Mini-Review

Improving Science by Overcoming Laboratory Pitfalls With Hormone Measurements

Jacquelien J. Hillebrand, Wjera V. Wickenhagen, and Annemieke C. Heijboer

Amsterdam UMC, Vrije Universiteit Amsterdam and University of Amsterdam, Endocrine Laboratory, Department of Clinical Chemistry, Amsterdam Gastroenterology Endocrinology & Metabolism, Amsterdam, Netherlands.

ORCiD number: 0000-0002-6712-9955 (A. C. Heijboer).

Abbreviations: CV, coefficient of variation; IGF1, insulin-like growth factor 1; LC-MS/MS, liquid chromatography–tandem mass spectrometry; SHBG, sex hormone–binding globulin.

Received: 11 September 2020; Editorial Decision: 4 December 2020; First Published Online: 31 December 2020; Corrected and Typeset: 11 January 2021.

Abstract

Despite all the effort taken, there is often surprisingly little attention paid to the hormone analyses involved in research studies. Thinking carefully about the quality of the hormone measurements in these studies is, however, of major importance, as this attention to methods may prevent false conclusions and inappropriate follow-up studies. We discuss issues regarding hormone measurements that one should consider, ideally prior to starting, or otherwise, as they arise during a scientific study: quality of the technique, expertise, matrices, timing and storage conditions, freeze-thaw cycles, lot-to-lot and day-to-day variation, analyses per batch or sample-wise, singlicate or duplicate measurements, combining methods, and standardization. This article and the examples mentioned herein aim to clarify the need to pay attention to the hormone analyses, and to help in making decisions. In addition, these examples help editors and reviewers of scientific journals to pay attention to the methods section in the submitted manuscripts and ask the right critical questions when needed.

Key words: quality, immunoassay, mass spectrometry, matrix, technique, pre-analysis

Scientific studies or trials on putative diagnostic markers, effects of treatment, or newly discovered mechanisms are important for expanding our knowledge about endocrine disease. These studies are often time- and money-consuming; involve patients or laboratory animals, with the accompanying need for ethical approval; and include many people/co-authors. Often, despite the great effort taken, surprisingly little attention is paid to the hormone analyses involved in such studies. Thinking carefully about the quality of the hormone measurements in these studies is, however, of major importance, as this may prevent false conclusions and inappropriate follow-up studies. In this paper, we discuss issues regarding hormone measurements that one should consider, ideally prior to starting, or otherwise, as they arise during a scientific study involving hormone measurements.
Quality of Technique

A range of techniques can be used to measure hormone concentrations, from bioassays to isotope dilution liquid chromatography-tandem mass spectrometry (ID-LC-MS/MS). Every technique has its advantages and disadvantages; however, not every technique is suitable for measuring the hormone of interest in the specific bodily fluid or medium that is collected during a study. At present, the most commonly used techniques for measuring hormone concentrations are immunoassays and liquid chromatography, the latter often coupled to tandem mass spectrometry. All immunoassays rely on antibody binding to the analyte (hormone) of interest and have inherent problems with specificity due to cross-reactivity. Steroid hormone immunoassays are particularly notorious for this problem. Problems of cross-reactivity in testosterone measurements are well-described in literature (1-6). For example, dehydroepiandrosterone sulfate (DHEAS) cross-reacts with testosterone concentrations, which is especially relevant dehydroepiandrosterone sulfate (DHEAS) cross-reacts with testosterone concentrations, which is especially relevant due to cross-reactivity. Steroid hormone immunoassays are particularly notorious for this problem. Problems of cross-reactivity in testosterone measurements are well-described in literature (1-6). For example, dehydroepiandrosterone sulfate (DHEAS) cross-reacts with several testosterone immunoassays, leading to falsely high testosterone concentrations, which is especially relevant in samples from women (6). Even the second-generation testosterone immunoassays show falsely high testosterone concentrations in neonates during the first month of life due to cross-reactivity (3). For other steroid hormones, similar cross-reactivity issues occur (7-11). Therefore, in theory for all steroid hormones, LC-MS/MS methods are superior to immunoassays (5). An additional advantage of measurement by LC-MS/MS is that, depending on assay design, multiple hormones (for instance cortisol and cortisone) can be measured in a single run, whereas separate immunoassays are necessary for measuring each hormone. LC-MS/MS may therefore be faster and requires less sample volume. However, in practice, LC-MS/MS methods are not always superior to immunoassays (12), as shown in a study from Legro et al (13). They sent serum samples from a study with women with polycystic ovary syndrome to the laboratories of Quest Diagnostics and the Mayo Clinic for LC-MS/MS testosterone measurements. The reported testosterone concentrations by these laboratories correlated as badly as immunoassays. This might be caused by, for instance, the experience of the laboratory with this complex technique, the time spent on development and validation, as well as the agreed quality criteria for LC-MSMS method validation.

In addition to the cross-reactivity problem, immunoassays may suffer from interference from other components in the sample (the matrix, see also below), for instance, problems due to differences in binding protein concentrations. All steroid hormones circulate in the serum mainly bound to binding proteins (eg, testosterone to sex hormone–binding globulin [SHBG], 25-hydroxy vitamin D to vitamin D binding protein, cortisol to cortisol binding globulin) and only in a small percentage as free, unbound, steroid in the serum. In most laboratories, the total steroid hormone concentration (which means bound + unbound hormone) is measured. For total steroid hormone measurements, the steroid hormone has to be extracted from the binding protein. In many automated immunoassays, the time and type of chemicals for extraction are fixed or limited, which may lead to analytical problems in serum of experimental subjects with relatively high or low binding protein concentrations (14-16). These relatively high or low binding protein concentrations are quite common; for instance, pregnant women and women using oral contraceptives have high levels of any binding proteins, whereas levels are low in patients in the intensive care unit and patients with liver disease. An immunoassay can therefore perform well in relatively healthy subjects but perform different or even poorly in a specific patient or study group. Using such a method may lead to incorrect conclusions (15, 17, 18). For example, in a Dutch study, serum testosterone concentrations, measured using a radioimmunoassay, decreased after 3 cycles of oral contraceptive use; however, when serum samples from this study were reanalyzed using an accurate LC-MS/MS method, the serum testosterone concentrations did not change upon the oral contraceptive use, indicating the previous conclusion was incorrect (15). The radioimmunoassay was ultimately proven to be influenced by SHBG concentrations. Like steroid hormones, thyroid hormone also circulates in the serum bound to binding proteins and total thyroid hormone levels are influenced as described above. For thyroid hormone, most often the free hormone concentration is measured, which is theoretically independent of changes in binding protein concentrations. In practice, it however turns out that several free thyroxine (fT4) immunoassays still have troubles coping with extreme differences in serum matrices of for instance increased thyroxine binding globulin (TBG) concentrations (19), or TBG deficiency (20) which has a prevalence of 1 in 4000 newborns (partial) or 1 in 15 000 male newborns (complete). Other free hormones, like free testosterone, are technically even more challenging to measure accurately (21). According to the (debatable) free hormone hypothesis (which suggests that, for example, only the free testosterone is biologically active), free hormones are preferred to be measured over total concentrations. As this is hindered by the complexity of the measurements, with evidence for inaccurate methods on the market, mathematical calculations of free hormones are often used (22, 23). These calculations depend on the quality of the measurements of total testosterone, SHBG, and albumin in case of free testosterone, and additionally, on correct estimates of the association.
constants and stoichiometry for binding of testosterone to SHBG and albumin, which makes the use of calculations debatable as well (21, 24).

At the moment, immunoassays are still the most frequently used methods for measuring peptide hormones, although gradually more LC-MS/MS methods for peptide hormone analysis are being developed. For peptide hormones, cross-reactivity also reduces immunoassay specificity but because of the higher molecular weight of peptides, the immunoassays are based on an immunometric (sandwich) principle, and therefore suffer less from cross-reactivity (relative to smaller steroid hormones measured by competitive immunoassays). In general, method comparisons between peptide immunoassays and LC-MS/MS assays show good agreement, this may however differ per peptide or immunoassay (25-28). A novel topic here is that (common) protein variants which are detected by (some) immunoassays, are not detected by LC-MS/MS (29). Hines et al (29) discovered a fairly common insulin-like growth factor 1 (IGF1) variant, A70T-IGF1, of unknown clinical significance and present in ~0.6% of the population, which is detected by the common IGF1 immunoassays yet leads to falsely low IGF1 concentrations when measured using their LC-MS/MS method. Whether this is a preferable characteristic of the LC-MS/MS technique, depends on whether the variant is pathogenic or not. As more LC-MS/MS methods to determine peptide hormones are developed, we will probably see more of these discrepancies. For example, a very common luteinizing hormone (LH)-β variant can be expected to show different results when using LC-MS/MS as it already shows different results in different immunoassays (30, 31). On the other hand, for instance, insulin analogues cross-react assay-dependently in insulin immunoassays, but can be separately analyzed using the LC-MS/MS technique, with main applications for doping control and factitious hyperinsulinism (32, 33).

**Expertise**

Manufacturers of assays suggest that it is straightforward to perform hormone analyses by simply buying their hormone assay kits, reading the manual, and performing the analyses. The kits are supplied with internal quality control samples, which would imply guaranteed quality of the analyses. This procedure, however, does not warrant high-quality measurements. For every new assay that is used in a laboratory, an assay verification should be performed on-site. This is common practice in diagnostic laboratories (according to ISO15189, the international standard that specifies requirements for quality and competence in medical laboratories (34)) but should also hold for assays performed for scientific studies (35). Some, certainly not all, manufacturers may provide assay validation data that are too good to be true. Validation data described in their kit insert may be generated with only control solutions with a different matrix than real human serum. Sometimes, assay precision as coefficients of variation (CVs) is given only for samples with extremely high concentrations, while the samples in the lower concentration range show a poor CV. Alternatively, without explanation, the sublime validation data of the kit insert are just not possible to repeat or not even available (36). It is therefore important that a verification of a new assay is performed on-site before measurements are done in valuable samples of a scientific study (see Table 1 for the parameters that should be verified). The specifics of the study participants should also be taken into account in this verification. The necessity to verify new assays before use specifically holds for multiplex analyses, which allow measurement of several hormones at once. The obvious advantages of these multiplex immunoassays (increased efficiency, low sample volume, high throughput) may distract from the sometimes-limited quality of these assays, especially regarding cross-reactivity and matrix effects. As soon as verification of an assay has been performed, the analysis of the study samples can be executed. Internal quality controls should always accompany the samples when the assay is performed. It is important that these controls of between-assay variability have concentrations that span the full range of likely results of the study and, in addition, are both independent (from a different manufacturer than the assay kit) and the same in every run. This is the only way to observe alterations in the assay performance over time, as the target of the kit controls may be changed with each new lot number of the assay kit by the manufacturer. Moreover, for those hormones that are frequently used in endocrine diagnostics, external quality assessment schemes are available. Participation in these schemes is essential (and mandatory according to ISO15189) for quality control and allows comparison of results of control samples with other laboratories. Smaller research-oriented laboratories that are running assays only once in a while often do not participate in these external quality assessment schemes. They can, however, take alternative approaches to monitor assay performance over time, such as periodically exchanging independent control samples with other laboratories using the same assay, or by storing internal independent control samples under optimal conditions for longer periods of time and measuring these samples each time the assay is run. The same solution can be used if no external quality assessment scheme is available.
Table 1. Checklist With Parameters That Should be Studied in Hormone Assays and Reported in a Publication

| Validation parameters for a hormone assay to be published (based on ISO 15189) | Check |
|---------------------------------------------------------------------------------|-------|
| Measurement precision<sup>a</sup>                                               |       |
| • Within-run                                                                    |       |
| • Between-run precision                                                          |       |
| Calibration standards used<sup>b</sup>                                            |       |
| Measurement trueness<sup>c</sup>                                                 |       |
| • Method comparison                                                              |       |
| • Reference material used                                                        |       |
| • Spike-recovery                                                                 |       |
| Linearity<sup>d</sup>                                                            |       |
| • Maximum dilution                                                               |       |
| • Measurement range                                                              |       |
| Detection limits (Sensitivity)<sup>e</sup>                                        |       |
| • Limit of quantitation (LOQ)                                                    |       |
| • Limit of detection (LOD)                                                       |       |
| Stability<sup>f</sup> (samples/standards/controls/reagents)                      |       |
| Specificity/selectivity<sup>g</sup>                                              |       |
| • Interferences/cross-reactivity                                                 |       |
| • Matrix effects                                                                 |       |
| • Suitable for patient groups                                                    |       |
| Carryover<sup>h</sup>                                                            |       |
| Diagnostic characteristics<sup>i</sup>                                            |       |
| • Reference interval                                                             |       |
| • Threshold values                                                               |       |

<sup>a</sup> Whether or not all of these performance characteristics have to be examined should be judged by someone with expertise regarding the measurements and who understands the goal and intention of the study and its intended use; in addition, examination should of course be statistically sound.

Additional explanation: <sup>b</sup> Within-run precision should always be performed and reported in a publication, between-run precision should be reported if samples were measured in 2 or more batches. <sup>c</sup> The standardization of the method should always be examined and reported in a publication; otherwise, it is not clear how it compares to other studies. <sup>d</sup> Dilution experiments are important if you need to dilute the study samples because of concentrations above the measurement range. The measurement range of the assay should always be reported in a publication. <sup>e</sup> LoQ/LoD is important if you have low concentrations in the study samples with a high (>15%) intra-assay variation. The measurement range of the assay should always be reported in a publication. <sup>f</sup> Important to validate the conditions of the study samples (duration and temperature): time stored in the fridge or in the freezer, at ~20 °C or ~80 °C; these results should also be reported in a publication. If using in house standards, controls, or reagents, the stability should also be checked. <sup>g</sup> The sample matrix should always be validated; is the method suitable to measure heparin plasma/serum/urine etcetera samples? Moreover, it should be validated whether the method is suitable for samples from your specific patient group. These data should be reported in a publication. In addition, interferences experiments are important, such as checking for cross-reactivity with similar hormones. Sometimes this information is available in literature. <sup>h</sup> Carryover experiments should be checked for, especially if there is a wide range in concentrations. <sup>i</sup> Diagnostic characteristics might be necessary to interpret measured concentration in the study; however, this depends on the study design. If important, these data including the origin need to be reported in a publication.

Lastly, experience in lab work improves precision. Although it might seem financially attractive to involve a relatively unexperienced student in performing manual assays, experience indicates that this leads to a higher imprecision of the results. This then may mean the difference between obtaining a significant result or just noise.

## Matrices

The matrix of a sample is defined as the components of the sample other than the analyte of interest. Several options of collection tubes exist when drawing blood. Many more than 50 types of blood collection tubes are available on the market to choose from. The blood collection tubes can be manufactured of plastic or glass, with or without a clot activator, with or without a (double) polymer gel, with K2 or K3 EDTA, with lithium heparin or with sodium heparin, and so on. Most used for hormone analyses are serum, EDTA plasma, and lithium heparin plasma containers. All these different tubes lead to different matrices, as the components will be different; for example, plasma does but serum does not contain clotting factors. Importantly, not all matrices will lead to the same results with regard to the hormone analysis and the choice of blood collection tube may thereby introduce pre-analytical variation. The experimental method may rule out a typical matrix; for instance EDTA plasma is not suitable for DELFIA assays based on Europium fluorescence. The matrix may also influence the stability of hormones and separator gels from blood containers can result in analytical problems. Hence, it is important to know beforehand what collection tube is needed for the specific hormone analysis and to process (centrifuge and aliquot) samples quickly. In addition, it is important to communicate clearly on the type of collection tube needed in a multicenter study. Examples are known of multicenter studies where it was communicated that “plasma” was needed; in some centers K3 EDTA plasma was collected, while other centers used lithium heparin plasma or citrate plasma. Even serum can be collected when the coworkers are not acquainted with the difference between plasma and serum. This mix-up may lead to unknown problems at the time of analysis or strange results if one wants to correct a mistake.

For instance, measuring leptin in citrate plasma leads to 20% to 25% lower concentrations compared with measuring it in EDTA plasma, an example that makes clear that it is crucial to only collect one type of plasma (the same matrix) in every center of a multicenter study. Apart from blood, also urine, saliva, cell lysates, hair, cerebrospinal fluid, or other bodily fluids can be used to measure hormone concentrations. It is important to know in advance whether it is possible to measure the hormones of interest in these matrices and a specific verification...
might be needed (42). Such a verification includes a spike-recovery study (addition of a known amount of the hormone of interest to the new matrix and measurement of what percent of the known amount is measured) with a sufficient recovery of 80% to 120% as a requirement. Some hormones, especially peptide hormones (see Table 2) are easily degraded following blood withdrawal. In such cases, blood samples must be processed quickly and unspecific protease inhibitors may be necessary to add to the blood immediately after collection (42, 43). This procedure should also be verified. Surprisingly, sometimes it may be demonstrated that it is no longer necessary to use protease inhibitors in newer, more specific, assays (44, 45).

**Timing and Storage Conditions**

Inappropriate timing of blood sampling may introduce unnecessary pre-analytical variation. Before taking samples, one should therefore consider whether circadian rhythms in hormones are involved and whether sampling should be performed early in the morning, possibly even in a fasted state, or randomly during the day (46). Samples are often collected during a substantial period of time and stored until analysis takes place (see also “Freeze-Thaw Cycles” section). Samples should be stored frozen at −20 °C—or preferably −70/−80 °C—since not all hormones are stable in serum but not in urine (37, 42). Before running the experiment, storage temperature and time should preferably be verified. For long storage time, this is often difficult; however, this is not always possible and freezer space should be checked beforehand.

Not only storage temperature but also storage time is of importance. The longer the storage time, the higher the change of degradation of the hormone. For instance, C-peptide is relatively stable following storage at −20 °C for 1 or several months, but Gislefoss et al showed complete loss of C-peptide in samples stored at −20 °C for 25 years (48). Prolactin, like C-peptide, is stable at −20 °C for several months, but not for several years; storage at −80 °C is here a better choice (48, 49). During storage sublimation (in this case the transition from ice (frozen serum) directly into water vapor without going through the liquid phase) may occur, especially if only a small sample volume is stored. This will influence the concentrations measured and result in incorrectly high concentrations. Once suspected, it is possible to check for sublimation by measuring the sodium concentration in the samples. If hypernatremia is seen in all samples, this indicates that samples were concentrated during storage in the freezer (50). Storage conditions and time of storage may depend on the matrix used. Steroid hormones are stable in blood serum or plasma as well as saliva (51, 52). However, the peptide hormone parathyroid hormone is more stable in EDTA plasma than serum, and for instance klotho, the co-receptor of FGF23, is stable when assayed in serum but not in urine (37, 42). Before running the experiment, storage temperature and time should preferably be verified. For long storage time, this is often difficult; evidence from literature may help here in deciding on optimal storage conditions.

### Table 2. Hormones That Are Easily Degraded Following Blood Withdrawal

| Hormone | Reference |
|---------|-----------|
| ACTH    | (68, 69)  |
| ADH     | (68, 70)  |
| Calcitonin | (71)     |
| PTH     | (72, 73)  |
| Insulin | (68, 73)  |
| (Acylated) Ghrelin | (74) |
| VIP     | (68)      |
| GLP1    | (75)      |
| Glucagon| (44, 75)  |
| PYY     | (74)      |
| CCK     | (74, 76, 77) |
| Osteocalcin | (72, 73, 78) |
| CTX     | (78)      |
| FGF23   | (45)      |

Abbreviations: ACTH, adrenocorticotropic hormone; ADH, antidiuretic hormone; CCK, cholecystokinin; CTX, C-terminal telopeptide (of type I collagen); FGF23, fibroblast growth factor 23; GLP1, glucagon-like peptide 1; PTH, parathyroid hormone; PYY, peptide YY; VIP, vasoactive intestinal peptide.

*Measurement of these hormones requires specific sampling conditions or prompt processing (separation of blood cells from plasma/serum) of the sample. The instability of these hormones may be variable, depending on, for instance, the type of (anticoagulant) blood containers and assay (measuring hormone fragments or not) used.

**Freeze-Thaw Cycles**

If samples are stored for batch analysis at the end of the data collection, it is important to store samples in several aliquots for different hormone analyses to prevent freeze-thaw cycles (every time a sample is frozen and thawed). If the samples must be thawed and again refrozen for every measurement, this may negatively influence the results, as some hormones are not resistant to freeze-thaw cycles (46, 53-55). Fortunately, steroid hormones generally are not affected by freeze-thaw cycles (56, 57). It often occurs that one wants to measure an additional hormone in the collected samples. If no aliquot is reserved for a specific hormone of interest and it is therefore necessary to use aliquots that have been thawed and again frozen before, and the impact of freeze-thaw cycles for the specific hormone is unknown, this issue should first be verified before study samples are analyzed.
Lot-to-Lot and Day-to-Day Variation

All methods and techniques show a certain day-to-day variation in their measurements, which limits the production of consistent results over time. This is due to varying temperature, humidity, reagents, equipment, technicians, and other, sometimes unknown, factors. Often reagents and calibrators are produced in lots (batches), which can differ substantially, leading to differences in measured hormone concentrations. Immunoassays are prone to this so-called lot-to-lot variation (58, 59). Both common day-to-day variation and lot-to-lot variation should be taken into account when performing hormone analyses for scientific purposes. Not every laboratory checks for this lot-to-lot variation, although it is important, as it can influence the results significantly. The earlier-mentioned manufacturer-independent human quality controls are needed to determine the imprecision, expressed as coefficient of variation (CV) of the method. When using these controls, increasing CVs warn for extra attention. The higher the CV of the measurement, the lower the correlation coefficient in association studies. In addition, with higher CVs, small (expected) differences between groups might not be noticed.

Analysis per Sample or Batches

The day-to-day and lot-to-lot variation will influence your choice of measuring your samples straight away per sample (flow) or after finishing the collection of all samples (batch). When a small number of samples is collected, it is often best to measure all samples in a single batch, which immediately dispenses with day-to-day or lot-to-lot variation. When samples need to be analyzed in a few batches, at least lot-to-lot variation can be avoided. When analyzing samples in more than one batch, take care that treatment and control groups are mixed over batches. The drawback of performing analyses in a batch after finishing the study is that sample storage time may influence the results (see storage temperature and time). The challenge here is to find the optimal conditions for the specific study.

Singlicate or Duplicate Analyses

Analyses can be performed in singlicate or duplicate (or even triplicate). It is common in diagnostic laboratories to perform automated analyses singlicate and manual analyses in duplicate, to reduce human errors. Duplicate measurements of manual analyses will lead to a higher precision of the reported value because the mean value of the 2 measurements will be reported and if the difference between the duplicate measurements is higher than the agreed criterion, measurements will be repeated. This procedure is necessary to report reliable results for individual patients in diagnostics. This is important in scientific studies as well, as a higher variation leads to undesirable noise in your data. However, depending on the study design, when using a robust method with a low intra-assay variation in combination with a study with a high number of subjects, you might consider running the manual assay with singular measurements (60). A power calculation can be used to help determining the measurement uncertainty that is acceptable.

Combining Methods or Standardization

Methods to measure hormones are well known for their differences in standardization (61-66), leading to difficulties in combining results from various laboratories using various methods. Preferably, all measurements of a certain hormone should be performed in one laboratory using one method. This is sometimes challenging in multicenter studies, which means transport of samples (under the right conditions), or in long-term studies with sample-wise (flow) measurements dealing with sometimes changing methods in the local laboratory. One should also pay attention for assay updates; assays may be improved and firms may release a newer, better version of the assay, which may lead to significantly different results (eg, Roche’s Elecsys second-generation serum cortisol assay, a few years ago). Although the international clinical chemistry community tries to standardize or harmonize methods (61, 66, 67), this is not an easy job. When method changes or assay updates occur during the course of a study, provisional solutions have to be found to deal with these changes for the current studies.

If insurmountable obstacles exist to using only one method in one study, method comparisons should be made to see whether combining data from samples measured using these different methods is possible. Differences in standardization can be solved in this way. Using the same quality controls can also help to cross-check the various methods. However, other problems, such as those mentioned in the “Quality of Technique” section, may still exist and prevent the possibility of combining results from different laboratories.

Money can play an important role in choosing how to perform the hormone analyses in studies. Simply going for the cheapest option, without thinking about the quality, will often lead to a “pig in a poke.” This is not beneficial for the researcher, not to the advantage of the scientific community, and furthermore not ethical with regard to
the patients participating in the study and the subsidizing party. We therefore urge all scientists to think carefully about the analyses performed in their studies. We hope that the issues mentioned in this article clarify the need and help in making decisions. A checklist is included (Table 3) to go through before collecting samples. We described some general and common problems that may be encountered and need to be addressed; however, we realize that each study is different and might need additional thought, leading to questions that can often be solved by contacting a laboratory specialist. Laboratory specialists may be based in local diagnostic laboratories or bigger commercial laboratories; either way, they can be consulted for questions related to experimental design and laboratory analyses.

In addition, we ask editors and reviewers of the scientific journals to pay attention to the methods section in the submitted manuscripts. To help editors and reviewers with this job, we added a checklist with parameters that should be studied in hormone assays and reported in a publication (Table 1). Too often the method used to measure the hormone concentration is not even mentioned in a published article. In the manuscripts it should be clear that the authors paid attention to the method used and its associated issues.

Table 3. Checklist of Issues to be Considered or Discussed With a Laboratory Specialist Before Starting a Study Containing Hormones Measurements

| Issue | Check |
|--------|-------|
| 1. What possible techniques can be used to measure the hormone of interest and what is the (expected) quality of these techniques in the specific subjects used in the study? |
| 2. What is the variation (precision) of the method? Is this good enough to observe the expected differences? |
| 3. Has the method of choice shown good performance for the study population? |
| 4. Will the medication used in the method influence the analyses? |
| 5. Which lab has experience with the method of choice? Is there a method validation/verification available? If not, will this be performed? |
| 6. What matrices (collection tubes) are needed for the measurements? (Or: can the measurements be performed in the matrix that was collected?) |
| 7. How do the samples need to be stored (time and temperature)? |
| 8. How many freeze-thaw cycles are allowed? |
| 9. Will the measurements be performed in singlicate or duplicate and is this justifiable? |
| 10. Will the measurements be performed in one batch, using one lot number of reagents? |
| 11. Will all measurements be performed using one method (preferable)? If not, can the methods be combined? |

Additional Information

Correspondence: Annemieke C. Heijboer, Amsterdam UMC, Endocrine Laboratory (K2-283), Meibergdreef 9, 1105 AZ, Amsterdam, The Netherlands. Email: a.heijboer@amsterdamumc.nl

Disclosures: The authors have nothing to disclose.

Data Availability: Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

References

1. Rosner W, Vesper H. Endocrine Society; American Association for Clinical Chemistry; American Association of Clinical Endocrinologists; Androgen Excess/P COS Society; American Society for Bone and Mineral Research; American Society for Reproductive Medicine; American Urological Association; Association of Public Health Laboratories; Endocrine Society; Laboratory Corporation of America; North American Menopause Society; Pediatric Endocrine Society. Toward excellence in testosterone testing: a consensus statement. J Clin Endocrinol Metab. 2010;95(10):4542-4548.

2. Herold DA, Fitzgerald RL. Immunoassays for testosterone in women: better than a guess? Clin Chem. 2003;49(8):1250-1251.

3. Hamer HM, Finken MJ, van Herwaarden AE, du Toit T, Swart AC, Heijboer AC. Falsely elevated plasma testosterone concentrations in neonates: importance of LC-MS/MS measurements. Clin Chem Lab Med. 2018;56(6):e141-e143.

4. 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(8):1381-1395.

5. Handelsman DJ, Wartofsky L. Requirement for mass spectrometry sex steroid assays in the Journal of Clinical Endocrinology and Metabolism. J Clin Endocrinol Metab. 2013;98(10):3971-3973.

6. Middel JG. Dehydroepiandrosterone sulphate interferes in many direct immunoassays for testosterone. Ann Clin Biochem. 2007;44(Pt 2):173-177.

7. 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.

8. Handelsman DJ, Newman JD, Jimenez M, McLachlan R, Sartorius G, Jones GR. Performance of direct estradiol immunoassays with human male serum samples. Clin Chem. 2014;60(3):510-517.

9. Debeljak Ž, Marković I, Pavela J, et al. Analytical bias of automated immunoassays for six serum steroid hormones assessed by LC-MS/MS. Biochem Med (Zagreb). 2020;30(3):30701.

10. 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. Ann Clin Biochem. 2011;48(Pt 5):441-446.

11. Krasowski MD, Drees D, Morris CS, Maakestad J, Blau JL, Ekins S. Cross-reactivity of steroid hormone immunoassays: clinical significance and two-dimensional molecular similarity prediction. BMC Clin Pathol. 2014;14:33.
43. Smith ER, Ford ML, Tomlinson LA, et al. Instability of fibroblast growth factor-23 (FGF-23): implications for clinical studies. Clin Chim Acta. 2011;412(11-12):1008-1011.
44. Emmen JM, Heijboer AC, de Jong SM, Endert E. Glucagon stability anno 2014. Clin Chim Acta. 2015;440:1-2.
45. Dirks NF, Smith ER, van Schoor NM, et al. Pre-analytical stability of FGF23 with the contemporary immunoassays. Clin Chim Acta. 2019;493:104-106.
46. Vlot MC, den Heijer M, de Jongh RT, et al. Clinical utility of bone markers in various diseases. Bone. 2018;114:215-225.
47. Goetze JP, Hansen LH, Terzic D, et al. Atrial natriuretic peptides. Clin Chem Lab Med. 2009;47(5):596-603.
48. Bolelli G, Mutti P, Micheli A, et al. Validity for epidemiological studies of long-term cryoconservation of steroid and protein hormones in serum and plasma. Cancer Epidemiol Biomarkers Prev. 1995;4(5):509-513.
49. Craft NE, Epler KS, Butler TA, May WE, Ziegler RG. Evaluation of Serum Volume Losses During Long-Term Storage. J Res Natl Inst Stand Technol. 1993;98(3):355-359.
50. Keevil BG, MacDonald P, Macdowall W, Lee DM, Wu FC; NATSAL Team. Salivary testosterone measurement by liquid chromatography tandem mass spectrometry in adult males and females. Ann Clin Biochem. 2014;51(Pt 3):368-378.
51. Handelmann DJ, Desai R, Seibel MJ, Le Couteur DG, Cumming RG. Circulating sex steroid measurements of men by mass spectrometry and molecular forms of circulating gut peptides. Clin Chim Acta. 2017;475:289-292.
52. Binkley N, Dawson-Hughes B, Durazo-Arvizu R, et al. Vitamin D status as an international issue: national surveys and the problem of standardization. Scand J Clin Lab Invest Suppl. 2012;243:32-40.
53. Heida JE, Boesten LSM, Ettema EM, et al. Comparison of ex vivo stability of copeptin and vasopressin. Clin Chem Lab Med. 2017;55(7):984-992.
54. Kratyczk J, Petzold A, Raue F, et al. Stability of calcitonin and procalcitonin by various assays in patients with and without medullary thyroid cancer. Clin Chem. 2011;57(3):467-474.
55. Koal T, Schmidtner D, Pham-Tuan H, Rohrning C, Rauh M. Standardized LC/MS/MS based steroid hormone profile-analysis. J Steroid Biochem Mol Biol. 2012;129(3-5):129-138.
56. Thompson S, Chesler D. Lot-to-lot variation. Clin Biochem Rev. 2018;39(2):51-60.
57. Algeciras-Schimich A, Bruns DE, Boyd JC, Bryant SC, LaFortune KA, Grebe SK. Failure of current laboratory protocols to detect lot-to-lot reagent differences: findings and possible solutions. Clin Chem. 2013;59(9):1187-1194.
58. Ye Z, Tu J, Midde K, Edwards M, Bennett P. Singlicate analysis: should this be the default for biomarker measurements using ligand-binding assays? Bioanalysis. 2018;10(12):909-912.
59. Vesper HW, Botelho JC, Shacklady C, Smith A, Myers GL. CDC project on standardizing steroid hormone measurements. Steroids. 2008;73(13):1286-1292.
60. Binkley N, Dawson-Hughes B, Durazo-Arvizu R, et al. Vitamin D measurement standardization: The way out of the chaos. J Steroid Biochem Mol Biol. 2017;173:117-121.
61. Nelson SM, La Marca A. The journey from the old to the new AMH assay: how to avoid getting lost in the values. Reprod Biomed Online. 2011;23(4):411-420.
62. Rehfled JF. How to measure cholecalciferol in tissue, plasma and cerebrospinal fluid. Regul Pept. 1998;78(1-3):31-39.
63. Eberlein GA, Eysselin VE, Hesse WH, Goebell H, Schaefer M, Reeve JR Jr. Detection of cholecalciferol-58 in human blood by inhibition of degradation. Am J Physiol. 1987;253(4 Pt 1):G477-G482.
64. Ferreira A, Alho I, Casimiro S, Costa L. Bone remodeling markers and bone metastases: From cancer research to clinical implications. Bonekey Rep. 2015;4:668.