Medicine Residents' Understanding of the Biostatistics and Results in the Medical Literature

Donna M. Windish, MD, MPH
Stephen J. Huot, MD, PhD
Michael L. Green, MD, MSc

Context Physicians depend on the medical literature to keep current with clinical information. Little is known about residents' ability to understand statistical methods or how to appropriately interpret research outcomes.

Objective To evaluate residents' understanding of biostatistics and interpretation of research results.

Design, Setting, and Participants Multiprogram cross-sectional survey of internal medicine residents.

Main Outcome Measure Percentage of questions correct on a biostatistics/study design multiple-choice knowledge test.

Results The survey was completed by 277 of 367 residents (75.5%) in 11 residency programs. The overall mean percentage correct on statistical knowledge and interpretation of results was 41.4% (95% confidence interval [CI], 39.7%-43.3%) vs 71.5% (95% CI, 57.5%-85.5%) for fellows and general medicine faculty with research training ($P < .001$). Higher scores in residents were associated with additional advanced degrees (50.0% [95% CI, 44.5%-55.5%] vs 40.1% [95% CI, 38.3%-42.0%]; $P < .001$); prior biostatistics training (45.2% [95% CI, 42.7%-47.8%] vs 37.9% [95% CI, 35.4%-40.3%]; $P = .001$); enrollment in a university-based training program (43.0% [95% CI, 41.0%-45.1%] vs 36.3% [95% CI, 32.6%-40.0%]; $P = .002$); and male sex (44.0% [95% CI, 41.4%-46.7%] vs 38.8% [95% CI, 36.4%-41.1%]; $P = .004$). On individual knowledge questions, 81.6% correctly interpreted a relative risk. Residents were less likely to know how to interpret an adjusted odds ratio from a multivariate regression analysis (37.4%) or the results of a Kaplan-Meier analysis (10.5%). Seventy-five percent indicated they did not understand all of the statistics they encountered in journal articles, but 95% felt it was important to understand these concepts to be an intelligent reader of the literature.

Conclusions Most residents in this study lacked the knowledge in biostatistics needed to interpret many of the results in published clinical research. Residency programs should include more effective biostatistics training in their curricula to successfully prepare residents for this important lifelong learning skill.

JAMA. 2007;298(9):1010-1022 www.jama.com

PHYSICIANS MUST KEEP CURRENT with clinical information to practice evidence-based medicine (EBM). In doing so, most prefer to seek evidence-based summaries, which give the clinical bottom line,[1] or evidence-based practice guidelines.[1-3] Resources that maintain these information summaries, however, currently include a limited number of common conditions.[4] Thus, to answer many of their clinical questions, physicians need to access reports of original research. This requires the reader to critically appraise the design, conduct, and analysis of each study and subsequently interpret the results.

Several surveys in the 1980s demonstrated that practicing physicians, particularly those with no formal education in epidemiology and biostatistics, had a poor understanding of common statistical tests and limited ability to interpret study results.[5-7] Many physicians likely have increased difficulty today because more complicated statistical methods are being reported in the medical literature.[8] They may be able to understand the analysis and interpretation of results in only 21% of research articles.[8]

Educators have responded by increasing training in critical appraisal and biostatistics throughout the continuum of medical education. Many medical schools currently provide some formal teaching of basic statistical concepts.[9] As part of the Accreditation Council for Graduate Medical Education's practice-based learning and improvement competency, residents must demonstrate ability in "locating, appraising, and assimilating evidence from scientific studies related to their patients' problems and apply knowledge of study designs and statistical methods to the appraisal of

Author Affiliations: Department of Internal Medicine, Yale University School of Medicine, New Haven, Connecticut.
Corresponding Author: Donna M. Windish, MD, MPH, Yale Primary Care Residency Program, 64 Robbins St, Waterbury, CT 06708 (donna.windish@yale.edu).

clinical studies."[10] Most residency programs address this competency through EBM curricula or journal clubs.[11-13] In 2000, the majority of these programs included training in appraisal of studies and study conduct, but fewer specifically addressed the selection and interpretation of statistical tests.[11,14] In addition, the majority of published assessments of residents' knowledge and skills in EBM were performed at single programs, were conducted in the context of determining the impact of a specific curriculum, evaluated critical appraisal skills more commonly than biostatistics, and found that residents scored well below EBM "experts" on evaluation instruments.[15] We performed a multiprogram assessment of residents' biostatistics knowledge and interpretation of study results using a new instrument developed for this study.

METHODS

Survey Development

We developed an instrument to reflect the statistical methods and results most commonly represented in contemporary research studies (APPENDIX). Thus, we reviewed all 239 original articles published from January to March of 2005 in each issue of 6 general medical journals (*American Journal of Medicine*, *Annals of Internal Medicine*, *BMJ*, *JAMA*, *Lancet*, and *New England Journal of Medicine*) and summarized the frequency of statistical methods described (TABLE 1). From this review, we developed questions that focused on identifying and interpreting the results of the most frequently occurring simple statistical methods (eg, χ^2, *t* test, analysis of variance) and multivariate analyses (eg, Cox proportional hazards regression, multiple logistic regression).

Survey Instrument

The survey (Appendix) contained 4 sets of questions: (1) 11 demographic questions that included age, sex, current training level, past training in biostatistics and EBM, and current journal-reading practices; (2) 5 attitude questions regarding statistics; (3) 4 confidence questions about interpreting and assessing statistical concepts; and (4) a 20-question biostatistics knowledge test that assessed understanding of statistical methods, study design, and interpretation of study results. Statistical attitudes and confidence questions were adapted from surveys on the Assessment Resource Tools for Improving Statistical Thinking (ARTIST) Web site, which is a resource for teaching statistical literacy, reasoning, and thinking.[16] Attitudes regarding statistics were rated on a 5-point Likert scale. Confidence questions were assessed using a 5-point scale in which 1 indicated no confidence and 5 indicated complete confidence. The remaining 20 knowledge test questions addressed understanding of statistical techniques, study design, and interpretation of study results most commonly represented in our journal review. These questions were multiple-choice, clinically oriented with a case vignette, and required no calculations. Two questions were adapted from a study of Danish physicians' statistical knowledge.[7] Seven questions were adapted from course materials used in statistics courses at the Johns Hopkins Bloomberg School of Public Health.[17] The remaining questions were developed by one of the study authors (D.M.W.). The knowledge questions addressed research variable types, statistical methods, confidence intervals, *P* values, sensitivity and specificity, power and sample size, study design, and interpretation of study results.

Pilot Testing of Biostatistics Knowledge Test

The original test contained 22 knowledge questions and was pilot tested with 5 internal medicine faculty with advanced training in epidemiology and biostatistics and 12 primary care internal medicine residents at 1 residency program. Faculty reviewed the instrument for content validity, completed the test, and provided feedback. Residents completed the test and provided written and oral feedback. Four of the 5 faculty answered 21 of 22 questions correctly and 1 faculty member correctly answered 19 questions. This resulted in an overall mean score of 94%. Incorrect responses did not favor any particular question. Residents answered 53% of questions correctly. Based on feedback, 1 question was modified to improve clarity, 3 questions were eliminated to avoid duplicating similar concepts, and 1 question was added to further assess interpretation of results. Therefore, the final version of the test consisted of 20 questions.

Table 1. Statistical Methods Used in 239 Original Research Articles in 6 General Medical Journals, 2005

Type of Test	No. (%)
Descriptive statistics[a]	219 (91.6)
Simple statistics	120 (50.2)
χ^2 Analysis	70 (29.3)
t Test	48 (20.1)
Kaplan-Meier analysis	48 (20.1)
Wilcoxon rank sum test	38 (15.9)
Fisher exact test	33 (13.8)
Analysis of variance	21 (8.8)
Correlation	16 (6.7)
Multivariate statistics	164 (68.6)
Cox proportional hazards	64 (26.8)
Multiple logistic regression	54 (22.6)
Multiple linear regression	7 (2.9)
Other regression analyses[b]	38 (15.9)
None	5 (2.1)
Other methods, techniques, or strategies	
Intention-to-treat analysis	42 (17.6)
Incidence/prevalence	39 (16.3)
Relative risk/risk ratio	29 (12.2)
Sensitivity analyses	21 (8.8)
Sensitivity/specificity	15 (6.3)

[a] Descriptive statistics included mean, median, frequency, standard deviation, and interquartile range.
[b] Other regression analyses included weighted logistic regression, unconditional logistic regression, conditional logistic regression, longitudinal regression, Poisson regression, pooled logistic regression, nonlinear regression, meta-regression, negative binomial regression, and generalized estimating equations.

Target Population and Survey Administration

We conducted an anonymous cross-sectional survey from February through July 2006 of 11 internal medicine residency programs in Connecticut, including 7 traditional internal medicine programs, 2 primary care medicine programs, 1 medicine/pediatrics program, and 1 medicine/preventive medicine program. We initially contacted all 15 internal medicine residency programs in Connecticut to ask for their participation in the study. All programs were successfully contacted and expressed interest. However, 3 programs could not accommodate the study because of scheduling conflicts, and 1 program was not included because its residents (medicine/pediatrics) were distributed to different training sites and therefore were not present at the conferences used for the survey.

Included residencies were both university affiliated (7 programs) and community based (4 programs). Residents at all postgraduate levels of training were invited to participate. Oral consent was obtained from each participant after providing a description of the survey's purpose. The survey was administered during the first 25 minutes of an inpatient noon conference lecture for current residents. After all questionnaires were collected, the remainder of the time was devoted to a seminar in statistical methods and interpretation of the literature. Four residency programs also allowed us to survey their entering intern classes during their orientations. To provide data for validity testing, an additional 10 faculty and fellows trained in clinical investigation also completed the final survey. The Yale University human investigation committee approved the study protocol.

Analysis

In addition to assessing the content validity, the psychometric properties of the 20-question knowledge test were determined by assessing internal consistency using Cronbach α. Discriminative validity was assessed by comparing the difference in mean scores obtained between residents and research-trained fellows and faculty using the *t* test.

The biostatistics knowledge test was scored by determining the percentage of questions correct, weighting each question equally. Missing values were counted as incorrect responses. The *t* test or a 1-way analysis of variance was used to compare survey scores by respondent characteristics. We calculated the percentage of residents who agreed or strongly agreed with each attitudinal question. We determined the percentage of respondents with fair to high confidence for each confidence question and the mean confidence score based on the sum of all 4 questions.

Correlation analyses were performed to test for multicollinearity between 3 sets of factors we hypothesized might be highly correlated (training outside of the United States and years since medical school graduation; training level and age; and past biostatistics training and past epidemiology training). Bivariate analyses were performed to identify factors that might be associated with knowledge scores. Candidate variables included sex, age, academic affiliation of residency program, advanced degrees, years since medical school graduation, training outside of the United States, current level of training, past biostatistics training, past epidemiology training, past EBM training, and currently reading medical journals. We also tested for effect modification for pairs of factors including past biostatistics training and past EBM training; past biostatistics training and past epidemiology training; and past biostatistics training and sex. The results of the correlation, bivariate, and effect modification analyses were used to determine which demographic variables to include in the multivariable model. Decisions to include factors in the multivariable regression analysis were based on the strength of correlated factors ($r < 0.75$) or a *P* value $< .05$ on bivariate analyses. Forward stepwise regression was subsequently used to identify which demographic factors were independently associated with biostatistics knowledge scores.

To adjust for multiple pairwise comparisons, a 2-sided level of statistical significance was set at $P < .01$ using a Bonferroni correction. With a sample size of 277 and a *P* value of .01, the study had 80% power to detect a 4.4% difference in mean knowledge scores. All analyses were performed using Stata release 8.2 (StataCorp, College Station, Texas).

RESULTS

Training Program Characteristics

The 11 targeted training programs had 532 residents, with a mean of 53.6 trainees (range, 12-118).[18,19] In comparison, the 388 internal medicine training programs in the United States have a total of 21 885 residents, with a mean of 56.4 trainees (range, 4-170) ($P = .76$ compared to targeted programs).[20] The study programs had 41.9% women residents, compared with 42.1% nationally ($P = .96$), and 49.9% of residents with training outside of the United States vs 52.3% nationally ($P = .51$).[19] Comparing targeted programs with all internal medicine programs, no statistically significant differences were seen for postgraduate year 1 trainees in mean duty hours per week (61.9 vs 65.2, $P = .13$), mean consecutive work hours (30 vs 27.5, $P = .09$), and mean number of days off per week (1.3 vs 1.2, $P = .31$).[18] Targeted programs also did not differ in these characteristics from the remaining 4 Connecticut training programs.

Respondent Characteristics

Three hundred sixty-seven residents in the 11 targeted programs were on rotations that would make them available to attend their respective noon conferences on the day of the survey. Of these, 309 (84.2%) were in attendance. Of the total available residents, 277 (75.5%) completed the assessment. The response rate for individual programs ranged from 28.1% to 80%. No differences in response rates or attendance were seen based on sex, level of training, or past training outside of the United States. **TABLE 2** lists the respondents' demographic characteristics. Approximately equal numbers of men and women were represented. Fifty-eight percent were enrolled in traditional internal medicine programs, 76.5% participated in university-based programs, 50.6% had some training outside of the United States, and 14.8% had advanced degrees. More than 68% of respondents had some training in biostatistics, with approximately 70% of this training occurring during medical school.

Psychometric Properties of the Knowledge Test

The survey instrument had high internal consistency (Cronbach $\alpha = 0.81$). Fellows and general medicine faculty with advanced training in biostatistics had a significantly higher score than residents (mean percentage correct, 71.5% [95% confidence interval {CI}, 57.5%-85.5%] vs 41.1% [95% CI, 39.7%-43.3%]; $P < .001$), indicating good discriminative validity.

Knowledge of Statistical Methods and Results

The overall mean resident knowledge score was 41.1% (SD, 15.2%; range, 10%-90%). Residents scored highest in recognition of double-blind studies (87.4% [95% CI, 83.5%-91.3%] answering correctly) and interpretation of relative risk (81.6% [95% CI, 77.0%-86.2%] answering correctly) (TABLE 3). They were least able to interpret the results of a Kaplan-Meier analysis, with 10.5% (95% CI, 6.9%-14.1%) answering correctly. Only 37.4% (95% CI, 31.9%-43.3%) understood how to interpret an adjusted odds ratio from a multivariate regression analysis, while 58.8% (95% CI, 53.0%-64.6%) could interpret the meaning of a *P* value.

Factors Associated With Statistical Knowledge

Training outside of the United States had moderate correlation with years since medical school graduation ($r = 0.59$), as did past epidemiology training with past biostatistics training ($r = 0.53$). Training level had a fair correlation with age ($r = 0.46$). No effect modification was seen for the 3 sets of factors assessed. In bivariate analyses, differences in scores were seen based on residency program type, with medicine/pediatric residents scoring the highest (TABLE 4). Residents with advanced degrees performed better than those without advanced training (50.0% [95% CI, 44.5%-55.5%] vs 40.1% [95% CI, 38.3%-42.0%]; $P < .001$). Statistically significant higher scores were also seen in residents who were just enter-

Table 2. Characteristics of the 277 Participants

Characteristic	No. (%)[a]
Sex	
Men	143 (52.0)
Women	134 (48.0)
Age range, y	
21-25	25 (9.2)
26-30	166 (60.8)
31-35	62 (22.7)
≥36	20 (7.3)
Other advanced degrees	41 (14.8)
Doctor of philosophy (PhD)	11 (4.0)
Master of public health (MPH)/master of health science (MHS)	16 (5.8)
Master of science (MSc)	12 (4.4)
Other	4 (1.5)
None	235 (85.1)
Years since medical school graduation	
<1	105 (35.0)
1-3	72 (26.8)
4-10	81 (30.1)
≥11	11 (4.1)
Training outside of the United States	139 (50.6)
College	36 (13.0)
Medical school	107 (38.6)
Residency	33 (11.9)
Other	7 (2.5)
None	138 (49.4)
Academic affiliation of program	
University based	212 (76.5)
Community based	65 (23.5)
Current residency training program type	
Traditional/categorical medicine	161 (58.3)
Primary care medicine	60 (21.7)
Medicine/pediatrics	12 (4.4)
Medicine/preventive medicine	7 (2.5)
Preliminary/transitional year	36 (13.0)
Current level of training	
Entering intern	103 (37.3)
Experienced intern[b]	72 (26.1)
Second-year resident	42 (15.2)
Third-year resident	45 (16.3)
Fourth-year resident	6 (6.2)
Chief resident	8 (2.9)
Previous training/coursework in biostatistics	190 (68.8)
Location of biostatistics training	
College	30 (15.9)
Medical school	132 (69.5)
Residency	6 (3.2)
Other	26 (13.7)
Previous training/coursework in epidemiology	190 (68.8)
Previous training/coursework in evidence-based medicine[c]	162 (58.5)
Regularly reads medical journals	187 (68.8)

[a] Percentages may not total 100% due to missing data or multiple responses.
[b] Trainees in month 8 to 12 of their intern year.
[c] Evidence-based medicine is defined as the integration of the best research evidence with patients' values and clinical circumstances in clinical decision making. This is in contrast to biostatistics, which is the scientific use of quantitative information to describe or draw inferences about natural phenomena, and epidemiology, which is the study of patterns, causes, and control of disease in groups of people.

Table 3. Percentages of Correct Answers for the Knowledge-Based Questions

Question No.[a]	Objective	Correct (95% CI), %
1a	Identify continuous variable	43.7 (37.8-49.5)
1b	Identify ordinal variable	41.5 (35.7-47.3)
1c	Identify nominal variable	32.9 (27.3-38.4)
2	Recognize a case-control study	39.4 (33.6-45.1)
3	Recognize purpose of double-blind studies	87.4 (83.5-91.3)
4a	Identify ANOVA	47.3 (41.4-53.2)
4b	Identify χ^2 analysis	25.6 (20.5-30.8)
4c	Identify *t* test	58.1 (52.3-63.9)
5	Recognize definition of bias	46.6 (40.7-52.4)
6	Interpret the meaning of *P* value >.05	58.8 (53.0-64.6)
7	Identify Cox proportional hazard regression	13.0 (9.0-17.0)
8	Interpret standard deviation	50.2 (42.3-56.1)
9	Interpret 95% CI and statistical significance	11.9 (8.0-15.7)
10	Recognize power, sample size, and significance-level relationship	30.3 (24.9-35.7)
11	Determine which test has more specificity	56.7 (50.8-62.5)
12	Interpret an unadjusted odds ratio	39.0 (33.3-44.7)
13	Interpret odds ratio in multivariate regression analysis	37.4 (31.9-43.3)
14	Interpret relative risk	81.6 (77.0-86.2)
15	Determine strength of evidence for risk factors	17.0 (12.6-21.4)
16	Interpret Kaplan-Meier analysis results	10.5 (6.9-14.1)

Abbreviations: ANOVA, analysis of variance; CI, confidence interval.
[a] See Appendix.

ing residency, had prior biostatistics training, were enrolled in a university-based training program, and were men (Table 4).

Using forward stepwise regression, 5 factors were found to be independently associated with knowledge scores (Table 4). An advanced degree was associated with an absolute increase of 9.2% questions correct after adjustment for other factors ($P<.001$). Successive years since medical school graduation were associated with decreasing knowledge scores, with 11 years or more postgraduation associated with a 12.3% absolute decrease in score compared with less than 1 year postgraduation. Male sex, belonging to a university-based program, and past biostatistics training were all associated with higher scores.

Attitudes and Confidence

The majority of residents agreed or strongly agreed that to be an intelligent reader of the literature it is necessary to know something about statistics (95%) and indicated they would like to learn more about statistics (77%). Seventy-five percent reported they did not understand all of the statistics they encountered in the literature, whereas only 15% felt that they do not trust statistics "because it is easy to lie." More than 58% of respondents indicated that they use statistical information in forming opinions or when making decisions in medical care.

The mean confidence score in understanding certain statistical concepts was 11.4 (SD, 2.7) (maximum possible confidence score, 20). The majority of residents reported fair to complete confidence in understanding *P* values (88%). Fewer were confident in interpreting results of statistical methods used in research (68%), identifying factors influencing a study's power (55%), or assessing if a correct statistical procedure was used (38%).

Respondents with higher confidence in their statistical knowledge (a score higher than the mean confidence score) performed better on the knowledge questions than those with lower confidence (43.6% [95% CI, 40.8%-46.3%] vs 39.3% [95% CI, 37.0%-41.6%]; $P=.02$). Those who reported fair to high confidence in interpreting a *P* value were more likely to correctly interpret its meaning (62.8% [95% CI, 56.8%-67.2%] vs 38.2% [95% CI, 24.3%-51.7%]; $P=.006$). No differences were seen in a resident's ability to appropriately identify the correct statistical procedure used based on their confidence to do so.

COMMENT

In this multiprogram survey of internal medicine residents' confidence in, attitudes toward, and knowledge of statistical methods and interpretation of research results, 95% believed that it was important to understand these concepts to be an intelligent reader of the literature, yet three-fourths of residents acknowledged low confidence in understanding the statistics they encounter in the medical literature. This lack of confidence was validated by their low knowledge scores, in which on average only 8 of 20 questions were answered correctly. Although past instruction in biostatistics and advanced degrees were associated with better performance, knowledge scores appeared to decline with progression through training.

The poor knowledge in biostatistics and interpretation of study results among residents in our study likely reflects insufficient training. Nearly one-third of trainees indicated that they never received biostatistics teaching at any point in their career. When training did occur, the majority of instruction took place during undergraduate medical education and was not reinforced in residency. The most recent comprehensive survey of medical school biostatistics teaching was conducted in the 1990s and found that more than 90% of medical schools focused their biostatistics teaching in the preclinical years without later instruction and that the depth and breadth of this education varied greatly among schools.[21] That review reported that familiar concepts such as *P* values, *t* tests, and χ^2 analyses were

Table 4. Knowledge Scores by Resident Characteristics[a]

Characteristic	Bivariate Analyses: Mean Correct, % (95% CI)	Bivariate Analyses: P Value	Multiple Linear Regression: Score Difference, %[b] (95% CI)	Multiple Linear Regression: P Value
Sex				
Women	38.8 (36.4 to 41.1)	.004[c]	1 [Reference]	.01
Men	44.0 (41.4 to 46.7)		4.4 (0.97 to 7.9)	
Age range, y				
21-25	44.4 (38.8 to 50.0)	.46[d]		
26-30	41.7 (39.5 to 44.0)			
31-35	41.0 (37.0 to 45.0)			
≥36	37.3 (30.4 to 44.2)			
Other advanced degrees				
No	40.1 (38.3 to 42.0)	<.001[c]	1 [Reference]	<.001
Yes	50.0 (44.5 to 55.5)		9.2 (4.2 to 14.3)	
Years since medical school graduation				
<1	45.2 (42.4 to 48.0)	<.001[d]	1 [Reference]	
1-3	42.2 (38.6 to 45.8)		−2.3 (−6.5 to 2.0)	.29
4-10	36.8 (33.6 to 40.0)		−4.7 (−9.4 to 0.01)	.05
≥11	34.5 (27.9 to 41.1)		−12.3 (−22.2 to −3.3)	.007
Training outside of the United States				
No	45.2 (42.7 to 47.8)	<.001[c]		
Yes	37.9 (35.4 to 40.3)			
Academic affiliation of program				
Community based	36.3 (32.6 to 40.0)	.002[c]	1 [Reference]	.02
University based	43.0 (41.0 to 45.1)		5.6 (0.93 to 10.2)	
Current level of training				
Entering intern	45.6 (42.8 to 48.4)	.01[d]		
Experienced intern[e]	39.2 (35.7 to 42.7)			
Second-year resident	39.3 (34.9 to 43.7)			
Third-year resident	38.4 (33.6 to 43.2)			
Fourth-year resident	43.3 (30.3 to 56.3)			
Chief resident	38.1 (31.0 to 45.2)			
Current residency training program type				
Traditional/categorical medicine	39.8 (37.5 to 42.1)	.003[d]		
Primary care medicine	42.4 (39.0 to 45.8)			
Medicine/pediatrics	54.6 (47.3 to 61.9)			
Medicine/preventive medicine	53.6 (37.8 to 69.5)[f]			
Preliminary/transitional year	41.0 (35.9 to 46.1)			
Previous training/coursework in biostatistics				
No	37.9 (35.4 to 40.3)	.001[c]	1 [Reference]	.04
Yes	45.2 (42.7 to 47.8)		4.5 (0.80 to 8.2)	
Location of biostatistics training				
College	38.1 (31.7 to 44.5)	.004[d]		
Medical school	42.2 (40.0 to 44.8)			
Residency	58.8 (38.2 to 79.4)			
Other	51.0 (44.8 to 57.2)			
Previous training/coursework in epidemiology				
No	37.5 (34.6 to 40.4)	.003[c]		
Yes	43.3 (41.1 to 45.6)			
Previous training/coursework in evidence-based medicine[g]				
No	39.0 (35.8 to 42.1)	.04[c]		
Yes	42.9 (40.7 to 45.0)			
Regularly reads medical journals				
No	42.3 (38.9 to 43.2)	.53[c]		
Yes	41.0 (38.9 to 45.7)			

[a] To adjust for multiple pairwise comparisons, $P < .01$ is considered statistically significant.
[b] Using forward stepwise regression, 5 factors (sex, advanced degree status, years since medical school graduation, program affiliation, and biostatistics training) were found to be associated with knowledge scores. The R^2 value for the final model was 0.18.
[c] Analysis by the *t* test.
[d] Analysis by 1-way analysis of variance.
[e] Trainee in month 8 to 12 of the intern year.
[f] The medicine/preventive medicine scores were not normally distributed. Median (interquartile range), 45% (35%-75%).
[g] Evidence-based medicine is defined as the integration of the best research evidence with patients' values and clinical circumstances in clinical decision making. This is in contrast to biostatistics, which is the scientific use of quantitative information to describe or draw inferences about natural phenomena, and epidemiology, which is the study of patterns, causes, and control of disease in groups of people.

frequently addressed (95%, 92%, and 88%, respectively), but advanced methods (such as Cox proportional hazards regression, multiple logistic regression, and Kaplan-Meier analyses) were not included in instruction.[21] If biostatistics teaching has continued at the same level in recent years, it would not be surprising that only a small percentage of residents in our survey (10.5%-37.6%) understood the results and use of these analyses.

The correlates of differences in knowledge scores might have been expected. Residents with prior biostatistical training and those with advanced instruction through a master's or PhD degree scored better than their counterparts. More senior residents performed worse than junior residents, potentially reflecting loss of knowledge over time, lack of reinforcement, or both. Although fourth-year residents were an exception to this pattern, these residents were part of a single medicine/pediatrics program that outperformed all other training programs. The higher scores in university-based residency programs may reflect exposure to faculty with more biostatistical training or teaching experience. In a survey study, community faculty considered EBM less important, were less confident in their EBM knowledge, and demonstrated poorer EBM skills than full-time faculty.[22]

Although sex was associated with a difference in scores, this finding is not supported by other literature. Studies of evidence-based practice knowledge and skills rarely report analyses by sex. In 2 studies, investigators found no sex differences in critical appraisal skills among family physicians[23] or in use of online evidence databases among public health practitioners.[24] Six studies assessing the biostatistics and epidemiology knowledge of physicians and trainees did not conduct comparisons by sex.[5-7,25-27] Furthermore, our result was not a confirmation of an a priori hypothesis and so should be interpreted with caution.

Our final regression model found 5 predictors of knowledge scores: advanced degrees, academic affiliation, prior biostatistics training, sex, and years since medical school graduation. The proportion of explained variation for the model was small, with $R^2 = 0.18$. This likely reflects in part the low variance in resident scores.

Our results suggest the need for more effective training in biostatistics in residency education. Such training has proven difficult, with systematic reviews showing only limited effectiveness of many journal clubs and EBM curricula.[14,28-32] Thus, it is not surprising that prior EBM experience, which in the past has not included biostatistics training,[11,14] was not associated with higher scores in our multivariable analysis. Interactive, self-directed, and clinically instructional strategies seem to stand the best chance of success.[33] Involvement in hypothesis-driven research during training that requires comprehensive reading of the literature may also enhance residents' knowledge and understanding.[34]

Faculty who are implementing biostatistics curricula can access several teaching resources. In internal medicine, the American College of Physicians' *ACP Journal Club* has presented a series of reports emphasizing basic study designs and statistics.[35] *CMAJ* has published a series of EBM "teaching tips" for learners and teachers.[36] A guide designed to help medical educators choose and interpret statistical tests when developing educational studies or when reading the medical literature is also available.[37]

Limitations of this study should be considered. First, while our instrument showed good content validity, internal consistency, and discriminative validity, these psychometric properties were not known in advance but were established in the current study. Second, our survey was purposely kept brief, thus limiting our ability to assess understanding of all biostatistical concepts and research results. Nonetheless, our questions focused on the most commonly used methods and results found in the contemporary literature. Third, we attempted to survey only those residents who were present at the time of their inpatient conference. Residents who did not attend, either by choice or by chance, might have scored differently. However, since we found no differences in demographic characteristics between responders and nonresponders, this is less likely. Fourth, our study was confined to internal medicine residents, limiting generalizability to other resident physicians. Nevertheless, we were able to assess multiple types of internal medicine training programs and found similar results.

Despite these limitations, this study also has several strengths. First, it was a multiprogram study that captured information on a wide range of internal medicine residents at different types of residency programs. Second, the residents in our survey, although limited to 1 state, possessed characteristics similar to all other trainees in internal medicine programs across the United States. Third, the 11 residency programs were similar in size and composition to the average US internal medicine program, and thus our study appears to be generalizable to internal medicine trainees and training programs in the United States.

Higher levels of statistical methods are being used in contemporary medical literature, but basic concepts, frequently occurring tests, and interpretation of results are not well understood by resident physicians. This inadequate preparation demonstrates lack of competence in meeting part of the Accreditation Council for Graduate Medical Education's practice-based learning and improvement requirement.[10] If physicians cannot detect appropriate statistical analyses and accurately understand their results, the risk of incorrect interpretation may lead to erroneous applications of clinical research. Educators should reevaluate how this information is taught and reinforced in order to adequately prepare trainees for lifelong learning, and further research should examine the effectiveness of specific educational interventions.

Author Contributions: Dr Windish had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.
Study concept and design: Windish, Huot, Green.
Acquisition of data: Windish.
Analysis and interpretation of data: Windish, Green.
Drafting of the manuscript: Windish, Green.
Critical revision of the manuscript for important intellectual content: Windish, Huot, Green.
Statistical analysis: Windish, Green.
Administrative, technical, or material support: Huot.
Study supervision: Green.
Financial Disclosures: None reported.

REFERENCES

1. McColl A, Smith H, White P, Field J. General practitioner's perceptions of the route to evidence based medicine: a questionnaire survey. *BMJ*. 1998;316(7128):361-365.
2. Young JM, Ward JE. Evidence-based medicine in general practice: beliefs and barriers among Australian GPs. *J Eval Clin Pract*. 2001;7(2):201-210.
3. Putnam W, Twohig PL, Burge FI, Jackson LA, Cox JL. A qualitative study of evidence in primary care: what the practitioners are saying. *CMAJ*. 2002;166(12):1525-1530.
4. Haynes RB. Of studies, syntheses, synopses, summaries, and systems: the "5S" evolution of information services for evidence-based healthcare decisions. *Evid Based Med*. 2006;11(6):162-164.
5. Berwick DM, Fineberg HV, Weinstein MC. When doctors meet numbers. *Am J Med*. 1981;71(6):991-998.
6. Weiss ST, Samet JM. An assessment of physician knowledge of epidemiology and biostatistics. *J Med Educ*. 1980;55(8):692-697.
7. Wulff HR, Andersen B, Brandenhoff P, Guttler F. What do doctors know about statistics? *Stat Med*. 1987;6(1):3-10.
8. Horton NJ, Switzer SS. Statistical methods in the journal. *N Engl J Med*. 2005;353(18):1977-1979.
9. Association of American Medical Colleges (AAMC). Curriculum directory. http://services.aamc.org/currdir/section4/start.cfm. Accessed April 14, 2007.
10. Accreditation Council for Graduate Medical Education (ACGME). Outcome project: enhancing residency education through outcomes assessment. http://www.acgme.org/Outcome/. Accessed January 12, 2007.
11. Green ML. Evidence-based medicine training in internal medicine residency programs: a national survey. *J Gen Intern Med*. 2000;15(2):129-133.
12. Dellavalle RP, Stegner DL, Deas AM, et al. Assessing evidence-based dermatology and evidence-based internal medicine curricula in US residency training programs: a national survey. *Arch Dermatol*. 2003;139(3):369-372.
13. Alguire PC. A review of journal clubs in postgraduate medical education. *J Gen Intern Med*. 1998;13(5):347-353.
14. Green ML. Graduate medical education training in clinical epidemiology, critical appraisal, and evidence-based medicine: a critical review of curricula. *Acad Med*. 1999;74(6):686-694.
15. Shaneyfelt T, Baum KD, Bell D, et al. Instruments for evaluating education in evidence-based practice: a systematic review. *JAMA*. 2006;296(9):1116-1127.
16. Assessment Resource Tools for Improving Statistical Thinking (ARTIST) Web site. https://ore.gen.umn.edu/artist/index.html. Accessed January 9, 2007.
17. Department of Biostatistics; Johns Hopkins Bloomberg School of Public Health. Course materials from Statistical Methods in Public Health II and III, 2003-2004 academic year. http://www.biostat.jhsph.edu/courses/bio622/index.html. Accessibility verified August 2, 2007.
18. American College of Physicians. Residency database search. http://www.acponline.org/residency/index.html?idxt. Accessed April 20, 2007.
19. American Medical Association. FREIDA online specialty statistics training statistics information. http://www.ama-assn.org/vapp/freida/spcstsc/0,1238,140,00.html. Accessed April 20, 2007.
20. American Medical Association. State-level data for accredited graduate medical education programs in the US: aggregate statistics on all resident physicians actively enrolled in graduate medical education during 2005-2006. http://www.ama-assn.org/ama/pub/category/3991.html#4. Accessed April 20, 2007.
21. Looney SW, Grady CS, Steiner RP. An update on biostatistics requirements in U.S. medical schools. *Acad Med*. 1998;73(1):92-94.
22. Beasley BW, Woolley DC. Evidence-based medicine knowledge, attitudes, and skills of community faculty. *J Gen Intern Med*. 2002;17(8):632-639.
23. Godwin M, Seguin R. Critical appraisal skills of family physicians in Ontario, Canada. *BMC Med Educ*. 2003;3:10.
24. Adily A, Westbrook J, Coiera E, Ward J. Use of on-line evidence databases by Australian public health practitioners. *Med Inform Internet Med*. 2004;29(2):127-136.
25. Estellat C, Faisy C, Colombet I, Chatellier G, Burnand B, Durieux P. French academic physicians had a poor knowledge of terms used in clinical epidemiology. *J Clin Epidemiol*. 2006;59(9):1009-1014.
26. Ambrosius WT, Manatunga AK. Intensive short courses in biostatistics for fellows and physicians. *Stat Med*. 2002;21(18):2739-2756.
27. Cheatham ML. A structured curriculum for improved resident education in statistics. *Am Surg*. 2000;66(6):585-588.
28. Coomarasamy A, Khan KS. What is the evidence that postgraduate teaching in evidence based medicine changes anything? a systematic review. *BMJ*. 2004;329(7473):1017.
29. Ebbert JO, Montori VM, Schultz HJ. The journal club in postgraduate medical education: a systematic review. *Med Teach*. 2001;23(5):455-461.
30. Taylor R, Reeves B, Ewings P, Binns S, Keast J, Mears R. A systematic review of the effectiveness of critical appraisal skills training for clinicians. *Med Educ*. 2000;34(2):120-125.
31. Parkes J, Hyde C, Deeks J, Milne R. Teaching critical appraisal skills in health care settings. *Cochrane Database Syst Rev*. 2001;(3):CD001270.
32. Norman GR, Shannon SI. Effectiveness of instruction in critical appraisal (evidence-based medicine) skills: a critical appraisal. *CMAJ*. 1998;158(2):177-181.
33. Khan KS, Coomarasamy A. A hierarchy of effective teaching and learning to acquire competence in evidenced-based medicine. *BMC Med Educ*. 2006;6:59.
34. Rogers LF. The "win-win" of research. *AJR Am J Roentgenol*. 1999;172(4):877.
35. ACP Journal Club. Evidence-based medicine for better patient care. http://www.acpjc.org/?hp. Accessed November 5, 2006.
36. Wyer PC, Keitz S, Hatala R, et al. Tips for learning and teaching evidence-based medicine: introduction to the series. *CMAJ*. 2004;171(4):347-348.
37. Windish DM, Diener-West M. A clinician-educator's roadmap to choosing and interpreting statistical tests. *J Gen Intern Med*. 2006;21(6):656-660.

Biostatistical Knowledge Test Survey Instrument

Please answer the following questions:

Gender: ☐ Male ☐ Female

Age:
☐ 21-25 ☐ 26-30 ☐ 31-35 ☐ 36-40 ☐ 41-45 ☐ ≥ 46

Advanced Degrees:
☐ MD ☐ DO ☐ PhD ☐ MPH/MHS ☐ MSc
☐ Other ______________

Years since medical school graduation:
☐ <1 ☐ 1-3 ☐ 4-10 ☐ 11-20 ☐ ≥ 21

Ever trained outside of the United States? ☐ Yes ☐ No
If yes, at what level(s)?
☐ College ☐ Medical school ☐ Residency
☐ Other ______________

Current level of training:
☐ Intern
☐ 2nd-year Resident
☐ 3rd-year Resident
☐ 4th-year Resident
☐ Chief Resident
☐ Fellowship/Scholarship Program
☐ Faculty

Residency training program type:
☐ Traditional/Categorical Medicine
☐ Primary Care
☐ Medicine/Pediatrics
☐ Preventive Medicine
☐ IM/Preventive Medicine
☐ Preliminary/Transitional

Ever taken a course in epidemiology? ☐ Yes ☐ No
If yes, where? ☐ College ☐ Medical school ☐ Residency
☐ Other ______________

Ever taken a course in biostatistics? ☐ Yes ☐ No
If yes, where? ☐ College ☐ Medical school ☐ Residency
☐ Other ______________

Ever had training in evidence-based medicine (EBM)?
☐ Yes ☐ No
If yes, where? ☐ College ☐ Medical school ☐ Residency
☐ Other ______________

Which of the following journals do you read regularly?
Check all that apply.
☐ I do not regularly read journals
☐ *JAMA*
☐ *New England Journal of Medicine*
☐ *The American Journal of Medicine*
☐ *Annals of Internal Medicine*
☐ *Lancet*
☐ *BMJ*
☐ *Mayo Clinic Proceedings*
☐ *ACP Journal Club*
☐ Other ______________

Please answer the following questions regarding statistics:

1. Given the chance, I would like to learn more about biostatistics.

1	2	3	4	5
strongly disagree		neutral		strongly agree

2. I can understand almost all of the statistical terms that I encounter in journal articles.

1	2	3	4	5
strongly disagree		neutral		strongly agree

3. Because it is easy to lie with statistics, I don't trust them at all.

1	2	3	4	5
strongly disagree		neutral		strongly agree

4. I often use statistical information in forming opinions or making decisions in medical care.

1	2	3	4	5
strongly disagree		neutral		strongly agree

5. To be an intelligent reader of the literature, it is necessary to know something about statistics.

1	2	3	4	5
strongly disagree		neutral		strongly agree

Please rate your confidence in your current level of ability in the following activities:

6. Interpreting the *P* value for a given result.

1	2	3	4	5
none	a little	a fair amount	a lot	complete confidence

7. Interpreting the results of a statistical method used in research.

1	2	3	4	5
none	a little	a fair amount	a lot	complete confidence

8. Assessing if the correct statistical procedure was used to answer a research question.

1	2	3	4	5
none	a little	a fair amount	a lot	complete confidence

9. Identifying the factors that influence a study's power.

1	2	3	4	5
none	a little	a fair amount	a lot	complete confidence

Please choose the best answer to each of the following questions:

1. A study wishes to assess birth characteristics in a population. Which of the following variables describes the appropriate measurement scale or type?
(Fill in each blank below with your answer. Use each letter as many times as is appropriate.)
A. discrete
B. continuous
C. ordinal
D. nominal
E. dichotomous

a. ______ Birthweight in grams
b. ______ Birthweight classified as low, medium, high
c. ______ Type of delivery classified as cesarean, natural, induced

2. To determine if fasting is associated with dengue fever, data from 40 patients with dengue fever were collected. These patients were matched for age, sex, and race to 40 patients without dengue fever. The hospital charts of these patients were then reviewed to determine whether they also fasted prior to their illness. This study type is known as:
a. Cross-sectional study
b. Concurrent cohort study
c. Case-control study
d. Retrospective cohort study
e. Randomized clinical trial

3. The purpose of a double-blind or double-masked study is to:
a. Achieve comparability of treated and untreated subjects
b. Reduce the effects of sampling variation
c. Avoid observer and subject bias
d. Avoid observer bias and sampling variation

4. A prospective study looked at obesity, diet, and exercise habits of individuals. Match the appropriate analytic method for each of the following hypotheses.
(Fill in each blank below with your answer. Use each letter as many times as is appropriate.)
A. T-test for comparing 2 population means
B. Analysis of Variance (ANOVA)
C. Correlation analysis
D. Chi-square test of homogeneity
E. Logistic regression

a. ______ Mean age does not vary across 4 groups of fat consumption.
b. ______ Multivitamin use does not vary across the 4 groups of fat consumption.
c. ______ Mean BMI is the same for the low fat and high fat consumption group.

5. Any systematic error in the design, conduct, or analysis of a study that results in a mistaken estimate of an exposure's effect on the risk of disease is called:
a. Confounding
b. Bias
c. Interaction
d. Stratification

6. In a placebo-controlled trial of the use of aspirin and dipyridamole to prevent arterial restenosis after coronary angioplasty, 38% of patients receiving the treatment had restenosis, and 39% of patients receiving placebo had restenosis. In reporting this finding, the authors stated that $P > .05$. This means:
a. The chances are greater than 1 in 20 that a difference would be found again if the study were repeated.
b. The probability is less than 1 in 20 that a difference this large could occur by chance alone.
c. The probability is greater than 1 in 20 that a difference this large could occur by chance alone.
d. The chance is 95% that the study is correct.

7. In the same aspirin vs dipyridamole study, the researchers wished to assess if there were any differences between groups over time with respect to the primary end point of restenosis while controlling for other potential risk factors. What analytic method would be most appropriate in assessing their question?
a. Kaplan-Meier analysis
b. Logistic regression
c. Linear regression
d. Cox proportional hazard regression
e. Chi-square test of homogeneity

8. In a research study, the age of the participants was 26 years ± 5 years (mean ± standard deviation). Which of the following statements is the most correct?
a. It is 95% certain that the true mean lies within the interval of 16-36 years.
b. Most of the patients were aged 26 years; the remainder were aged between 21 and 31 years.
c. Approximately 95% of the patients were aged between 16 years and 36 years.
d. No patients were younger than age 16 or older than age 36.

9. Researchers measure cholesterol levels in a sample of patients in New Zealand and Asia and find the following results:

Region	Sample Size	Mean Cholesterol Level (mmol/L)	Standard Deviation (mmol/L)
New Zealand	100	5.4	1.2
Asia	150	4.9	1.3

They calculate the mean and 95% CI for the true difference in mean cholesterol levels between the 2 populations and find: Mean Difference 0.5 mmol/L and 95% CI (0.18-0.82).

The 95% confidence interval for the true difference in mean cholesterol levels between 2 populations suggests that:
a. There is no statistically significant difference in mean cholesterol levels between the 2 populations.
b. There is a statistically significant higher mean cholesterol level in the Asian population as compared to the New Zealand population.
c. There is a statistically significant higher mean cholesterol level in the New Zealand sample as compared to the Asian sample.
d. There is a statistically significant higher mean cholesterol level in the New Zealand population as compared to the Asian population.

10. Researchers designed a study looking at cardiovascular deaths comparing a new drug to placebo. They determined that they would need 200 patients in each group to detect a 15% difference in cardiovascular end points given 90% power and a significance level of .01.
Which of the following changes would require the researchers to increase their sample size?
  a. Aim to detect a difference of 20%.
  b. Specify a power of 80%.
  c. Use a significance level of .05.
  d. Aim to detect a difference of 10%.

11. In a diabetes detection program, the cut-off level of blood sugar for Test A is set at 130 mg/100 mL and for Test B at 160mg/100 mL. This means:
  a. The sensitivity of Test B is greater than that of Test A.
  b. The specificity of Test B is greater than that of Test A.
  c. The sensitivity and specificity are the same for both tests.
  d. The number of false positives is greater with Test B than with Test A.

12. The Third National Health and Nutrition Examination Survey was conducted in the United States in the 1990s to examine the relationship between obesity and depression. The authors investigated the association between major depression and body mass index (BMI) for males and females. (*American Journal of Epidemiology*. 2003;158:1139-1147.)

Table. Unadjusted Odds Ratios of Major Depression During the Past Month

BMI Category, kg/m^2	Unadjusted Odds Ratio	95% Confidence Interval
Normal weight (BMI 18.5-24.9)	1.00	
Underweight (BMI < 18.5)	1.17	0.49-2.80
Overweight (BMI 25.0-29.9)	0.86	0.53-1.41
Obese (BMI ≥ 30)	1.88	1.02-3.46
Class 1 (BMI 30-34.9)	1.28	0.64-2.56
Class 2 (BMI 35-39.9)	1.76	0.78-3.95
Class 3 (BMI ≥ 40)	4.98	2.07-11.99

From the table above, what is the correct interpretation of the overweight values of 0.86?
  a. Overweight individuals' odds of having major depression are 14% lower than the odds of having major depression for individuals with normal weight.
  b. Overweight individuals' odds of having major depression are 14% higher than the odds of having major depression for individuals with normal weight.
  c. An overweight individual has a 0.86 probability of having major depression.
  d. An overweight individual has a 0.86 odds of having major depression.

13. The authors in the previous study also looked at the adjusted odds ratios for major depression during the previous month with the same population and found the following results:

Table. Adjusted Odds Ratios of Major Depression During the Past Month

	Adjusted Odds Ratio	95% Confidence Interval
BMI Category, kg/m^2		
Normal weight (BMI 18.5-24.9)	1.00	
Underweight (BMI < 18.5)	1.13	0.43-3.01
Overweight (BMI 25.0-29.9)	0.96	0.57-1.64
Obese (BMI ≥ 30)	1.84	0.95-3.55
Class 1 (BMI 30-34.9)	1.33	0.57-3.13
Class 2 (BMI 35-39.9)	1.90	0.79-4.60
Class 3 (BMI ≥ 40)	4.63	2.06-10.42
Gender		
Male	1.00	
Female	2.62	1.76-3.92
Age (years)		
15-19	1.00	
20-24	0.80	0.36-1.76
25-29	0.61	0.22-1.69
30-34	0.64	0.30-1.39
Race/ethnicity		
White	1.00	
African American	0.80	0.48-1.32
Hispanic/other	1.02	0.53-1.94

Based on the table above, which of the following statements is true?
  a. The odds of depression is statistically significantly increased in individuals with BMI exceeding 40 as compared to that of individuals in all other BMI categories, controlling for all other covariates.
  b. The odds of depression in individuals of normal weight is statistically significantly lower than that of individuals with BMI exceeding 40, controlling for all other covariates.
  c. The odds of depression statistically significantly decrease with increasing age, controlling for all other covariates.
  d. There is no statistically significant difference in the odds of depression between genders, controlling for all other covariates.

14. As part of the Women's Health Study, researchers wished to investigate the role of systemic inflammation in predicting cardiovascular disease in women (*N Engl J Med*. 2002;347[20]:1557-1565).

Researchers used a prospective study design with a large sample size. Blood levels of C-reactive protein (CRP) were measured at baseline and women were followed for an average of 8 years. The following table shows the relative risk estimates of a cardiovascular event (heart attack or stroke) within 5 years by quintile of CRP level for the 30,000 women. The first quintile is used as the reference category.

	Quintile of CRP Level				
	1	2	3	4	5
	0.49 mg/dL	>0.49-1.08 mg/dL	>1.08-2.09 mg/dL	>2.09-4.19 mg/dL	>4.19 mg/dL
Relative Risk	1.0	1.8	2.3	3.2	4.5
Number of Women	6000	6000	6000	6000	6000

Based on the relative risk data above, one can conclude:

a. There is no risk of heart attack/stroke for women with CRP levels in the first quintile.
b. Decreasing CRP level appears to increase the risk of heart attack/stroke.
c. Increasing CRP level appears to increase the risk of heart attack/stroke.
d. There appears to be no association between CRP levels and heart attack/stroke.

15. The National Osteoporosis Risk Assessment study evaluated 200,160 postmenopausal women aged 50 years or older in the United States. At baseline, 14,412 of these women had osteoporosis as defined by a bone mineral density T score ≤−2.5 (*JAMA*. 2001;286[22]:2815-2822).

The table on the right provides results of a multivariable logistic regression model of correlates with osteoporosis.

Other than age, the strongest association with osteoporosis is:

a. Cigarette smoking
b. Years since menopause
c. BMI
d. Estrogen use
e. Maternal history of fracture

Table 3. Multivariable Logistic Regression Model of Statistically Significant Correlates of T Score ≤−2.5

Risk Factors	Odds Ratio (95% Confidence Interval)
Age group, y	
50-54	1.00 (Referent)
55-59	1.79 (1.56-2.06)
60-64	3.84 (3.37-4.37)
65-69	5.94 (5.24-6.74)
70-74	9.54 (8.42-10.81)
75-79	14.34 (12.64-16.26)
≥80	22.56 (19.82-25.67)
Years since menopause	
≤5	1.00 (Referent)
6-10	0.79 (0.70-0.89)
11-15	0.83 (0.76-0.91)
16-20	0.96 (0.89-1.03)
21-25	1.01 (0.95-1.08)
26-30	1.02 (0.95-1.09)
31-35	1.10 (1.03-1.19)
36-40	1.14 (1.05-1.24)
≥41	1.24 (1.14-1.35)
College education or higher	0.91 (0.87-0.94)
Self-rated health status	
Excellent	1.00 (Referent)
Very good	1.04 (0.97-1.13)
Good	1.23 (1.14-1.33)
Fair/poor	1.62 (1.50-1.76)
Fracture history	
Hip	1.96 (1.75-2.20)
Wrist	1.90 (1.77-2.03)
Spine	1.34 (1.17-1.54)
Rib	1.43 (1.32-1.56)
Maternal history of osteoporosis	1.08 (1.01-1.17)
Maternal history of fracture	1.16 (1.11-1.22)
Race/ethnicity	
White	1.00 (Referent)
African American	0.55 (0.48-0.62)
Native American	0.97 (0.82-1.14)
Hispanic	1.31 (1.19-1.44)
Asian	1.56 (1.32-1.85)
Body mass index, kg/m^2	
<23	1.00 (Referent)
23.01-25.99	0.46 (0.44-0.48)
26.00-29.99	0.27 (0.26-0.28)
≥30	0.16 (0.15-0.17)
Current medication use	
Cortisone	1.63 (1.47-1.81)
Diuretics	0.81 (0.76-0.85)
Estrogen use	
Former	0.77 (0.73-0.80)
Current	0.27 (0.25-0.28)
Cigarette smoking	
Former	1.14 (1.10-1.19)
Current	1.58 (1.48-1.68)
Regular exercise	0.86 (0.82-0.89)
Alcohol use, drinks/wk	
None	1.00 (Referent)
1-6	0.85 (0.80-0.90)
7-13	0.76 (0.69-0.83)
≥14	0.62 (0.54-0.71)
Technology	
Heel single-energy x-ray absorptiometry	1.00 (Referent)
Forearm peripheral dual-energy x-ray absorptiometry	2.86 (2.75-2.99)
Finger peripheral dual-energy x-ray absorptiometry	4.86 (4.56-5.18)
Heel ultrasonography	0.79 (0.70-0.90)

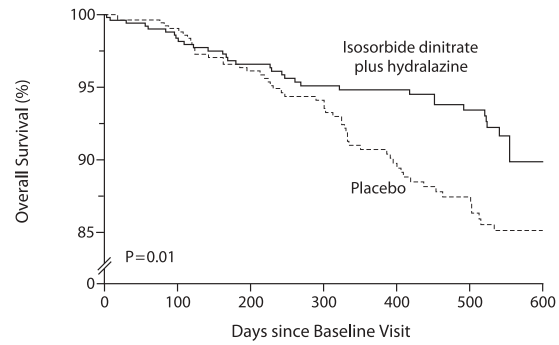
JAMA. 2001;286(22):2815-2822. © American Medical Association

16. A randomized controlled trial was conducted to determine whether the combination of isosorbide dinitrate and hydralazine together in African Americans with heart failure was superior to placebo with respect to overall survival (*N Engl J Med*. 2004;351[20]:2049-2057).

A Kaplan-Meier analysis was performed and is shown on the right.

Based on this information we can conclude:

a. The overall risk of death is statistically significantly lower in the treatment group vs placebo comparing all time points between day 0 to day 600.
b. The overall risk of death is statistically significantly lower in the treatment group vs placebo at day 600.
c. The overall risk of death is statistically significantly higher in the treatment group vs placebo at day 600.
d. The overall risk of death is approximately the same in both groups.



No. at Risk							
Placebo	532	466	401	340	285	232	24
Isosorbide dinitrate plus hydralazine	518	463	407	359	313	251	13

Figure 1. Kaplan–Meier Estimates of Overall Survival.

N Engl J Med. 2004;351(20):2049-2057. © Massachusetts Medical Society

Answers to Biostatistical Knowledge Test Questions: 1a. B, 1b. C, 1c. D, 2. C, 3. C, 4a. B, 4b. D, 4c. A, 5. B, 6. C, 7. D, 8. C, 9. D, 10. D, 11. B, 12. A, 13. B, 14. C, 15. C, 16. A

