The Trojan Lifetime Champions Health Survey: Development, Validity, and Reliability

Shawn C. Sorenson, PhD, CSCS*; Russell Romano, MA, ATC†; Robin M. Scholefield, PhD†; E. Todd Schroeder, PhD*; Stanley P. Azen, PhD‡; George J. Salem, PhD*

*Division of Biokinesiology and Physical Therapy; Departments of †Athletic Medicine and ‡Preventive Medicine, University of Southern California, Los Angeles

Context: Self-report questionnaires are an important method of evaluating lifespan health, exercise, and health-related quality of life (HRQL) outcomes among elite, competitive athletes. Few instruments, however, have undergone formal characterization of their psychometric properties within this population.

Objective: To evaluate the validity and reliability of a novel health and exercise questionnaire, the Trojan Lifetime Champions (TLC) Health Survey.

Design: Descriptive laboratory study.

Setting: A large National Collegiate Athletic Association Division I university.

Patients or Other Participants: A total of 63 university alumni (age range, 24 to 84 years), including former varsity collegiate athletes and a control group of nonathletes.

Intervention(s): Participants completed the TLC Health Survey twice at a mean interval of 23 days with randomization to the paper or electronic version of the instrument.

Main Outcome Measure(s): Content validity, feasibility of administration, test-retest reliability, parallel-form reliability between paper and electronic forms, and estimates of systematic and typical error versus differences of clinical interest were assessed across a broad range of health, exercise, and HRQL measures.

Results: Correlation coefficients, including intraclass correlation coefficients (ICCs) for continuous variables and κ agreement statistics for ordinal variables, for test-retest reliability averaged 0.86, 0.90, 0.80, and 0.74 for HRQL, lifetime health, recent health, and exercise variables, respectively. Correlation coefficients, again ICCs and κ , for parallel-form reliability (ie, equivalence) between paper and electronic versions averaged 0.90, 0.85, 0.85, and 0.81 for HRQL, lifetime health, recent health, and exercise variables, respectively. Typical measurement error was less than the a priori thresholds of clinical interest, and we found minimal evidence of systematic test-retest error. We found strong evidence of content validity, convergent construct validity with the Short-Form 12 Version 2 HRQL instrument, and feasibility of administration in an elite, competitive athletic population.

original research

Conclusions: These data suggest that the TLC Health Survey is a valid and reliable instrument for assessing lifetime and recent health, exercise, and HRQL, among elite competitive athletes. Generalizability of the instrument may be enhanced by additional, larger-scale studies in diverse populations.

Key Words: athletes, exercise, quality of life, questionnaires, reproducibility of results

Key Points

- The Trojan Lifetime Champions Health Survey is a valid and reliable instrument for assessing lifetime and recent health, exercise, and health-related quality of life among university alumni, including National Collegiate Athletic Association Division I athletes.
- The survey is distinguished from previous questionnaires used to evaluate lifetime health and exercise in elite athletes by the formal characterization of its psychometric properties.
- Larger-scale studies in diverse populations are needed to enhance the generalizability of the instrument, and the electronic form of this survey is a suitable and feasible means.

P articipation in competitive sports presents unique and important health considerations. Elite athletes are celebrated for their extraordinary physical achievements and possess superior physiologic characteristics that are associated positively with health (eg, cardiorespiratory fitness, strength, and power) compared with the general population.^{1,2} However, competitive sports also are recognized as a potential health risk. Concern for athlete health and safety has led to recent public scrutiny of the long-term consequences of sport participation, notably regarding orthopaedic injury,³ cardiovascular disease,⁴ head injury,⁵ and related psychosocial effects.⁶ Consideration of these risks has been substantial enough to prompt formal inquiries by the US Congress^{7,8} and legal action⁹ and to drive extensive changes to health policy.¹⁰ A sense of responsibility is growing among institutions (eg, colleges and universities) and organizations (eg, National Collegiate Athletic Association [NCAA] and professional sports leagues) to understand and promote athletes' health. Despite this attention, data on health and exercise outcomes across the lifespan of elite athletes remain sparse, and a critical need for additional research in this area exists.

Table 1.	Trojan Lifetime	Champions	Health Survey	Content

Survey Section	Description	Scoring	Items, No.
I: Demographics	Age, sex, ethnicity, height, and mass	Open response (age, height, mass), 2- and 6-level multiple choice (sex, ethnicity)	6
II: Experience in Competitive Sports	Ages of participation, sport or sports, professional experience, currently in season or out of season, and qualitative perception of sports' influence on health	Open response (age, professional experience), 2-, 3-, 5-, and 12-level multiple choice (sport, in season/out of season, and qualitative perceptions)	8
III: Health Assessment	Lifetime history (59 items), recent history (59 items), and age of first symptoms (59 items) across 59 different health concerns in 6 domains (Joints, Bone & Muscle, Cardiopulmonary, Neurological, Other Clinical, and Psychosocial)	0–3 For each item (lifetime and recent history), 0–39 (summary domain scores), open response (age of first symptoms)	177
IV: Health-Related Quality of Life	Short-Form 12 Version 2 Health Survey	3- and 5-Level multiple choice, 0–100 (summary scores)	12
V: Current Exercise Behavior	Previous week's frequency and average duration of cardiovascular, resistance, and mixed exercise	Open response	7
Health & Exercise Attitudes	Perceptions about exercise importance, body image, relation of physical and mental health, and health care	5- and 6-Level multiple choice	6
Total			216

Self-report questionnaires are perhaps the most practical way of gathering such data among large, diverse populations. These instruments have been used to describe health and exercise outcomes among select groups of elite athletes.^{11–16} However, health, exercise, and health-related quality of life (HRQL) rarely have been comprehensively addressed together. An additional limitation is that most of these instruments have not undergone formal characterization of their psychometric properties in an elite, competitive athletic population. Assessing the validity, test-retest reliability, and precision/error of any instrument in relation to clinically meaningful effects is critical to establishing confidence in study results and their implications for health care and policy.¹⁷

As part of the collaborative research and education program Trojan Lifetime Champions (TLC), our research team developed a comprehensive instrument to measure lifetime and recent health, exercise, and HRQL among current and former university students, including collegiate athletes. Our ultimate aim is to use the questionnaire to better understand the unique influence of elite competitive sports on lifespan health and well-being. Therefore, the purpose of our study was to evaluate the validity and reliability of a novel health and exercise questionnaire: the TLC Health Survey. We describe the development and formal psychometric assessment of that instrument, with the goal of providing an accumulation of evidence to support its use.¹⁷

METHODS

The TLC Health Survey

Overview. The TLC Health Survey includes 216 unique items in 5 sections (Demographics, Experience in Competitive Sports, Health Assessment, Health-Related Quality of Life, and Current Exercise Behavior / Health & Exercise Attitudes) as presented in Table 1. It incorporates both an established instrument and novel measures of

interest. The survey initially was a paper questionnaire, with a Web-based electronic version developed in parallel. The electronic survey contained identical content presented in a series of Web pages that matched the pages of the paper survey and was administered via a dedicated Web site. The TLC Health Survey is available online (see Supplemental Material found at http://dx.doi.org/10.4085/1062-6050-50.2.10.S1).

Development Process. We sought an instrument to provide comprehensive data on lifetime and recent holistic health, exercise behavior, and HRQL among a diverse population of current and former University of Southern California (USC) students, including elite, competitive athletes. After conducting a literature review and a series of informal, open-ended interviews with former athletes and consulting with knowledge experts, we determined that no existing validated instrument provided a suitable means to do so. In particular, health assessments were focused too narrowly on specific regions (eg, knee) or conditions (eg, osteoarthritis), and exercise instruments did not capture important features of exercise behavior (eg, resistance training). Therefore, we developed a novel questionnaire for our study.

Development of the TLC Health Survey was a collaborative process involving an interdisciplinary team of health and exercise scientists and diverse experts affiliated with the USC Department of Athletics (Table 2). Through a series of group meetings, 6 major survey revisions, and 4 pilot administrations in different populations, we built an instrument that the study team deemed to possess appropriate content validity and suitability for large-scale administration. This development process is detailed in Table 3.

At each stage of the development process, we revised the survey to include comprehensive content, improve clarity and usability for the survey taker, and minimize opportunities for bias. We incorporated color-coded instructions and examples, along with careful language and the specific criteria used to evaluate the prevalence and severity of

Table 2. Trojan Lifetime Champions Health Survey Development Team

Team Member	Qualifications	Related Experience, y
Primary authors ^a		
Shawn C. Sorenson, PhD, CSCS	Health and exercise scientist	10
Russell Romano, MA, ATC	USC head athletic trainer	22
Robin M. Scholefield, PhD	USC clinical sports psychologist	20
E. Todd Schroeder, PhD	Exercise physiologist	18
Stanley P. Azen, PhD	Biostatistician and epidemiologist	50
George J. Salem, PhD	Exercise and aging biomechanist	35
Contributing advisors ^b		
USC Athletics Health & Wellness Committee	Interdisciplinary panel including representatives from athletic administration, academic support, athlete development, strength and conditioning, athletic medicine, counseling, coaching, and athlete leadership	>100
USC Athletics coaches	Coaching staff representing each of the 20 varsity athletic teams at USC	>100
USC Athletics Student-Athlete Advisory Council	Athletes representing each of the 20 varsity athletic teams at USC	>100
Additional advisors	USC epidemiologist, athletic director, athletics dietitian, physical therapist, team physician, and athletic training staff	>100

Abbreviation: USC, University of Southern California.

^a Primary authors were responsible for the authorship, revision, and approval of the survey instrument.

^b Contributing advisors reviewed, provided feedback, and suggested edits to the survey.

various health concerns. A standardized set of guidelines was developed for interpretating paper-survey responses. These guidelines provided specific rules that ensured consistency of data entry when responses were ambiguous, contradictory, or missing. For example, if age was reported as a decimal or fraction (eg, 19.5), it was rounded down to the nearest integer (eg, 19). Interrater reliability was assessed by repeated entry of a random 10% sample of surveys, with excellent (99.96%) repeatability. The complete set of guidelines is available online (see additional Supplemental Material found at http://dx.doi.org/10.4085/1062-6050-50.2.10.S2).

Feedback from the interdisciplinary study team, and particularly the expertise of a licensed clinical psychologist (R.M.S.), was essential to developing a user-friendly instrument. The collective expertise of this group, their knowledge and experience with an elite athletic population, and systematic documentation of the development process provided substantial evidence for content validity.^{17,18}

Section I: Demographics. Demographic information, consisting of age, sex, ethnicity, height, and mass, was recorded in a series of straightforward multiple-choice and short-response questions typical to epidemiologic instruments.

Section II: Experience in Competitive Sports. History of participation in Division I collegiate athletics, consisting of ages at times of participation, sport or sports played, and in-season status, was recorded in a series of yes/no, multiple choice, and short-response questions. In addition, self-rated perception of the influence of competitive sports on overall health and the record of any postcollegiate professional sports participation were documented.

Section III: Health Assessment. A lifetime and recent health inventory was developed uniquely for this instrument.

	Table 3.	Trojan Lifetime Cham	pions Health Survey	Development Tir	meline, Major Revisio	ns, and Pilot Studies
--	----------	----------------------	---------------------	-----------------	-----------------------	-----------------------

Year	Month	Survey Revision	Review Process and Major Revisions	Pilot Study
2007	July	1	Former athlete interviews, literature review, content and item development, and first draft by survey development team	NA
	August		Review by health and exercise science research experts	Pilot 1 (5 graduate students)
	September		Review by health and exercise science research experts	NA
	November	2	Introduction of Section III: Health Assessment matrix and detailed instructions	NA
	December		Presentation to USC athletic administration	NA
2008	January	3	Text edits	NA
	February		Review by USC Athletics Health & Wellness Committee, coaches, and staff	NA
	March			Pilot 2 (4 graduate students)
	April	4	Section II: USC sport list, new Section III: Health Assessment items, Section V: Health & Exercise Attitudes perceptions	NA
	May		Review by USC Athletics Student-Athlete Advisory Council	Pilot 3 (13 athletes)
	June		Review by survey development team	NA
	July	5	Added matrix and intensity definitions for Section V: Exercise Behavior; added new Section III: Health Assessment items; and incorporated simplified, color-coded instructions and examples	NA
	August			Pilot 4 (23 athletes)
	September	6	Final formatting, approval by survey development team	NA

Abbreviations: NA, not applicable; USC, University of Southern California.

JOINTS	Lifetime Concerns	Age of First Symptoms	Recent Concerns (Last Three Years)
Ankle			
Knee			
Нір			
Lower Back			
Upper Back			
Neck			
Shoulder			
Elbow			
Wrist			
Fingers & Toes			
Jaw / TMJ			

	Lif	etime Concerns		<u>Age of</u> First Symptoms		<u>Recent Concerns</u>
	Use these cate at any	gories to describe concerns y time in your life :		How old were		Have you had a concern of this kind in the last three years ?
0 1 2	No Concerns Some Concerns	Never had any concerns. Some pain or discomfort; not serious enough to seek attention from a professional.		first experienced symptoms of this kind?		(Use the same categories as Lifetime Experience)
3	Concerns Major Concerns	doctor, physical therapist, or other medical professional. Surgery or hospital stay.	[If you have experienced mult report the earliest age you h extent of any concern	ltiple conc ad any co	cerns of the same kind, please ncern, and the most severe

Figure 1. Trojan Lifetime Champions Health Survey: health inventory matrix, Joint domain sample. Abbreviation: TMJ, temporomandibular joint.

The Section III health inventory assessed the prevalence of health concerns across 6 domains: Joints, Bone & Muscle, Cardiopulmonary, Neurological, Other Clinical, and Psychosocial. Other Clinical included items that did not clearly fall into other domains (eg, liver, kidney, and lymphatic system). A total of 59 unique health items were identified (Joints = 11, Bone & Muscle = 9, Cardiopulmonary = 8, Neurological = 9, Other Clinical =9, and Psychosocial = 13). The health inventory was structured as a set of matrixed tables, with a table for each health domain. The Joints domain table is provided as an example in Figure 1. For each item in the table, respondents described lifetime and recent (3 years before the study) health concerns. We quantified concerns using a 4-point ordinal scale calibrated to the degree of professional treatment, as well as the age at which individuals first experienced symptoms. Choices were No Concerns, Some Concerns (without attention from a professional), Serious Concerns (treated by a medical professional), and Major Concerns (with surgery or hospital stay). Concerns were defined with respect to a specific body part (eg, knee) and without discrimination of specific pathologic conditions (eg, ligament versus cartilage injury). We adopted this structure to establish straightforward, patient-centered quantitative criteria for describing health concerns, thereby simplifying the analysis while reducing the opportunities for response bias and outcome misclassification.

Summary scores for each domain of the health inventory (domain scores) were calculated by summing individual item scores from the domain and were computed for both lifetime and recent experience. Lifetime domain scores reflect cumulative lifetime experience with health concerns in a given domain, and recent domain scores reflect cumulative experience for the 3 years before the survey. A summary domain score of 0 indicates *perfect health* in that domain, and progressively higher scores indicate greater evidence of concerns. Maximal possible scores ranged from 24 for the 8-item Cardiopulmonary domain to 39 for the 13-item Psychosocial domain.

Section IV: Health-Related Quality of Life. The Short-Form 12 Version 2 Health Survey (SF-12v2) was adapted¹⁹ for Section IV (HRQL) and is referred to as SF-12 throughout this document. The TLC Health Survey used identical text, but we reformatted fonts and layout to match the remainder of the questionnaire. The SF-12 and its parent instrument (the Short Form-36) have been widely used to measure HRQL across diverse populations, including collegiate^{15,20-22} and former professional athletes.²³ The inclusion of an established, validated, and recognized instrument, such as the SF-12, enhances content validity.^{17,18} The SF-12 Physical Component scores (PCS) and Mental Component scores (MCS) were computed from a proprietary algorithm for summary physical and mental health.¹⁹ A score of 50 reflects the approximate population mean for all US adults, with higher scores indicating better HRQL.

Section V: Current Exercise Behavior / Health & Exercise Attitudes. Exercise behavior questions were designed to mirror weekly exercise guidelines for healthy adults as published by the American College of Sports Medicine (ACSM).²⁴ We were unaware of a previously validated instrument specifically designed to measure acute exercise behavior, including resistance exercise, relative to these guidelines. Specifically, the TLC Health Survey

Table 4. Demographic Characteristics of Participants in the Trojan Lifetime Champions Phase 2 Validity and Reliability Study, 2010–2011, Compared With the University of Southern California Varsity Athlete Alumni Source Population

		Reliabilit	y Population ^b
Characteristic	Source Population ^a	Athletes (n = 33)	Nonathletes (n = 30)
Age, y, mean ± SD	46.0	45.3 ± 16.2	47.4 ± 15.7
Age, y, range	22–72	24–77	24–84
Men, %	73	70	67
Women, %	27	30	33

^a Estimated from roster sizes over the 50 years before the study using media guides for each current varsity sport.

^b Reliability population included study participants completing both test and retest surveys.

measures the previous week's frequency and average duration of exercise across 3 activity subtypes: cardiovascular, resistance, and mixed. Cardiovascular exercise included activities such as brisk walking, hiking, running, swimming, using an elliptical, and cycling. Resistance exercise included weight lifting with or without machines. Mixed exercise encompassed all other exercise activities, including sports (eg, basketball, soccer, volleyball) and heavy yard work. A minimal exercise threshold of moderate intensity, which was defined as "working hard enough to raise your heart rate and break a sweat, yet still being able to carry on a conversation," was specified for all forms of exercise. Exercise volume (in minutes) from the previous week was calculated for each form of exercise as the reported number of sessions multiplied by the average session duration. Total exercise volume from the previous week was calculated as the sum of exercise volumes across the 3 exercise subtypes. Given our definition of the minimal exercise threshold of moderate intensity, we reasoned that all reported exercise (cardiovascular, resistance, and mixed) contributes to target cardiorespiratory exercise training volume as specified by the ACSM guidelines.²⁴ Self-reported exercise importance was measured using a multiple-choice item with a 4-point ordinal scale. The question read, "In general, how important would you say exercise is in your life?" Response choices were Very Important, Important, Somewhat Important, and Not Important. Several multiple-choice questions regarding health-and-wellness perceptions and exercise attitudes also were incorporated into Section V.

Administration of the TLC Health Survey

Survey administration took a phased approach, allowing for progressive testing and in-the-field validation of the instrument before full-scale implementation in larger study populations. Phase 1 (September 2008 to November 2011) comprised administration of the survey to 20 USC varsity athletic teams and 3 USC undergraduate classes, including students who were not varsity athletes. Briefly, surveys were collected in dedicated team or class meetings wherein prospective participants were informed of the content, goals, risks, and benefits of the study and were provided with an invitation to participate. The same investigator (S.C.S.) provided identical instructions to each group. Meetings lasted from 20 to 30 minutes. All participants completed surveys in approximately 10 to 20 minutes and then received a \$5 coffeehouse gift card, as approved by the Office of Athletic Compliance. A total of 444 surveys were distributed, and 423 (95%) were completed. This provides evidence of feasibility for efficient administration of the survey among current university students, including athletes, and preliminary data for subsequent outcomes studies.

Phase 2 (April 2010 to September 2011), which is reported in this article, was designed to formally assess the psychometric properties of the TLC Health Survey, including test-retest reliability and parallel-form reliability between the paper and electronic versions, among a group of USC alumni who were representative of the larger alumni source population. Phase 2 used the same version of the TLC Health Survey as in phase 1. Assessing the psychometric properties of the instrument provides a foundation for larger-scale studies (phase 3) of all USC alumni or other populations of interest.

Study Population and Recruitment

We divided USC alumni into 2 subgroups: athlete alumni and nonathlete alumni. To be eligible for participation, athlete alumni must have practiced or competed in Division I athletics at USC. Nonathletes were former undergraduate students at USC who never practiced or competed in Division I athletics at any university.

Prospective study participants were recruited through university records, student and alumni organizations, and personal referrals. Study participants were selected to provide a convenience sample representative of the USC athlete alumni source population and an age- and sexmatched control group. For each current varsity sport, we examined media guides to estimate roster sizes over the previous 50 years and defined accordingly the demographic characteristics of the source population, including sex, age, and sport. An estimated 4500 USC varsity athlete alumni were identified. Characteristics of the source population are presented in Tables 4 and 5. Athlete alumni were targeted to proportionally match these demographic characteristics.

Nonathlete alumni were recruited to mirror the athlete alumni based on age (primarily) and sex (secondarily). Nonathletes were excluded from participation if they indicated any experience in varsity or organized club sports during college. Those with experience in high school or collegiate intramural sports were allowed to participate. Control participants who were matched only on age (but not sex) were also included.

Study recruitment and participation are shown in Table 6. A total of 109 participants (58 athletes and 51 nonathletes) were recruited for the study (recruit population). Of these, 86 (79%) completed the first survey (study population), and 63 (58%) completed both surveys (reliability population). The 63 participants in the reliability population were

Table 5. Athlete Participants' Sports in the Trojan LifetimeChampions Phase 2 Validity and Reliability Study, 2010–2011,Compared With the University of Southern California VarsityAthlete Alumni Source Population

	Population, %				
Sport	Source ^a	Reliability ⁱ			
Men's baseball	8	6			
Men's basketball	4	0°			
Women's basketball	3	9			
Women's cross-country	1	0			
Men's football	24	3°			
Men's golf	2	9			
Women's golf	1	0 ^d			
Women's rowing	2	6			
Women's soccer	2	3			
Men's swimming and diving	7	21			
Women's swimming and diving	4	3			
Men's tennis	2	9			
Women's tennis	1	0 ^d			
Men's track and field	12	3			
Women's track and field	9	3			
Men's volleyball	5	6			
Women's volleyball	3	3			
Men's water polo	9	9			
Women's water polo	2	0			
Men's multisport	Not applicable ^e	3°			
Women's multisport	Not applicable ^e	3 ^d			

^a Estimated from roster sizes over the 50 years before the study using media guides for each current varsity sport.

^b Reliability population included study participants completing both test and retest surveys.

- ^c Men's multisport athlete reported competing in basketball and football.
- ^d Women's multisport athlete reported competing in golf and tennis.

^e Multisport athlete proportions were not estimated in the source population.

included in the subsequent survey validity and reliability analysis. All participants provided informed consent, and the experimental protocol was approved by the USC Health Sciences Institutional Review Board.

Study Design and Data-Collection Methods

On condition of anonymity, participants were invited to complete the TLC Health Survey twice at a self-directed, 1week retest interval. Anonymity of responses is an important characteristic of the TLC Health Survey and is designed to encourage open and honest reporting of health information, including conditions that may be perceived as sensitive. Accordingly, each survey was identified by a study identification number dissociated from the unique identity of the participant. Participants were assigned randomly to 1 of 4 survey types: paper survey followed by paper survey (PP), electronic survey followed by electronic survey (EE), paper survey followed by electronic survey (PE), or electronic survey followed by paper survey (EP). The 4 survey types allowed us to assess both testretest and parallel-form reliability (ie, paper and electronic survey equivalence).

Each participant received a single postal mailing containing an introductory letter, 2 surveys based on survey-type group assignment with electronic survey access code or codes when applicable, anonymous return envelope

	Source Population					
Recruitment and Participation	Athletes	Nonathletes	Total			
Recruited to study, No. Completed first survey, No. (%) Completed second survey, No. (%)	58 44 (76) 33 (57)	51 42 (82) 30 (59)	109 86 (79) 63 (58)			

or envelopes, an instructional DVD, and a coffeehouse gift card. The DVD provided instructions emulating those given to phase 1 study participants during live administrations of the survey by the same investigator (S.C.S.). A single reminder e-mail was sent to all participants approximately 12 weeks after the initial mailing.

Data Analysis

We used χ^2 proportion tests to compare responders and nonresponders on sex, age, collegiate sport participation, and survey type. Demographic characteristics (ie, age, sex, and sport distribution) were evaluated between the source and reliability populations to examine external validity. A 1-sample *t* test was used for age, a binomial test for sex distribution, and a χ^2 proportion test for sport distribution. The actual retest interval was compared with the selfdirected 1-week interval using a 1-sample *t* test. We assessed the effects of collegiate sport participation and sex on the retest interval via independent-samples *t* tests, including the Levene test for unequal variances, and the effects of survey type via 1-way analysis of variance.

Analytic Plan. Dependent variables of interest were the TLC lifetime and recent domain scores, SF-12 summary HRQL scores (PCS and MCS), weekly exercise volume (total and for each exercise type), and self-rated exercise importance. Independent variables were age, sex, collegiate sport participation (ie, athlete versus nonathlete), and survey type (ie, PP, EE, EP, or PE).

Data Screening. Before statistical analysis, we screened all data for integrity, including identification of spurious and outlier values. Spurious data included miscoded responses (eg, calendar birth year reported for age or ambiguous descriptive text). A 2-stage outlier-screening process was used for continuous variables, including visual inspection of a scatter plot and assessment of statistical variance versus the group mean. *Outliers* were defined as data points that demonstrated apparent perturbation from combined group data via visual inspection and that deviated by more than 3 standard deviations from the mean. For categorical and ordinal variables, reported values out of the specified range were considered outliers. All spurious and outlier data were excluded from subsequent analysis.

Test-Retest and Parallel-Form Reliability. Test-retest reliability was evaluated using combined data from the PP and EE survey types, whereas parallel-form reliability was evaluated using combined data from the PE and EP types. We used the 2-way mixed-effects intraclass correlation coefficient (ICC) [3,1] for continuous dependent variables, and the κ agreement statistic was used for ordinal dependent variables. Within each analysis, we reviewed correlation coefficients for consistency between the groups (ie, PP versus EE and PE versus EP). Substantive differences occurred where correlation coefficients differed by more than 0.20, suggesting a difference in interpretation.²⁵

Error Estimates. Estimates of error magnitude, including systematic and random error, were evaluated using combined data across the survey types. This method provided an error estimate for desired comparisons of outcomes variables across all study participants, irrespective of survey type. Systematic error was assessed via paired *t* tests, with a difference between test and retest considered evidence of a systematic change. Random error was evaluated according to the observed typical error using 95% confidence limits.²⁶

A Priori Clinical Thresholds. Error magnitude was compared with minimal a priori thresholds of substantial clinical meaning as follows: (1) TLC Health Survey domain scores (2 units) corresponding to a single health concern in that domain requiring medical treatment or 2 subclinical concerns; (2) SF-12 HRQL summary scores (5 units) corresponding to a minimal clinically important difference suggested by the previous literature, using a standardized effect-size benchmark²⁷ of 0.50; (3) total weekly exercise volume (150 minutes) corresponding to minimal ACSM guidelines for healthy adults²⁴; and (4) self-reported exercise importance (1 unit) corresponding to the minimal precision for the 4-point TLC Health Survey exercise-importance scale. Threshold values were not evaluated for individual exercise subtypes.

Convergent Construct Validity. We evaluated convergent construct validity via linear regression of TLC domain summary scores versus SF-12 HRQL summary scores for first survey data with Pearson product moment correlation coefficients, using interpretive guidelines as recommended by Hopkins et al²⁵ (*small* = 0.1, *moderate* = 0.3, *large* = 0.5, *very large* = 0.7, *extremely large* = 0.9). We anticipated that SF-12 PCS scores would correlate with the TLC Health Survey physical domain summary scores (Joints, Bone & Muscle, Cardiopulmonary, Neurological, and Other Clinical) and that SF-12 MCS scores would correlate with the TLC Health Survey Psychosocial domain summary score based on shared constructs of physical and mental health.

All statistical analyses were conducted with 2-sided tests using SPSS (version 16; SPSS Inc, Chicago, IL). The α level was set at .05.

RESULTS

Response rates were similar between athlete and nonathlete alumni. Responders and nonresponders were similar in sex ($\chi^2_2 = 0.91$, P = .64), age ($\chi^2_{10} = 10.0$, P =.44), collegiate sport participation ($\chi^2_2 = 0.83$, P = .66), and survey type ($\chi^2_6 = 3.51$, P = .74). Age ($t_{32} = -0.259$, P =.80) and sex (P = .40) distributions were similar between the athlete source and reliability populations (Table 4). We found a difference in sport distribution between the athlete source and reliability populations ($\chi^2_{18} = 43.6$, P < .001; Table 5). Specifically, the reliability population overrepresented athletes in women's basketball, men's golf, women's rowing, men's swimming and diving, and men's tennis and underrepresented athletes in women's crosscountry, men's football, men's and women's track and field, and women's water polo. Two multisport athletes were part of the reliability population, including a man who played both basketball and football and a woman who played both golf and tennis.

The mean test-retest interval was 23.1 ± 30.2 days. This was longer than the self-directed interval of 1 week ($t_{62} = 4.22$, P < .001). We found no effect of collegiate sport participation ($t_{61} = 1.071$, P = .29) or survey type ($F_{3,59} = 0.41$, P = .75) on test-retest interval. Women (14.5 ± 13.0 days) had a shorter test-retest interval than men (27.1 ± 34.9 days; $t_{59} = -2.07$, P = .04), with unequal variances per the Levene test ($F_{1,61} = 7.67$, P = .007).

Subsequent to outlier screening, data from 4 participants were excluded from analysis in the exercise-behavior variables. Two participants were excluded from analysis for each of the Bone & Muscle and Neurological domains. Figure 2 illustrates a sample of outlier data.

Results of the test-retest and parallel-form reliability assessments are summarized in Table 7. Test-retest reliability coefficients (ICC) averaged 0.86, 0.90, 0.80, and 0.74 for HRQL, lifetime domain scores, recent domain scores, and exercise variables, respectively, indicating very large to extremely large agreement.²⁵ The Bone & Muscle recent domain score and mixed exercise variables were notable exceptions, with ICCs of 0.58 and 0.49, respectively. These values reflect moderate to large agreement.²⁵ The Bone & Muscle recent domain score showed a substantial difference between ICCs for PP (0.78) and EE (0.50) survey types.

Parallel-form reliability coefficients (ICC) averaged 0.90, 0.85, 0.85, and 0.81 for HRQL, lifetime domain scores, recent domain scores, and exercise variables, respectively, indicating very large to extremely large agreement. Substantial differences between ICCs for PE and EP survey types were noted for the SF-12 PCS score (0.65 versus 0.92), Joints lifetime domain score (0.67 versus 0.90), and Psychosocial recent domain score (0.70 versus 0.92). Each of these correlation coefficients, however, demonstrated large (or greater) agreement. For the Bone & Muscle lifetime domain score, the PE survey type had an ICC of 0.27, which was not statistically significant, whereas the EP survey type had an ICC of 0.91, which was statistically significant.

Systematic error was detected for the Other Clinical lifetime domain score, with an average test-retest difference of -0.48 ± 1.39 units. No other variable demonstrated evidence of a systematic change.

Typical errors were 2.85, 1.12, and 0.93 units for HRQL, lifetime domain scores, and recent domain scores, respectively. For exercise volume of the previous week, typical errors were 58.7, 17.2, 74.5, and 79.7 minutes for cardiovascular, resistance, mixed, and total exercise, respectively. Typical error for self-rated exercise importance, modeled as a continuous variable, was 0.30 units. In all cases, 95% confidence limits for typical error were smaller than respective minimal a priori thresholds for substantial clinical meaning.

Results of the convergent construct validity assessment are summarized in Table 8. Negative Pearson product moment correlations were observed between 5 of the 6 lifetime and recent domain scores, excluding the Psychosocial domain, and the SF-12 PCS HRQL score. Coefficients ranged from -0.33 to -0.65 for lifetime domain scores and from -0.49 to -0.69 for recent domain scores, suggesting moderate to large associations.²⁵ The Psycho-



Figure 2. Trojan Lifetime Champions Health Survey validity and reliability study, 2010-2011, sample exercise outlier data.

social lifetime (r = -0.58) and recent domain (r = -0.62) scores were correlated with the SF-12 MCS HRQL score, reflecting a large association.

DISCUSSION

We conducted a systematic validity and reliability assessment of a novel health-and-exercise questionnaire (the TLC Health Survey) designed to measure lifetime and recent health, exercise, and HRQL among elite, competitive athletes. The study was conducted in a population of former varsity collegiate athletes and an age- and sex-matched control group of alumni at a large NCAA Division I university. The TLC Health Survey demonstrated excellent test-retest reliability and parallel-form reliability (ie, equivalence) between the paper and electronic versions across a diverse range of measures, including 6 domains of lifetime and recent health, several forms of exercise, and HRQL. Bone & Muscle domain scores were less reliable than other data and should be interpreted with caution. Quantitative estimates of measurement error were less than a priori thresholds of clinical interest. The instrument demonstrated strong evidence of convergent validity in relation to the SF-12. The development and implementation process of the instrument, as well as inclusion of an established HRQL instrument, reflect evidence for content validity and feasibility of administration in the populations of interest.

These findings provide strong evidence of test-retest reliability and parallel-form reliability between the paper and electronic versions of the TLC Health Survey. In general, ICCs for TLC domain scores exceeded 0.70. Lifetime domain score ICCs generally exceeded 0.80, whereas ICCs for recent domain scores were somewhat lower. Reliability and validity for these novel measures were comparable with those of the widely used SF-12 PCS and MCS HRQL indices.

Lower correlation coefficients and evidence of substantial differences between survey types were found within the Bone & Muscle domain; thus, these scores should be interpreted with caution. Instructions for this domain possibly lacked clarity when compared with instructions for the other domains. The Bone & Muscle instructions directed participants to "describe bone & muscle concerns [including] fractures, pulled muscles or tendons, pain, stiffness, or weakness" in various regions of the body (eg, lower leg, upper torso, and hand). In retrospect, these instructions may be less specific and, therefore, more subject to misclassification than the reference to a particular joint (eg, knee) or medical condition (eg, high blood pressure) described in other domains. In future studies, researchers might consider identifying specific bones and muscles (eg, clavicle, hamstrings) or combining joints, bones, and muscles into a single musculoskeletal domain within the instrument. Bone and muscle concerns also could be more transient or episodic,²⁸ leading to differential reporting between the test and retest that appears as measurement error.

We found some evidence of an influence of test order in the parallel-form reliability assessment. The ICCs differed substantially (>0.2) between the PE and EP survey types for the SF-12 PCS, Joints lifetime domain score, and Psychosocial recent domain score. However, the overall interpretation of large (or greater) agreement between the test and retest was unaffected, and we do not believe this presents a substantial threat to reliability for these measures. Authors of future studies in expanded sample populations might explore whether this effect has a systematic explanation. Clinicians and researchers using multiple iterations of the TLC Health Survey should

	Correlation Coefficient ^a			Magni			
Measure	Test-Retest Reliability	Parallel-Form Reliability	All	Retest-Test Change, Mean \pm SD	Typical Error	95% Confidence Interval	Clinically Important Difference
Health-related quality of life							
Short-Form 12 Version 2 Health Survey physical component score	0.92 ^b	0.90 ^{b,c}	0.91 ^b	0.20 ± 3.81	2.69	2.21, 3.28	5
Short-Form 12 Version 2 Health Survey mental component score	0.79 ^b	0.90 ^b	0.87 ^b	0.98 ± 4.24	3.00	2.46, 3.66	5
Lifetime domain summary scores							
Joints	0.94 ^b	0.86 ^{b,c}	0.90 ^b	-0.10 ± 2.09	1.48	1.21, 1.80	2
Bone & Muscle	0.91 ^b	0.78 ^{b,c}	0.86 ^b	-0.28 ± 1.32	0.93	0.77, 1.14	2
Cardiopulmonary	0.98 ^b	0.90 ^b	0.89 ^b	-0.05 ± 1.38	0.98	0.80, 1.19	2
Neurological	0.95 ^b	0.82 ^b	0.86 ^b	0.00 ± 1.18	0.83	0.68, 1.02	2
Other Clinical	0.83 ^b	0.89 ^b	0.86 ^b	-0.48 ± 1.39^{b}	0.98	0.80, 1.20	2
Psychosocial	0.79 ^b	0.84 ^b	0.83 ^b	-0.12 ± 2.16	1.52	1.25, 1.86	2
Recent domain summary scores							
Joints	0.76 ^b	0.90 ^b	0.85 ^b	-0.41 ± 1.98	1.40	1.15, 1.71	2
Bone & Muscle	0.58 ^{b,c}	0.75 ^b	0.65 ^b	-0.41 ± 1.51	1.07	0.88, 1.30	2
Cardiopulmonary	0.93 ^b	0.87 ^b	0.92 ^b	-0.20 ± 0.94	0.67	0.55, 0.81	2
Neurological	0.86 ^b	0.98 ^b	0.94 ^b	0.02 ± 0.57	0.40	0.33, 0.49	2
Other Clinical	0.84 ^b	0.78 ^b	0.80 ^b	-0.12 ± 1.40	0.99	0.81, 1.21	2
Psychosocial	0.83 ^b	0.83 ^{b,c}	0.83 ^b	-0.07 ± 1.52	1.08	0.88, 1.31	2
Previous week's exercise							
Cardiovascular	0.71 ^b	0.88 ^b	0.78 ^b	5.3 ± 82	58.7	48.1, 71.6	Not applicable
Resistance	0.94 ^b	0.91 ^b	0.93 ^b	-3.2 ± 24	17.2	14.1, 20.9	Not applicable
Mixed	0.49 ^b	0.65 ^b	0.61 ^b	12 ± 105	74.5	61.1, 90.9	Not applicable
Total	0.87 ^b	0.85 ^b	0.86 ^b	14 ± 113	79.7	65.4, 97.3	150
Exercise importance	0.69 ^b	0.77 ^b	0.73 ^b	0.02 ± 0.42^{d}	0.30 ^d	0.25, 0.37	1

Table 7. Test-Retest and Parallel-Form Reliability Results for Participants in the Trojan Lifetime Champions Health Survey Phase 2 Validity and Reliability Study, 2010–2011

^a Correlation coefficients are intraclass correlation coefficients (3,1) for all variables except exercise importance (κ agreement statistic).

 $^{\text{b}}$ $P \leq .05.$

° Substantive difference between subgroups.

^d Exercise importance was modeled as a continuous variable for estimates of error magnitude.

consider using the same form (paper or electronic) for all administrations.

Validity and reliability of exercise-outcome variables generally were excellent. Mixed Exercise, which by its nature is perhaps more ambiguous than Cardiovascular or Resistance Exercise, had relatively lower ICCs but still demonstrated moderate agreement. Exercise behavior is more likely to vary from week to week than measures of health over the lifespan or the previous 3 years (ie, recent domain scores). Thus, changes in exercise outcomes between the test and retest that appear as measurement error actually may be attributable to true changes in behavior. Exercise data constituted the most common exclusion in the outlier screening process. Figure 2 depicts 2 examples of outliers for weekly Cardiovascular Exercise. One participant reported nearly 19 hours at both the test and retest and approximately 50 total hours of exercise per week. These values are questionable and well outside the group data. Another participant reported more than 13 hours in the first survey and just 8 minutes in the second. This likely reflects week-to-week variability as opposed to measurement error. In the practical effort to measure exercise using self-report questionnaires, these challenges remain an important limitation for this and any comparable instrument.

Whereas no universally accepted consensus for acceptable validity and reliability measures exists,^{25,29} the testretest and parallel-form reliability coefficients that we report are comparable with those in previous studies of the SF-12^{30–33} and SF-36³⁴ HRQL instruments. Similarly, reliability for the exercise measures was comparable with that of the highest-quality physical activity instruments identified in a recent systematic review.³⁵ Nonetheless, given that this is one of the first formal assessments of validity and reliability for a health-and-exercise outcomes instrument in an elite, competitive athletic population, additional larger-scale studies in similar populations are warranted to provide further validity evidence.

Review of the test-retest differences demonstrated minimal evidence of systematic error. One exception (Other Clinical lifetime domain score) demonstrated a reduction, but the magnitude of this change was less than one-fourth of the minimal a priori threshold for substantial clinical meaning. In all cases, typical error was less than the respective threshold of interest. More powerfully than a correlation coefficient, these data provide evidence that observed differences in outcome variables between groups or over time reflect true differences of clinical interest as opposed to measurement error. This "magnitude-based"

Table 8. Convergent Construct Validity Correlation Coefficients for Participants in the Trojan Lifetime Champions Health Survey Phase 2 Validity and Reliability Study, 2010–2011^a

	Short-Form 12 Version 2 Health Survey Physical Component Score		Short-Form 12 Version 2 Health Survey Mental Component Score	
Measure	r Value	P Value	r Value	P Value
Health-related quality of life				
Short-Form 12 Version 2 Health Survey physical component score Short-Form 12 Version 2 Health Survey mental component score	NA -0.22	NA .09	-0.22 NA	.09 NA
Lifetime domain summary scores				
Joints	-0.33 ^b	.009	-0.09	.48
Bones & Muscle	-0.35 ^b	.006	0.13	.34
Cardiopulmonary	-0.37 ^b	.003	0.16	.21
Neurological	-0.65 ^b	<.001	-0.14	.29
Other Clinical	-0.38 ^b	.002	-0.17	.18
Psychosocial	-0.01	.94	-0.58 ^b	<.001
Recent domain summary scores				
Joints	-0.49 ^b	<.001	-0.20	.13
Bones & Muscle	-0.55 ^b	<.001	0.15	.24
Cardiopulmonary	-0.52 ^b	<.001	0.09	.51
Neurological	-0.69 ^b	<.001	-0.04	.74
Other Clinical	-0.58^{b}	<.001	-0.22	.08
Psychosocial	-0.14	.29	-0.62 ^b	<.001

Abbreviation: NA, not applicable

^a Correlation coefficients are Pearson product moment correlation values.

^b *P* < .05.

approach is a preferred alternative to traditional hypothesis testing versus the null value.²⁵

We find it interesting that the mean test-retest interval was more than 3 times the prescribed 1-week interval and was different between men and women. Although this illustrates a limitation to study compliance, an interval of 3 weeks is not uncommon for reliability studies of comparable instruments. The SF-12, for example, has undergone test-retest reliability assessment at intervals ranging from 1 week to 3 months with comparable results.³⁰⁻³² Whereas women had a shorter test-retest interval than men in this study (approximately 2 versus 4 weeks), reliability coefficients were similar. A longer test-retest interval increases the potential for true changes in health status or exercise behavior that appear as measurement error. Oneweek reliability, therefore, might be presumed superior to that measured at 3 weeks. However, a reliability study for a musculoskeletal-symptoms questionnaire showed no differences between retest intervals of 2 and 4 weeks.³⁶

The correlation of questionnaire items sharing the same theoretic construct provides evidence of construct validity.¹⁷ Our findings indicate moderate to large associations among each of the TLC Health Survey physical health domain scores and the SF-12 PCS physical HRQL index, as well as the Psychosocial health domain scores and the MCS mental HRQL index. Correlations were somewhat larger for recent domain scores than for lifetime domain scores. This likely is attributable to the fact that SF-12 scores also reflect recent (past 4 weeks) health experience. Although health domain scores share broad constructs of physical and mental health, they do not measure exactly the same components of health. Thus, a strong correlation would not be expected. However, the reported associations demonstrate both validity evidence for the novel health domain scores and the relevance of association with HRQL.

The primary limitation of this study was the relatively small sample population compared with source populations of interest. More than 450 000 individuals actively participate in NCAA sports, a figure that has increased more than 60% over the past 20 years.³⁷ Whereas this study population was representative of the age and sex distribution of former athletes at 1 Division I university, it does not necessarily represent the overall athlete population or that of other elite competitive athletes (eg, professionals and Olympians). Furthermore, we found a difference in sport distribution between the source and study populations, but the sample size was insufficient to evaluate sport-specific influences on the findings. Together, these factors limit the generalizability of study results and support the need for additional studies with larger and more diverse sample populations. Additionally, this instrument was developed to assess health and exercise in athletic populations, but it could be readily adapted to other populations (eg, soldiers, firefighters) in whom demands for optimal physical performance present comparable lifespan health challenges. Independent validity assessments would be required to confirm the suitability of the instrument in these populations. In general, these findings should be considered a substantial first step in the ongoing accumulation of evidence necessary to support the utility, validity, reliability, and generalizability of this novel instrument.¹⁷

Evaluation of psychometric properties using the robust measures presented here distinguishes the TLC Health Survey from previous questionnaires used to assess health outcomes and exercise behaviors in comparable populations and, indeed, from most epidemiologic instruments. Bennett et al³⁸ reported that of 117 recently published reports of self-administered surveys, less than 20%

provided validity or reliability data for the instrument. Similarly, less than 30% of physical activity instruments reviewed by van Poppel et al^{35} met the highest level of reliability evidence. Whereas such evidence typically is given in the form of test-retest correlation coefficients, researchers^{25,29} have argued that the typical error and change in the mean between trials are more important measures. Our study provides both correlation coefficients and estimates of error magnitude relative to thresholds of substantive meaning.²⁵ It provides evidence of content validity as determined by an interdisciplinary team of experts and through inclusion of an established instrument in the SF-12. Furthermore, it provides assessments of feasibility and external validity in relation to source, study, and target populations of interest. Finally, parallel-form reliability between the paper and electronic versions supports the use of the electronic survey as an equivalent instrument in large-scale studies, offering an additional feasibility benefit.

CONCLUSIONS

Our study provides strong evidence for the validity and reliability of the TLC Health Survey in assessing lifetime and recent health, exercise, and HRQL among university alumni, including NCAA Division I athletes. The formal characterization of its psychometric properties, including the evaluation of error magnitude relative to thresholds of substantial clinical meaning, distinguishes the TLC Health Survey from previous questionnaires used to evaluate lifetime health and exercise in elite athletes. Additional, larger-scale studies in diverse populations are necessary to enhance the generalizability of the instrument. The electronic form of the TLC Health Survey is a suitable and feasible means to do so.

ACKNOWLEDGMENTS

We thank Doyle Lee, DPT, for his substantial contributions to data processing.

REFERENCES

- Faulkner JA, Davis CS, Mendias CL, Brooks SV. The aging of elite male athletes: age-related changes in performance and skeletal muscle structure and function. *Clin J Sport Med.* 2008;18(6):501–507.
- Haskell WL, Lee IM, Pate RR, et al. Physical activity and public health: updated recommendation for adults from the American College of Sports Medicine and the American Heart Association. *Med Sci Sports Exerc.* 2007;39(8):1423–1434.
- Vecsey G. Accepting the costs of a life in football. New York Times. May 12, 2008. http://www.nytimes.com/2008/05/12/sports/football/ 12vecsey.html?_r=0. Accessed March 3, 2014.
- Brown C. Former players dealing with lingering health issues. *New York Times*. February 1, 2007. http://www.nytimes.com/2007/02/01/sports/football/01disability.html. Accessed March 3, 2014.
- Schwarz A. Dark days follow hard-hitting career in N.F.L. *The New York Times*. February 2, 2007. http://query.nytimes.com/gst/fullpage. html?res=9501E7DE123FF931A35751C0A9619C8B63. Accessed March 3, 2014.
- Agostino J. Once an athletic star, now an unheavenly body. *The New York Times*. July 6, 2006. http://www.nytimes.com/2006/07/06/fashion/06Fitness.html?ei=5087&en=5330174e14431ce4&ex=1152417600&adxnnl=1&adxnnlx=1393864376-s/B29J5/QIeM/d9cYYjoeQ. Accessed March 3, 2014.

- Jensen M. Congressional panel hears testimony on concussions. *The Philadelphia Inquirer*. September 24, 2010. http://articles.philly.com/ 2010-09-24/sports/24977128_1_effects-of-multiple-concussions-mildconcussions-cte. Accessed March 3, 2014.
- Johnson G. NFL is focus of Senate hearing. Los Angeles Times. September 18, 2007. http://articles.latimes.com/2007/sep/18/sports/ sp-nflhearing18. Accessed March 3, 2014.
- Fendrich H. NFL Concussion saga moves to new phase: litigation. Associated Press. October 20, 2011. http://www.sportingnews.com/ nfl/story/2011-10-20/nfl-concussion-saga-moves-to-new-phase-litigation. Accessed March 3, 2014.
- Zarda B. Lawsuit prompts NCAA to screen athletes for sickle cell. USA Today. July 2, 2010. http://usatoday30.usatoday.com/sports/college/ 2010-06-30-sickle-cell-ncaa-cover_N.htm. Accessed March 3, 2014.
- Kujala U, Orava S, Parkkari J, Kaprio J, Sarna S. Sports careerrelated musculoskeletal injuries: long-term health effects on former athletes. *Sports Med.* 2003;33(12):869–875.
- Golightly YM, Marshall SW, Callahan LF, Guskiewicz K. Earlyonset arthritis in retired National Football League players. *J Phys Act Health*. 2009;6(5):638–643.
- Weir DR, Jackson JS, Sonnega A. National Football League Player Care Foundation: study of retired NFL players. University of Michigan Institute for Social Research Web site. http://www.ns. umich.edu/Releases/2009/Sep09/FinalReport.pdf. Published 2009. Accessed November 14, 2013.
- Turner AP, Barlow JH, Heathcote-Elliott C. Long term health impact of playing professional football in the United Kingdom. *Br J Sports Med.* 2000;34(5):332–336.
- Huffman GR, Park J, Roser-Jones C, Sennett BJ, Yagnik G, Webner D. Normative SF-36 values in competing NCAA intercollegiate athletes differ from values in the general population. *J Bone Joint Surg Am.* 2008;90(3):471–476.
- 16. Horn S, Gregory P, Guskiewicz KM. Self-reported anabolicandrogenic steroids use and musculoskeletal injuries: findings from the center for the study of retired athletes health survey of retired NFL players. Am J Phys Med Rehabil. 2009;88(3):192–200.
- Denegar CR, Vela LI, Evans TA. Evidence-based sports medicine: outcomes instruments for active populations. *Clin Sports Med.* 2008; 27(3):339–351, vii.
- Haynes SN, Richard DCS, Kubany ES. Content validity in psychological assessment: a functional approach to concepts and methods. *Psychol Assess*. 1995;7(3):238–247.
- 19. Ware JE, Kosinski M, Turner-Bowker DM, Gandeck B. User's Manual for the SF-12v2 Health Survey With a Supplement Documenting SF-12 Health Survey. Lincoln, RI: QualityMetric Incorporated; 2002.
- McAllister DR, Motamedi AR, Hame SL, Shapiro MS, Dorey FJ. Quality of life assessment in elite collegiate athletes. *Am J Sports Med.* 2001;29(6):806–810.
- Kuehl MD, Snyder AR, Erickson SE, McLeod TC. Impact of prior concussions on health-related quality of life in collegiate athletes. *Clin J Sport Med.* 2010;20(2):86–91.
- Evans TA. The Reliability, Validity, Sensitivity, and Standard Values of the Athletic Training Outcome Assessment Self-Report Instrument [dissertation]. State College, PA: The Pennsylvania State University; 2002.
- Kerr ZY, Marshall SW, Harding HP Jr, Guskiewicz KM. Nine-year risk of depression diagnosis increases with increasing self-reported concussions in retired professional football players. *Am J Sports Med.* 2012;40(10):2206–2212.
- 24. Garber CE, Blissmer B, Deschenes MR, et al. American College of Sports Medicine position stand: quantity and quality of exercise for developing and maintaining cardiorespiratory, musculoskeletal, and neuromotor fitness in apparently healthy adults. Guidance for prescribing exercise. *Med Sci Sports Exerc*. 2011;43(7):1334–1359.

- Hopkins WG, Marshall SW, Batterham AM, Hanin J. Progressive statistics for studies in sports medicine and exercise science. *Med Sci Sports Exerc*. 2009;41(1):3–13.
- Hopkins WG. Measures of reliability in sports medicine and science. Sports Med. 2000;30(1):1–15.
- Samsa G, Edelman D, Rothman ML, Williams GR, Lipscomb J, Matchar D. Determining clinically important differences in health status measures: a general approach with illustration to the Health Utilities Index Mark II. *Pharmacoeconomics*. 1999;15(2):141–155.
- Orchard J, Best TM. The management of muscle strain injuries: an early return versus the risk of recurrence. *Clin J Sport Med.* 2002; 12(1):3–5.
- 29. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. J Strength Cond Res. 2005; 19(1):231–240.
- Hurst NP, Ruta DA, Kind P. Comparison of the MOS short form-12 (SF12) health status questionnaire with the SF36 in patients with rheumatoid arthritis. *Br J Rheumatol*. 1998;37(8):862–869.
- Lenert LA. The reliability and internal consistency of an Internetcapable computer program for measuring utilities. *Qual Life Res.* 2000;9(7):811–817.
- 32. Salyers MP, Bosworth HB, Swanson JW, Lamb-Pagone J, Osher FC. Reliability and validity of the SF-12 health survey among people with severe mental illness. *Med Care*. 2000;38(11):1141–1150.
- Cheak-Zamora NC, Wyrwich KW, McBride TD. Reliability and validity of the SF-12v2 in the medical expenditure panel survey. *Qual Life Res.* 2009;18(6):727–735.

- 34. Gudbergsen H, Bartels EM, Krusager P, et al. Test-retest of computerized health status questionnaires frequently used in the monitoring of knee osteoarthritis: a randomized crossover trial. BMC Musculoskelet Disord. 2011;12:190.
- van Poppel MN, Chinapaw MJ, Mokkink LB, van Mechelen W, Terwee CB. Physical activity questionnaires for adults: a systematic review of measurement properties. *Sports Med.* 2010;40(7):565–600.
- Rosecrance JC, Ketchen KJ, Merlino LA, Anton DC, Cook TM. Test-retest reliability of a self-administered musculoskeletal symptoms and job factors questionnaire used in ergonomics research. *Appl Occup Environ Hyg.* 2002;17(9):613–621.
- 1981–1982–2011–2012 NCAA Sports Sponsorship and Participation Rates Report. Indianapolis, IN: National Collegiate Athletic Association; 2012.
- Bennett C, Khangura S, Brehaut JC, et al. Reporting guidelines for survey research: an analysis of published guidance and reporting practices. *PLoS Med.* 2010;8(8):e1001069.

SUPPLEMENTAL MATERIAL

Trojan Lifetime Champions Survey, found at DOI: http:// dx.doi.org/10.4085/1062-6050-50.2.10.S1

Trojan Lifetime Champions Survey – Data Entry Guidelines, found at DOI: http://dx.doi.org/10.4085/1062-6050-50.2.10.S2

Address correspondence to Shawn C. Sorenson, PhD, CSCS, Division of Biokinesiology and Physical Therapy, University of Southern California, 1540 East Alcazar Street, CHP 155, Los Angeles, CA 90033-9006. Address e-mail to sorenson@usc.edu.