm	43 (34)

Assessed on the basis of the patient's photograph (see text for details).
†*Based on self-report, selecting from a list of eight specified conditions: (1) heart disease, such as angina, heart attack, or congestive heart failure (CHF); (2) lung disease, such as asthma, emphysema, chronic bronchitis, or chronic obstructive pulmonary disease (COPD); (3) liver disease, such as cirrhosis or chronic hepatitis; (4) intestinal or stomach ulcers, or other bowel disorders, such as Crohn's disease or ulcerative colitis; (5) kidney failure or chronic kidney disease; (6) diabetes; (7) arthritis; and (8) stroke*
‡*Assessed by the SF-12 and classified using four different criteria based on number of standard deviations below the age group norm (see text for details)*

associated with being perceived as appearing older than actual age.

Across all physician ratings of patients' apparent age, 11% of assessments were that the patient looked ≥5 years younger

than actual age, 72% of assessments were that the patient looked within 5 years of actual age, and 17% of assessments were that the patient looked ≥5 years older than their age. In 4% of assessments, the patient was felt to look ≥10 years older than actual age. There was only a moderate level of agreement among physicians regarding the difference between how old each patient looked and the patient's actual age, as indicated by an intra-class correlation coefficient of 0.43 (95% confidence interval, 0.37 to 0.50).

Because physicians differed widely in their assessments of apparent age difference, the sensitivity and specificity of this measure for the detection of poor health status varied across physicians. Table 2 shows median sensitivity and specificity at the thresholds of looking ≥1 year, ≥5 years, and ≥10 years older than actual age. These values varied only slightly depending on which of the four different criteria were used to define poor health status. At the threshold of looking ≥1 year older than actual age, median sensitivity was 48–56% and median specificity was 57–60%. At the threshold of looking ≥5 years older than actual age, median sensitivity fell to 23–29% and median specificity rose to 82–83%. At the threshold of looking ≥10 years older than actual age, median sensitivity and specificity were 4–5% and 98–99%, respectively. Figure 1 shows boxplots of sensitivity and specificity with poor health status defined as SF-12 scores ≥2.0 SD below the age group norm.

Table 3 shows median positive and negative likelihood ratios at the thresholds of looking ≥1 year, ≥5 years, and ≥10 years older than actual age. Looking ≥1 year older than actual age was associated with a median negative likelihood ratio of 0.78–0.92. Figure 2 shows boxplots of likelihood ratios with poor health status defined as SF-12 scores ≥2.0 SD below the age group norm. Likelihood ratios associated with looking ≥10 years older than actual age are not shown because the positive likelihood ratio was infinity for 18 physicians (31%); among the 40 physicians for whom the positive likelihood ratio could be calculated, the median was 2.0 (interquartile range, 1.6 to 3.6).

The area under the receiver-operator curve, which ranges from 0.50 for a test that performs no better than chance to 1.00 for a perfect test, can be interpreted in this context as the probability that a physician will assign a larger apparent age difference to a patient who is in poor health than to a patient who is not in poor health. Among 58 physicians, the mean area under the receiver-operator curve was 0.59, which is considered poor for a diagnostic test. The median area under the receiver-operator curve was 0.58, with a minimum of 0.50 and a maximum of 0.71 (Fig. 3). When faculty and resident physicians were compared, mean sensitivity, specificity, and positive and negative likelihood ratios did not differ significantly, but the mean area under the receiver-operator curve was slightly higher among faculty than residents (0.60 vs. 0.57, p=0.04).

DISCUSSION

The diagnostic value of apparent age depends on how many years older than his or her actual age a patient looks. Physicians' assessments that a patient looked ≥1 year older than his or her actual age had low sensitivity and specificity for the detection of poor health status. Using the higher threshold of appearing ≥5 years older than actual age, the median

Table 2. Sensitivity and Specificity of Apparent Age for the Detection of Poor Health. Poor Health Status Was Assessed by SF-12 Physical and/or Mental Health Scores, Using Four Different Criteria Based on Number of SD Below the Age Group Norm. IQR Denotes Interquartile Range

	Apparent age					
	Looking ≥1 year older than actual age		Looking ≥5 years older than actual age		Looking ≥10 years older than actual age	
Criterion to define poor health status	Sensitivity, median (IQR)	Specificity, median (IQR)	Sensitivity, median (IQR)	Specificity, median (IQR)	Sensitivity, median (IQR)	Specificity, median (IQR)
≥0.5 SD below norm	48.0 (36.7–56.0)	57.4 (48.7–68.6)	23.4 (16.5–30.7)	83.2 (76.3–89.4)	3.8 (1.3–6.3)	97.9 (95.7–100.0)
≥1.0 SD below norm	47.8 (37.1–55.7)	57.1 (48.2–68.3)	23.9 (17.1–31.5)	82.1 (76.1–89.0)	4.3 (1.4–6.2)	98.2 (96.4–100.0)
≥1.5 SD below norm	52.8 (43.4–60.4)	59.6 (50.7–70.2)	24.8 (20.3–34.0)	82.2 (76.7–87.7)	3.8 (1.9–7.5)	98.6 (95.9–100.0)
≥2.0 SD below norm	55.8 (43.6–62.8)	60.2 (51.5–69.0)	29.1 (18.6–35.5)	81.9 (76.8–88.0)	4.7 (2.3–9.3)	98.7 (96.4–100.0)

specificity increased to 82–83% with a median positive likelihood ratio of 1.4–1.7, indicating that this finding is associated with an only modestly increased likelihood of poor health status. For a few physicians, this finding had somewhat higher diagnostic value, with positive likelihood ratios in the range of four to seven. The assessment that a patient looked ≥10 years older than his or her actual age had high specificity but poor sensitivity for the detection of poor health. The impression that a patient did not look older than his or her actual age had almost no value in ruling out poor health (median negative likelihood ratios of 0.8–0.9).

Previous research on the association between facial appearance and health has not examined the specific question addressed by the present study. The facial attractiveness of young adults is positively associated with their current physical health and subsequent longevity[5,6], but these observations have limited applicability to the field of medicine because physicians typically assess how old their patients look rather than their facial attractiveness. In one study that did examine apparent age, physicians, medical students, nurses, and non-medical staff (12 in each group) viewed photographs of 27 healthy outpatients and estimated how old they looked, without knowledge of their actual age[7]. Patients who were rated as appearing 5 or more years older than their actual age were more likely to be tobacco users, but there was no significant association with alcohol use, sun exposure, marital status, or health status[7]. In a Danish study of 387 sets of same-sex twins age 70 years and older, 20 nurses and 21 individuals who were not health care providers were asked to estimate the age of each twin based on their photographs[8,9]. Apparent age was significantly associated with survival over the 7-year period after the photographs were taken, even after adjustment for age, sex, and physical and cognitive functioning, and the twin who was rated as looking older was more likely to die first[9].

These previous studies each have certain features that limit their relevance to clinical practice: a study design restricted to the assessment of twin pairs, the use of non-physicians as raters, and/or the fact that raters estimated the apparent age of individuals without any knowledge of their actual age. In our study, physicians were informed of each patient's actual age and asked how old the patient looked, rather than being asked to estimate the patient's age. This aspect of our study is an important strength because physicians are almost always aware of their patient's actual age before the clinical encounter, and they subsequently form an impression of whether the patient appears younger or older than his or her true age. Thus, our procedure more closely approximates actual prac-

tice than a study in which physicians are asked to guess the age of a patient.

A similar but somewhat different clinical approach to evaluating the general appearance of a patient is to assess whether the individual "appears chronically ill" rather than "older than his/her stated age." Our study did not examine whether the overall impression that a patient is chronically ill appearing is more sensitive or specific than apparent age for the detection of poor health. Future research could evaluate this interesting question.

This study has certain limitations. First, because the age of patient participants ranged from 30 to 70 years, our findings may not be generalizable to patients outside this age range. The exclusion of patients less than 30 years of age is reasonable, since differences in apparent age among younger individuals may represent variation in rates of maturation rather than the effects of poor health. In light of the positive findings of the Danish twin study, further research is needed to determine whether apparent age is a predictor of health status and/or survival among non-twin older individuals. Second, the patients who participated in this study were
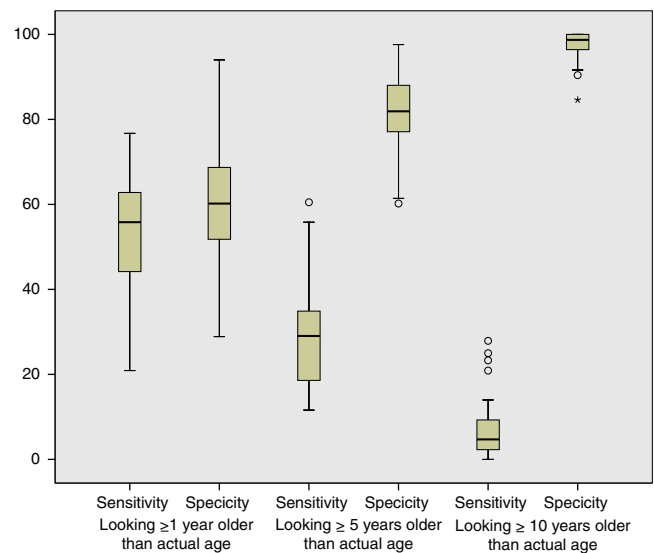
Figure 1. Sensitivity and specificity of apparent age for the detection of poor health (using the criterion of SF-12 scores ≥2.0 SD below the age group norm). The median for all physicians is denoted by the horizontal line within each box; the interquartile range (IQR) is indicated by the box. Physicians with outlier values (1.5–3.0 IQRs from the upper or lower edge of the box) are denoted by circles.

Table 3. Positive and Negative Likelihood Ratios of Apparent Age for the Detection of Poor Health. Poor Health Status Was Assessed by SF-12 Physical and/or Mental Health Scores, Using Four Different Criteria Based on Number of SD Below the Age Group Norm. IQR Denotes Interquartile Range
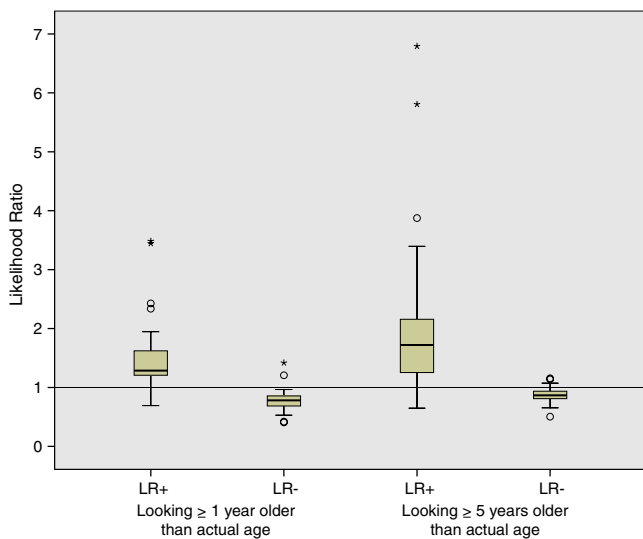
	Apparent age					
	Looking ≥1 year older than actual age		Looking ≥5 years older than actual age		Looking ≥10 years older than actual age	
Criterion used to define poor health status	LR+, Median (IQR)	LR−, Median (IQR)	LR+, Median (IQR)	LR−, Median (IQR)	LR+, Median (IQR)	LR−, Median (IQR)
≥0.5 SD below norm	1.15 (1.00–1.32)	0.92 (0.81–0.99)	1.39 (1.04–1.93)[*]	0.92 (0.85–0.99)	1.49 (0.83–1.78)[†]	0.98 (0.96–1.00)
≥1.0 SD below norm	1.13 (1.04–1.35)	0.92 (0.82–0.98)	1.37 (1.10–1.83)	0.92 (0.88–0.96)	1.60 (0.87–2.40)[‡]	0.98 (0.96–1.00)
≥1.5 SD below norm	1.30 (1.19–1.64)	0.80 (0.70–0.86)	1.54 (1.26–2.00)	0.90 (0.84–0.94)	1.63 (1.38–2.75)[§]	0.98 (0.94–1.00)
≥2.0 SD below norm	1.29 (1.20–1.64)	0.78 (0.68–0.86)	1.72 (1.25–2.16)	0.87 (0.81–0.94)	1.96 (1.50–3.81)[‖]	0.97 (0.94–0.99)

[*]*LR+ undefined for n=1 physician since specificity was 100%*
[†]*LR+ undefined for n=27 physicians since specificity was 100%*
[‡]*LR+ undefined for n=26 physicians since specificity was 100%*
[§]*LR+ undefined for n=19 physicians since specificity was 100%*
[‖]*LR+ undefined for n=18 physicians since specificity was 100%*

recruited in outpatient clinics; thus, our findings may not be applicable to acutely ill or hospitalized patients. Third, patients were not asked to provide information on their human immunodeficiency virus status because this information was felt to be too sensitive to obtain in the context of a brief self-administered survey. As a result, data on patients' chronic conditions do not account for possible human immunodeficiency virus infection. Fourth, physicians assessed each patient's apparent age based on a single frontal-view photograph; an assessment based on an in-person clinical interaction or a standardized video might be more robust. Finally, our sample included too few patients in excellent physical or mental health to allow us to test the association between being rated as appearing younger than one's actual age and being in excellent health. This limitation may be considered relatively minor, however, since physicians are usually focused on the

identification of individuals in poor health rather than those in excellent health.

In summary, the common practice of assessing whether or not a patient looks older than his or her actual age can be useful for identifying individuals with poor health status, but with certain important limitations. The impression that "the patient appears his/her stated age" has little or no diagnostic value, since this finding does not substantially reduce the likelihood that the patient is in poor health. The assessment that "the patient appears older than his/her stated age" provides more diagnostic information if one specifies how many years older than his or her actual age the patient appears. Physicians may wish to focus on patients who look 10 or more years older than their actual age, as this finding has very high specificity for the detection of poor health. When physicians encounter such patients, a detailed inquiry into the patient's physical and mental health status is justified.

Figure 2. Positive (LR+) and negative (LR−) likelihood ratios of apparent age for the presence of poor health (using the criterion of SF-12 scores ≥2.0 SD below the age group norm). The median for all physicians is denoted by the horizontal line within each box; the interquartile range (IQR) is indicated by the box. Physicians with outlier values (1.5–3.0 IQRs from the upper or lower edge of the box) or extreme values (>3.0 IQRs from the upper or lower edge of the box) are denoted by circles and asterisks, respectively.
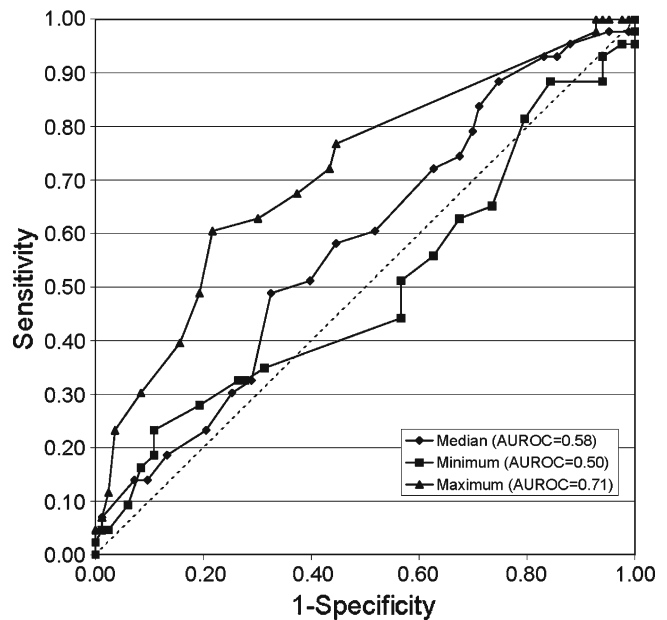
Figure 3. Receiver-operator curves showing the median, minimum, and maximum values for the area under the receiver-operator curve (AUROC) among physicians.

REFERENCES

1. **Krakauer EL, Zhu AX, Bounds BC, et al.** Case 6-2005: a 58-year-old man with esophageal cancer and nausea, vomiting, and intractable hiccups. N Engl J Med. 2005;352:817–25.
2. **Wolf M, Rose H, Smith RN.** Case 28-2005: A 42-year-old man with weight loss, weakness, and a rash. N Engl J Med. 2005;353:1148–57.
3. **Ware JE, Kosinski M, Keller SD**. SF-12: How to Score the SF-12 Physical and Mental Health Summary Scales. 2nd ed. Boston, MA: The Health Institute, New England Medical Center; 1995.
4. **Nordström REA.** Classification of androgenetic alopecia. In: **Unger WP, Shapiro R, eds.** Hair transplantation. 4th ed. New York, NY: Marcel Dekker; 2004:49–51.
5. **Shackelford TK, Larsen RJ.** Facial attractiveness and physical health. Evol Hum Behav. 1999;20:71–6.
6. **Henderson J, Anglin J.** Facial attractiveness predicts longevity. Evol Hum Behav. 2003;24:351–6.
7. **Sheretz EF, Hess SP.** Stated Age. N Engl J Med. 1993;329:281–2.
8. **Christensen K, Iachina M, Rexbye H, et al.** "Looking old for your age": genetics and Mortality. Epidemiology. 2004;15:251–2.
9. **Christensen K, Thinggaard M, McGue M, et al.** Perceived age as clinically useful biomarker of ageing: cohort study. Br Med J. 2009;339:b5262.

APPENDIX

Examples of actual patient photographs, as displayed to physicians.
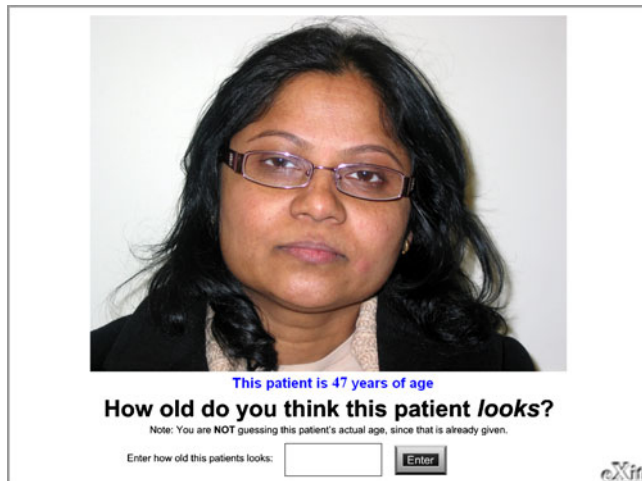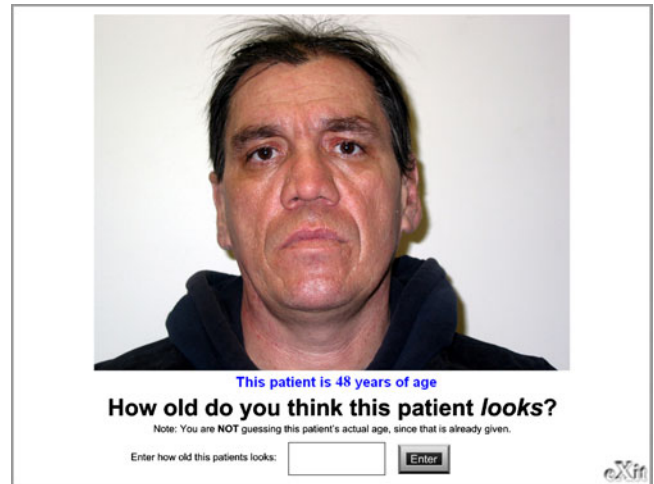
Figures 4, 5, 6, and 7

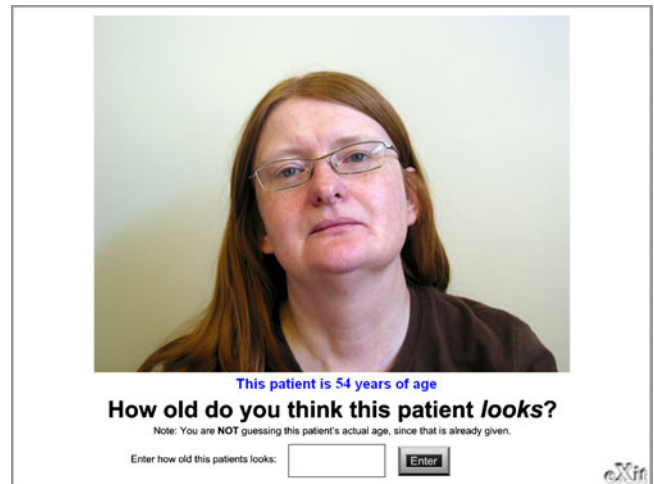Figure 5. Actual patient photograph shown to physicians

Figure 6. Actual patient photograph shown to physicians
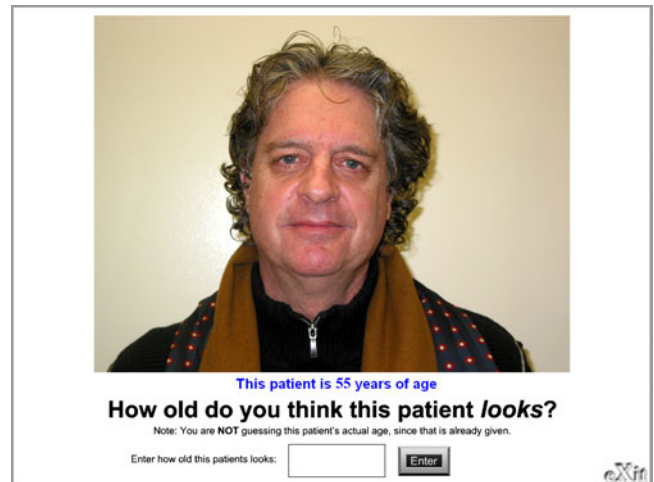

Figure 4. Actual patient photograph shown to physicians

Figure 7. Actual patient photograph shown to physicians