

## Predicting low testosterone in aging men: a systematic review

Adam C. Millar MD MScCH, Adrian N.C. Lau MD, George Tomlinson PhD, Alan Kraguljac MSc, David L. Simel MD MHS, Allan S. Detsky MD PhD, Lorraine L. Lipscombe MD MSc

### ABSTRACT

**Background:** Physicians diagnose and treat suspected hypogonadism in older men by extrapolating from the defined clinical entity of hypogonadism found in younger men. We conducted a systematic review to estimate the accuracy of clinical symptoms and signs for predicting low testosterone among aging men.

**Methods:** We searched the MEDLINE and Embase databases (January 1966 to July 2014) for studies that compared clinical features with a measurement of serum testosterone in men. Three of the authors independently reviewed articles for inclusion, assessed quality and extracted data.

**Results:** Among 6053 articles identified, 40 met the inclusion criteria. The prevalence of low testosterone ranged between 2% and 77%. Threshold testosterone levels used for reference standards also varied substantially. The summary likelihood ratio associated with decreased

libido was 1.6 (95% confidence interval [CI] 1.3–1.9), and the likelihood ratio for absence of this finding was 0.72 (95% CI 0.58–0.85). The likelihood ratio associated with the presence of erectile dysfunction was 1.5 (95% CI 1.3–1.8) and with absence of erectile dysfunction was 0.83 (95% CI 0.76–0.91). Of the multiple-item instruments, the ANDROTEST showed both the most favourable positive likelihood ratio (range 1.9–2.2) and the most favourable negative likelihood ratio (range 0.37–0.49).

**Interpretation:** We found weak correlation between signs, symptoms and testosterone levels, uncertainty about what threshold testosterone levels should be considered low for aging men and wide variation in estimated prevalence of the condition. It is therefore difficult to extrapolate the method of diagnosing pathologic hypogonadism in younger men to clinical decisions regarding age-related testosterone decline in aging men.

#### Competing interests:

David Simel receives honoraria for work submitted to JAMAEvidence.com. No other competing interests were declared.

This article has been peer reviewed.

**Accepted:** Mar. 14, 2016

**Online:** June 20, 2016

#### Correspondence to:

Allan Detsky, adetsky@mtsinai.on.ca

CMAJ 2016. DOI:10.1503/cmaj.150262

Male hypogonadism is defined as the presence of low serum testosterone and spermatozoa levels, accompanied by clinical signs and symptoms.<sup>1</sup> The Endocrine Society divides the symptoms and signs of androgen deficiency into 2 groups, based on expert consensus.<sup>1</sup> The first group, which is considered more specific, includes incomplete or delayed sexual development; eunuchoidism; reduced sexual desire (libido); erectile dysfunction; gynecomastia; decreased axillary, facial and pubic hair; small testes (i.e., volume < 5 mL); infertility; low-trauma fracture; low bone mineral density; and hot flashes.<sup>1</sup> The second group includes less specific signs and symptoms, such as decreased energy and motivation, depressed mood, poor concentration and memory, sleep disturbance, mild anemia, reduced muscle bulk and strength, increased body fat or body mass index, and diminished physical performance.<sup>1</sup> Similar definitions have recently been developed by the Canadian Men's Health Foundation Multidisciplinary Guidelines Task Force on Testosterone Deficiency.<sup>2</sup>

In young men, hypogonadism is more commonly characterized by signs and symptoms from the first group, such as reduced libido and erectile dysfunction. This condition is most often caused by testicular or pituitary pathology, including hyperprolactinemia, pituitary or hypothalamic disorders, testicular disease, radiation exposure or genetic diseases such as Klinefelter syndrome.<sup>3</sup> Testosterone replacement is indicated in these cases of “classic hypogonadism,” as it ameliorates the clinical symptoms.<sup>4</sup>

In contrast, although these entities exist in older men too, they are less frequent causes of low testosterone than age-related changes. There is evidence that testosterone levels decline with age in all men, regardless of symptoms, at an estimated rate of 1%–3% per year.<sup>5,6</sup> One study found that serum testosterone levels were below the normal range in 20% of men in their 60s and in close to 50% of men in their 80s.<sup>7</sup> However, the prevalence of symptomatic low testosterone (hypogonadism) is estimated by some to be much lower in this population, at about 2%.<sup>8</sup> Given the high prevalence of low testosterone

and more limited correlation with symptoms in aging men, it is uncertain to what extent this represents a physiologic or pathologic event.<sup>6,7</sup> Moreover, symptoms typically associated with low testosterone are less specific in older men and may be caused by other comorbidities. For example, erectile dysfunction can be the result of vascular insufficiency, neurologic impairment, psychogenic causes or substance use.<sup>9</sup> Conditions such as diabetes mellitus and atherosclerosis are more common in older men, with up to 40% of men over 50 years of age having evidence of vascular insufficiency as the primary cause of their erectile dysfunction.<sup>10</sup> Low libido similarly can result from psychiatric or medical conditions that are more common in older men.<sup>11</sup>

Currently, many clinicians diagnose hypogonadism in older men on the basis of low serum testosterone levels, with or without symptoms, largely on the assumption that this is a pathologic condition requiring treatment. The purpose of this study was to systematically review the available literature to estimate the accuracy and operating characteristics of signs and symptoms for predicting low testosterone in aging men.

## Methods

### Literature search and quality assessment

Using MEDLINE and Embase (January 1966 to July 2014), 3 of the authors (A.C.M., A.N.C.L., A.K.) retrieved articles on patient history or physical examination findings used in the diagnosis of hypogonadism in aging men. Medical Subject Headings and keywords included “hypogonadism,” “androgen deficiency” and relevant terms for various signs and symptoms of male hypogonadism (see Appendix 1, available at [www.cmaj.ca/lookup/suppl/doi:10.1503/cmaj.150262/-/DC1](http://www.cmaj.ca/lookup/suppl/doi:10.1503/cmaj.150262/-/DC1)). This search was supplemented with a manual review of the bibliographies of the identified articles, as well as additional articles that have been used to develop recent guidelines on treatment of male hypogonadism. We included studies that compared clinical findings with a measurement of testosterone (total, bioavailable or free), with a defined range of normal testosterone values. We excluded review articles, as well as those that were nonclinical in context, that focused on therapy for hypogonadism, that had no raw data available for calculations of likelihood ratios or that were limited to specific disease conditions.

### Reference standard for low testosterone

The reference standard for diagnosing low testosterone is 2 values for morning serum total, free or

bioavailable testosterone below a defined normal limit, determined with an accurate and reliable assay.<sup>1</sup> The free or bioavailable testosterone measure is preferred for cases in which alterations in sex hormone binding globulin are suspected; for example, obesity, diabetes and glucocorticoids may lower this protein, whereas cirrhosis, HIV and anticonvulsants may increase it.<sup>1</sup> There is no universally accepted threshold value for low serum testosterone in older men. Testosterone thresholds vary greatly across studies, with many studies not stating how their threshold values were determined. We therefore based the reference standard for low serum testosterone on study-specific thresholds. When available, the bioavailable testosterone measurement was the preferred basis for diagnosis because it is considered by many to be most accurate.<sup>12</sup> If bioavailable testosterone was unavailable, we used total testosterone, and if neither of these was available we used free testosterone. Although free testosterone measured by equilibrium dialysis is considered highly accurate, calculated free testosterone values vary depending on the formula used,<sup>13</sup> and analogue methods of measurement are recognized as having poor accuracy.<sup>14</sup>

### Data extraction and analysis

Three authors (A.C.M., A.N.C.L., A.K.) independently reviewed the selected articles for inclusion and quality. The potential for bias in all studies was assessed using the Quality Assessment of Diagnostic Accuracy tool (Appendix 2, available at [www.cmaj.ca/lookup/suppl/doi:10.1503/cmaj.150262/-/DC1](http://www.cmaj.ca/lookup/suppl/doi:10.1503/cmaj.150262/-/DC1)), adapted to the topic of hypogonadism.<sup>15,16</sup> Two of the authors (A.C.M., A.N.C.L.) independently extracted the data from each of the selected studies and compared the electronic tables of results for any discrepancies, which were resolved by discussion. A third author (G.T.) reviewed the results from each study to ensure internal consistency.

We used the extracted data to calculate sensitivity, specificity and likelihood ratios associated with signs, symptoms or combinations thereof.<sup>17</sup> Additionally, for each study, we calculated the kappa values measuring agreement between low testosterone and the clinical variables examined. We considered signs and symptoms with a positive likelihood ratio greater than 2.0 or a negative likelihood ratio less than 0.5 to be clinically useful.<sup>18</sup> For clinical variables that were examined in only 2 studies, we reported the range. For findings from 3 or more studies, we derived the summary sensitivity, specificity, likelihood ratios and 95% confidence intervals (CIs) using the DerSimonian and Laird random-effects

approach.<sup>19</sup> We used the usual method for ratios of proportions to estimate variances of log-likelihood ratios and used their reciprocals as study weights; we pooled sensitivity and specificity on the logit scale. We assessed heterogeneity between multiple studies examining the same clinical variable with the  $I^2$  statistic and used a test of heterogeneity based on Cochran's  $Q$  statistic.<sup>20</sup> Where findings were reported in 4 or more studies, we also calculated summary measures of diagnostic accuracy from a bivariable model<sup>21,22</sup> using the *mada* package of R (Meta-analysis of diagnostic accuracy, R package, version 0.5.7/r79). Where there were 8 or more studies, we used bivariable meta-regression to assess the dependence of diagnostic accuracy on the mean age of men in the study. All analyses were performed using R version 3.1.

## Results

In total, 6053 articles were identified by the search strategy, of which 40 met the inclusion criteria and were included in the analysis (Figure 1 and Appendices 3 and 4, available at [www.cmaj.ca/lookup/suppl/doi:10.1503/cmaj.150262/-/DC1](http://www.cmaj.ca/lookup/suppl/doi:10.1503/cmaj.150262/-/DC1)). Overall, these articles accounted for a total of 37 565 patients. In the 27 studies that reported mean age, the range was 43 to 82 years. Seven of the studies included men under age 40 years, but in all of these, the mean age was over 40. According to the Quality Assessment of Diagnostic Accuracy system, the most frequent causes of suspected bias were lack of justification for the cut-off used to define hypogonadism, use of nonconsecutive patients and lack of explanation for patients withdrawn from studies (Table 1).

### Prevalence of low testosterone

The prevalence of low testosterone differed widely among the studies, ranging between 2%<sup>62</sup> and 77%.<sup>49</sup> Testosterone thresholds and approaches to measuring testosterone also varied considerably across the studies (total testosterone, 29 studies, range 200–433 ng/dL [6.9–15 nmol/L]; bioavailable testosterone, 9 studies, range 69.4–198.4 ng/dL [2.4–6.9 nmol/L]; free testosterone, 4 studies, range 4.6–7.0 ng/dL [0.16–0.24 nmol/L]) (see Appendices 3 and 4 for details).<sup>23,60</sup> Whereas some researchers assumed a normal distribution of testosterone values to define their threshold level, others used the cut-off values provided by the manufacturer of the testosterone kit. Some investigators relied on testosterone thresholds proposed by consensus guideline statements, whereas others provided no reasoning as to why they selected a specific threshold value.

### Accuracy of signs and symptoms

Of the 40 studies included in this review, 26 examined individual signs and symptoms and their relation to low testosterone in aging men (Table 2). We did not evaluate “classic” signs, such as testicular volume and gynecomastia, because of a lack of adequate data in the included studies. The description of how each sign and symptom was defined can be found in Appendix 3.

The positive likelihood ratio was less than 2.0 for all tests, except for the following: hot flushes (positive likelihood ratio 2.0); decreased pubic hair (2.4); inability to complete chair stands, defined as the ability to stand from a seated position at least 5 times without support from the arms of a chair (2.4); and inability to perform Nottingham power rigs, which involves a device used to measure leg extension power (2.1). No negative likelihood ratio was lower than 0.5, except for decreased vigour (negative likelihood ratio 0.32). The specific test characteristics of each sign and symptom are presented in Appendix 5 (available at [www.cmaj.ca/lookup/suppl/doi:10.1503/cmaj.150262/-/DC1](http://www.cmaj.ca/lookup/suppl/doi:10.1503/cmaj.150262/-/DC1)). For most studies, the test characteristics were close to the line of identity on the receiver operating characteristic curve, which is the diagonal line where sensitivity equals the complement of specificity (i.e., 1 – specificity), indicating that these clinical features added little change to the diagnostic probability of low testosterone (Figure 2). The agreement between low testosterone and individual signs and symptoms was weak, with only 3 of the 70 kappa values being larger than 0.3 and only one having an upper 95% CI above 0.5 (Figure 3). Additionally, we created forest plots of the sensitivity and specificity of each individual sign and symptom (Appendix 6, available at [www.cmaj.ca/lookup/suppl/doi:10.1503/cmaj.150262/-/DC1](http://www.cmaj.ca/lookup/suppl/doi:10.1503/cmaj.150262/-/DC1)).

### Accuracy of multiple-item instruments

Of the 40 included studies, 16 measured the accuracy of prespecified questionnaires of signs and symptoms to identify low testosterone (Table 2). Five instruments to identify low testosterone in older men have been studied. The ANDROTEST appears to have both the most favourable positive likelihood ratio (range 1.9–2.2) and negative likelihood ratio (range 0.37–0.49), but not all instruments have undergone head-to-head comparisons to determine which is the most accurate. Their specific test characteristics can be found in Appendix 7 (available at [www.cmaj.ca/lookup/suppl/doi:10.1503/cmaj.150262/-/DC1](http://www.cmaj.ca/lookup/suppl/doi:10.1503/cmaj.150262/-/DC1)). In addition, we created forest plots of the sensitivity and specificity of the multiple-item instruments (Appendix 6). None of the multiple-item instruments had clinically useful (as defined in this paper) positive or negative likelihood ratios.

### Bivariable models and meta-regression

For all variables, summary estimates of sensitivity and specificity from bivariable models were within 0.015 of values from univariable models. Similarly, values of likelihood ratios from the 2 approaches

were within 0.1 of each other. Age was not a statistically significant predictor of sensitivity or specificity in any of the bivariable meta-regression models for Androgen Deficiency in Aging Males (ADAM) score, erectile dysfunction or libido.

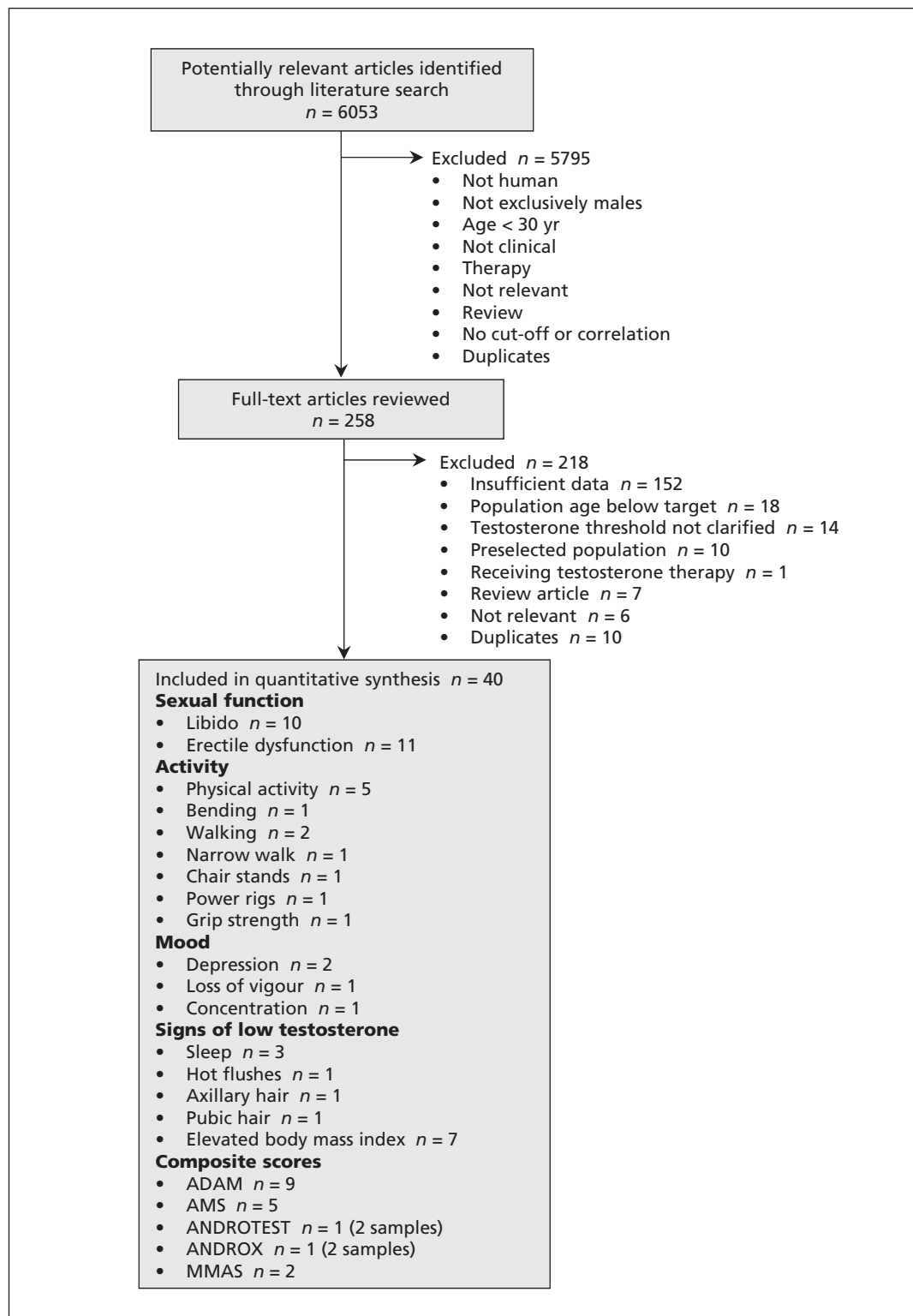


Figure 1: Literature search and study selection. ADAM = Androgen Deficiency in Aging Males, AMS = Aging Males' Symptoms, MMAS = Massachusetts Male Aging Study.

**Table 1:** Assessment of bias for included studies

Study	Quality Assessment of Diagnostic Accuracy tool item*														
	1	2	3	3a	4	5	6	7	8	9	10	11	12	13	14
<b>Studies with individual variables</b>															
Zitzmann et al., 2006 <sup>23</sup>	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-
Ansong et al., 1999 <sup>24</sup>	-	+	+	-	+	+	+	+	+	+	+	+	+	+	+
Fillo et al., 2012 <sup>25</sup>	-	+	+	+	+	+	+	+	+	+	+	+	+	+	-
Mulligan et al., 2006 <sup>26</sup>	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Shi et al., 2014 <sup>27</sup>	-	+	+	+	+	+	+	+	+	+	+	+	+	+	-
Hintikka et al., 2009 <sup>28</sup>	-	+	+	-	+	+	+	+	+	+	+	+	+	+	-
Khaw et al., 2007 <sup>29</sup>	-	+	+	-	+	+	+	+	+	+	+	+	+	+	-
Orwoll et al., 2006 <sup>30</sup>	-	+	+	-	+	+	+	+	+	+	+	+	+	+	-
Ponholzer et al., 2005 <sup>31</sup>	-	+	+	-	+	+	+	+	+	+	+	+	+	+	-
Araujo et al., 2007 <sup>32</sup>	-	+	+	+	+	+	+	+	+	+	+	+	+	+	-
Clapauch et al., 2008 <sup>33</sup>	-	+	+	+	+	+	+	+	+	+	+	+	+	+	-
Ghazi et al., 2012 <sup>34</sup>	-	+	+	+	+	+	+	+	+	+	+	+	+	+	-
Paick et al., 2007 <sup>35</sup>	-	+	+	-	+	+	+	+	+	+	+	+	+	+	+
Müezzinoglu et al., 2007 <sup>36</sup>	-	+	+	-	+	+	+	+	+	+	+	+	+	+	+
Liu et al., 2009 <sup>37</sup>	-	+	+	+	+	+	+	+	+	+	+	+	+	+	-
Arrabal-Polo et al., 2012 <sup>38</sup>	-	+	+	-	+	+	+	+	+	+	+	+	+	+	-
Tajar et al., 2010 <sup>39</sup>	-	+	+	+	+	+	+	+	+	+	+	+	+	+	-
Kratzik et al., 2007 <sup>40</sup>	-	+	+	+	+	+	+	+	+	+	+	+	+	+	-
Hall et al., 2008 <sup>41</sup>	-	+	+	+	+	+	+	+	+	+	+	+	+	+	-
Acar et al., 2004 <sup>42</sup>	-	+	+	-	+	+	+	+	+	+	+	+	+	+	-
Drinka et al., 1993 <sup>43</sup>	-	+	+	+	+	+	+	+	+	+	+	+	+	+	-
Hyde et al., 2012 <sup>44</sup>	-	+	+	+	+	+	+	+	+	+	+	+	+	+	-
Maggio et al., 2011 <sup>45</sup>	-	+	+	-	+	+	+	+	+	+	+	+	+	+	-
Travison et al., 2006 <sup>46</sup>	-	+	+	-	+	+	+	+	+	+	+	+	+	+	-
Rhoden et al., 2002 <sup>47</sup>	-	+	+	+	+	+	+	+	+	+	+	+	+	+	-
Allan et al., 2006 <sup>48</sup>	-	+	+	+	+	+	+	+	+	+	+	+	+	+	-
<b>Studies with combination of variables</b>															
Kratzik et al., 2005 <sup>49</sup>	-	+	+	+	+	+	+	+	+	+	+	+	+	+	-
Fillo et al., 2012 <sup>25</sup>	-	+	+	+	+	+	+	+	+	+	+	+	+	+	-
Chu et al., 2008 <sup>50</sup>	-	+	+	+	+	+	+	+	+	+	+	+	+	+	-
Tancredi et al., 2004 <sup>51</sup>	-	+	+	+	+	+	+	+	+	+	+	+	+	+	-
Blümel et al., 2009 <sup>52</sup>	-	+	+	+	+	+	+	+	+	+	+	+	+	+	-
Chen et al., 2013 <sup>53</sup>	-	+	+	+	+	+	+	+	+	+	+	+	+	+	-
Martínez-Jabaloyas et al., 2007 <sup>54</sup>	-	+	+	+	+	+	+	+	+	+	+	+	+	+	-
Morley et al., 2000 <sup>55</sup>	-	+	+	+	+	+	+	+	+	+	+	+	+	+	-
Goel et al., 2009 <sup>56</sup>	-	+	+	+	+	+	+	+	+	+	+	+	+	+	-
Chueh et al., 2012 <sup>57</sup>	-	+	+	+	+	+	+	+	+	+	+	+	+	+	-
Clapauch et al., 2008 <sup>33</sup>	-	+	+	+	+	+	+	+	+	+	+	+	+	+	-
Smith et al., 2000 <sup>58</sup>	-	+	+	-	+	+	+	+	+	+	+	+	+	+	-
Rabah et al., 2009 <sup>59</sup>	-	+	+	+	+	+	+	+	+	+	+	+	+	+	-
Corona et al., 2006 <sup>60</sup>	+	+	+	-	+	+	+	+	+	+	+	+	+	+	-
Zengerling et al., 2012 <sup>51</sup>	-	+	+	-	+	+	+	+	+	+	+	+	+	+	-
Araujo et al., 2004 <sup>62</sup>	-	+	+	+	+	+	+	+	+	+	+	+	+	+	-

Note: + = bias assessment adequately addressed; - = bias assessment inadequately addressed.

\*A detailed description of the tool items is provided in Appendix 2, available at [www.cmaj.ca/lookup/suppl/doi:10.1503/cmaj.150262/-DC1](http://www.cmaj.ca/lookup/suppl/doi:10.1503/cmaj.150262/-DC1).

## Interpretation

Our review of 40 studies showed weak associations between signs and symptoms and serum testosterone levels in aging men. The unimpressive positive likelihood ratios may be because many symptoms and signs of low testosterone are non-specific — the result of other comorbid conditions that commonly occur in older men. In addition,

weak negative likelihood ratios may be because a high proportion of older men — many of whom are asymptomatic — have lower levels of testosterone than the currently proposed thresholds derived from younger men. The true threshold below which serum testosterone is abnormal may be lower in older men, and thresholds at which different signs and symptoms occur may also vary.<sup>63</sup>

This review raises the following important ques-

**Table 2:** Diagnostic accuracy of symptoms, signs and composite measures for hypogonadism\*

Finding	No. of studies (no. of patients)†	Sensitivity (95% CI)	Specificity (95% CI)	Positive LR (95% CI)	Negative LR (95% CI)
<b>Symptoms</b>					
Hot flushes	1 (434)	0.35 (0.25–0.46)	0.83 (0.79–0.87)	2.0 (1.4–3.0)	0.79 (0.66–0.93)
Decreased concentration	1 (434)	0.54 (0.46–0.62)	0.70 (0.64–0.75)	1.8 (1.4–2.3)	0.66 (0.54–0.80)
Erectile dysfunction	11 (6918)	0.52 (0.39–0.65)	0.67 (0.50–0.80)	1.5 (1.3–1.8) <i>I</i> <sup>2</sup> = 74%	0.83 (0.76–0.91) <i>I</i> <sup>2</sup> = 44%
Depression	2 (550)	0.26–0.54	0.69–0.91	1.7–2.8	0.66–0.82
Limited walking	2 (5712)	0.11–0.23	0.88–0.93	1.7–2.0	0.87–0.95
Decreased libido	10 (8676)	0.51 (0.35–0.67)	0.68 (0.54–0.79)	1.6 (1.3–1.9) <i>I</i> <sup>2</sup> = 72%	0.72 (0.58–0.85) <i>I</i> <sup>2</sup> = 83%
Inability to bend	1 (3120)	0.09 (0.06–0.12)	0.94 (0.93–0.95)	1.5 (1.1–2.1)	0.97 (0.94–1.0)
Decreased vigour	1 (434)	0.88 (0.84–0.91)	0.38 (0.30–0.46)	1.4 (1.2–1.6)	0.32 (0.22–0.46)
Decreased physical activity	5 (7971)	0.36 (0.26–0.48)	0.74 (0.65–0.82)	1.4 (1.3–1.6) <i>I</i> <sup>2</sup> = 10%	0.85 (0.79–0.91) <i>I</i> <sup>2</sup> = 36%
Impaired sleep	3 (2971)	0.25 (0.12–0.49)	0.78 (0.66–0.87)	1.1 (0.63–2.1) <i>I</i> <sup>2</sup> = 92%	0.94 (0.75–1.2) <i>I</i> <sup>2</sup> = 92%
<b>Signs</b>					
Decreased pubic hair	1 (99)	0.63 (0.28–0.87)	0.74 (0.64–0.82)	2.4 (1.3–4.5)	0.51 (0.21–1.3)
Inability to complete chair stands	1 (2587)	0.05 (0.04–0.07)	0.98 (0.97–0.98)	2.4 (1.5–3.8)	0.97 (0.95–0.99)
Inability to perform power rigs	1 (2587)	0.03 (0.02–0.04)	0.99 (0.98–0.99)	2.1 (1.1–3.9)	0.99 (0.97–1.0)
Decreased axillary hair	1 (100)	0.50 (0.20–0.80)	0.73 (0.63–0.81)	1.8 (0.85–4.0)	0.69 (0.34–1.4)
Elevated BMI	7 (6517)	0.50 (0.34–0.66)	0.71 (0.62–0.80)	1.8 (1.6–1.9) <i>I</i> <sup>2</sup> = 0%	0.71 (0.62–0.83) <i>I</i> <sup>2</sup> = 85%
Impaired balance	1 (2587)	0.13 (0.11–0.16)	0.93 (0.91–0.94)	1.8 (1.4–2.3)	0.94 (0.91–0.97)
Decreased grip strength	1 (2587)	0.02 (0.01–0.04)	0.98 (0.97–0.98)	1.0 (0.56–1.8)	1.00 (0.99–1.0)
<b>Combinations of findings</b>					
ANDROTEST	2 (879)	0.68–0.76	0.65–0.66	1.9–2.2	0.37–0.49
ANDROX	2 (1387)	0.48–0.70	0.64–0.74	1.8–2.0	0.47–0.70
ADAM	9 (8327)	0.84 (0.81–0.87)	0.29 (0.22–0.38)	1.2 (1.0–1.4) <i>I</i> <sup>2</sup> = 93%	0.57 (0.37–0.85) <i>I</i> <sup>2</sup> = 86%
MMAS	2 (3330)	0.53–0.71	0.53–0.59	1.3–1.5	0.55–0.79
AMS	5 (1853)	0.74 (0.57–0.85)	0.49 (0.28–0.69)	1.5 (0.94–2.4) <i>I</i> <sup>2</sup> = 93%	0.59 (0.27–1.1) <i>I</i> <sup>2</sup> = 89%

Note: ADAM = Androgen Deficiency in Aging Males, AMS = Aging Males' Symptoms, BMI = body mass index, CI = confidence interval, *I*<sup>2</sup> = heterogeneity, LR = likelihood ratio, MMAS = Massachusetts Male Aging Study.

\*For the results from individual studies, see Appendices 5 and 7, available at [www.cmaj.ca/lookup/suppl/doi:10.1503/cmaj.150262/-/DC1](http://www.cmaj.ca/lookup/suppl/doi:10.1503/cmaj.150262/-/DC1).

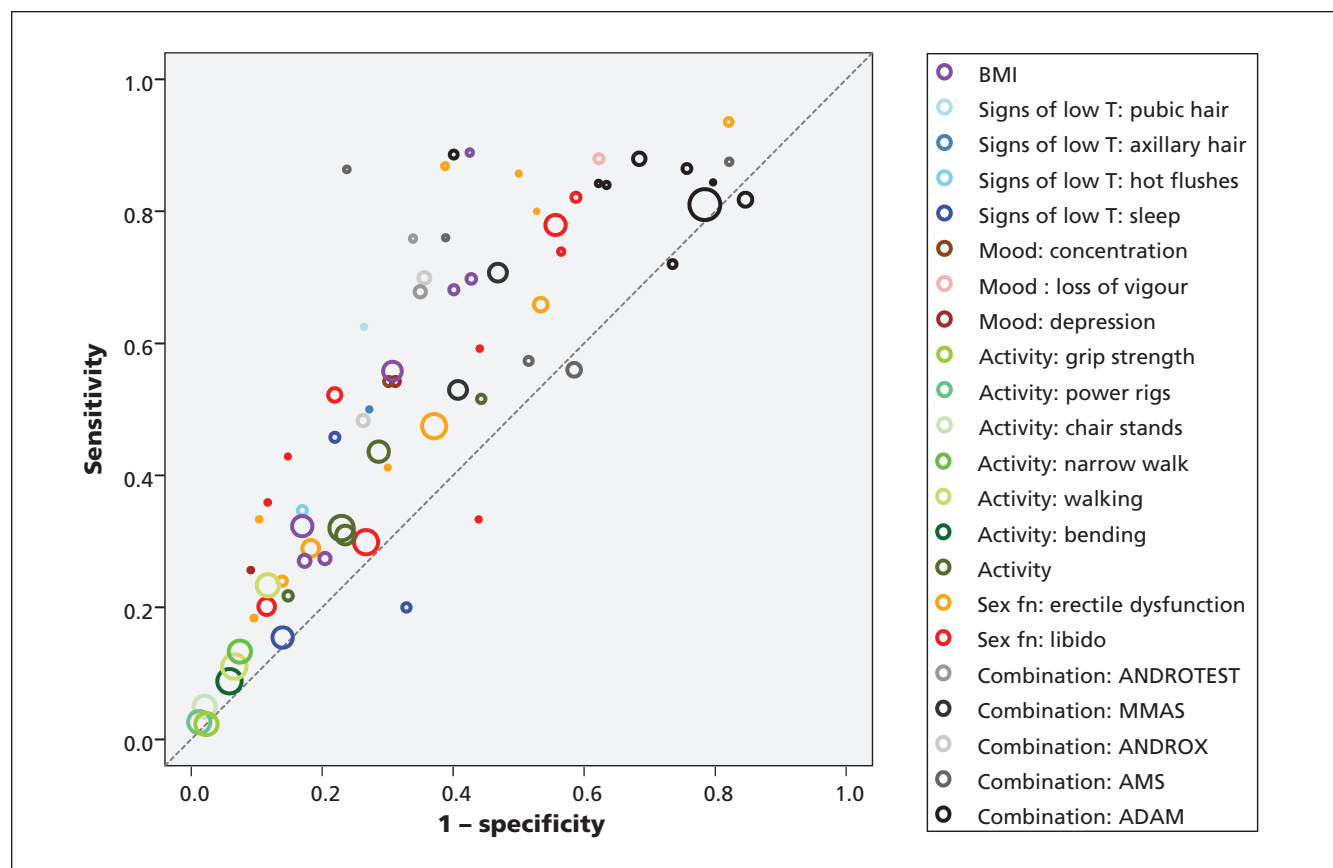
†For variables with 4 or more studies, a bivariable model was used; for variables with 3 studies, a univariable model was used; and for variables with 2 studies, a range is displayed.

tion: In the face of a low correlation between symptoms and biochemical testosterone levels, how should low testosterone in older men be defined and interpreted? To answer this question, we first require rigorously performed studies comparing the signs and symptoms of hypogonadism in aging men to the results of standardized testosterone assays to determine whether a correlation exists. Next, we need to determine the threshold testosterone level that discriminates those with the syndrome from those without. Third, we need rigorously performed large-scale trials to determine the benefits and risks of testosterone replacement in men categorized as having testosterone deficiency.

Because the likelihood ratios of the clinical findings were mostly between 0.5 and 2.0, the estimate of prevalence becomes the main determinant of post-test probability. In other words, the post-test probability is not altered from the pretest probability in any meaningful way. This highlights the importance of generating better information on the actual prevalence of clin-

ically significant low testosterone in older men.

A high-quality study by Wu and associates<sup>8</sup> could not be included in the present study because published raw data necessary for the calculation of sensitivity, specificity and likelihood ratios were lacking. That study of 3369 older men suggested that, compared with the absence of any symptoms, combined symptoms of poor morning erection, low sexual desire and erectile dysfunction were associated with a modest odds ratio of 1.7 (95% CI 1.1–2.6) for androgen deficiency, based on a serum total testosterone of less than 11 nmol/L (317 ng/dL) in men between the ages of 40 and 79 years.<sup>8</sup> Similar to our findings, Wu and associates<sup>8</sup> noted a “weak overall association between symptoms and testosterone levels in this population.” They also stated that there was “substantial overlap between late-onset hypogonadism and nonspecific symptoms of aging.” They concluded that applying their criteria could “guard against the excessive diagnosis of hypogonadism and curb the injudicious use of testosterone therapy in older men.”<sup>8</sup>



**Figure 2: Receiver operating characteristics of individual findings.** The figure has a point for each of the individual reported sensitivities and specificities for the studies included in the review. Findings for combinations are plotted in greys and black, for sexual function in red and orange, for activities in greens, for mood in browns, for signs of low testosterone in blues and for body mass index in purple. The area of each circle is proportional to the sample size used to calculate the findings. The diagonal line shows where sensitivity equals the complement of specificity (i.e.,  $1 - \text{specificity}$ ), which indicates that a clinical feature added little change to the diagnostic probability of low testosterone. The particular sign or symptom for any point can be located by referring to Figure 3, which uses the same colour coding. ADAM = Androgen Deficiency in Aging Males, AMS = Aging Males' Symptoms, BMI = body mass index, MMAS = Massachusetts Male Aging Study, Sex fn = sexual function, T = testosterone.

The diagnosis of low testosterone in older men is complicated by controversies surrounding the potential benefits and harms of testosterone replacement therapy. A joint US Food and Drug Administration advisory committee recently

stated that “both safety and efficacy of testosterone replacement in older men has not been established.”<sup>64</sup> It recommended that a potential signal regarding cardiovascular risk be included in labelling and that “the use of testosterone

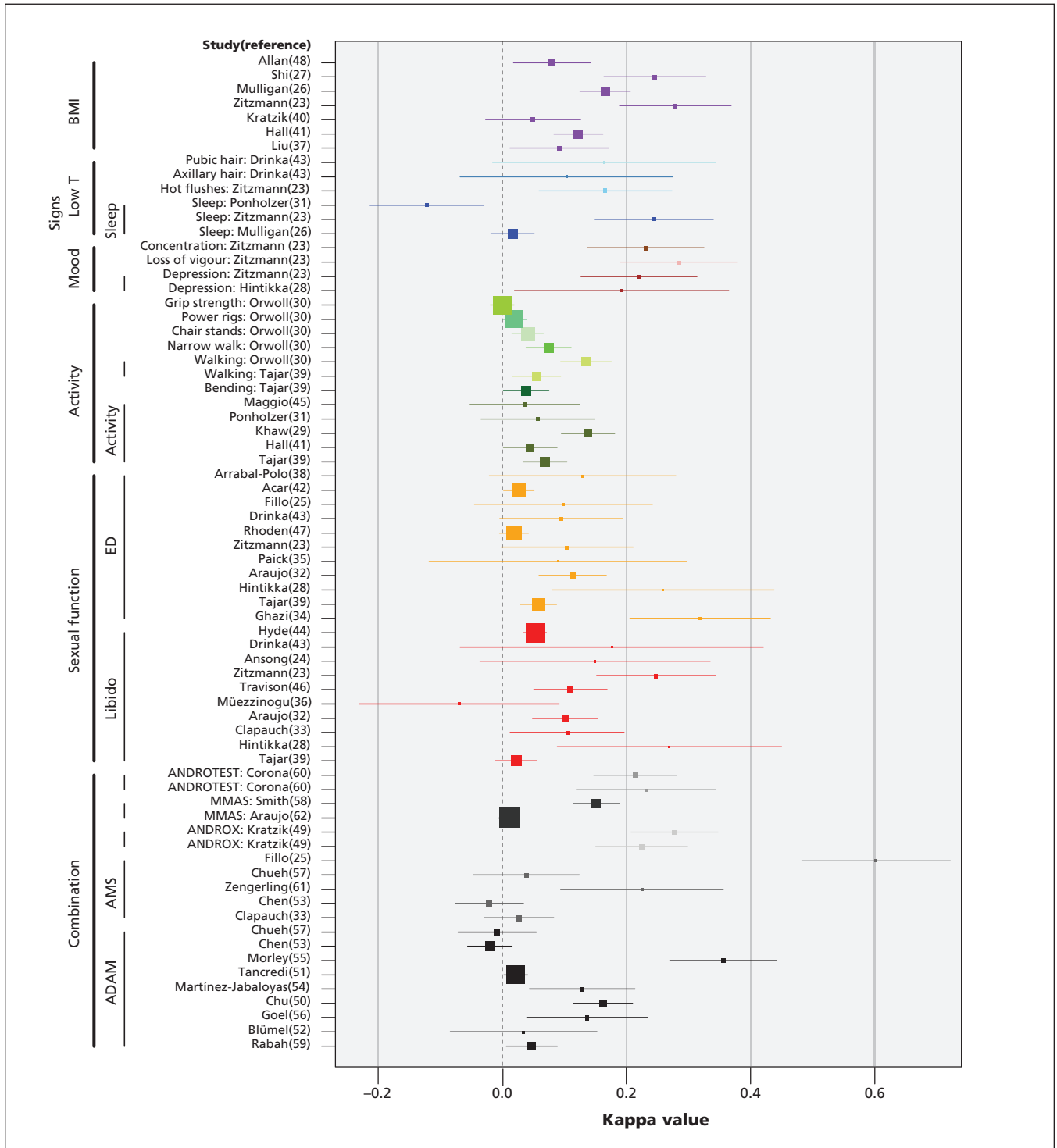


Figure 3: Display of kappa values for each study. For each finding, the square shows the estimated kappa value between the 2 dichotomous variables “low testosterone” and “positive sign or symptom”; horizontal lines show the 95% confidence intervals, and the area of each square is proportional to the estimated variance of the kappa value. Each sign or symptom is plotted in a different colour, with similar signs and symptoms sharing the same colour family (see Figure 2). ADAM = Androgen Deficiency in Aging Males, AMS = Aging Males’ Symptoms, BMI = body mass index, ED = erectile dysfunction, MMAS = Massachusetts Male Aging Study, T = testosterone.



replacement should exclude men with age-related testosterone decline.”<sup>64</sup>

The recently published Testosterone Trials consisted of 3 trials that examined the effects of testosterone therapy in symptomatic men 65 years of age and older with total testosterone levels less than 275 ng/dL (9.54 nmol/L).<sup>65</sup> These trials showed modest improvements in measures of sexual function, although these effects declined over time. Small improvements in mood, depressive symptoms and walking distance were also reported. As noted in the accompanying editorial,<sup>66</sup> the clinical significance of these treatment responses remains unclear, and no benefits for overall vitality were noted. In addition, the sample sizes were too small to determine potential risks of testosterone therapy in this population.<sup>65</sup> These findings further support the lack of clarity regarding how to define and treat low testosterone in older men.

### Limitations

This review had several important limitations. First, the studies included in the review used different assays for measuring testosterone and different thresholds for defining abnormal values. Recognizing the substantial variability among testosterone assays, the US Centers for Disease Control and Prevention is leading the Hormone Standardization Project, which will help to standardize testosterone measurement in the United States.<sup>67</sup> Second, many of the studies had modest sample sizes and used nonconsecutive patients, which limits the quality of their data. Third, there were no data on the accuracy of physical findings such as gynecomastia (no studies) and testicular size (one study,<sup>68</sup> which was ultimately excluded because raw data were unavailable). Fourth, many of the studies had heterogeneous definitions for terms relating to signs and symptoms, such as libido, which can be difficult to quantify objectively. Fifth, many of the studies assessed in this paper did not examine patients with other major comorbidities.

### Conclusion

This systematic review, based on all relevant existing data, highlights the current lack of clarity regarding the definition and management of age-related declines in testosterone levels. Weak correlations between signs, symptoms and testosterone levels, uncertainty about what threshold testosterone levels should be considered low for aging men and wide variation in estimated prevalence of the condition make it difficult to extrapolate the method of diagnosing pathologic hypogonadism in younger men to clinical decisions regarding age-related testosterone decline in aging men.

### References

1. Bhasin S, Cunningham GR, Hayes FJ, et al. Testosterone therapy in men with androgen deficiency syndromes: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab* 2010;95:2536-59.
2. Morales A, Bebb RA, Manjoo P, et al. Diagnosis and management of testosterone deficiency syndrome in men: clinical practice guideline. *CMAJ* 2015;187:1369-77.
3. Yin A, Swerdloff R. Treating hypogonadism in younger males. *Expert Opin Pharmacother* 2010;11:1529-40.
4. Wang C, Cunningham G, Dobs A, et al. Long-term testosterone gel (AndroGel) treatment maintains beneficial effects on sexual function and mood, lean and fat mass, and bone mineral density in hypogonadal men. *J Clin Endocrinol Metab* 2004;89:2085-98.
5. Wu FC, Tajar A, Pye SR, et al. Hypothalamic-pituitary-testicular axis disruptions in older men are differentially linked to age and modifiable risk factors: the European Male Aging Study. *J Clin Endocrinol Metab* 2008;93:2737-45.
6. Feldman HA, Longcope C, Derby CA, et al. Age trends in the level of serum testosterone and other hormones in middle-aged men: longitudinal results from the Massachusetts Male Aging Study. *J Clin Endocrinol Metab* 2002;87:589-98.
7. Harman SM, Metter EJ, Tobin JD, et al. Longitudinal effects of aging on serum total and free testosterone levels in healthy men. Baltimore Longitudinal Study of Aging. *J Clin Endocrinol Metab* 2001;86:724-31.
8. Wu FC, Tajar A, Beynon JM, et al. Identification of late-onset hypogonadism in middle-aged and elderly men. *N Engl J Med* 2010;363:123-35.
9. McVary KT. Clinical practice. Erectile dysfunction. *N Engl J Med* 2007;357:2472-81.
10. Kaiser FE, Viosca SP, Morley JE, et al. Impotence and aging: clinical and hormonal factors. *J Am Geriatr Soc* 1988;36:511-9.
11. Meuleman EJ, van Lankveld JJ. Hypoactive sexual desire disorder: an underestimated condition in men. *BJU Int* 2005;5:291-6.
12. Morales A, Bella AJ, Chun S, et al. A practical guide to diagnosis, management and treatment of testosterone deficiency for Canadian physicians. *Can Urol Assoc J* 2010;4:269-75.
13. Ly LP, Sartorius G, Hull L, et al. Accuracy of calculated free testosterone formulae in men. *Clin Endocrinol (Oxf)* 2010;73:382-8.
14. Rosner W, Auchus RJ, Azziz R, et al. Position statement: utility, limitations, and pitfalls in measuring testosterone: an Endocrine Society position statement. *J Clin Endocrinol Metab* 2007;92:405-13.
15. Whiting P, Rutjes AW, Reitsma JB, et al. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol* 2003;3:25.
16. Whiting PF, Weswood ME, Rutjes AW, et al. Evaluation of QUADAS, a tool for the quality assessment of diagnostic accuracy studies. *BMC Med Res Methodol* 2006;6:9.
17. Sackett DL. The rational clinical examination. A primer on the precision and accuracy of the clinical examination. *JAMA* 1992;267:2638-44.
18. Jaeschke R, Guyatt GH, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? The Evidence-Based Medicine Working Group. *JAMA* 1994;271:703-7.
19. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986;7:177-88.
20. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med* 2002;21:1539-58.
21. Reitsma JB, Glas AS, Rutjes AW, et al. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol* 2005;58:982-90.
22. Simel DL, Bossuyt PM. Differences between univariate and bivariate models for summarizing diagnostic accuracy may not be large. *J Clin Epidemiol* 2009;62:1292-300.
23. Zitzmann M, Faber S, Nieschlag E. Association of specific symptoms and metabolic risks with serum testosterone in older men. *J Clin Endocrinol Metab* 2006;91:4335-43.
24. Ansong KS, Punwaney RB. An assessment of the clinical relevance of serum testosterone level determination in the evaluation of men with low sexual drive. *J Urol* 1999;162:719-21.
25. Fillo J, Breza J, Levckikova M, et al. Occurrence of erectile dysfunction, testosterone deficiency syndrome and metabolic syndrome in patients with abdominal obesity. Where is a sufficient level of testosterone? *Int Urol Nephrol* 2012;44:1113-20.
26. Mulligan T, Frick MF, Zuraw QC, et al. Prevalence of hypogonadism in males aged at least 45 years: the HIM study. *Int J Clin Pract* 2006;60:762-9.

27. Shi MD, Chao JK, Ma MC, et al. Factors associated with sex hormones and erectile dysfunction in male Taiwanese participants with obesity. *J Sex Med* 2014;11:230-9.
28. Hintikka J, Niskanen L, Koivumaa-Honkanen H, et al. Hypogonadism, decreased sexual desire, and long-term depression in middle-aged men. *J Sex Med* 2009;6:2049-57.
29. Khaw KT, Dowsett M, Folkard E, et al. Endogenous testosterone and mortality due to all causes, cardiovascular disease, and cancer in men: European prospective investigation into cancer in Norfolk (EPIC-Norfolk) Prospective Population Study. *Circulation* 2007;116:2694-701.
30. Orwoll E, Lambert LC, Marshall LM, et al. Endogenous testosterone levels, physical performance, and fall risk in older men. *Arch Intern Med* 2006;166:2124-31.
31. Pohnholzer A, Plas E, Schatzl G, et al. Relationship between testosterone serum levels and lifestyle in aging men. *Aging Male* 2005;8:190-3.
32. Araujo AB, Esche GR, Kupelian V, et al. Prevalence of symptomatic androgen deficiency in men. *J Clin Endocrinol Metab* 2007;92:4241-7.
33. Clapauch R, Braga DJ, Marinheiro LP, et al. Risk of late-onset hypogonadism (andropause) in Brazilian men over 50 years of age with osteoporosis: usefulness of screening questionnaires. *Arq Bras Endocrinol Metabol* 2008;52:1439-47.
34. Ghazi S, Zohdy W, Elkhat Y, et al. Serum testosterone levels in diabetic men with and without erectile dysfunction. *Andrologia* 2012;44:373-80.
35. Paick JS, Yang JH, Kim SW, et al. Severity of erectile dysfunction in married impotent patients: interrelationship with anthropometry, hormones, metabolic profiles and lifestyle. *Int J Urol* 2007;14:48-53.
36. Müezzinoğlu T, Gümiş B, Temeltaş G, et al. A relationship of sex hormone levels and erectile dysfunction: Which tests should be done routinely? *Yonsei Med J* 2007;48:1015-9.
37. Liu CC, Wu WJ, Lee YC, et al. The prevalence of and risk factors for androgen deficiency in aging Taiwanese men. *J Sex Med* 2009;6:936-46.
38. Arrabal-Polo MÁ, Arias-Santiago S, López-Carmona Pintado F, et al. Metabolic syndrome, hormone levels, and inflammation in patients with erectile dysfunction. *Sci World J* 2012;2012:272769.
39. Tajar A, Forti G, O'Neill TW, et al. Characteristics of secondary, primary, and compensated hypogonadism in aging men: evidence from the European Male Ageing Study. *J Clin Endocrinol Metab* 2010;95:1810-8.
40. Kratzik CW, Schatzl G, Lackner JE, et al. Mood changes, body mass index and bioavailable testosterone in healthy men: results of the Androx Vienna Municipality Study. *BJU Int* 2007;100:614-8.
41. Hall SA, Esche GR, Araujo AB, et al. Correlates of low testosterone and symptomatic androgen deficiency in a population-based sample. *J Clin Endocrinol Metab* 2008;93:3870-7.
42. Acar D, Cayan S, Bozlu M, et al. Is routine hormonal measurement necessary in initial evaluation of men with erectile dysfunction? *Arch Androl* 2004;50:247-53.
43. Drinka PJ, Voeks S, Bauwens S, et al. Sensitivity and positive predictive value of clinical signs of hypogonadism in elderly men. *South Med J* 1993;86:1264-5.
44. Hyde Z, Flicker L, Hankey GJ, et al. Prevalence and predictors of sexual problems in men aged 75-95 years: a population-based study. *J Sex Med* 2012;9:442-53.
45. Maggio M, Ceda GP, Lauretani F, et al. Gonadal status and physical performance in older men. *Aging Male* 2011;14:42-7.
46. Travison TG, Morley JE, Araujo AB, et al. The relationship between libido and testosterone levels in aging men. *J Clin Endocrinol Metab* 2006;91:2509-13.
47. Rhoden EL, Teloken C, Mafessoni R, et al. Is there any relation between serum levels of total testosterone and the severity of erectile dysfunction? *Int J Impot Res* 2002;14:167-71.
48. Allan CA, Strauss BJ, Burger HG, et al. The association between obesity and the diagnosis of androgen deficiency in symptomatic ageing men. *Med J Aust* 2006;185:424-7.
49. Kratzik C, Heinemann LA, Saad F, et al. Composite screener for androgen deficiency related to the Aging Males' Symptoms Scale. *Aging Male* 2005;8:157-61.
50. Chu LW, Tam S, Kung AW, et al. A short version of the ADAM Questionnaire for androgen deficiency in Chinese men. *J Gerontol A Biol Sci Med Sci* 2008;63:426-31.
51. Tancredi A, Reginster JY, Schleich F, et al. Interest of the Androgen Deficiency in Aging Males (ADAM) questionnaire for the identification of hypogonadism in elderly community-dwelling male volunteers. *Eur J Endocrinol* 2004;151:355-60.
52. Blümel JE, Chedraui P, Gili SA, et al. Is the Androgen Deficiency of Aging Men (ADAM) questionnaire useful for the screening of partial androgenic deficiency of aging men? *Maturitas* 2009;63:365-8.
53. Chen W, Liu ZY, Wang LH, et al. Are the Aging Male's Symptoms (AMS) scale and the Androgen Deficiency in the Aging Male (ADAM) questionnaire suitable for the screening of late-onset hypogonadism in aging Chinese men? *Aging Male* 2013;16:92-6.
54. Martínez-Jabaloyas JM, Queipo-Zaragoza A, Rodríguez-Navarro R, et al. Relationship between the Saint Louis University ADAM questionnaire and sexual hormonal levels in a male outpatient population over 50 years of age. *Eur Urol* 2007;52:1760-7.
55. Morley JE, Charlton E, Patrick P, et al. Validation of a screening questionnaire for androgen deficiency in aging males. *Metabolism* 2000;49:1239-42.
56. Goel A, Sinha RJ, Dalela D, et al. Andropause in Indian men: a preliminary cross-sectional study. *Urol J* 2009;6:40-4, discussion 44-6.
57. Chueh KS, Huang SP, Lee YC, et al. The comparison of the Aging Male Symptoms (AMS) scale and Androgen Deficiency in the Aging Male (ADAM) questionnaire to detect androgen deficiency in middle-aged men. *J Androl* 2012;33:817-23.
58. Smith KW, Feldman HA, McKinlay JB. Construction and field validation of a self-administered screener for testosterone deficiency (hypogonadism) in ageing men. *Clin Endocrinol (Oxf)* 2000;53:703-11.
59. Rabah DM, Arafa MA. Validation of an Arabic ADAM questionnaire for androgen deficiency screening in the Arab community. *Aging Male* 2009;12:95-9.
60. Corona G, Mannucci E, Petrone L, et al. ANDROTEST: a structured interview for the screening of hypogonadism in patients with sexual dysfunction. *J Sex Med* 2006;3:706-15.
61. Zengerling F, Schrader AJ, Cronauer MV, et al. The "Aging Males' Symptoms" Scale (AMS): predictive value for lowered circulating androgens. *Aging Male* 2012;15:253-7.
62. Araujo AB, O'Donnell AB, Brambilla DJ, et al. Prevalence and incidence of androgen deficiency in middle-aged and older men: estimates from the Massachusetts Male Aging Study. *J Clin Endocrinol Metab* 2004;89:5920-6.
63. Finkelstein JS, Lee H, Burnett-Bowie SA, et al. Gonadal steroids and body composition, strength, and sexual function in men. *N Engl J Med* 2013;369:1011-22.
64. Garnick MB. Testosterone replacement therapy faces FDA scrutiny. *JAMA* 2015;313:563-4.
65. Snyder PJ, Bhasin S, Cunningham GR, et al. Effects of testosterone treatment in older men. *N Engl J Med* 2016;374:611-24.
66. Orwoll ES. Establishing a framework — Does testosterone supplementation help older men? *N Engl J Med* 2016;374:682-3.
67. Standardizing hormone measurements. Atlanta: Centers for Disease Control and Prevention, National Center for Environmental Health Division of Laboratory Sciences; 2014. Available: [www.cdc.gov/labstandards/pdf/hs/HoSt\\_Brochure.pdf](http://www.cdc.gov/labstandards/pdf/hs/HoSt_Brochure.pdf) (accessed 2014 Apr. 13).
68. Mahmoud AM, Goemaere S, El-Garem Y, et al. Testicular volume in relation to hormonal indices of gonadal function in community-dwelling elderly men. *J Clin Endocrinol Metab* 2003;88:179-84.

**Affiliations:** Department of Medicine (Millar, Tomlinson, Detsky, Lipscombe) and Institute of Health Policy, Management and Evaluation (Tomlinson, Detsky, Lipscombe), University of Toronto, Toronto, Ont.; Department of Medicine (Millar, Lau, Tomlinson, Kraguljac, Detsky), Mount Sinai Hospital and University Health Network, Toronto, Ont.; Durham Veterans Affairs Medical Center (Simel), Durham, NC; Department of Medicine (Simel), Duke University, Durham, NC; Department of Medicine (Lipscombe), Women's College Research Institute of Women's College Hospital, Toronto, Ont.

**Contributors:** Adam Millar, Allan Detsky, Adrian Lau, David Simel and Lorraine Lipscombe contributed to the study concept and design. George Tomlinson was responsible for the data analysis, and all of the authors contributed to the interpretation of the data. Adam Millar, Adrian Lau and Alan Kraguljac systematically reviewed and rated the studies. Adam Millar and Adrian Lau drafted the manuscript, and all of the authors revised it critically for important intellectual content. All of the authors gave final approval of the version to be published and agreed to act as guarantors of the work.

**Funding:** Lorraine Lipscombe is supported by a Canadian Institutes of Health Research New Investigator Award.

**Acknowledgements:** The authors thank Paul Shekelle, Sheri A. Keitz, Matthew J. Crowley, and Cathleen Colon-Emeric, for their thoughtful comments on earlier drafts of the manuscript.