Requirement for Mass Spectrometry Sex Steroid Assays in the *Journal of Clinical Endocrinology and Metabolism*

D. J. Handelsman and L. Wartofsky

ANZAC Research Institute (D.J.H.), Concord Hospital, University of Sydney, Sydney, NSW 2139, Australia; and Department of Medicine, Washington Hospital Center (L.W.), Washington, DC 20010

The accurate measurement of hormones is the pivot of modern endocrinology. Although tracing its roots to ancient writings, the modern scientific discipline came into existence after a series of mid-20th century Nobel Prize-winning achievements in identifying and measuring hormones. Arguably the most decisive step was the invention of immunoassay in 1959 (1), originally for peptide hormones but extended a decade later to nonimmunogenic steroids (2). This armed the nascent discipline with revolutionary impact through the ability to measure virtually every hormone at orders of magnitude lower than had ever been possible before with earlier cumbersome, insensitive, whole-animal bioassays.

Yet the pioneers of steroid immunoassay understood its limitations as that decade's delay was consumed in finding the additional workarounds required to create valid formats for immunoassay of small, subimmunogenic molecules like steroids. These essentials comprised solvent extraction, chromatography, and structurally authentic tracers, a triplet of validity criteria. Employing these required skills originally confined steroid measurement to manual methods in specialized laboratories; however, in the 1980–1990s, the steeply increasing demand for steroid measurement in the clinic and laboratory drove assay simplification to adapt steroid immunoassays into popular one-step kits and multiplex assay platform formats. This commodification resulted in "direct" steroid immunoassays, which bypassed all the original triplet of validity criteria (most fundamentally extraction), sacrificing accuracy and specificity for throughput speed and lower cost. Although problems with validity (analytical specificity, accuracy) of direct immunoassays were soon identified and were reported over 25 years ago (3, 4), the fundamental significance of violating these validity criteria resurfaced prominently in the last decade as the accuracy of direct hormone assays came under more stringent scrutiny—and as a solution, affordable and accessible mass spectrometry (MS) steroid assays came into view.

The method-specific bias and analytical nonspecificity of direct T immunoassays are worst at the low circulating T levels in children, women, and men with pathological, functional, or experimentally induced T deficiency (5–7) where such direct immunoassays have been famously likened to random number generation (8). By modern standards, putative findings using such suboptimal methods would require verification by more specific MS-based assays. This renders direct T immunoassays methodologically inadequate for high-quality clinical reproductive medicine research. Not surprisingly, aiming to measure circulating (picomolar) estradiol concentrations 10- to 100-fold lower than circulating (nanomolar) T levels that render direct T immunoassays unreliable is clearly problematic. Thus, direct estradiol immunoassays also lack accuracy in children (9), men (10, 11), and postmenopausal (3, 12–15) or aromatase inhibitor-treated (16) women who all share comparably low circulating estradiol concentrations. It is beyond doubt now that direct immunoassays of sex steroids are methodologically inadequate for high-quality clinical research.

MS-based steroid measurements have been available for decades, indeed as long as immunoassays (17). Although they always remained the reference method for steroid specificity and structure, MS-based steroid assays were always too expensive and unavailable for routine use in the laboratory to support clinical research because they required highly skilled operators using complex, difficult-

Abbreviation: MS, mass spectrometry.

ISSN Print 0021-972X ISSN Online 1945-7197
Printed in U.S.A.

Received September 4, 2013. Accepted September 4, 2013.

to-access equipment. Hence, while remaining the reference method for steroids, MS was too insensitive, expensive, and unavailable for routine clinical research use. All of this changed over the last decade with the advent of bench-top mass spectrometers with sensitivity matching the best immunoassays while retaining reference level specificity and introducing multianalyte capability. Not only does the availability of MS-based steroid assays resolve the problems of direct steroid immunoassays, but multianalyte profiling also ushers in a new era of snapshot profiling of the overlooked importance of steroid pathway fluxes that will broaden the interpretation of clinical studies beyond the standard restricted single analyte-by-analyte analysis.

In this context, it is salient that the nonspecificity and single analyte limitations of steroid immunoassays always rendered them inadequate for antidoping testing where MS-based profiling has been standard for decades. Where an elite athlete's pursuit of his or her profession could be ended by a positive antidoping test, detection of banned steroids had to be undertaken at the highest standards of scientific rigor to withstand the most vigorous medico-legal challenge. It is time that clinical research and practice operates under equally high standards that have now become feasible.

Recognizing the limitations of direct steroid immunoassays, The Endocrine Society commissioned two Position Papers written by experienced immunoassayists on the challenges to T (18) and estradiol (19) measurement and supported a renewed effort to standardize T assays (20). Although all assays require rigorous quality control to minimize random errors, additional methodological limitations of steroid immunoassay technology constituting systematic errors are impervious to such performance-based standardization. For example, the method-specific bias of immunoassays, a function of antibody epitope, allows for cross-reactivity in all steroid assays with steroid precursors, metabolites, and/or conjugates as well as matrix effects for unextracted direct assays. These limitations are aggravated by regular changes in proprietary assay reagents, especially, but not only, primary antibodies causing changed assay cross-reactivity profiles. Typically, validation of such modified assays via correlation of new with old assays (rather than with the reference method) allows for subtle, gradual "drift" in assay performance from reference standards. The net effect of method-specific bias is that steroid immunoassays do not constitute durable, definitive chemical measurements, independent of method, like hemoglobin, cholesterol, urea, and creatinine for which measurements are comparable in studies separated widely in time, place, and/or laboratory. Such long-term assay "drift" and inaccuracy is troublesome in epidemiology (19) and in evaluating temporal trends in population-based studies (21, 22) where the noncomparability of different steroid immunoassays over time hinders interpretation. In addition, method-specific bias is incompatible with research studies that promulgate clinical decision-making thresholds (23). The wider ramifications of the limitations of direct steroid immunoassays and challenges arising from them are well discussed in The Endocrine Society's Position Papers (18, 19).

Similarly, derived sex steroid measures ("free," "bioavailable") are often calculated rather than actually measured (24). However, such calculational shortcuts use inaccurate formulae for "free" (25–27) or "bioavailable" (28, 29) measurements with wide deviations from measured values due to mistaken stoichiometry and arbitrary binding affinity constants in the formulae. Furthermore, such questionable calculations purport to be hormonal measurements, yet they have no reproducibility measures like measured hormone variables. Such unreliable calculations are misleading if absolute measurements are projected, inferred, or adopted for clinical use outside the study, with the potential for confusion and misunderstanding, rather than the enlightenment intended by high-quality clinical research.

It is timely to recognize that for high-impact clinical research, the steroid immunoassay era is gradually drawing to a close, and for direct immunoassays it is effectively over. Over recent years an increasing number of manuscripts submitted to the *JCEM* have been rejected, largely for their reliance on direct steroid immunoassays as major endpoints. Because the *JCEM* remains a leader in publishing clinical endocrinology research, reflected in consistently high bibliometrics (total citations, Eigenfactor score), the *Journal* is taking the next step (24) in setting acceptable assay standards for the field in upgrading its submission requirements for publication of studies using sex steroid measurements. Effective January 1, 2015, manuscripts reporting sex steroid assays as important endpoints must use MS-based assays including reporting or citing their methods with sufficient detail to allow them to be reproduced together with standard quality control, specificity, and reproducibility metrics. Some limited editorial discretion may be exercised for rare circumstances, such as where a study reports hormone measurements made long ago and where samples no longer exist; however, that would not extend to the sort of high-quality, well-designed clinical studies that the *JCEM* expects to publish where it is expected that extra serum samples would be stored for key endpoints. It is anticipated that this requirement for MS-based assays will extend to adrenal steroids and vitamin D in the near future.

In introducing the new standard, the *Journal* makes provision for a realistic timescale for the switchover. This allows for changes in planning or conduct of clinical research to comply with the new requirement. Although re-

solving the problems of direct sex steroid immunoassays is at hand, it is natural to expect that convenience of a familiar, old but failing technology may foster rationalization and resistance to necessary progressive change. When *Endocrinology* recently instituted the requirement for more rigorous reporting of antibody specificity, a few months were allowed for authors to come into compliance (30). When a consortium of general medical journals required a momentous change to combat publication bias by requiring advance registration of all clinical trials before they commenced patient intake (31), a grace period of 12 months was allowed for studies in progress to become compliant (32). Hence, the *JCEM* is adopting a generous grace period to allow ample time for investigators to ensure that future submissions for which sex steroid assays are an important endpoint have suitably rigorous assay formats. The best clinical research will always feature standards that are rising, usually gradually but sometimes more abruptly. Ultimately, the need for the *Journal* to foster the highest standards in clinical research and practice will drive us— or we will be driven out of that top echelon by failing to do what is necessary and timely.

Acknowledgments

Address all correspondence and requests for reprints to: Professor David J. Handelsman, M.B., B.S., F.R.A.C.P., Ph.D., ANZAC Research Institute, University of Sydney–Andrology, Hospital Road, Concord Hospital, Sydney, New South Wales 2139, Australia. E-mail: djh@anzac.edu.au.

Disclosure Summary: The authors have nothing to declare.

References

1. **Yalow RS, Berson SA.** Assay of plasma insulin in human subjects by immunological methods. *Nature*. 1959;184(suppl 21):1648–1649.
2. **Abraham GE.** Solid-phase radioimmunoassay of estradiol-17β. *J Clin Endocrinol Metab*. 1969;29:866–870.
3. **Schiøler V, Thode J.** Six direct radioimmunoassays of estradiol evaluated. *Clin Chem*. 1988;34:949–952.
4. **Thomas CM, van den Berg RJ, Segers MF.** Measurement of serum estradiol: comparison of three "direct" radioimmunoassays and effects of organic solvent extraction. *Clin Chem*. 1987;33:1946–1947.
5. **Taieb J, Mathian B, Millot F, et al.** Testosterone measured by 10 immunoassays and by isotope-dilution gas chromatography-mass spectrometry in sera from 116 men, women, and children. *Clin Chem*. 2003;49:1381–1395.
6. **Wang C, Catlin DH, Demers LM, Starcevic B, Swerdloff RS.** Measurement of total serum testosterone in adult men: comparison of current laboratory methods versus liquid chromatography-tandem mass spectrometry. *J Clin Endocrinol Metab*. 2004;89:534–543.
7. **Sikaris K, McLachlan RI, Kazlauskas R, de Kretser D, Holden CA, Handelsman DJ.** Reproductive hormone reference intervals for healthy fertile young men: evaluation of automated platform assays. *J Clin Endocrinol Metab*. 2005;90:5928–5936.
8. **Herold DA, Fitzgerald RL.** Immunoassays for testosterone in women: better than a guess? *Clin Chem*. 2003;49:1250–1251.
9. **Ankarberg-Lindgren C, Norjavaara E.** A purification step prior to commercial sensitive immunoassay is necessary to achieve clinical usefulness when quantifying serum 17β-estradiol in prepubertal children. *Eur J Endocrinol*. 2008;158:117–124.
10. **Huhtaniemi IT, Tajar A, Lee DM, et al.** Comparison of serum testosterone and estradiol measurements in 3174 European men using platform immunoassay and mass spectrometry; relevance for the diagnostics in aging men. *Eur J Endocrinol*. 2012;166:983–991.
11. **Khosla S, Amin S, Singh RJ, Atkinson EJ, Melton LJ 3rd, Riggs BL.** Comparison of sex steroid measurements in men by immunoassay versus mass spectroscopy and relationships with cortical and trabecular volumetric bone mineral density. *Osteoporos Int*. 2008;19:1465–1471.
12. **Lee JS, Ettinger B, Stanczyk FZ, et al.** Comparison of methods to measure low serum estradiol levels in postmenopausal women. *J Clin Endocrinol Metab*. 2006;91:3791–3797.
13. **Dowsett M, Folkerd E.** Deficits in plasma oestradiol measurement in studies and management of breast cancer. *Breast Cancer Res*. 2005;7:1–4.
14. **Yang DT, Owen WE, Ramsay CS, Xie H, Roberts WL.** Performance characteristics of eight estradiol immunoassays. *Am J Clin Pathol*. 2004;122:332–337.
15. **Rothman MS, Carlson NE, Xu M, et al.** Reexamination of testosterone, dihydrotestosterone, estradiol and estrone levels across the menstrual cycle and in postmenopausal women measured by liquid chromatography-tandem mass spectrometry. *Steroids*. 2011;76:177–182.
16. **Santen RJ, Demers L, Ohorodnik S, et al.** Superiority of gas chromatography/tandem mass spectrometry assay (GC/MS/MS) for estradiol for monitoring of aromatase inhibitor therapy. *Steroids*. 2007;72:666–671.
17. **Shackleton C.** Clinical steroid mass spectrometry: a 45-year history culminating in HPLC-MS/MS becoming an essential tool for patient diagnosis. *J Steroid Biochem Mol Biol*. 2010;121:481–490.
18. **Rosner W, Auchus RJ, Azziz R, Sluss PM, Raff H.** Position statement: utility, limitations, and pitfalls in measuring testosterone: an Endocrine Society position statement. *J Clin Endocrinol Metab*. 2007;92:405–413.
19. **Rosner W, Hankinson SE, Sluss PM, Vesper HW, Wierman ME.** Challenges to the measurement of estradiol: an Endocrine Society position statement. *J Clin Endocrinol Metab*. 2013;98(4):1376–1387.
20. **Vesper HW, Botelho JC.** Standardization of testosterone measurements in humans. *J Steroid Biochem Mol Biol*. 2010;121:513–519.
21. **Travison TG, Araujo AB, Hall SA, McKinlay JB.** Temporal trends in testosterone levels and treatment in older men. *Curr Opin Endocrinol Diabetes Obes*. 2009;16:211–217.
22. **Andersson AM, Jensen TK, Juul A, Petersen JH, Jørgensen** T, Skakkebaek NE. Secular decline in male testosterone and sex hormone binding globulin serum levels in Danish population surveys. *J Clin Endocrinol Metab*. 2007;92:4696–4705.
23. **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;95:2536–2559.
24. **Wartofsky L, Handelsman DJ.** Standardization of hormonal assays for the 21st century. *J Clin Endocrinol Metab*. 2010;95:5141–5143.
25. **Hackbarth JS, Hoyne JB, Grebe SK, Singh RJ.** Accuracy of calculated free testosterone differs between equations and depends on gender and SHBG concentration. *Steroids*. 2011;76:48–55.
26. **Ly LP, Sartorius G, Hull L, et al.** Accuracy of calculated free testosterone formulae in men. *Clin Endocrinol (Oxf)*. 2010;73:382–388.
27. **Sartorius G, Ly LP, Sikaris K, McLachlan R, Handelsman DJ.** Predictive accuracy and sources of variability in calculated free testosterone estimates. *Ann Clin Biochem*. 2009;46:137–143.
28. **Egleston BL, Chandler DW, Dorgan JF.** Validity of estimating non-sex hormone-binding globulin bound testosterone and oestradiol from total hormone measurements in boys and girls. *Ann Clin Biochem*. 2010;47:233–241.
29. **Giton F, Fiet J, Guéchot J, et al.** Serum bioavailable testosterone: assayed or calculated? *Clin Chem*. 2006;52:474–481.
30. **Gore AC.** Editorial: antibody validation requirements for articles published in endocrinology. *Endocrinology*. 2013;154:579–580.
31. **Simes RJ.** Publication bias: the case for an international registry of clinical trials. *J Clin Oncol*. 1986;4:1529–1541.
32. **De Angelis CD, Drazen JM, Frizelle FA, et al.** Is this clinical trial fully registered? A statement from the International Committee of Medical Journal Editors. *N Engl J Med*. 2005;352:2436–2438.