Mass spectrometry and immunoassay: How to measure steroid hormones today and tomorrow

3 authors, including:

Brian G Keevil
University Hospital Of South Manchester NHS Foundation Trust
287 PUBLICATIONS **7,129** CITATIONS

SEE PROFILE

Ilpo Huhtaniemi
Imperial College London
992 PUBLICATIONS **38,224** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

European Male Ageing Study (EMAS) View project

SHBG and the free hormone hypothesis for sex steroids View project

Mass spectrometry and immunoassay: how to measure steroid hormones today and tomorrow

Angela E Taylor*, **Brian Keevil**[1],* and **Ilpo T Huhtaniemi**[2,3],*

School of Medicine, Centre for Endocrinology, Diabetes and Metabolism, University of Birmingham, Birmingham, UK, [1]Department of Clinical Biochemistry, Manchester Academic Health Science Centre, University Hospital of S Manchester, The University of Manchester, Manchester, UK, [2]Department of Surgery and Cancer, Imperial College London, Institute of Reproductive and Developmental Biology, Hammersmith Campus, London W12 0NN, UK and [3]Department of Physiology, Institute of Biomedicine, University of Turku, 20520 Turku, Finland
*(A E Taylor, B Keevil and I T Huhtaniemi contributed equally to this work)

Correspondence
should be addressed to
B Keevil or I T Huhtaniemi
Email
Brian.Keevil@UHSM.NHS.UK
or ilpo.huhtaniemi@
imperial.ac.uk

Summary

The recent onslaught of mass spectrometry (MS) to measurements of steroid hormones, including demands that they should be the only acceptable method, has confused clinicians and scientists who have relied for more than 40 years on a variety of immunoassay (IA) methods in steroid hormone measurements. There is little doubt that MS methods with their superior specificity will be the future method of choice in many clinical and research applications of steroid hormone measurement. However, the majority of steroid measurements are currently, and will continue to be, carried out using various types of IAs for several reasons, including their technical ease, cost and availability of commercial reagents. Speedy replacement of all IAs with MS is an unrealistic and unnecessary goal, because the availability of MS measurements is limited by cost, need of expensive equipment, technical demands and lack of commercial applications. Furthermore, IAs have multiple well-known advantages that vindicate their continuing use. The purpose of this article is to elucidate the advantages and limitations of the MS and IA techniques from two angles, i.e. promotion of MS and defence of IA. The purpose of the text is to give the reader an unbiased view about the current state and future trends of steroid analysis and to help him/her choose the correct assay method to serve his/her diagnostic and research needs.

Mass spectrometry provides the best approach for steroid measurement (Angela E Taylor and Brian Keevil)

Introduction

'Start with what is right rather than what is acceptable'. – Franz Kafka

As early as the 1930s, steroids were investigated using mass spectrometry (MS) (1). In the 1960s, gas chromatography–MS (GC–MS) was advanced enough to investigate steroid metabolism and urinary steroid profiles were defined including androgen and cortisol metabolites (2, 3). From the 1980s to 1990s, introduction of the quadrupole mass analyser, HPLC and reliable sample extraction techniques allowed more sensitive steroid assays to be developed (4).

GC–MS remains unrivaled for the unequivocal identification of nearly all steroid metabolic disorders (5) and has contributed greatly to our current knowledge on inborn errors of steroid metabolism. Urinary steroid profiling by GC–MS is an ideal steroid metabolomic tool which still leads to discoveries of novel steroids, even for such well characterised conditions as 21-hydroxylase deficiency (6).

Liquid chromatography–tandem mass spectrometry (LC–MS/MS) is much simpler to use in routine laboratories than GC–MS because of shorter sample preparation methods (the analytes do not need derivatising in order

to make them volatile in the gas phase). Despite the increased complexity of sample preparation, it is recognised that GC–MS still has an important part to play in determining the steroid metabolome, but for the routine determination of individual or panels of steroids LC–MS/MS is now the method of choice for use in the clinical laboratory, while in research laboratories the combination of GC–MS and LC–MS/MS provides a cutting-edge research tool for endocrinology. It is therefore not correct to think of MS as the 'new guy' in steroid analysis, as there is a wealth of research available looking at panels of steroids for more than 50 years.

Being developed in parallel to MS were immunoassays (IAs), and it was IA that was widely adopted as the method of choice for steroid analysis in clinical chemistry laboratories because of its sensitivity, cost and simplicity. However, problems associated with this technique are well known. They include cross reactivity with similar analytes, standardisation issues between labs and sensitivity issues (7). It is now recognised that performance of some IA techniques may be suboptimal, this recently prompted a recommendation from the editors of *Journal of Clinical Endocrinology and Metabolism* to avoid using such assays and instead use MS for the measurement of sex steroids (8). This sparked a vigorous debate which resulted in the Endocrine Society Council establishing a 'Sex Steroid Assays Reporting Task Force'. Whilst originally focussing on sex steroid assays it became evident that other steroid assays also suffered from poor performance (9). The task force, reporting recently, recommended necessary performance criteria that should be met for any method that is used in clinical and non-clinical studies. Although recommendations have fallen short of making LC–MS/MS mandatory for steroid analysis it was recognized that high quality well validated assays are required to improve the quality of work (10). Improving the quality of assays by making them traceable to a certified standard has already started with initiatives such as the Hormone Standardization Program (HoSt), which show that LC–MS/MS can be used as a gold standard when used in the appropriate manner under highly regulated conditions (11). This makes LC–MS/MS is now the go to method for ensuring the quality of steroid results (12).

These initiatives clearly demonstrate that LC–MS/MS is already taking a lead role in improving the quality of steroid measurement and is delivering the necessary targets by which all routine assays will be assessed including IA methods. If a laboratory follows good calibration practice, there is the potential to develop reference ranges applicable to all assays that can be

shown to align to the reference method. This will standardise the interpretation of test results by making results comparable between laboratories no matter which method they are using.

Here we explain why we think IA will eventually become obsolete for steroid analysis and high throughput LC–MS/MS will become the method of choice in both the clinical chemistry and research laboratories.

GC–MS and LC–MS/MS basic principles

GC/MS permits separation of liquid analytes which are vaporised in the GC injector. These (now gaseous) steroids travel through the heated column and arrive at the mass spectrometer at different times due to differences in column liquid phase interactions. GC is widely used in steroid chemistry usually after chemical derivatisation of the steroids to increase the volatility and stability of the steroid which produces optimum sensitivity and chromatographic resolution (4). The method for steroid profiling by GC/MS has been developed over 40 years by Shackleton & Snodgrass (13), Sjövall (14) and others, and is the gold standard for urinary steroid profiling (15). GC–MS has been largely superseded in the past 10 years by LC–MS/MS due to improvements in HPLC systems and column design as well as the introduction of MS/MS. LC permits separation of a mixture of liquid analytes into separate analyte fractions. This is achieved with the use of a column packed with a stationary phase which retains the analytes.

Steroids are retained on the column stationary phase at differing degrees based on their polarity, increasing the polarity of the mobile phase elutes the steroids sequentially. MS offers positive compound identification to enable the unequivocal identification of a compound free from interference, thus enabling accurate quantitation. In the case of GC–MS, this is achieved using a very long, typically >10 m, chromatography column to separate out interfering substances. In the case of LC–MS/MS, a much shorter, typically <20 cm chromatography column is used, but the specificity is augmented by the use of two quadrupole mass analysers. The molecular ion is selected in the first quadrupole after which molecular ion fragments, selectively formed in a collision cell, are selected using the second quadrupole. There are therefore three ways to gain increased specificity using LC–MS/MS and this is achieved with analytical run times of <10 min compared to the 30 min run times of GC–MS methods. This makes LC–MS/MS especially attractive for the fast turnaround times and high sample throughput

required in clinical laboratories. In addition, it is becoming standard practice to monitor a second qualifier fragment ion as a quality check, in order to improve the specificity of the technique even further and check for interfering compounds/matrix effects.

LC–MS/MS method validation

In both IA and MS, the key to good reproducible data is validation of assays. A full validation takes a few days to perform and system performance can be continuously monitored throughout the mass spectrometer's lifetime. Validation of a LC–MS/MS assay requires investigation into a number of parameters such as selectivity, sensitivity, stability, reproducibility and matrix effects, which should be completed during method development for each steroid (16). Over the past decade, there have been huge improvements in the performance and reliability of MS equipment. The equipment is often portrayed as being difficult to use but this notion probably arises from the deficiencies in classical analytical skills which accompanied the move to greater laboratory automation. It is perfectly possible to run MS instruments in a busy routine laboratory; it is just a question of training and perhaps changing laboratory culture. In fact, all of the shift workers at the Clinical Chemistry department in the University Hospital of South Manchester can operate the MS instruments to a good standard. They are rotated through the MS section regularly to maintain skill levels and are capable of running MS assays during the weekend.

Inter-assay performance by LC–MS/MS and IA

Things are far from perfect in the world of MS. There are good and bad LC–MS/MS methods and this has been shown in interlaboratory surveys for a number of steroids. This is because LC–MS/MS is a relatively recent introduction into clinical laboratories and some may struggle initially with the increased complexity. An important point to make is that it is generally sample preparation and not the MS measurement technique which is at fault. Better standardisation of sample clean up and chromatography conditions will improve results whereas with IA the differences between antibodies and kit formulations from different suppliers make assay standardisation difficult. LC–MS/MS methods can be improved by care and attention to detail whereas IAs need better antibodies. This is not so easy to achieve. Many laboratories find calibrators difficult to make and studies have shown that using the same calibration material can significantly

improve inter-laboratory performance (17). This problem has already been solved for vitamin D assays by the introduction of commercially available calibrator sets and this is also thankfully now being addressed for steroids.

The enhanced specificity of LC–MS/MS should be compared to specificity problems with IA methods encountered since the move to perform 'direct' IA (without any extraction step) routinely on automated analysers. The extraction steps were discarded to streamline the process but the resultant direct methods, although convenient, could suffer significant interference from other steroids. For example, the most abundant serum steroid DHEAS (μM concentration) will interfere with IAs looking for low level androgens such as testosterone (<10 nM) (18). Before the introduction of large analyser systems, there was scope to change the kit supplier if the assay performed badly, but this is now virtually impossible because reagents are tied into the analyser and kits are not interchangeable between different manufacturer systems. In the early days of IA automation, it was not uncommon for laboratories to have several different analysers to cover the best available assays from each supplier. This is not possible due to financial and operational reasons and the only options now, if performance is dire, are to use alternative methods such as small niche IA analysers, ELISA or LC–MS/MS.

Esoteric tests

A significant drawback of IA is that tests are generally developed to fit a global market and generally for only high throughput assays for steroids such as cortisol, oestradiol (E_2) and testosterone. Development programmes are driven by a combination of clinical demand and the ability to recoup development costs and generate profits in a reasonable time frame. Because of this, IA does not easily lend itself to the development of niche assays which are performed less frequently and in much smaller numbers. This includes the less commonly measured steroid assays such as 11-deoxycorticosterone, aldosterone, 17-hydroxyprogesterone (17OHP), DHEA and dihydrotestosterone (DHT) which are all readily measured using LC–MS/MS but are not available on main laboratory IA platforms. Market forces and economics can also intervene, and there is always the possibility of IA tests being withdrawn fairly abruptly from the market as was recently seen with the withdrawal of RIA methods for aldosterone.

The demand for niche assays will only increase as the results of metabolomics programmes generate new

candidate diagnostic markers which will need proper validation. The task of validating these new biomarkers will be ideally suited to the accuracy, precision and analytical flexibility provided by LC–MS/MS techniques, especially as many of the candidates will have been discovered using MS methods.

Multiplexing

Improved analytical performance aside, the great advantage that LC–MS/MS has over IA is the capability to simultaneously measure several different steroids to produce multiplexed test panels. Performing multiplexed analysis is often at no greater expense because the sample preparation and LC–MS/MS conditions are the same, whereas using IA one test at a time, adding in extra tests becomes more expensive. Some have advocated the analysis of completely unrelated steroids (19) but selective testing panels are becoming popular and have been developed to investigate different clinical conditions such as second tier testing of congenital adrenal hyperplasia to supplement newborn screening programmes (20, 21) and for the investigation of hypoadrenalism (22).

Commercially available kits have been developed for the measurement of steroid panels and one in particular claims to measure 12 steroids in two differently run panels, one for adrenal steroids and one for sex steroids, although there is some overlap between the two panels (www.Chromsystems.com). The introduction of easy-to-use kits will undoubtedly simplify the introduction and service delivery of LC–MS/MS in a routine laboratory.

In the last few years, LC–MS/MS has allowed a more thorough investigation of conditions such as PCOS and studies have shown that just measuring one analyte, total testosterone, is inadequate. For example, androstenedione is now thought to be just as important as testosterone in the assessment of PCOS (23), but androstenedione is not routinely available on main clinical chemistry analysers and the smaller niche IA analysers typically give results twice the concentration of those reported by LC–MS/MS methods (UK NEQAS, Birmingham Quality, Birmingham, UK, www.birminghamquality.org.uk). It has also been recently shown, using LC–MS/MS, that the testosterone to DHT ratio in PCOS subjects identifies an adverse metabolic phenotype (24). In an attempt to improve the workflow, the Manchester laboratory for the assessment of hirsute women has recently developed a simple method for the simultaneous measurement of female androgens including testosterone, androstenedione, 17OHP and DHEAS.

Commonly requested steroids

Cortisol ▶ Serum cortisol IA suffers with interference from cortisol precursors when patients are given metyrapone (25). Cortisol IAs also shows differing abilities to remove cortisol from cortisol binding globulin, as demonstrated by under recovery in female subjects taking the oral contraceptive pill (OCP) (26). In a recent survey, up to 33% of cortisol results were unreportable because of patients receiving prednisolone. Prednisolone, prednisone, cortisol and cortisone can all be selectively analysed by LC–MS/MS whereas there would be a significant amount of cross reactivity on an IA platform. From a practical standpoint, rapid MS assays for the routine measurement of serum cortisol are now available with the added benefit of identifying samples that show evidence of exogenous steroid use (27).

On the strength of this, we now measure all samples for serum cortisol by LC–MS/MS. In addition, the ingestion of commonly prescribed exogenous steroids can cause hypoadrenalsim and lead to diagnostic problems when this is unknown to the physician, but using LC–MS/MS it is now possible to screen for these steroids (28). Salivary cortisol is a suitable alternative to urinary free cortisol for the investigation of hypercortisolism but many IAs struggle to measure cortisol at the lower concentrations found in saliva. Salivary cortisol is a useful surrogate for circulating free cortisol, but its concentration is determined both by serum free cortisol and parotid metabolism to cortisone. Salivary cortisone closely reflects free serum cortisol after adrenal stimulation and hydrocortisone (cortisol) administration, and is unaffected by changes in corticosteroid binding globulin. Salivary cortisone has potential as a useful surrogate for serum free cortisol in research and clinical assessment and may be a better marker than cortisol because of interference in cortisol measurement from exogenous hydrocortisone (29). This was further investigated by Raff (30) who also found that dramatically increased salivary cortisol value by IA could be caused by cortisol contamination. In these cases, contamination with topical hydrocortisone was suggested by a normal salivary cortisone concentration as measured by LC–MS/MS and a high cortisol-to-cortisone ratio (31).

Testosterone ▶ Despite the acknowledged superiority of LC–MS/MS for measuring female testosterone, there are still only a very small percentage of laboratories (18%) using this technology that participate in the UKNEQAS scheme. This shows just how popular IA still is despite its many limitations.

Oestradiol ▶ It has been known for many years that E_2 IA methods behave poorly for the measurement of samples from men, children and postmenopausal women (32). Achieving the necessary sensitivity using LC–MS/MS without resorting to complex derivatization methods has been a challenge but the discovery of using trace amounts of ammonium fluoride in the mobile phase has enabled the development of assays capable of measuring as low as 10 pmol/l (33, 34), which is suitable for the vast majority of samples. However, we have still not reached the required sensitivity of 1 pg/ml (3.7 pmol/l) advocated in the Endocrine Society position statement (35) and because these methods require expensive high sensitivity MS instruments, it may be some time before they are used universally.

Vitamin D ▶ Laboratories are becoming inundated with vitamin D requests and whilst LC–MS/MS is still regarded as the best method available, there is real pressure to put these tests on IA platforms to save time and money. It is generally accepted that 25-hydroxy vitamin D3 (25(OH) vitD3) and 25(OH) vitD2 should be reported together, and this is stipulated by the US Food and Drug Administration. Some IAs have differing cross reactivity with 25(OH) vitD2 and this may cause problems where supplementation with 25(OH) vitD2 is widespread. There is common agreement that 25-OHD is a difficult analyte (36). All IAs show high cross reactivity with 24,25-dihydroxyvitamin D (24,25(OH)2D) which can be present in serum at concentrations of up to 12 nmol/l (37), and are likely to be more susceptible to variations in vitamin D binding protein concentrations (38). Although performance of IA is known to be variable between manufacturers and is poor below 20 nmol/l, in some cases it is accepted that some IAs are adequate for clinical use for concentrations above this (39). The consequences of systemic differences between 25-OHD IAs are not confined to clinical laboratories. Epidemiologic studies are an important source of normative 25-OHD data, but differences in methods used by national surveys have limited the extent to which results could be pooled (40).

MS in the research laboratory

While implementation of LC–MS/MS in the clinical laboratory has been slow, use in research laboratories has been expanding more rapidly. Advances in MS design have produced a boom in research not only in endocrinology but in other biological applications such as proteomics and metabolomics. In most academic institutions, there

will be multiple MS platforms (the University of Birmingham currently has more than 60 instruments). Assay requirements are much the same as in the clinical laboratory, although more time can be allocated to method development and validation. Access to both GC/MS and LC–MS/MS platforms allows development of novel assays which translate to the clinic.

Conclusions

Laboratories with established LC–MS/MS services are in a fortunate position that remains an aspiration for many and we accept that this situation will continue for many years. MS is roughly at the same stage in its clinical development that IA was 30 years ago when labour intensive manual testing changed within several years to high throughput testing on large analyser platforms. Manufacturers are actively trying to develop mass spectrometers that are more user friendly to the high throughput clinical laboratory. In the meantime, many laboratories will still be reliant on IA for steroid analysis and the manufacturers must raise their game and improve the IA kits that they sell. Undoubtedly, the move to direct IA has hurt performance but improvements in antibody production and kit design are needed to improve assay specificity. We are constantly hearing how IA provides clinically valid results, provided various caveats are taken into account. These are expensive tests and perhaps we should not be so accepting of substandard performance. Some manufacturers are making valiant efforts to improve the quality of their kits but the problems have been known for years and one suspects that it is only since the introduction of LC–MS/MS that increased efforts are being belatedly made. Steps in the right direction are however being taken, and the manufacturers of IA kits are now using LC–MS/MS to calibrate their assays, which is a positive development and should be applauded. There is still scope to use IA assays in the clinical laboratory because clinicians will still ask for them (typically as this is what they are used to). If your steroid of interest is a high concentration, meaning your sample is high volume or unique, than a well validated AI method is completely reliable (for example, testosterone in men). If you want to look at lower level, low volume or structurally similar steroids then LC–MS/MS is the only platform that can offer accurate, reliable quantitation.

LC–MS/MS will continue to grow in popularity and the larger laboratories will make more use of multiplexed steroid panels to aid diagnosis and monitoring of different clinical conditions. We believe that MS is here to stay

and because of its many advantages over IA will remain the method of choice for ensuring the quality of endocrine science.

Immunoassays are still useful in the measurement of steroid hormones (Ilpo T Huhtaniemi)

Introduction

'*The best is the enemy of good*'. – Voltaire

Steroid hormones have been measured for nearly the last 50 years mainly using immunoassays (IAs) (41). For the first time, this offered the opportunity to measure the active hormone molecule directly in biological samples, instead of measuring indirectly their inactive metabolites in urine, e.g. 17-ketosteroids or 17-ketogenic steroids, with colorimetric methods. There were also cumbersome *in vivo* bioassays where the hormone was quantified on the basis of its functional response in experimental animals. Hence, IAs, enabling the measurement of hormones in a small volume of peripheral serum represented a real paradigm shift, inaugurating a new era in endocrine research and clinical practice. In hindsight, it is understandable that large proteins, such as insulin, growth hormone or gonadotropins, are immunogenic and evoke antibody generation upon immunization in experimental animals, e.g. the rabbit. It was more surprising that also smaller non-immunogenic molecules, such as steroids, could be immunogenic as haptens, when coupled to a macromolecule, such as albumin or thyroglobulin. In this way, it became possible to develop IAs also for steroid hormones (42). The small molecular size and cross reactivities between structurally similar steroid molecules posed a special challenge for the specificity of steroid IAs. The advent of monoclonal antibodies represented an improvement, but still the physical fact of structural similarity of steroid molecules remained a challenge. Moreover, the low affinity of steroid antibodies and the competitive (as opposed to immunometric) assay principle limited the sensitivity of steroid IAs.

In the early days of steroid RIAs, the lack of antibody specificity was compensated for by solvent extractions and chromatographic fractionation (e.g. Celite and Lipidex), and the sensitivity was improved by extraction and concentration. In the newer direct platform assays (using chemiluminescence, fluorescence or enzymatic colour reaction as detection method), these pre-purification steps have been eliminated, and are only partly compensated for by more specific antibodies. Technological advancements made it possible to develop simplified and rapid platform assays with high capacity and speed, which are widely available as commercialized automated applications at tolerable cost. Unfortunately, the various shortcuts taken in the methodology introduced to routine steroid assays exhibit such disadvantages as reduced sensitivity, hard-to-control matrix effects, interference with plasma steroid binding proteins and non-specific competition with the assay antibody. We are now in a situation where all commercial steroid kits do not have the desired specificity, sensitivity and traceability. The problematic interlaboratory variability of assay results is not specific for the IA principle but caused by poor assay calibration and standardisation.

An increasing number of original publications are today convinced that the future of steroid analysis is in applications of the mass spectrometry (MS) technology. However, when one reads the articles comparing the performance of IA vs MS methods, one often finds the conclusion that both are equally good in clinical diagnostics when assessed in the light of intended use. Likewise, clinicians seem to be much more content with the information provided with IAs than clinical biochemists and chemical pathologists. The former have learned to live with the limitations of these assays and can use the information provided for diagnostics. The latter strive to improve laboratory methods, are developing novel in-house MS assays and are understandably enthusiastic about the new technology. For this reason some level of publication bias may exist for the benefit of MS. In fact, the majority of novel development in the field of steroid analytics occurs in MS technology. However, we have to remember that also IAs undergo continuous developmental work and new assays are likely to be devoid of some of the pitfalls of the earlier generation assays. Furthermore, we cannot let down those clinicians and scientists who do not have access to steroid measurements by MS.

An interesting episode in the IA/MS duel was the message published about a year ago by *Journal of Clinical Endocrinology and Metabolism* (8), where it was demanded that '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'. The decision of the journal apparently met with such strong objections that it was suspended in

a Letter of Concern a few months later subject to further scientific review (43). Following this U-turn the Endocrine Society established a 'Sex Steroid Assay Reporting Task Force (SSATF)', whose first report was recently published (44). They stipulated that 'new instructions to authors on the reporting of steroid hormone assay measurements addresses important requirements for minimal analytical validity including standards of accuracy, precision, specificity, sensitivity, reproducibility and stability'. Importantly, the instructions do not mandate a specific hormone assay technique, but clearly state the criteria for a valid assay for clinical and non-clinical studies and give examples of where some widely available steroid hormone assays are not sufficient to meet these criteria. The words 'immunoassay' or 'mass spectrometry' were not mentioned in the report, implying that common sense had won, and that the authors now realised that it is the assay quality and not the specific method that is important. There are good and bad IAs, but the same also applies to MS, which is better called 'comparator method' than 'gold standard'.

Advantages and disadvantages of IAs

Broadly speaking, IA and MS are complementary methods, because the advantages of one are the same as the limitations of the other. IA measurements can be conducted even in primitive laboratory conditions, and they do not need special skills of the personnel. Numerous commercial applications exist and the reagents are easily available. With automation, the assays are high throughput at reasonable price. If quality control measures are in place, the results are reliable, provided that the pitfalls and limitations of the assay are known. We also have to remember that although MS, in general, is more specific and has lower intra- and interassay variability than IA, it faces similar inter-laboratory variability issues as IA (45, 46).

The disadvantages of IAs include the potential lack of specificity, usually due to problems with the antiserum/antibody. Sensitivity may be another problem because of the low affinity of anti-steroid antibodies. The commercial reagents are usually proprietary and all necessary assay details are not available for the end user. Manufacturers may also change reagents without providing users with sufficient information about their impact on assay results. In the interest of speed and high throughput, the reagent manufacturers may take shortcuts that impinge on quality of the data. In general, however, the commercial IA reagents do measure reliably the analyte they have been developed for. Proper quality control and standardization

in the user laboratory is important, as with all assay methods. Furthermore, it is important to know the limitations of the IAs, in particular their sensitivity and specificity and to use them only for the purposes where the latter problems do not apply.

Reliability of commercial platform IAs

The literature is full of studies comparing various steroid platform assays with the 'gold standard' of liquid chromatography–tandem mass spectrometry (LC–MS/MS). The former are often criticized for lack of sensitivity and specificity, which are inherent problems of the IA technique, due to matrix effects, cross-reactivities with structurally related compounds and interference with steroid binding proteins in the sample. These problems were formerly largely eliminated in the 'indirect IAs' by sample extraction, concentration and fractionation (chromatography). However, such additional steps are eliminated in more recent 'direct IAs' of commercial platform assays, in the interest of speed, simplicity and high throughput. Rather than the method *per se*, poor calibration and validation are similar problems in IA and MS (47). In principle, the performance of a good IA can be as good as that of a good MS. However, in real life, there are differences in the performance of steroid IAs and it is important to know when an IA is reliable and when MS offers a real practical advantage. In the following paragraphs, we will overview the performance characteristics and applicability of the commonest steroid IAs in clinical diagnostics. Admittedly, it is difficult to decipher consensus about the real performance of the different IAs from the existing literature. Conclusions may often be influenced by different types of author bias, including source of funding. The authors developing MS methods seem to underestimate – understandably – the performance of IAs, and many reports about novel developments in IA are from laboratories supported by reagent manufacturers.

Cortisol

Cortisol secretion is measured in clinical settings from serum, 24 h urine sample or mid-night saliva. Serum-based IAs are usually used, with variable modifications, validations and success, for urine and saliva. The direct assays, without extraction and pre-purification, often have insufficient specificity, leading to under- and overestimation of true concentration and may lead to clinically significant misclassifications. Other problems of the direct serum assays are the competition of plasma cortisol

binding proteins with the assay antibody, as well as interference with antisteroidogenic drugs. In particular, the general standard of urinary cortisol IAs is unsatisfactory with grossly overestimated urinary free cortisol concentrations due to cross reactivity of cortisol metabolites (48, 49). Some experienced clinicians consider the low specificity of urinary free cortisol assays an advantage, because in this way they better monitor the overall adrenal glucocorticoid production. However, opinions from laboratory scientists suggest that they should not be used for routine clinical diagnostics at all (50). The between-assay agreement of serum and salivary cortisol measurements with commercial reagent kits has been found unsatisfactory (51, 52), but there are also platform IAs that show excellent agreement with the reference MS method (53, 25). However, the majority of recent publications demonstrate major non-linearity between cortisol measurements by IA and MS and lack of accuracy of different IAs (48, 51, 52). Therefore, if possible, research and clinical diagnostics of cortisol secretion should move towards the MS based methodology (50). Until this is achievable, the clinician has to be familiar with the limitations of the cortisol IA at his/her disposal, in order to interpret the results correctly.

Testosterone

The poor quality of testosterone assays has been a long-time target of complaints (54). The accuracy and sensitivity of the IAs is not sufficient for samples from women and children. How much of this is due to poor assay quality and how much to insufficient standardization and validation is not completely clear. In any case, concerning testosterone IAs for male samples, several recent studies on large sample numbers have demonstrated that the new platform testosterone assays are fully adequate for measurement of this hormone in male samples at all concentrations. We carried out recently a large comparative study on testosterone measurements by a well-standardized IA and a comparator MS method in serum samples of over 3000 men (55). The salient conclusions from this study were that testosterone measurements by IA offer good accuracy at all concentrations found in eugonadal and hypogonadal men. Our findings do not support a mandatory requirement, on either analytical or clinical grounds, to switch from good-quality IAs to MS in the measurements of testosterone in male subjects. Importantly, we found out that our platform IA for testosterone had sensitivity and specificity also in the hypogonadal male range, though with somewhat lower, but still acceptable, accuracy than in the eugonadal range.

Using MS as the comparator method, the sensitivity (% of true positives) and specificity (% of true negatives) of IA to detect testosterone concentrations determined by MS as <8 nmol/l were 76.9 and 98.3% respectively. These figures are probably underestimates because the comparator MS method is not free of variability either (45, 46). It has to be emphasized that the testosterone IA used was of good quality, having passed with acceptable accuracy a rigorous standardization procedure (http://www.cdc.gov/lab standards/hs.html). Not all IAs used in clinical testosterone measurements are likely to have the same high quality.

A very recent study comparing total testosterone assays in women concluded that the results obtained by IA and MS were comparable, but there was significant variability and poor precision also between various MS methods at low levels (46). Hence, switching from IA to MS is not a guaranteed solution to improve the quality of sex steroid measurements at low concentrations. With the current knowledge it seems prudent to conclude that a good-quality IA is able to detect reliably all testosterone concentrations of male samples, but they do not have sufficient sensitivity to detect the normal low levels in samples of females and children. However, pathological testosterone concentrations in women and children are typically higher than normal (e.g. in the common conditions of polycystic ovary syndrome and non-classical congenital adrenal hyperplasia, and rarer, more serious pathologies such as adrenal or ovarian tumours), and such levels may be detectable by IA. Testosterone measurements are an example where the applicability of IA depends critically on the clinical application.

Oestradiol

According to an Endocrine Society position statement (35), both modern IAs and LC–MS/MS-based methods are reasonably well suited for the measurement of serum oestradiol (E_2) in clinical diagnosis and follow-up of management of infertility as well as for the detection of oestrogen-secreting tumours, but the situation with measurement of E_2 in situations with expected low levels is problematic. In another recent review (32), although no distinction was made between IA and MS as long as they are well standardized, detection of low levels by IAs poses a challenge that is best answered by LC–MS/MS techniques. A good clinical assay has to be specific and should detect E_2 levels down to 5 pmol/l in children, men, postmenopausal women and women receiving aromatase inhibitors for breast cancer. Such sensitivity is not achieved even with the best platform IAs (47, 56), or even with most of the MS

methods (35). Numerous concerns about the current IA and MS methods have been reported over the years (35), concerning their limited sensitivity, specificity, variation of interference, inconsistent assay accuracy and precision at low concentrations. Male serum E_2 levels are below the limit of sensitivity of direct IAs (57). Our own data on comparison of IA and MS results on >3000 male serum samples demonstrated poor correlation at low E_2 levels ($r = 0.32$ at concentration <40.8 pmol/l) and acceptable at higher levels ($r = 0.74$ at >40.8 pmol/l) (55). Several ultrasensitive in-house E_2 IAs or *in vitro* bioassays have been developed in research laboratories but they are labour-intensive and not generally available (58, 59).

In IVF settings, E_2 levels are supraphysiological in response to ovarian stimulation, and IAs continue to be the method of choice because of their speed. In one study on sex steroid measurements of men upon examination of bone mineral density, it was confirmed that MS measurements provide more accurate measure of serum E_2 than IA (60). Despite this, IA E_2 results allowed clinically relevant conclusions to be made for the association between E_2 levels and bone mineral density. Hence, although there is a consensus that MS measurements are a preferred method for E_2 measurement in men (60), also IA can produce clinically meaningful data, especially when elevated levels are expected to confirm diagnosis, e.g. in oestrogen producing tumours and gynaecomastia. Altogether we can conclude that IA measurements perform well in premenopausal women (including the follow-up of IVF treatments), but do not have the necessary sensitivity and specificity in men, children and postmenopausal women, unless the expected concentration is elevated. Because even most of the current MS methods are not able to detect adequately low E_2 levels, switching to MS does not offer a solution for the measurement of low levels. IA and MS have similar problems in assay specificity (47). One important future application of E_2 measurements will be the assessment of osteoporosis risk in men, but for this none of the current IAs or MS assays are sensitive enough (61). Moreover, recent epidemiologic studies in men and women have demonstrated associations between low sex hormone concentrations (including E_2) and the risk of cardiovascular disease in both sexes (62, 63, 64, 65), suggesting another indication where more reliable methods for E_2 measurement would be needed.

Vitamin D

Testing for vitamin D is now in great demand because of the increased awareness of frequency of vitamin D deficiency and possible health benefits of vitamin D supplementation. Therefore there is pressure on clinical laboratories to provide results of large numbers of samples in a timely fashion. The available methods include manual IAs, automated platform IAs, HPLC with u.v. detection and LC–MS/MS. A number of assays from diagnostic companies are available for automated IA platforms. The most common and recommended way of measuring vitamin D is to assess combined 25-hydroxy-vitamin D3 (the endogenous molecule) and 25-hydroxy-vitamin D2 (the food-derived molecule) (66). High quality 25-hydroxyvitamin D (25-OHD) assay poses challenges because of the highly lipofilic nature of vitamin D, the presence of multiple cross-reacting vitamin D metabolites in the circulation, and the high affinity of vitamin D binding protein. The latter, in particular, results in variable underestimation of low levels and has caused an epidemic of misdiagnosed vitamin D deficiency, the frequency of which is assay dependent (67). There has been lots of development work in the field, with numerous automated assays being withdrawn from the market because of poor performance, and being replaced by newly improved ones. The field has now developed to such a state that recent authoritative reviews conclude that automated vitamin D assays, if validated (traceable) can be used for the diagnosis of vitamin D nutritional status in the general population at the clinically significant concentrations of 20–375 nmol/l (67, 68). Some of them even offer a sensitivity of 3–4 nmol/l, which should be sufficient to detect vitamin D deficiency. The current automated vitamin D assays appear in general to show acceptable correlation with reference LC–MS/MS methods (68, 69, 70, 71). Nevertheless, LC–MS/MS is considered by experts the best available technique to measure 25-OHD and there are strong recommendations to replace IAs whenever possible with the former (72). All in all, it seems that platform IAs for 25-OHD are marginally sufficient for clinical diagnostics of vitamin D deficiency, the main problem being their insufficient sensitivity to classify a patient reliably as vitamin D deficient. The clinicians can with some reservation rely on their results. However, the future trend of leading laboratories is to move to MS methods.

Conclusions and future directions

There is little doubt that MS measurements will be the future winner in the current steroid analysis race. However, it is unrealistic to think that all IAs could be

replaced overnight with MS, for various reasons as previously described. Until the MS methods attain better availability and become more attractive in price, speed, capacity and technical simplicity we have to give some practical advice to end-users of the IA data on their reliability in clinical diagnostics and research. First, it is important to know the limitations of the available IAs, whether they can provide reliable results throughout their whole measuring range. Also the most common sources of error have to be known, many of them at the pre-analytical level, e.g. interfering medications. Improvements in performance and standardization in platform IAs are feasible alternatives that are already being implemented by manufacturers. It is a major investment to abandon IA technology in favour of MS, and the reasons for this must be tangible and supported by evidence rather than conjecture.

It is important that IA kit manufacturers check their results against traceable MS assays and calibrators, and provide this information in their package inserts, which would provide credence for their methods. Realistic information should be given about the applicability and limitations of the assays in clinical diagnostics, the critical point being the sensitivity and situations where interference is expected (e.g. drugs).

The between laboratory comparison of results, even with the same method, is a perennial challenge for quality control, and the clinician should be aware of this.

Despite the 'vogue' of MS, we should not forget the fact that it also is a method subject to errors and does not automatically deserve the epithet 'gold standard'. The choice and development of steroid hormone measurements must be assay quality-driven and not method-driven. IAs will remain in clinical diagnostics because in small laboratories MS is not cost effective. A good and well-validated IA is not inferior to MS. It seems prudent to conclude that the selection of an assay should be driven by the measurement performance dictated by the clinical need and not by assay technology. We just have to know the limitations of IAs and use the information wisely. MS methods represent a major technical advancement but do not provide a 'patent solution' to all problems in steroid analysis. The majority of experts in the field emphasize the importance of assay validation and quality control, which go beyond the importance of methodological aspects.

References

1 Schoenheimer R & Rittenberg D. Deuterium as an indicator in the study of intermediary metabolism. *Science* 1935 **82** 156–157. (doi:10.1126/science.82.2120.156)

2 Eneroth P, Hellström K & Ryhage R. Identification and quantification of neutral fecal steroids by gas–liquid chromatography and mass spectrometry: studies of human excretion during two dietary regimens. *Journal of Lipid Research* 1964 **5** 245–262.

3 Rhyage R. Use of a mass spectrometer as a detector and analyzer for effluent emerging from high temperature gas liquid chromatography columns. *Analytical Chemistry* 1964 **36** 759–764. (doi:10.1021/ac60210a019)

4 Shackelton C. Clinical steroid mass spectrometry: a 45-year history culminating in HPLC–MS/MS becoming an essential tool for patient diagnosis. *Journal of Steroid Biochemistry and Molecular Biology* 2010 **121** 481–490. (doi:10.1016/j.jsbmb.2010.02.017)

5 Krone N, Hughes BA, Lavery G, Stewart PM, Arlt W & Shackelton CH. Gas chromatography/mass spectrometry (GC/MS) remains a pre-eminent discovery tool in clinical chemistry investigations even in the era of fast liquid chromatography tandem mass spectrometry (LC/MS/MS). *Journal of Steroid Biochemistry and Molecular Biology* 2010 **121** 196–504. (doi:10.1016/j.jsbmb.2010.04.010)

6 Christakoudi S, Cowan DA & Taylor NF. Steroids excreted in urine by neonates with 21-hydroxylase deficiency: characterization, using GC–MS and GC–MS/MS, of the D-ring and side chain structure of pregnanes and pregnenes. *Steroids* 2010 **75** 34–52. (doi:10.1016/j.steroids.2009.09.011)

7 Zavada M, Safarcik K & Topolcan O. Some problems of radio immunoassay control. *Journal of Radioanalytical Chemistry* 1978 **46** 57–66. (doi:10.1007/BF02519728)

8 Handelsman DJ & Wartofsky L. Requirement for mass spectrometry sex steroid assays in the journal of clinical endocrinology and metabolism. *Journal of Clinical Endocrinology and Metabolism* 2013 **98** 3971–3973. (doi:10.1210/jc.2013-3375)

9 Monaghan PJ, Keevil BG, Stewart PM & Trainer PJ. Case for the wider adoption of mass spectrometry-based steroid testing, and beyond. *Journal of Clinical Endocrinology and Metabolism* 2014 **99** 4434–4437. (doi:10.1210/jc.2014-2258)

10 Weirman ME, Auchus RJ, Haisenleder DJ, Hall JE, Handelsman D, Hankinson S, Rosner W, Singh RJ, Sluss PM & Stanczyk FZ. Editorial: the new instructions to authors for the reporting of steroid hormone measurements. *Journal of Clinical Endocrinology and Metabolism* 2014 **99** 4375. (doi:10.1210/jc.2014-3424)

11 Botelho JC, Shacklady C, Cooper HC, Tai SS, Van Uytanghe K, Theinport LM & Vesper HW. Isotope-dilution liquid chromatography–tandem mass spectrometry candidate reference method for total testosterone in human serum. *Clinical Chemistry* 2013 **59** 372–380. (doi:10.1373/clinchem.2012.190934)

12 Hoofnagle AN & Wener MH. The fundamental flaws of immunoassay and potential solutions using tandem mass spectrometry. *Journal of Immunological Methods* 2009 **347** 3–11. (doi:10.1016/j.jim.2009.06.003)

13 Shackleton CH & Snodgrass GH. Steroid excretion by an infant with an unusual salt-losing syndrome: a gas chromatographic–mass spectrometric study. *Annals of Clinical Biochemistry* 1974 **11** 91–99. (doi:10.1177/000456327401100134)

14 Sjövall J. Analysis of steroids by liquid–gel chromatography and computerized gas chromatography–mass spectrometry. *Journal of Steroid Biochemistry* 1975 **6** 227–232.

15 Arlt W, Biehl M, Taylor AE, Hahner S, Libe R, Hughes BA, Schneider P, Smith DJ, Stiekema H, Krone N *et al*. Urine steroid metabolomics as a biomarker tool for detecting malignancy in adrenal tumors. *Journal of Clinical Endocrinology and Metabolism* 2011 **96** 3775–3784. (doi:10.1210/jc.2011-1565)

16 Honour J. Development and validation of a quantitative assay based on tandem mass spectrometry. *Annals of Clinical Biochemistry* 2011 **48** 97–111. (doi:10.1258/acb.2010.010176)

17 Yates AM, Bowron A, Calton L, Heynes J, Field H, Rainbow S & Keevil B. Interlaboratory variation in 25-hydroxyvitamin D2 and 25-hydroxyvitamin D3 is significantly improved if common calibration material is used. *Clinical Chemistry* 2008 **54** 2082–2084. (doi:10.1373/clinchem.2008.110452)

18 Middle JG. Dehydroepiandrosterone sulphate interferes in many direct-infusion immunoassays for testosterone. *Annals of Clinical Biochemistry* 2007 **44** 173–177. (doi:10.1258/000456307780118082)

19 Guo T, Taylor RL, Singh RJ & Soldin SJ. Simultaneous determination of 12 steroids by isotope dilution liquid chormatography–photospray ionization tandem mass spectrometry. *Clinica Chimica Acta* 2006 **372** 76–82. (doi:10.1016/j.cca.2006.03.034)

20 Scwartz E, Liu A, Randall H, Keune F, Murray M, Longo N & Pasquali M. Use of steroid profiling by UPLC-MS/MS as a second tier test in newborn screening for congenital adrenal hyperplasia: the Utah experience. *Pediatric Research* 2009 **66** 230–235. (doi:10.1203/PDR.0b013e3181aa3777)

21 Hicks R, Yee JK, Mao CS, Graham S, Kharrazi M, Lorey F & Lee WP. Precursor to product ratios reflect biochemical phenotype in congenital adrenal hyperplasia. *Metabolomics* 2014 **10** 123–131. (doi:10.1007/s11306-013-0558-1)

22 Peitzsch M, Dekkers T, Haase M, Sweep FC, Quack I, Antoch G, Siegert G, Lenders JW, Deinum J, Willenberg HS & Eisenhofer G. An LC-MS/MS method for steroid profiling during adrenal venous sampling for investigation of primary aldosteronism. *Journal of Steroid Biochemistry and Molecular Biology* 2015 **14** 75–84. (doi:10.1016/j.jsbmb.2014.10.006)

23 O'Reilly M, Taylor AE, Crabtree NJ, Hughes BA, Capper F, Crowley RK, Stewart PM, Tomlinson JW & Arlt W. Hyperandrogenemia predicts metabolic phenotype in polycystic ovary syndrome: the utility of serum androstenedione. *Journal of Clinical Endocrinology and Metabolism* 2014 **99** 1027–1036. (doi:10.1210/jc.2013-3399)

24 Münzker J, Hofer D, Trummer C, Ulbing M, Harger A, Pieber T, Owen L, Keevil B, Brabant G, Lerchbaum E *et al*. Testosterone to dihydrotestosterone ratio as a new biomarker for an adverse metabolic phenotype in the polycystic ovary syndrome. *Journal of Clinical Endocrinology and Metabolism* 2015 **100** 653–660. (doi:10.1210/jc.2014-2523)

25 Monaghan PJ, Owen LJ, Trainer PJ, Brabant G, Keevil BG & Darby D. Comparison of serum cortisol measurement by immunoassay and liquid chromatography–tandem mass spectrometry in patients receiving the 11β-hydroxylase inhibitor metyrapone. *Annals of Clinical Biochemistry* 2011 **48** 441–446. (doi:10.1258/acb.2011.011014)

26 Jung C, Ho JT, Torpy DJ, Rogers A, Doogue M, Lewis JG, Czajko RJ & Inder WJ. A longitudinal study of plasma and urinary cortisol in pregnancy and *postpartum*. *Journal of Clinical Endocrinology and Metabolism* 2011 **96** 1533–1540. (doi:10.1210/jc.2010-2395)

27 Owen LJ, Adaway JE, Davies S, Neale S, El-Farhan N, Ducroq D, Evans C, Rees DA, MacKenzie F & Keevil BG. Development of a rapid assay for the analysis of serum cortisol and its implementation into a routine service laboratory. *Annals of Clinical Biochemistry* 2013 **50** 345–352. (doi:10.1177/0004563212473448)

28 Taylor RL, Grebe SK & Singh RJ. Quantitative highly sensitive liquid chromatography–tandem mass spectrometry method for detection of synthetic corticosteroids. *Clinical Chemistry* 2004 **50** 2345–2352. (doi:10.1373/clinchem.2004.033605)

29 Perogamvros B, Keevil BG, Ray DW & Trainer PJ. Salivary cortisone is a potential biomarker for serum free cortisol. *Journal of Clinical Endocrinology and Metabolism* 2010 **95** 4951–4958. (doi:10.1210/jc.2010-1215)

30 Raff H. Cushing sydrome: update on testing. *Endocrinology and Metabolism Clinics of North America* 2015 **44** 43–50. (doi:10.1016/j.ecl.2014.10.005)

31 Hershel R & Singh RJ. Measurement of late-night salivary cortisol and cortisone by LC–MS/MS to assess preanalytical sample contamination with topical hydrocortisone. *Clinical Chemistry* 2012 **58** 947–948. (doi:10.1373/clinchem.2012.182717)

32 Middle JG & Kane JW. Oestradiol assays: fitness for purpose? *Annals of Clinical Biochemistry* 2009 **46** 441–456. (doi:10.1258/acb.2009.009102)

33 Fiers T, Casetta B, Bernaert B, Vandersypt E, Debock M & Kaufman J-M. Development of a highly sensitive method for the quantification of oestrone and oestradiol in serum without derivatisation. *Journal of Chromatography. B, Analytical Technologies in the Biomedical and Life Sciences* 2012 **15** 893–894. (doi:10.1016/j.jchromb.2012.02.034)

34 Owen LJ, Wu FC & Keevil BG. A rapid direct assay for the routine measurement of oestradiol and oestrone by liquid chromatography–tandem mass spectrometry without derivatisation. *Annals of Clinical Biochemistry* 2014 **51** 360–367. (doi:10.1177/0004563213501478)

35 Rosner W, Hankinson SE, Sluss PM, Vesper HW & Wierman ME. Challenges to the measurement of estradiol: an endocrine society position statement. *Journal of Clinical Endocrinology and Metabolism* 2013 **98** 1376–1387. (doi:10.1210/jc.2012-3780)

36 Carter GD. 25-hydroxyvitamin D: a difficult analyte. *Clinical Chemistry* 2012 **58** 486–488. (doi:10.1373/clinchem.2011.180562)

37 Coldwell RD, Trafford DJ, Makin HL, Varely MJ & Kirk DN. Specific estimation of 24,25-dihydroxyvitaminD in plasma by gas chromatography–mass spectrometry. *Clinical Chemistry* 1984 **30** 1193–1198.

38 Heijboe AC, Blankenstein MA, Kema IP & Bujis MM. Accuracy of 6 routine 25-hydroxyvitamin D assays: influence of vitamin D binding protein concentration. *Clinical Chemistry* 2012 **58** 543–548. (doi:10.1373/clinchem.2011.176545)

39 Farrell CJ, Martin S, McWhinney B, Straub I, Williams P & Herrmann M. State-of-the-art vitamin D assays: a comparison of automated immunoassays with liquid chromatography–tandem mass spectrometry methods. *Clinical Chemistry* 2012 **58** 531–543. (doi:10.1373/clinchem.2011.172155)

40 Notice of vitamin D standardization program. *Federal Register* 2011 **76** 11502.

41 Yalow RS & Berson SA. Assay of plasma insulin in human subjects by immunological methods. *Nature* 1959 **184** (Suppl 21) 1648–1649. (doi:10.1038/1841648b0)

42 Abraham GE. Solid-phase radioimmunoassay of estradiol-17β. *Journal of Clinical Endocrinology and Metabolism* 1969 **29** 866–870. (doi:10.1210/jcem-29-6-866)

43 Letter of concern. *Journal of Clinical Endocrinology and Metabolism* 2014 **99** 1499. (doi:10.1210/jc.2014-1873)

44 Wierman ME, Auchus RJ, Haisenleder DJ, Hall JE, Handelsman D, Hankinson S, Rosner W, Singh RJ, Sluss PM & Stanczyk FZ. Editorial: the new instructions to authors for the reporting of steroid hormone measurements. *Journal of Clinical Endocrinology and Metabolism* 2014 **28** 1917. (doi:10.1007/s12672-014-0202-1)

45 Vesper HW, Bhasin S, Wang C, Tai SS, Dodge LA, Singh RJ, Nelson J, Ohorodnik S, Clarke NJ, Salameh WA *et al*. Interlaboratory comparison study of serum testosterone [corrected] measurements performed by mass spectrometry methods. *Steroids* 2009 **74** 498–503. (doi:10.1016/j.steroids.2009.01.004)

46 Legro RS, Schlaff WD, Diamond MP, Coutifaris C, Casson PR, Brzyski RG, Christman GM, Trussel JC, Krawetz SA, Snyder PJ *et al*.

Total testosterone assays in women with polycystic ovary syndrome: precision and correlation with hirsutism. *Journal of Clinical Endocrinology and Metabolism* 2010 **95** 5305–5313. (doi:10.1210/jc.2010-1123)

47 Vesper HW, Botelho JC, Vidal ML, Rahmani Y, Thienport LM & Caudill SP. High variability in serum estradiol measurements in men and women. *Steroids* 2014 **82** 7–13. (doi:10.1016/j.steroids.2013.12.005)

48 Wood L, Ducroq DH, Fraser HL, Gillingwater S, Evans C, Pickett AJ, Rees DW, John R & Turkes A. Measurement of urinary free cortisol by tandem mass spectrometry and comparison with results obtained by gas chromatography–mass spectrometry and two commercial immunoassays. *Annals of Clinical Biochemistry* 2008 **45** 380–388. (doi:10.1258/acb.2007.007119)

49 Dodd AJ, Ducroq D, Neale SM, Wise MP, Mitchem KL, Armston A, Barth JH, El-Farhan N, Rees DA & Evans C. The effect of serum matrix and gender on cortisol measurements by commonly used immuno-assays. *Annals of Clinical Biochemistry* 2014 **51** 379–385. (doi:10.1177/0004563213514567)

50 Turpeinene U & Hamalainen E. Determination of cortisol in serum, saliva and urine. *Best Practice & Research. Clinical Endocrinology & Metabolism* 2013 **27** 795–801. (doi:10.1016/j.beem.2013.10.008)

51 El-Farhan N, Pickett AJ, Ducroq D, Bailey C, Mitchem K, Morgan N, Armston A, Jones L, Evans C & Rees DA. Method-specific serum cortisol responses to the adrenocorticotrophin test: comparison of gas chromatography–mass spectrometry and five automated immuno-assays. *Clinical Endocrinology* 2013 **78** 673–680. (doi:10.1111/cen.12039)

52 Miller R, Plessow F, Rauh M, Gröschl M & Kirschbaum C. Comparison of salivary cortisol as measured by different immunoassays and tandem mass spectrometry. *Psychoneuroendocrinology* 2013 **38** 50–57. (doi:10.1016/j.psyneuen.2012.04.019)

53 Briegel J, Sprung CL, Annane D, Singer M, Keh D, Moreno R, Möhnle P, Weiss Y, Avidan A, Brunkhorst FM et al. Multicenter comparison of cortisol as measured by different methods in samples of patients with septic shock. *Intensive Care Medicine* 2009 **35** 2151–2156. (doi:10.1007/s00134-009-1627-9)

54 Rosner W, Auchus RJ, Azziz R, Sluss PM & Raff H. Position statement: utility, limitations, and pitfalls in measuring testosterone: an Endo-crine Society position statement. *Journal of Clinical Endocrinology and Metabolism* 2007 **92** 405–413. (doi:10.1210/jc.2006-1864)

55 Huhtaniemi IT, Tajar A, Lee DM, O'Neill TW, Finn JD, Bartfai G, Boonen S, Casanueva FF, Giwercman A, Han TS 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. *European Journal of Endocrinology/European Federation of Endocrine Societies* 2012 **166** 983–991. (doi:10.1530/EJE-11-1051)

56 Chen Y, Kinney L & Soldin SJ. Performance evaluation of Siemens ADVIA Centaur enhanced estradiol assay and a split sample comparison with liquid chromatography–tandem mass spectrometry. *Clinical Biochemistry* 2012 **45** 811–815. (doi:10.1016/j.clinbiochem.2012.02.016)

57 Hsing AW, Stanczyk FZ, Bélanger A, Schroeder P, Chang L, Falk RT & Fears TR. Reproducibility of serum sex steroid assays in men by RIA and mass spectrometry. *Cancer Epidemiology, Biomarkers & Prevention* 2007 **16** 1004–1008. (doi:10.1158/1055-9965.EPI-06-0792)

58 Norjavaara E, Ankarberg C & Albertsson-Wikland K. Diurnal rhythm of 17β-estradiol secretion throughout pubertal development in healthy girls: evaluation by a sensitive radioimmunoassay. *Journal of Clinical Endocrinology and Metabolism* 1996 **81** 4095–4102.

59 Klein KO, Baron J, Colli MJ, McDonnell DP & Cutler GB Jr. Estrogen levels in childhood determined by an ultrasensitive recombinant cell bioassay. *Journal of Clinical Investigation* 1994 **94** 2475–2480. (doi:10.1172/JCI117616)

60 Khosla S, Amin S, Singh RJ, Atkinson EJ, Melton LJ III & Riggs BL. Comparison of sex steroid measurements in men by immunoassay versus mass spectroscopy and relationships with cortical and trabecular volumetric bone mineral density. *Osteoporosis International* 2008 **19** 1465–1471. (doi:10.1007/s00198-008-0591-5)

61 Watts NB, Adler RA, Bilezikian JP, Drake MT, Eastell R, Orwoll ES, Finkelstein JS & Endocrine Society . Osteoporosis in men: an Endocrine Society clinical practice guideline. *Journal of Clinical Endocrinology and Metabolism* 2012 **97** 1802–1822. (doi:10.1210/jc.2011-3045)

62 Callou de Sá EQ, Feijó de Sá FC, e Silva Rde S, de Oliveira KC, Guedes AD, Feres F & Verreschi IT. Endogenous oestradiol but not testosterone is related to coronary artery disease in men. *Clinical Endocrinology* 2011 **75** 177–183. (doi:10.1111/j.1365-2265.2011.04017.x)

63 Jeon GH, Kim SH, Yun SC, Chae HD, Kim CH & Kang BM. Association between serum estradiol level and coronary artery calcification in postmenopausal women. *Menopause* 2010 **17** 902–907. (doi:10.1097/gme.0b013e3181d76768)

64 Chen Y, Zeleniuch-Jacquotte A, Arslan AA, Wojcik O, Toniolo P, Shore RE, Levitz M & Koenig KL. Endogenous hormones and coronary heart disease in postmenopausal women. *Atherosclerosis* 2011 **216** 414–419. (doi:10.1016/j.atherosclerosis.2011.01.053)

65 Ahmed B, Bairey Merz CN, Johnson BD, Bittner V, Berga SL, Braunstein GD, Hodgson TK, Smith K, Shaw L, Kelsey SF et al. Diabetes mellitus, hypothalamic hypoestrogenemia, and coronary artery disease in premenopausal women (from the National Heart, Lung, and Blood Institute sponsored WISE study). *American Journal of Cardiology* 2008 **102** 150–154. (doi:10.1016/j.amjcard.2008.03.029)

66 Heaney RP, Armas LA, Shary JR, Bell NH, Binkley N & Hollis BW. 25-hydroxylation of vitamin D3: relation to circulating vitamin D3 under various input conditions. *American Journal of Clinical Nutrition* 2008 **87** 1738–1742.

67 Granado Lorencio F, Blanco-Navarro I & Pérez-Sacristán B. Critical evaluation of assays for vitamin D status. *Current Opinion in Clinical Nutrition and Metabolic Care* 2013 **16** 734–740. (doi:10.1097/MCO.0b013e328364ca96)

68 Farrell CJ & Herrmann M. Determination of vitamin D and its metabolites. *Best Practice & Research. Clinical Endocrinology & Metabolism* 2013 **27** 675–688. (doi:10.1016/j.beem.2013.06.001)

69 Chen H, McCoy LF, Schleicher RL & Pfeiffer CM. Measurement of 25-hydroxyvitamin D3 (25OHD3) and 25-hydroxyvitamin D2 (25OHD2) in human serum using liquid chromatography–tandem mass spectrometry and its comparison to a radioimmunoassay method. *Clinica Chimica Acta* 2008 **391** 6–12. (doi:10.1016/j.cca.2008.01.017)

70 Koivula MK, Turpeinen U, Laitinen P & Risteli J. Comparison of automated 25-OH vitamin D immunoassays with liquid chromato-graphy isotope dilution tandem mass spectrometry. *Clinical Laboratory* 2012 **58** 1253–1261.

71 Shin SY, Kwon MJ, Song J, Park H & Woo HY. Measurement of serum total vitamin D (25-OH) using automated immunoassay in comparison [corrected] with liquid chromatography tandem–mass spectrometry. *Journal of Clinical Laboratory Analysis* 2013 **27** 284–289. (doi:10.1002/jcla.21598)

72 Caillet P & Schott AM. Vitamin D testing-author reply. *Lancet* 2012 **379** 5–11. (doi:10.1016/S0140-6736(12)60710-7)