ORIGINAL PAPER

IJCP THE INTERNATIONAL JOURNAL OF CLINICAL PRACTICE

# Psychometric testing of two new patient-reported outcome instruments for the evaluation of treatment for hypogonadism

R. P. Hayes, X. Ni,* D. E. Heiselman, K. Kinchen

**SUMMARY**

**Aim:** The aim of this study was to perform psychometric testing and estimate minimal important change (MIC) of two new patient-reported outcome (PRO) instruments – Sexual Arousal, Interest and Drive Scale (SAID) and Hypogonadism Energy Diary (HED). **Methods:** New PRO instruments were administered immediately after screening (Time 1, test–retest subset only) and immediately prior to both randomisation (Time 2) and end-point (Time 3) to men participating in a randomised clinical trial comparing the effect of testosterone solution 2% (TS) and placebo on serum total testosterone. Psychometric analyses included reliability, validity and responsiveness. Total scores for both PRO instruments were transformed to a 0–100 scale. **Results:** Study participants ($n = 694$) were 80% age $\leq 65$ years, 79% White, with mean baseline testosterone = 202 ng/dl. Clinicians identified 86% subjects as having low sex drive, 86% with low energy and 76% with both symptoms. Reliability analyses for SAID and HED yielded reliability coefficients $> 0.70$. SAID scores discriminated between men having low sex drive ($n = 553$) and those who did not ($n = 80$) (34.5 vs. 42.8, $p < 0.001$). HED scores discriminated between men having low energy ($n = 541$) and those who did not ($n = 64$) (48.9 vs. 60.2, $p < 0.001$). In the men randomised to TS (vs. placebo), SAID and HED detected effect sizes of 0.61 (vs. 0.39) and 0.68 (vs. 0.48), respectively. MIC estimates for SAID and HED were approximately 10 and 8, respectively. **Conclusions:** This study provided evidence of the reliability, validity and responsiveness of SAID and HED as measures of sex drive and energy, respectively, making them potentially useful for evaluation of hypogonadal treatment.

**What's known**

Loss of sex drive and loss of energy are symptoms frequently reported by men with hypogonadism. To adequately evaluate the benefit of hypogonadal treatment, these symptoms should be assessed using patient-reported outcome (PRO) instruments. The content validity of two new PRO instruments assessing sex drive (SAID) and energy (HED) has been established in men with both early and late onset hypogonadism.

**What's new**

This article describes psychometric testing of SAID and HED in men with hypogonadism participating in a clinical trial. Psychometric testing suggests that these two instruments are reliable, valid, and responsive to treatment and that score improvement of 10 points for SAID and 8 for HED may represent an important change to patients. These instruments add value to the evaluation of treatment of hypogonadism by providing an assessment of patient-reported benefit.

Eli Lilly and Company, Indianapolis, IN, USA
***Present address:** Alcon Laboratories, Inc. (a Novartis company), Cambridge, MA, USA

**Correspondence to:**
Risa P. Hayes, PhD
Eli Lilly and Company (Retired)
6441 Bastani Place
Indianapolis, IN 46237, USA
Tel: + 1 317 331 9648
Email: hayesrisa2013@gmail.com

**Disclosures**

Risa P. Hayes is a retired full-time employee and stockholder of Eli Lilly and Company. Xiao Ni is a former employee and current stockholder of Eli Lilly and Company. Darell E. Heiselman and Kraig Kinchen are full-time employees and stockholders of Eli Lilly and Company.

## Introduction

Many men with hypogonadism report decreased sex drive, decreased energy or both as symptoms of their hypogonadism (1–7). There are multidimensional instruments that assess either or both symptoms but there are few, if any, instruments focused on accurately and reliably assessing sex drive and energy as unidimensional concepts. More importantly, there are few, if any, existing instruments that assess these symptoms and have been developed according to regulatory guidance that stresses the importance of patient input in the item development process (8).

In a previous article, Hayes et al. described the development of the Sexual Arousal, Interest, and Drive Scale (SAID) and the Hypogonadism Energy Diary (HED) (9). Results of a large qualitative study involving men with both late and early onset hypogonadism led to the identification of concepts that were deemed important and relevant to men with hypogonadism (9). These concepts served as the basis for the SAID and HED item generation. Cognitive interviewing provided evidence that items were comprehensive and potential respondents' interpretations of items were consistent with intended meanings. Cognitive interviewing also established the equivalency of the paper-based SAID and HED, with an electronic (ePRO) version appearing on an electronic handheld device.

The purpose of these two new patient-reported outcome PRO instruments was to serve as key secondary end-points in clinical trials of hypogonadism treatment. However, before the value of the SAID and HED can be determined, evidence of their psychometric properties is needed. The aim of this study was to address the following psychometric research


*Int J Clin Pract*, July 2016, **70**, 7, 587–595. doi: 10.1111/ijcp.12828

questions for each of the PRO instruments in a sample of men with hypogonadism:

- What are the factor structure and reliability (internal consistency and test–retest or reproducibility) of the SAID and HED?
- How well do the SAID and HED correlate with similar existing instruments (i.e. have convergent validity)?
- How well do the SAID and HED discriminate between those men clinically identified as having symptoms of loss of sex drive or loss of energy, respectively, and those who are not clinically identified as having these symptoms (known group validity)?
- How responsive (i.e. ability to detect change) are the HED and SAID to symptom improvement as a result of treatment for hypogonadism?
- If the SAID and HED are responsive to change in self-reported loss of sex drive or loss of energy, what is the change in SAID and HED scores that represents patient-perceived benefit [i.e. minimal important change (MIC)] (10,11).

## Methods

This study was conducted in the context of a multicenter, randomised, double-blind, parallel, placebo-controlled trial designed to assess the effect of testosterone solution on total testosterone (primary end-point), sex drive, and energy in men with hypogonadism (ClinicalTrials.gov, number NCT01816295) (12). This study included a protocol addendum of which the primary objective was to evaluate test–retest reliability and perform an exploratory factor analysis (EFA) of the SAID and HED with a subset of study participants. To obtain data for the addendum, an interim blinded data snapshot was created.

### Study participants

Eligibility criteria for this study were the same as those for the primary clinical trial (TSAT) and included the following: (i) being male and at least 18 years of age; (ii) having two total testosterone levels < 300 ng/dl (10.4 nmol/l) taken at least 1 week apart and (iii) having been identified by a clinician as having at least one symptom of testosterone deficiency, including decreased energy or decreased sex drive (12).

### Procedure

Under the protocol addendum, ePRO versions of the SAID and HED were administered twice within the 3-week period between Visit 2 (screening) and Visit 3 (randomisation) to a subset of study participants. That is, the first 7-day administration (Time 1) occurred during the 7-day period immediately after Visit 2, and the second 7-day administration (Time 2) occurred during approximately the 7-day period prior to Visit 3 or randomisation. Men completed the HED three times per day for seven consecutive days. On the seventh day, the SAID was administered only once and prior to the administration of the HED for that day. In addition, the ePRO version of the Psychosexual Daily Questionnaire (PDQ) was administered for all 7 days after the last administration of the HED for the day. The SAID, HED and PDQ were administered on the same handheld ePRO device. The decision to administer the three PROs as ePROs was based on the findings that electronic administration is the optimal approach for ensuring data integrity of patient diaries (i.e. HED and PDQ) (13). For the remaining psychometric analyses (e.g. responsiveness), the same procedure as was used in the protocol addendum was used with all study participants at Time 2 and during approximately the 7-day period prior to Visit 12 or study end-point (Time 3, approximately 12 weeks after Time 2). The Sexual Desire domain of the International Index of Erectile Function Questionnaire (IIEF), also included in the psychometric testing, was administered at Visit 3.

This study was conducted in accordance with the ethical principles originating in the Declaration of Helsinki, good clinical practices and all applicable laws and regulations. The institutional review board at each site approved the primary study and the protocol addendum, and all men provided written informed consent before participating in the study or addendum.

### Patient-reported outcome instruments

*Sexual Arousal, Interest and Drive Scale (SAID)*

The SAID is a five-item self-administered instrument intended to assess the following in men with hypogonadism: the level of thinking about sex (2 items), arousal (1 item), and rating the level of interest in sex and sex drive (2 items). Men with hypogonadism are asked to recall the past 7 days and respond to all five items on 5-point Likert-type scales scored from 1 to 5, with 5 corresponding to greater levels of sexual arousal, interest or drive (9). A total score is obtained by summing item scores, and dividing by the number of items ($n = 5$) with higher scores corresponding to greater sex drive. When used with other PROs, the SAID total score is linearly transformed to a scale of 0–100 to facilitate comparisons.

*Hypogonadism Energy Diary (HED)*

The HED is a self-administered instrument intended to assess real-time energy levels, including the extent

to which a respondent feels energetic or has feelings of tiredness/exhaustion. There are two unique items administered three times during the day: approximately 2-h after waking (proposed time for administration begins at 8 a.m., with three reminder alarms over a 2-h period), late afternoon (proposed time for administration begins at 3 p.m., with three reminders over a 2-h period), and late evening (proposed time of administration begins at 8 p.m., with three reminders over a 2-h period). Men with hypogonadism respond to both energy items using an 11-point numerical rating scale, with 10 corresponding to full of energy or extreme tiredness. The tiredness item is then reverse scored so that higher total scores correspond to greater levels of energy (9). For each of the six HED daily items (two questions for three times per day), a weekly score is derived as the 7-day average for that item. The total HED scale score is the sum of each of the six item weekly scores, with higher scores corresponding to greater energy level. When used with other PRO instruments, the HED total score is linearly transformed to a scale of 0–100 to facilitate comparisons.

*Psychosexual Daily Questionnaire (PDQ)*

The PDQ was developed to assess sexual function and mood on a daily basis. The PDQ has been used in previous studies of men with hypogonadism receiving testosterone replacement therapy (14). Although the entire PDQ was administered at Times 1–3, only one item pertaining to the overall level of sexual desire [rated on a 7-point numerical rating scale from 1 (none) to 7 (very high)] and two items pertaining to the extent to which full of energy or tired describes the respondent [rated on a 7-point numerical rating scale from 1 (not at all true) to 7 (very true)] administered at Time 2 were used in the psychometric analyses.

*International Index of Erectile Function Questionnaire (IIEF)*

The IIEF was developed to assess levels of erectile dysfunction (ED) in men who are sexually active with the same female partner. The IIEF is commonly used to assess therapeutic efficacy of ED therapy (15). Although the entire IIEF was administered at Visit 3, only the Sexual Desire domain (Questions 11 and 12 pertaining to the feeling and rating of sexual desire in the past 4 weeks as rated on a 5-point Likert-type scale) was used for the psychometric analyses.

*Patient Global Impression and Improvement (PGI-I) Scales*

The PGI-I, adapted from trials of other conditions (16,17), was used to develop a measure of a patient's perception of changes in both their sexual drive (PGI-I-S) and energy level (PGI-I-E) at end of the double-blind treatment phase (Visit 8). Each questionnaire asks the patient to rate, on a 7-point numerical rating scale from 1 (very much better) to 7 (very much worse), how his PGI-I-S or PGI-I-E is now, compared with how it was before he began taking medication.

## Statistical analysis

With the exception of the analysis for known group validity, the SAID was psychometrically tested only in men who were identified by their clinician as having a reported history of decreased sex drive (Low Sex Drive analysis set). Correspondingly, with the exception of the analysis for known group validity, the HED was psychometrically tested only in men who were identified by their clinician as having a reported history of decreased energy (Low Energy Analysis set). It was anticipated that data from some men would be included in both the Low Sex Drive and the Low Energy analysis sets. All psychometric analyses were performed using the final data from all TSAT participants with the exception of test–retest reliability and exploratory factor analysis (EFA), which was based on a subset of study participants through a protocol addendum (and an interim blinded data snapshot). For the protocol addendum, sample size considerations included: (i) a minimum of 75 men for the test–retest reliability analysis based on Bonett (18), assuming an intraclass correlation coefficient (ICC) of 0.75 with a 95% confidence interval (CI) width of 0.2 for each of the analysis sets (Low Sex Drive and Low Energy) and (ii) a minimum of 100 men for the EFAs performed for SAID and HED items.

Unless otherwise specified, all psychometric analyses were based on *usable* questionnaires defined as follows at each visit: for the SAID, subjects had to provide at least four complete items of five (i.e. $\geq$ 80%, or no more than one item was missing) and for each item of the HED, subjects had to provide data for at least 5 of 7 days (no more than two missing responses per item). Missing responses in *usable* questionnaires were imputed using the average of the non-missing responses. Note that for the HED, imputations were performed as necessary for each item to obtain the item weekly scores, which were then summed to derive the total score as previously described.

*Descriptive statistics*

Frequency distributions, inter-item correlations, and item-total correlations were evaluated for individual items of the SAID and HED using data from Time 2 for all TSAT study participants.

*Factor analysis*

Both the SAID and HED were designed to measure a single (i.e. unidimensional) concept: sex drive and energy, respectively. To evaluate the factor structure of each PRO instrument so as to make decisions regarding the use of a SAID and HED total score as sum of item scores, an EFA was performed for the SAID and the HED. For the HED, the item weekly scores (7-day average) were used for the factor analysis. Factors were extracted using the principal component analysis with varimax rotation. Factors associated with Eigenvalue $\geq 1$ were retained. Data for these analyses were obtained at Time 2 from the subset of men participating in the protocol addendum.

*Reliability*

Test–retest reliability for the 2 PRO instruments was assessed by calculating the ICC and 95% CI between the Time 1 (7-day period immediately after Visit 2) and Time 2 (7-day period prior to Visit 3 or randomisation) administrations based on a one-way random analysis of variance (ANOVA) model, according to Shrout and Fleiss (19). Data for these analyses were obtained at Time 2 from the subset of men participating in the protocol addendum. Internal consistency of the SAID and HED was assessed by calculating Cronbach's alpha coefficient using data from Time 2 for all study participants. ICC coefficients > 0.60 and Cronbach's alpha > 0.70 were considered acceptable.

*Construct validity (convergent and known group validity)*

For convergent validity, it was hypothesised that the Pearson correlation coefficients calculated between SAID total scores (administered at Time 2) and both the IIEF Sexual Desire domain scores and the weekly average of the PDQ overall level of sexual desire item (both administered at Visit 3) would be significant and > 0.60. Similarly, it was hypothesised that the Pearson correlation coefficients calculated between the HED scores and the weekly average of the PDQ 'full of pep/energetic?' and 'tired?' items would be significant and > 0.60. The 'tired' item was reverse scored to be consistent with the scoring of HED in which a higher score corresponds to more positive outcome.

For known group validity, it was hypothesised that significant differences would be observed in SAID and HED scores between men clinician-identified with decreased sex drive or decreased energy, respectively, and those who had not been clinician-identified as having the symptom. Student's *t*-tests were used to detect the hypothesised differences.

*Responsiveness*

Responsiveness of the SAID and HED to changes in sex drive and energy, respectively, were tested by performing a paired t-test using data from all study participants from Time 2 (baseline) and Time 3 (endpoint). Responsiveness indices, including effect size (mean change divided by the standard deviation of the baseline score) and standardised response mean [(SRM), mean change divided by the standard deviation of the change score], were calculated.

*MIC*

To estimate MIC, Crosby et al. (20) recommend an integrated approach that includes both distribution- and anchor-based methods. Distribution-based methods use a statistical property [e.g. standard deviation (SD)] of the sample or PRO measure to estimate MIC, while anchor-based methods compare changes in the PRO measure with, for example, patient impression of change or other clinically relevant variables. For this study, distribution-based methods were 0.5 SD and 1 standard error (SE) of measurement [i.e. 1 SEM (square root of 1 minus reliability multiplied by baseline SD)]. These two statistics have been shown to be good approximations of MIC determined by other methods (10,11,21,22). For an anchor-based method, the SAID and HED scores were compared with the corresponding PGI-I-S and PGI-I-E, respectively. Respondents to the PGIs were categorised as either reporting 'no change to very much worse' or 'a little better to very much better' regardless of treatment group. The average SAID and HED scores were then compared between the groups.

All statistical tests were based on a two-side alpha level of 0.05. All analyses were performed using SAS version 9.2 software (SAS Institute, Inc., Cary, NC).

## Results

### Study participants

The majority of men for whom data were included in one or more of the psychometric analyses ($n = 694$) were < 65 years of age (80%) and white (79%) with a mean total testosterone level (average of testosterone levels taken at Visits 1 and 2 at least 7 days apart) of 202 ng/dl. Clinicians identified 86% subjects as having low sex drive, 86% as having low energy, and 76% as having both (Table 1). Of the 595 men in the Low Sex Drive analysis set (usable SAID questionnaire), 97 participated in and met criteria for (complete data for Time 1 and Time 2) the test–retest analysis. Of the 599 men in the Low Energy analysis set (usable HED questionnaire), 127 met criteria for the test–retest analysis.

**Table 1** Baseline characteristics for subjects in all psychometric analysis dataset with usable ePRO

| Patient characteristics | Low sex drive (*N* = 595) | | Low energy (*N* = 599) | | Low sex drive or low energy or both (*N* = 694) | |
|---|---|---|---|---|---|---|
| | Mean (SD) | Min–max | Mean (SD) | Min–max | Mean (SD) | Min–max |
| Age (years) | 55.7 (10.7) | 19–85 | 55.1 (11.0) | 19–92 | 55.4 (11.0) | 19–92 |
| | *n* | % | *n* | % | *n* | % |
| Age groups | | | | | | |
| < 65 years | 473 | 80 | 484 | 81 | 554 | 80 |
| ≥ 65 years | 122 | 21 | 115 | 19 | 140 | 20 |
| Race (% White) | 468 | 79 | 472 | 79 | 545 | 79 |
| Region (% North America) | 382 | 64 | 403 | 67 | 452 | 65 |
| | Mean (SD) | Min–max | Mean (SD) | Min–max | Mean (SD) | Min–max |
| Body mass index | 30.6 (4.2) | 11–39 | 30.7 (4.2) | 11–39 | 30.6 (4.1) | 11–39 |
| | Mean (SD) | Min, Max | Mean (SD) | Min, Max | Mean (SD) | Min, Max |
| Total testosterone (average of Visit 1 and Visit 2) (ng/dL) | 202.5 (67.4) | 4.3, 296.8 | 200.5 (68.0) | 4.3, 296.8 | 201.5 (67.1) | 4.3, 296.8 |
| | *n* | % | *n* | % | *n* | % |
| Clinician-identified symptom | | | | | | |
| Decreased sexual drive | 73 | 12 | 0 | 0 | 74 | 11 |
| Low energy | 0 | 0 | 89 | 15 | 91 | 13 |
| Both | 522 | 88 | 510 | 85 | 529 | 76 |

ePRO, electronic patient-reported outcome.

## Descriptive analysis

For the five SAID items, frequency distributions showed mean and median item scores ranging from 2.0 to 2.5 (on a scale of 1–5, with higher scores corresponding to greater sex drive) (Table 2). Inter-item correlations for SAID items ranged from 0.48 to 0.70 (all $p < 0.001$). For the six HED items, frequency distributions showed mean and median item scores for HED ranging from 4.4 to 5.2 (on a 0–10 numerical rating scale, with higher scores corresponding to more energy) (Table 3). Inter-item correlations for all six HED items ranged from 0.38 to 0.88 (all $p < 0.001$).

**Table 2** Item and Total Scale Statistics for SAID at Time 2 (prior to Visit 3 or randomisation)

| Abbreviated items | Mean (SD) (*n* = 553) | Median (*n* = 553) | Min–max (*n* = 553) | Inter-item correlations (*n* = 553) | Factor loadings (*n* = 175) | Item-total correlation (*n* = 553) |
|---|---|---|---|---|---|---|
| . . .THINK about sexual activity? | 2.5 (0.9) | 2.0 | 1–5 | 0.56–0.70 | 0.88 | 0.76 |
| . . .FANTASIZE about sexual activity? | 2.4 (0.9) | 2.0 | 1–5 | 0.48–0.70 | 0.85 | 0.72 |
| . . .PHYSICALLY FEEL a sense of sexual arousal, or a PHYSICAL stirring or tingling of arousal? | 2.3 (1.0) | 2.0 | 1–5 | 0.49–0.59 | 0.77 | 0.65 |
| . . .rate your level of interest in sex? | 2.5 (0.8) | 2.0 | 1–5 | 0.56–0.67 | 0.86 | 0.76 |
| . . .rate your sex drive? | 2.2 (0.7) | 2.0 | 1–5 | 0.48–0.67 | 0.74 | 0.65 |
| SAID Scale Transformed Total Score (0–100) | 34.5 (17.8) | 30.0 | 0–90 | – | – | – |

Scoring: Higher item and total scores correspond to greater sex drive, scores range from 1 to 5 and 0–100, respectively. Factor analysis (performed with data from Protocol Addenda participants only): eigenvalue = 3.4, Variance explained = 67% (*n* = 175). Internal consistency: Cronbach's alpha = 0.87 (*n* = 553). Test–retest Reliability: Intraclass correlation coefficient = 0.75 (*n* = 97, 95% CI = 0.65–0.83). SAID, Sexual Arousal, Interest and Drive; SD, standard deviation.

**Table 3** Item and Total Scale Statistics for HED

| Abbreviated Items (7-day average) | Mean (SD) ($n$ = 541) | Median ($n$ = 541) | Min–max ($n$ = 541) | Inter-item correlations ($n$ = 541) | Factor loadings ($n$ = 204) | Item-total correlation ($n$ = 541) |
|---|---|---|---|---|---|---|
| Full of energy (morning) | 4.8 (1.9) | 4.8 | 0.0–10.0 | 0.38–0.76 | 0.79 | 0.69 |
| Not at all tired (morning) | 5.2 (2.0) | 5.1 | 0.0–10.0 | 0.44–0.73 | 0.77 | 0.71 |
| Full of energy (afternoon) | 5.0 (1.8) | 5.0 | 0.0–10.0 | 0.53–0.88 | 0.91 | 0.82 |
| Not at all tired (afternoon) | 5.2 (1.8) | 5.0 | 0.0–10.0 | 0.48–0.86 | 0.88 | 0.81 |
| Full of energy (evening) | 4.6 (1.8) | 4.6 | 0.1–10.0 | 0.44–0.88 | 0.83 | 0.78 |
| Not at all tired (evening) | 4.6 (1.9) | 4.4 | 0.0–10.0 | 0.38–0.86 | 0.79 | 0.74 |
| HED Transformed Total Score (0–100) | 48.9 (15.6) | 48.8 | 2.9–95.0 | – | – | – |

Scoring: Higher item and total scores correspond to greater energy, scores range from 1 to 10 and 0–100, respectively. Factor analysis (performed with data from Protocol Addenda participants only): Eigenvalue = 4.1, Variance explained = 69% ($n$ = 204). Internal consistency: Cronbach's alpha = 0.91 ($n$ = 541). Test–retest reliability: Intraclass correlation coefficient = 0.88 ($n$ = 127, 95% CI = 0.83–0.91). HED, Hypogonadism Energy Diary; SD, standard deviation.

## Factor analysis

Both EFAs performed yielded only one factor with an eigenvalue > 1.0. For the five SAID items, this factor had an eigenvalue of 3.4 (67% variance explained) and factor loadings ≥ 0.74 (Table 2). For the six HED items, this factor had an eigenvalue of 4.1 (69% variance explained) and factor loadings ≥ 0.77 (Table 3). Both factor analyses suggested that the SAID and HED assess unidimensional concepts; therefore, SAID and HED item scores were summed for SAID and HED total scores, respectively. The mean and median SAID transformed total score (0–100 scale) was 34.5 and 30.0, respectively (Table 2). The mean and median HED transformed total score (0–100 scale) was 48.9 and 48.8 (Table 3).

## Evaluation of reliability

Cronbach's alpha (internal consistency) calculated for the SAID items was 0.87, with item-total correlations ranging from 0.65 to 0.76 (Table 2). Cronbach's alpha calculated for the HED items was 0.91, with item-total correlations ranging from 0.69 to 0.82 (Table 3). The ICC coefficient (test–retest reliability) was 0.75 (95% CI = 0.65–0.83) for the SAID (Table 2) and 0.88 (CI = 0.83–0.91) for the HED (Table 3).

## Construct validity

SAID total scores were significantly ($p < 0.05$) positively correlated with both the IIEF Sexual Desire domain scores ($r$ = 0.64, $n$ = 515) and the PDQ sexual desire question ($r$ = 0.68, $n$ = 519). HED total scores were significantly ($p < 0.05$) positively correlated with both the PDQ 'full of pep/energetic?' ($r$ = 0.76, $n$ = 535) and reversed 'tired?' items ($r$ = 0.66, $n$ = 534) (Table 4).

The mean SAID total scores of men who were identified by their clinicians as having low or

**Table 4** Convergent validity for SAID and HED

| Patient-reported outcome instrument | IIEF Sexual Desire domain ($n$ = 515) $r$ (95% CI) | PDQ Weekly average of 'overall level of sexual desire' item ($n$ = 519) $r$ (95% CI) | PDQ Weekly average of 'full of pep/energetic?' item ($n$ = 535) $r$ (95% CI) | PDQ Weekly average of 'tired?' item* ($n$ = 534) $r$ (95% CI) |
|---|---|---|---|---|
| SAID Scale | 0.64 (0.58–0.69) | 0.68 (0.64–0.73) | | |
| HED | | | 0.76 (0.72–0.79) | 0.66 (0.71–0.61) |

*PDQ 'Tired' item was reverse scored to be consistent with the scoring of the HED in which a higher score corresponds to more positive outcome. SAID, Sexual Arousal, Interest and Drive; HED, Hypogonadism Energy Diary; IIEF, International Index of Erectile Function; PDQ, Psychosexual Daily Questionnaire; $r$, Pearson correlation coefficient; CI, confidence interval.

decreased sex drive ($n = 553$) were significantly different from the mean scores of those men not identified ($n = 80$) (34.5 vs. 42.8, respectively; $p < 0.001$). The mean HED total scores of men who were identified by their clinicians as having low or decreased energy ($n = 541$) were significantly different from the mean scores of those men who were not identified ($n = 64$) (48.9 vs. 60.2, respectively; $p < 0.001$; Table 5).

### Responsiveness

From baseline to end-point, SAID scores significantly ($p < 0.001$) improved in the men with low sex drive and randomised to testosterone solution ($n = 244$), with effect size and SRM of 0.61 and 0.63, respectively (Table 6). From baseline to end-point, HED scores significantly ($p < 0.001$) improved in men with low energy and randomised to testosterone solution ($n = 230$), with effect size and SRM of 0.68 and 0.74, respectively.

### MIC

The two distribution-based methods used to estimate MIC in SAID and HED scores were 0.5 SD and 1 SEM. For SAID in men randomised to testosterone solution ($n = 244$), these were calculated as 9.1 and 5.1, respectively. Corresponding values for HED in men randomised to testosterone solution ($n = 230$), were calculated as 7.9 and 3.7, respectively. MIC was also estimated by 'anchoring' SAID and HED scores to PGI-I-S and PGI-I-E, respectively. For SAID, patients who improved according to the PGI-I-SD scored 9.6 points higher on average than those patients who reported no change or worsening ($p < 0.001$). For HED, patients who improved according to the PGI-I-E scored 8.3 points higher on average than those patients who reported no change or worsening ($p < 0.001$).

## Discussion

The SAID and HED are two new PRO instruments that have been developed to assess two common symptoms of hypogonadism – loss of (or decreased) sex drive and loss of (or decreased) energy. These two instruments were developed according to regulatory guidance that stresses the need for qualitative study to ensure patient input into the item development process (8). The overall goal of this study was to provide the initial step in the extensive and continued psychometric testing required to establish the reliability, validity and responsiveness of all new evaluation instruments (8). Through the addendum to a clinical trial, we explored the factor structure and evaluated the test–retest reliability. In general, the SAID and HED appear to have acceptable psychometric properties when administered to men with hypogonadism with decreased sex drive and decreased energy, respectively.

**Table 5** Known-group validity of SAID and HED

| Patient-reported outcome instrument | Decreased or low sexual drive Mean (SD) | No decreased or low sexual drive Mean (SD) | p |
|---|---|---|---|
| SAID Scale | $n = 553$<br>34.5 (17.8) | $n = 80$<br>42.8 (17.9) | < 0.001 |
| HED | $n = 541$<br>48.9 (15.6) | $n = 64$<br>60.2 (18.3) | < 0.001 |

SAID, Sexual Arousal, Interest and Drive; HED, Hypogonadism Energy Diary; SD, standard deviation.

**Table 6** Evaluation of responsiveness of SAID and HED from baseline to end-point for treatment arm and placebo

| Instrument | *n* | Mean change | p-value | Effect size | SRM |
|---|---|---|---|---|---|
| **SAID** | | | | | |
| Testosterone solution | 244 | 11.0 | < 0.001 | 0.61 | 0.63 |
| Placebo | 257 | 6.7 | < 0.001 | 0.39 | 0.40 |
| **HED** | | | | | |
| Testosterone solution | 230 | 10.6 | < 0.001 | 0.68 | 0.74 |
| Placebo | 243 | 7.4 | < 0.001 | 0.48 | 0.46 |

SAID, Sexual Arousal, Interest and Drive; HED, Hypogonadism Energy Diary; SRM, standardised response mean.

For a PRO instrument to be useful as a treatment evaluation measure, total score ceiling effects (endorsement of the highest possible positive outcome score) should be small in untreated populations, thereby indicating room for positive movement in scores with an intervention or treatment (23). SAID and HED item score distributions showed little to no ceiling effects. In addition, the mean item and total scores for both instruments were in the middle of their respective score ranges. These results suggest that the SAID and HED scores have the potential for positive movement (i.e. improved score) with treatment.

Results of the factor analyses performed through an addendum protocol suggest that total scores derived from summing of SAID or HED item scores (i.e. one domain) are productive for measurement. The internal consistency reliability analyses also suggested one domain for both the SAID and HED with item-total correlations for all items for both $\geq 0.65$. Construct validity was demonstrated for the SAID and HED both in terms of the correlations between their total scores and items or domains measuring similar constructs (i.e. convergent validity) and their ability to discriminate between men who were identified by their clinicians as having low or decreased sex drive or energy as a symptom of their hypogonadism from those who were not identified by their clinicians (i.e. known group validity).

The intent of the SAID, as well as the HED, is to serve as an instrument to evaluate treatment benefit. The most important type of reliability for an evaluative instrument is to demonstrate test–retest reliability or stability between two time points in which no known change in the construct being measured has occurred. Both the SAID and HED demonstrated good test–retest reliability. Following test–retest reliability, the true test of an evaluative instrument is the extent to which the change in scores from baseline to end-point detects an improvement in the construct being measured. Both SAID and HED detected significant improvements in sex drive and energy, respectively, for men randomised to the testosterone treatment group. Moderate to large effect sizes (0.60–0.70) were observed. In the primary study (12), efficacy analyses comparing testosterone solution to placebo for the SAID was significant at alpha level of 0.05 and reached the prespecified more stringent alpha level of 0.01. For the subset with low energy at baseline, participants assigned testosterone solution also showed a statistically significant baseline to end-point improvement in HED scores compared with those assigned placebo ($p = 0.02$); however, the difference did not reach the prespecified significance level of $p < 0.01$ (12). Further research may be needed to clarify whether administration of the HED as a diary three times a day for a 7-day period at baseline and end-point is the optimal timing to capture improvements in the fluctuations in energy reported by men with hypogonadism (9).

The distribution-based methods for determining MIC indicated an approximately 5–9 point change in the SAID and an approximately 4–8 point change in the HED. An anchor-based approach, which was based on patient input, suggested a MIC of approximately 10 points for the SAID and eight points for the HED. Thus, for sample size estimation aimed at demonstrating a change in which patients actually perceive a difference in their sexual drive or their energy level, the MIC may be closer to nine points for the SAID and eight points for the HED on a 0–100 scale. It should be noted, however, that, as with other psychometric properties of instruments, the MIC is sample-specific. Additional research is needed to provide support for the MIC estimates reported in this study.

The results of the previous manuscript by Hayes et al. suggest that the SAID and HED meet current regulatory standards (8) for content validity (9). The research reported here provides preliminary psychometric evidence to support the use of these instruments in the evaluation of treatment for hypogonadism. However, the suitability of PRO instruments as end-points for inclusion in product labelling is determined on a case-by-case basis by regulatory agencies and takes into account study population and desired indications. Moreover, all psychometric analyses are sample-specific and will need to be repeated for each unique patient population of interest. Future studies will increase our understanding of the reliability and validity of the SAID and HED and their potential for aiding clinicians in the assessment of treatment benefit.

## Acknowledgements

This study was funded by Eli Lilly and Company. For permission to use the SAID or HED, please contact copyright@lilly.com. The authors acknowledge Teresa Tartaglione, PharmD (ClinGenuity, LLC, Cincinnati, Ohio) for medical writing assistance during the preparation of this article.

## Author contributions

RPH contributed to the development of the study concept and design, interpreted the results and drafted the manuscript. XN contributed to the study concept and design, supervised the statistical analysis, interpreted the results and critically revised the manuscript for important intellectual content. DEH

and KK contributed to study concept and design, supervised the study, and critically revised the manuscript for important intellectual content. All authors had full access to the data and take full responsibility for the integrity of the data and the accuracy of the data analysis.

## References


1 Morales A, Lunenfeld B; International Society for the Study of the Aging Male. Investigation, treatment and monitoring of late-onset hypogonadism in males: official recommendations of ISSAM. *Aging Male* 2002; **5**: 74–86.

2 Bhasin S, Cunningham GR, Hayes FJ et al. Testosterone therapy in adult men with androgen deficiency syndromes: an endocrine society clinical practice guideline. *J Clin Endocrinol Metab* 2010; **95**: 2536–59.

3 American Association of Clinical Endocrinologists. Medical guidelines for clinical practice for the evaluation and treatment of hypogonadism in adult male patients–2002 update. *Endocr Pract* 2002; **8**: 439–56.

4 Novák A, Brod M, Elbers J. Andropause and quality of life: findings from patient groups and clinical experts. *Maturitas* 2002; **43**: 231–7.

5 Rosen RC, Araujo AB, Conner MK et al. Assessing symptoms of hypogonadism by self-administered questionnaire: qualitative findings in patients and controls. *Aging Male* 2009; **12**: 77–85.

6 Moncada I. Testosterone and men's quality of life. *Aging Male* 2006; **9**: 189–93.

7 Morley JE, Perry HM 3rd, Kevorkian RT, Patrick P. Comparison of screening questionnaires for the diagnosis of hypogonadism. *Maturitas* 2006; **53**: 424–9.

8 Food and Drug Administration. *Guidance for Industry; Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims*. Silver Spring, MD: FDA, 2009.

9 Hayes RP, Henne J, Kinchen KS. Establishing the content validity of the Sexual Arousal, Interest, and Drive Scale (SAID) and the Hypogonadism Energy Diary (HED). *Int J Clin Pract* 2014; **69**: 454–65.

10 Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. *Med Care* 2003; **41**: 582–92.

11 Terwee CB, Roorda LD, Knol DL, De Boer MR, De Vet HC. Linking measurement error to minimal important change of patient-reported outcomes. *J Clin Epidemiol* 2009; **62**: 1062–7.

12 Brock G, Heiselman D, Maggi M et al. Effect of testosterone solution 2% on testosterone concentration, sex drive and energy in hypogonadal men: results of a placebo-controlled study. *J Urol* 2016; **195**: 699–705.

13 Coons SJ, Gwaltney CJ, Hays RD et al. Recommendations on evidence needed to support measurement equivalence between electronic and paper-based patient-reported outcome (PRO) measures: ISPOR ePRO good research practices task force report. *Value Health* 2009; **12**: 419–29.

14 Lee KK, Berman N, Alexander GM, Hull L, Swerdloff RS, Wang C. A simple self-report diary for assessing psychosexual function in hypogonadal men. *J Androl* 2003; **24**: 688–98.

15 Rosen RC, Riley A, Wagner G, Osterloh IH, Kirkpatrick J, Mishra A. The international index of erectile function (IIEF): a multidimensional scale for assessment of erectile dysfunction. *Urology* 1997; **49**: 822–30.

16 Viktrup L, Hayes RP, Wang P, Shen W. Construct validation of patient global impression of severity (PGI-S) and improvement (PGI-I) questionnaires in the treatment of men with lower urinary tract symptoms secondary to benign prostatic hyperplasia. *BMC Urol* 2012; **12**: 30.

17 Yalcin I, Bump RC. Validation of two global impression questionnaires for incontinence. *Am J Obstet Gynecol* 2003; **189**: 98–101.

18 Bonett DG. Sample size requirements for estimating intraclass correlations with desired precision. *Stat Med* 2002; **21**: 1331–5.

19 Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 1979; **83**: 420–8.

20 Crosby RD, Kolotkin RL, Williams GR. An integrated method to determine meaningful changes in health-related quality of life. *J Clin Epidemiol* 2004; **57**: 1153–60.

21 Sloan J, Symonds T, Vargas-Chanes D, Fridley B. Practical guidelines for assessing the clinical significance of health related quality of life changes within clinical trials. *Drug Inf J* 2003; **37**: 23–31.

22 Rejas J, Pardo A, Ruiz MA. Standard error of measurement as a valid alternative to minimally important difference for evaluating the magnitude of changes in patient-reported outcomes measures. *J Clin Epidemiol* 2008; **61**: 350–6.

23 Kirshner B, Guyatt G. A methodological framework for assessing health indices. *J Chronic Dis* 1985; **38**: 27–36.


*Paper received February 2016, accepted April 2016*