Assessment of the hormonal milieu

Summary

The hormonal milieu has been hypothesized to play a role in a range of human diseases, and therefore has been a topic of much epidemiologic investigation. Hormones of particular interest include: sex steroids; growth hormones; insulin-like growth factors; stress hormones, such as cortisol; and hormones produced by the adipose tissue, termed adipokines. Depending on the hormone, levels may be measured in plasma or serum, urine, saliva, tissue, or by assessing genetic variation in the hormone or hormone metabolizing genes. Sample collection, processing, and storage requirements vary according to the type of sample collected (e.g. blood or urine) and the hormone of interest. Laboratory analysis of hormones is frequently complex, and the technology used to conduct the assays is constantly evolving. For example, direct or indirect radioimmunoassay, bioassay or mass spectrometry can be used to measure sex steroids, each having advantages and disadvantages. Careful attention to laboratory issues, including close collaboration with laboratory colleagues and ongoing quality control assessments, is critical. Whether a single hormone measurement, as is frequently collected in epidemiologic studies, is sufficient to characterize the hormonal environment of interest (e.g. long-term adult hormone exposure) is also an important issue. While the assessment of hormones in epidemiologic studies is complex, these efforts have, and will continue to, add importantly to our knowledge of the role of hormones in human health.

Introduction

The study of hormones and their involvement in human health has been considered for many years, and their measurement has increasingly become an important part of many epidemiologic studies. Examining how various endogenous and exogenous hormones are related to disease increases our understanding of disease etiology, which may ultimately lead to improved prevention recommendations for both high-risk groups and the general population. Issues surrounding the appropriate use of hormone measures in epidemiologic studies are complex and require careful planning by study investigators. Many choices must be made, including the type of biospecimen to collect from participants, the timing and conduct of sample collection, the choice of hormones and assay modalities,
and ultimately how to interpret the results. While several of these issues are dealt with in previous chapters (sample collection and processing in Chapter 3 and interpretation of assay results in Chapter 8), here the focus is on examples and concerns in measuring hormones. Because of the broad range of hormones found in humans, this chapter cannot cover every aspect of hormone measurement. However, general issues are addressed that should be considered when designing epidemiologic studies of hormones, such as the importance of hormones in medical research, their measurement in human samples, and issues regarding assay development and interpretation.

**Context and public health significance**

Hormones are chemicals produced by living cells that act as chemical messengers or signal molecules (1). The hormonal environment is a critical regulator of many physiologic processes, including growth, energy metabolism, fertility and the stress response. A wide range of well-documented diseases are linked to changes, either increases or decreases, in the hormonal milieu (Table 12.1). The role of hypoinsulinemia and insulin resistance in diabetes, excess insulin-like growth factor (IGF I in acromegaly and deficiency in dwarfism, excessive production of thyroid hormone in Graves disease, and overproduction of cortisol in Cushing syndrome are all examples. In addition to these well-established causal relationships, the hormonal milieu has been hypothesized to play a role in a range of other diseases, which have been the focus of many epidemiologic studies. Several examples, described further below, include the association of sex steroid hormones and breast cancer, the IGF system and cognitive function, and the role of adipokines such as adiponectin in both diabetes and heart disease.

This chapter focuses on the measurement of endogenous hormones (i.e. hormones produced by the body). The role of exogenous hormones (i.e. originating outside the body), particularly the use of oral contraceptives and postmenopausal hormones by millions of women worldwide, also has been the subject of substantial scientific study. The evaluation of exogenous hormones is not addressed here, largely because characterizing exposure to these agents is routinely accomplished via administration of questionnaires or tallying of pharmacy prescription records. However, any epidemiologic study of endogenous hormones must take into account sources of exogenous exposure that may influence endogenous hormone levels.

**Examples/case studies**

### Sex steroids and breast cancer risk in postmenopausal women

Substantial data support a role of hormones, particularly sex steroids, in the etiology of breast cancer. There are consistent associations with reproductive factors, and increased risks associated with postmenopausal obesity and use of postmenopausal hormones (2). Further, drugs that either block estrogen binding to the estrogen receptor (selective estrogen receptor modulators, such as tamoxifen) or prevent the production of estradiol (aromatase inhibitors) are effective both in preventing breast cancer and improving survival of women with the disease (3–5). Considerable data assessing circulating sex steroids in postmenopausal women and breast cancer risk have accrued from prospective epidemiologic studies, where circulating levels of endogenous hormones are measured in study subjects before

<table>
<thead>
<tr>
<th>Hormone</th>
<th>Conditions known or hypothesized to be related</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex steroids (e.g. estradiol and testosterone)</td>
<td>Infertility, osteoporosis, cancer (e.g. breast, endometrial and prostate cancers)</td>
</tr>
<tr>
<td>Vitamin D metabolites</td>
<td>Hypertension, osteoporosis, cancer (e.g. colon and breast cancer)</td>
</tr>
<tr>
<td>Insulin</td>
<td>Diabetes, heart disease, cancer (e.g. colon and endometrial cancers), cognitive function</td>
</tr>
<tr>
<td>Insulin-like growth factor/Growth hormone axis</td>
<td>Cancer (e.g. colon, prostate and breast cancers), heart disease, cognitive function, osteoporosis</td>
</tr>
<tr>
<td>Prolactin</td>
<td>Immunologic diseases (e.g. rheumatoid arthritis and systemic lupus erythematosus), breast cancer</td>
</tr>
<tr>
<td>Adiponectin and other adipokines</td>
<td>Diabetes, heart disease, cancer (e.g. colon and breast cancers)</td>
</tr>
<tr>
<td>Stress hormones (e.g. cortisol)</td>
<td>Heart disease, cancer (e.g. breast cancer)</td>
</tr>
<tr>
<td>Melatonin</td>
<td>Breast cancer, sleep disorders</td>
</tr>
</tbody>
</table>
disease diagnosis (6). Overall, a strong positive association exists between breast cancer risk and circulating levels of both estrogens and androgens. Women in the top versus bottom 20% of estrogen levels have a two- to three-fold higher risk of breast cancer (7). The associations are similar for several forms of estrogen (e.g. estradiol, estrone, estrone sulfate). Although additional confirmation is required, the association appears strongest for estrogen receptor-positive breast tumours and is robust across groups of women at varying risk of breast cancer (e.g. defined by family history). Also, a single blood estrogen measure predicts subsequent breast cancer risk for at least 8–10 years. Generally, the more limited data available on urinary estrogens suggests similar predictive ability. For testosterone, a commonly measured androgen, the data are very consistent, with a significant positive association between circulating levels and postmenopausal breast cancer; the magnitude of the association is similar to that observed for estrogens (7). Most studies also noted a similar, although somewhat modest, positive association with other androgens, such as androstenedione, dehydroepiandrosterone (DHEA) and DHEA sulfate.

The insulin-like growth factor (IGF) axis and cognitive function

Insulin-like growth factor I (IGF-I) is a protein hormone that mediates many actions of growth hormone and plays a key regulatory role in cell growth and proliferation (1). Tissue bioavailability of IGF is regulated in large part by its binding to six known IGF binding proteins. Insulin-like growth factor binding protein 3 (IGFBP-3) is the most abundant of these, and it substantially prolongs the circulating half-life of IGF-I. Most circulating IGF is produced by the liver, although it can be produced locally in other body tissues. IGF is known to play a role in brain development and function (8); it is produced in the brain and can pass through the blood-brain barrier. In animal studies, IGF improves memory and learning (8), and raising IGF levels was found to decrease formation of amyloid β (9), a major constituent of the neural plaques that are a hallmark of Alzheimer’s disease. In several recent cross-sectional and prospective studies, the association between circulating levels of IGF-I, IGFBP-3, or free (unbound) IGF-I have been assessed in relation to cognitive function. Although results have not been entirely consistent, it appears that older adults with higher levels of IGF, the IGF-1:IGFBP-3 ratio, or free IGF-1 tended to have better cognitive function as assessed by several cognitive tests (10–13).

Adiponectin and risk of diabetes and heart disease

In recent years, adipose tissue has been recognized as an active endocrine organ that secretes many biologically active substances, termed adipokines. Although adipokine research is relatively new, accruing laboratory and human data on adiponectin, one of the most abundant adipokines, support a role in diabetes and possibly heart disease (14–16). Adiponectin functions as an insulin-sensitizer, and also has important anti-inflammatory and anti-atherogenic actions. For example, adiponectin increases insulin sensitivity in animal models of insulin resistance, and reverses diet-induced insulin resistance in adiponectin knockout mice. The protein is inversely associated with body mass index (BMI) and insulin, and positively associated with serum lipids. Several cross-sectional, case–control, and prospective studies have reported either a strong significant inverse association for diabetes, or a modest inverse or no association for heart disease (17–21).

Predictors of hormone levels

In addition to the role of hormones in human health and disease, many studies have evaluated how the external environment influences the endogenous production of hormones in an effort to determine potential modes of action in causing (or preventing) disease. Examples include assessments of body size, physical activity and diet in relation to circulating hormone levels. The influence of alcohol intake on the hormonal milieu provides a good example. Both small randomized trials (22,23) and cross-sectional studies (24,25) have confirmed that alcohol intake increases estrogen levels in women, providing one potential mechanism for the positive association between alcohol intake and breast cancer risk. Further, alcohol intake has been found to increase insulin sensitivity (26) and HDL levels (27) and decrease fibrinogen levels (27), providing several mechanisms for the well-confirmed inverse association between alcohol intake and heart disease risk.

Strengths, limitations and lessons learned

Biologic samples for hormone evaluation

Many types of biologic specimens have been collected in epidemiologic studies where hormones are of interest (Table 12.2). In this section is a brief discussion of common sample types, and their advantages
<table>
<thead>
<tr>
<th>Sample Type</th>
<th>Collection, processing and storage requirements</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blood, venipuncture</td>
<td>+Trained phlebotomist +Needles, tubes, biohazard waste, tourniquet, gloves, storage vials +Primarily clinic-based +Spin and aliquot plasma or serum +Delay processing for 48 hours okay for some hormones +Mechanical or liquid nitrogen freezers</td>
<td>+Can measure broad range of hormones +Collect large volumes for future assays +For plasma tubes, can collect white blood cells for DNA and red blood cells +Can collect multiple types of samples (e.g. serum, plasma) simultaneously</td>
<td>+Less feasible for cost and logistic reasons for large studies, especially over a large geographic area +Invasive technique may lower participation rates +Requires laboratory and storage space +Biohazard potential</td>
</tr>
<tr>
<td>Blood, finger stick (filter paper)</td>
<td>+Finger lance, filter paper, instructions +Clinic or home-based +OK at room temperature for ~1 week for some hormones +Standard 4°C refrigerator or −20°C freezer</td>
<td>+Can measure some hormones +Feasible to collect on a large-scale +Allows serial sampling +Relatively non-invasive +Minimal storage space needed +Correlated with plasma levels for women; less clear for men +Not a biohazard after sample dries</td>
<td>+Low sample volume limits number of assays +Individual haematocrit levels can alter measured concentrations +Requires low humidity during transport and storage +Certain filter papers may inhibit some assays</td>
</tr>
<tr>
<td>Urine, cup</td>
<td>+Collection cup, antiseptic wipe, instructions, storage vials +For infants put a pad in the diaper +Clinic or home-based +May require addition of acid +Delay processing/ freezing for 48 hours okay for some hormones +Standard −20°C freezer</td>
<td>+Can measure broad range of hormone metabolites +Feasible to collect on a modest scale +Non-invasive and painless +Allows serial sampling +Generally correlated with plasma levels +Collect large volumes for future assays</td>
<td>+Assays often require larger volume than for blood +Must measure creatinine to determine urine concentration +Requires laboratory and storage space +Has biohazard potential +Can collect 24-hour, overnight, first-morning, or spot urine</td>
</tr>
<tr>
<td>Urine, filter paper</td>
<td>+Collection cup, antiseptic wipe, filter paper, instructions +Clinic or home-based +OK at room temperature for ~1 week for some hormones +Standard −20°C freezer</td>
<td>+Can measure some hormone metabolites +Feasible to collect on a large scale +Non-invasive and painless +Minimal storage space needed +Allows serial sampling +Not a biohazard after sample dries</td>
<td>+Has been used only in a limited number of studies + Low sample volume limits number of assays +Certain filter papers may inhibit some assays</td>
</tr>
<tr>
<td>Saliva, spot collection</td>
<td>+Several collection types: Unstimulated, salivette, cotton rolls, gum-stimulated +Need instructions, storage vials +Clinic or home-based +OK at room temperature for ~1 week feasible for some hormones +Standard −20°C freezer</td>
<td>+Can measure some hormones +Feasible to collect on a large-scale +Non-invasive and painless +Allows serial sampling +May reflect tissue exposure, as hormone must diffuse across salivary gland cells +Correlated with plasma levels for women; less clear for men</td>
<td>+Can only detect free (not bound) hormones, which often have a low concentration +Can be contaminated with blood, which has high concentrations of total hormone +May not reflect some circulating hormones in men +Use of stimulation or salivettes increases some hormone values</td>
</tr>
<tr>
<td>Saliva, oral diffusion sink</td>
<td>+Diffusion sink (hormones diffuse across it), instructions, storage vials +Collect continuously for long periods +Home-based +OK at room temperature for ~1 week feasible for some hormones +Standard −20°C freezer</td>
<td>+Measures integrated levels over time +May reflect tissue exposure of hormones under natural conditions +Separates bound and unbound hormones to reduce blood contamination +Participants find it acceptable</td>
<td>+Device cost can be high +Only available for a limited number of hormones +Not as feasible in a large, population-based study</td>
</tr>
</tbody>
</table>
and disadvantages, with respect to measuring the hormonal milieu.

**Blood**

Blood specimens, collected by venipuncture, are the most common and flexible sample type collected in epidemiologic studies. Levels of many hormones can be determined in blood, primarily because assay development traditionally has focused on this sample type. Various kinds of blood samples can be collected, including serum and plasma (e.g. EDTA, sodium heparin, citrate); each has advantages and disadvantages depending on the biomarker(s) of interest (28,29). For example, sex hormone levels generally are similar when comparing serum with EDTA or heparin plasma (30–33). However, most studies suggest slightly higher levels of sex hormones in plasma versus serum; despite this, both are acceptable (30,31,33).

The primary advantages of blood collected using venipuncture include the capability to measure many hormones, as well as the ability to collect multiple blood specimens simultaneously (e.g. plasma and serum) (34). Additionally, if collecting plasma, both red and white blood cells can be saved; the latter can be used to isolate DNA. Also, relatively large sample volumes can be collected at one time, allowing for many assays to be conducted per participant.

Despite these advantages, venipuncture is expensive, as it requires a trained phlebotomist and extensive equipment (e.g. needles, appropriate collection tubes, gloves, etc.) (34). This method has been used for years to test for uncommon, but treatable, genetic conditions in infants (38). Recently, extraction methods have improved such that many hormones can be assayed from blood spots, including thyroid hormones, prolactin, sex hormones, gonadotropins, growth factors, leptin and stress hormones (37,39–43). The primary advantages include that the method is relatively non-invasive, can be conducted serially, presents a reduced biohazard potential when dry, and is feasible to collect on a large-scale (40–43). One study reported that women found repeated blood spot sampling to be less troublesome than venipuncture or saliva collection (40).
However, there are several disadvantages of blood collection via finger prick. The most important is that only a limited number of hormones can be assayed from one collection. In addition, differences in haematocrit between participants can introduce systemic or random measurement error in hormone levels, since the sample is whole blood (37,41). Some evidence suggests, however, that the filter paper can partially ameliorate this problem, as blood with high haematocrit tends to impregnate a smaller volume on the filter paper (37). Further, some studies have reported that while correlations for sex hormones between venipuncture and filter paper are high for women, the correlations appear lower in men (40,41); the mechanism behind this is unclear. When measuring estradiol, testosterone and progesterone in men and premenopausal women, blood spot hormone levels explained 89% of the variance of serum levels in women, but only 46% in men (41). Other issues to note are that certain filter paper types may inhibit some assays (38), clear participant instructions are important for obtaining reliable samples (34), and the filter paper must be kept at a relatively low humidity (38).

**Urine**

Urine specimens are another commonly collected biological specimen, particularly because their collection is non-invasive (34). It also can be easily collected in infants and toddlers by putting a pad in the child’s diaper. In general, urine contains hormone metabolites, rather than primary hormones, and reflects excretion over the period of the collection. Therefore, it can be difficult to determine over what time period to collect the urine (e.g. 24 hours, overnight, first morning sample, or spot collection (34). The timing of urine collections depends on the hormone of interest. For example, the intraclass correlations (ICC) for a morning-spot versus 24-hour urine were 0.78 for estrone-3-glucuronide and 0.46 for pregnandiol-3-glucuronide; similar ICCs were observed for overnight versus 24-hour urines. This suggests that a morning spot urine was acceptable for the estrone metabolite, but neither the spot nor overnight urine appear to capture all the circadian variation for pregnandiol. Morning urines are acceptable for assessing nocturnal urinary melatonin production (44). Urine can be collected serially and in large volumes, allowing for multiple assays to be run. It can also be collected on filter paper to minimize storage needs and reduce the biohazard potential (34).

A primary disadvantage of urine is that its concentration varies substantially both between persons and within the same person over time (45). Creatinine is commonly used to measure urine concentration. Most studies have calculated the analyte/creatinine ratio to adjust for volume. However, it has been reported that creatinine levels should be included in the regression model as an independent variable. This adjusts for concentration while allowing one to assess the significance of other predictors in the model independently of creatinine levels (45). One other disadvantage of urine is that most hormones, or their metabolites, exist in low concentrations, often necessitating large volumes to conduct the assays.

**Saliva**

More recently, saliva has been used for measuring hormones. Generally, plasma and salivary hormone values were highly correlated, including for cortisol, androgens, estrogens and progesterones (46). These hormones enter the saliva by passive diffusion, thus the levels specifically reflect the free, unbound, circulating fraction (47,48). Melatonin levels also correlate well between plasma and saliva (49). However, for several protein hormones, such as thyroid hormones, prolactin, or IGF-1, the salivary level bears little relationship to plasma levels and is unlikely to be of any research value (47).

Advantages of saliva sampling include: it is non-invasive, painless, easily performed, relatively inexpensive, has higher rates of compliance, and can provide quantitative data of biologically active hormone levels in circulation (40,46). Salivary sampling also avoids stress sometimes associated with venipuncture, which can elevate some hormones, particularly cortisol (50). Another benefit is that saliva samples can be collected at home with minimal training (40,50); further serial collections are easily conducted.

Despite these advantages, a major disadvantage of saliva is that hormone concentrations are much lower than in plasma, because salivary hormone levels reflect the free levels, which is typically 1–10% of the total plasma level (40). Therefore, saliva assays often require large sample volumes and highly sensitive (and thus expensive) assays (50). Further, blood contamination in the oral mucosa can lead to substantial measurement error by increasing measured levels. Stimulated saliva collection methods, additionally, can bias hormone assays. For example, the use of cotton-based absorbent materials, or chewing gum, to stimulate saliva flow can artificially elevate assay results (40). Therefore, it is important to...
pilot sample collection techniques to ascertain whether such procedures interfere with the assays of interest.

One lesser known method of saliva sampling is a diffusion-sink device (51). This device is a small ring that the participant places orally for some set time period. The sink has a membrane that allows diffusion of free hormones, and thus measures the average freely diffusing concentration of an analyte over time. These devices reject artefacts arising from blood plasma contamination of saliva and provide a time-averaged sample, without requiring the subject to adhere to a frequent-sampling schedule (51).

Other sample types

Other specimen types, such as breast milk, breast nipple aspirate fluid and other tissues (e.g. tumour, adipose, colon polyps) can be obtained (34,52); however, they often are difficult to collect or require invasive procedures. The greatest disadvantage, though, is that it can be hard to conduct assays on these specimen types. Despite this, measuring hormones in these specimens may reflect true exposure at the tissue level better than from other sampling types.

Collection, processing and storage

As discussed in Chapter 3, the collection, processing, and storage of samples may affect the ability to accurately measure the hormonal milieu (36,53–56). Several factors for sample collection must be considered, including the study population, timing and location of the collection. Depending on these factors, the samples may need to be processed in a non-standard manner. A common issue in epidemiologic studies is that of delayed processing or delayed freezing. The effects of such protocols on the hormone of interest must be evaluated before assaying. Finally, storage of study specimens, particularly long-term storage in prospective studies, is an important and complex issue requiring appropriate acquisition of space and resources to maintain freezers and other related equipment.

Sample collection

Selection of the appropriate study population for any epidemiologic study is important. However, when studying hormone levels, careful consideration of the participants is often necessary to reduce bias. For example, in a study of predictors of estrogen levels, it would be inappropriate to combine men and women, premenopausal and postmenopausal women, or postmenopausal women taking hormones (PMH) versus not, as these groups have different mean estrogen levels. In this case, if gender, menopausal or PMH status was associated with the exposure, the observed association will be biased. Statistical adjustment alone generally cannot correct for this strong bias, particularly if the association varies across these subgroups.

The following example of the relationship between adiponectin and estradiol levels illustrates how combining inappropriate populations can alter study results. Experimental data suggest that adiponectin may, in part, regulate estradiol levels. But because adiponectin and estradiol are both derived primarily from adipose tissue in postmenopausal women, it is important to study this association on a population level, including adjustment for body mass index (BMI). One study observed no relationship after adjustment for BMI in postmenopausal women not using PMH (57). This contradicted two previous studies, which reported that additional adjustment for BMI did not attenuate the relationship (58,59). However, one study (59) combined PMH users and non-users, likely biasing the results, since PMH users had higher estrogen and lower adiponectin levels than non-users. The other study combined premenopausal and postmenopausal women (58). Given that premenopausal women have higher estradiol levels, and that the primary sources of estradiol are the ovaries in premenopausal women and body fat in postmenopausal women (1), the results of this study likely were biased as well.

Also important is the timing of the sample collection, since some hormones can fluctuate yearly, seasonally, monthly, daily, hourly and even from minute to minute (Figure 12.1). Other factors that can influence some hormones include fasting, alcohol intake, physical activity and medications. Understanding the underlying biology of the hormone(s) of interest is important to determining the optimal timing of sample collection. Three examples of this issue are elaborated upon.

Estrogen and progesterone are known to vary widely during the menstrual cycle in premenopausal women (1). Sample collection in this population, therefore, should either standardise the day(s) in the cycle on which samples are drawn, or collect detailed information about menstrual cycle start dates before and after the collection. In the NHSII, premenopausal women were asked to collect two blood samples, one in the early follicular phase and one in the mid-luteal phase, times when sex hormones are relatively stable from day to day (60). Women also returned a postcard with the
date their next menstrual cycle began. This information, along with the date of their previous cycle, allowed calculation of the cycle day on which the blood samples were drawn. One disadvantage of this method is that the investigator is reliant on women to remember when to collect the blood samples. Another method is to have women use home-based ovulation kits to time sample collection for a certain number of days after ovulation (61). The major disadvantages of this method are cost and the need to train women to use the kits appropriately. An alternative approach is to ask women to collect a sample on any day of their menstrual cycle and then provide specific dates of their cycle before (and if possible, after) collection and their average cycle length (62). This allows estimation of the cycle day. The main disadvantage of this method is that it reduces power to examine menstrual phase-specific associations. Ultimately, the collection method is dependent on the population and resources.

Several hormones have a circadian rhythm, with the most well characterized being melatonin. Levels are high at night while sleeping, and decrease during the daylight hours (44). In small, laboratory-based studies, the most common method of assessing melatonin has been to collect serial blood samples over a 24-hour period. However, this method is too labour-intensive for large epidemiologic studies. One alternative is to collect a blood sample at the same time of day for each participant. The utility
of this method is limited, though, since the circadian pattern is not entrained to the same time of day for everyone (63). Another option is to collect urine to assay melatonin metabolites. Studies indicate that urinary levels, from either a 24-hour or first morning urine, or sequential saliva samples, are highly correlated with plasma melatonin (44,64,65).

A third example is that of plasma, or serum insulin, which is strongly affected by the number of hours since last eating (1). At minimum, time of last food consumption should be collected at the blood draw. However, if these hormones are important biomarkers, investigators should instruct participants to not consume any food or drink for at least 8 to 10 hours before the blood draw. Study staff should carefully ask participants about their food intake during that time and reschedule the collection if necessary. Of note is that for some diseases an alternative hypothesis exists: that the postprandial insulin response is most relevant to disease risk (66). If so, collecting a blood sample soon after eating would be preferred.

**Sample processing**

Sample collection and its processing should be conducted in a rigorous and standardized manner. In general, certain methods are preferred (e.g. immediate processing of samples); however this may not be feasible in some studies, particularly when participants are dispersed geographically. While this topic is covered in detail in Chapter 3, two issues are highlighted which are often faced when assessing hormones: delayed processing and delayed freezing.

Extensive pilot testing has shown that many hormones, including estrogens, androgens, prolactin, IGFs and gonadotropins, are not substantially affected by blood remaining unprocessed for 24 to 48 hours, while others, such as adrenocorticotropic hormone (ACTH), arginine vasopressin and free PSA, cannot be assessed with this protocol (31–33,67–70). Interestingly, several studies reported increasing testosterone levels with delayed processing, likely due to ex vivo conversion of precursor hormones (32,68,70). However, levels across delayed processing times remained highly correlated, suggesting that this approach is acceptable. Thus, each hormone of interest should be pilot tested for stability of the analyte over increasing time of delayed processing.

A delay in freezing can occur if there is delayed processing or if samples are processed immediately and frozen at a later time. For blood specimens, most hormones are stable if kept chilled (−4 °C) for up to 3 days before freezing; however, others, such as free PSA and ACTH, should be processed and frozen immediately (31,33,67,69). Urinary catecholamine levels appear to be stable when stored for 24 hours at room temperature or chilled, provided that samples are acidified at once (71). Saliva hormones tend to be stable, and can be stored at room temperature for at least one week without degradation (47). However, these samples can mould after 4–7 days; thus they should be frozen or refrigerated if possible (34). In general, filter paper collections need to be kept refrigerated at a low humidity to maintain hormone stability (38).

**Storage options**

Freezing and refrigeration are the most commonly used storage modalities; the merits of these options are enumerated in Chapter 3. In general, liquid nitrogen freezers (≤ 130 °C) are the best choice for long-term storage of samples, since temperatures in mechanical freezers can vary widely (72). Unfortunately it is difficult to directly assess the effects of long-term storage on hormone degradation. Two study designs can be used. One method is to collect samples at one time point and then measure the hormone(s) of interest several times over a period of years. Thus, baseline biomarker levels are the same for each person, but laboratory drift can make comparison of assay results over time difficult, especially if the assay changes. Interpretability strongly depends on the reliability of the assay. The second method is to collect samples from the same individuals, or population, over a period of years, storing them at each time point. Then, assay the samples together at the end of the study, reducing issues with assay variability. However, within-person changes in levels over time means that it is unclear whether the levels at each time are the same. Despite this, degradation of samples stored for long periods is an important issue.

There is some evidence that storage at −20 °C may not be acceptable for sex hormones (73,74). In particular, sex hormone binding globulin (SHBG) may dissociate from estradiol and testosterone, decreasing measurable non-bound levels of these hormones (73). However, long-term storage for blood at −70 °C or colder appears to be acceptable for estradiol, testosterone, DHEAS, prolactin, IGFs, TGF-β1 and urinary 6-sulfatoxymelatonin, among others (30,74–79). If degradation is at issue, samples should be transferred to a colder storage modality. Given that modest levels of degradation
can be difficult to detect, another approach is to match the samples being compared (e.g. cases and controls) on storage time. This will reduce the effect of measurement error. Statistical modeling can also adjust for storage time or freezer temperature.

Overall issues surrounding sample collection, processing and storage are vital considerations when using hormone samples in epidemiologic studies. In particular, pilot studies that test sample collection and processing procedures are needed to determine feasibility and participant acceptability. Furthermore, if any non-standard protocols are used, it is important to test the effect of this on the hormones of interest before sending study samples for assay. In studies with long-term storage of samples, it is important to be aware of possible sample degradation over time and how that may affect the study design, analysis and interpretation.

**Laboratory measurement issues**

Three common sources of error are introduced when using biomarkers: issues related to specimen collection, processing, and storage (discussed in the previous section) (30,53,80); laboratory error and variability (36); and within-person variability over time (81). Since Chapter 8 discusses these latter two issues in detail, this chapter focuses specifically on hormone assays. In particular, there are often multiple methods for assaying a hormone, each with advantages and disadvantages. For example, one assay may require a large volume but have a higher sensitivity, while another uses a smaller volume but has lower sensitivity. Two examples are illustrated below.

The first example relates to measurement of sex hormones such as estradiol and testosterone. Three classes of assays are available to measure these hormones: mass spectrometry (MS); indirect radioimmunoassay (RIA), including a pre-extraction step; and direct immunoassays using chemiluminescent, colorimetric or fluorescent markers (82). When choosing which assay to use in a study, several factors should be taken into account, such as the amount of sample used, cost per sample, ease of assay, comparability with previous studies, and most importantly the assay reliability, validity and sensitivity in the hormone value range of the population under study (especially of concern if the values are low). Other issues are the abundance of structurally similar hormone metabolites that can cross-react with assay antibodies, and the binding of some hormones by SHBG, which can interfere with antibody binding (83,84). Differences in assays can be observed merely by noting the very different median levels of sex hormones measured in postmenopausal women across nine studies of breast cancer risk (7).

For sex hormones, the MS method obtains the highest marks for reliability, sensitivity and cross-reactivity (83). This method can measure multiple hormone metabolites simultaneously with a moderate amount of serum, plasma, or urine (~0.5mL). While MS is thought to be the gold standard for hormone measurement, recent analysis of inter-laboratory variation suggests that further standardization across laboratories is needed (83). This assay also is not widely available due to the expensive equipment and the need for highly-trained personnel to run the assays. The indirect assay methodology employs an extraction step before RIA to remove hormone metabolites that can cross-react with the antibody. This method has a high correlation with MS measures, although values tend to be slightly higher than those measured by MS, and uses a similar sample volume (85,86). The primary disadvantage of this method is that it cannot be easily automated and is thus labour-intensive and expensive (83). Direct assays, in general, do not have an extraction step before antibody binding. While these assays are high-throughput, easily available, inexpensive and use low volume, they may have only modest correlations with MS and indirect assays, and substantially overestimate hormone values, as well as a poor sensitivity for samples with low hormone concentrations (82–86). In general, these assays are not useful for clinical applications where precise levels must be determined, and are likely a major source of variability in epidemiologic study results (for a thorough review of this topic, see (83)).

The second example exemplifies the importance of understanding the biology of the hormone being measured, in this case prolactin. One limitation of the prolactin assay used in most epidemiologic studies to date (an immunoassay) is that it measures multiple forms of prolactin circulating in plasma (87). However, these forms likely have different biological activities (88,89). For example, glycosylated prolactin appears to have a higher metabolic clearance rate and lower biologic activity than the nonglycosylated form (89,90). Assays to specifically measure particular prolactin isoforms are difficult, time-intensive and require large amounts of plasma, and hence are not feasible in epidemiologic studies. The Nb2 lymphoma cell bioassay, however, is a sensitive measure of
overall somatolactogenic activity in biological fluids. This assay measures the activity of both prolactin and growth hormone (91), although a modification of the assay, including anti-growth hormone antibodies, allows for specific evaluation of prolactin bioactivity. This measure and the ratio between the prolactin bioassay and immunoassay have been evaluated in several studies of systemic lupus erythematosus and found to be of importance (90,92,93). A breast cancer case–control study reported that prolactin levels measured by bioassay, but not immunoassay, were significantly higher in cases versus controls (94). The correlation between the immunoassay and bioassay is about 0.50, suggesting that the bioassay provides additional information beyond the immunoassay (94).

Another important issue arises when pooling data from multiple different studies, such as in the sex hormone-breast cancer study mentioned above (7). Since each study conducts assays at various laboratories and times, frequently the distribution of the analyte (e.g. hormone levels) differs across studies (95). Various analytic techniques are available to deal with this problem (96). One method is to use study-specific quantile cutpoints to determine risk estimates comparing high versus low values for each study, which can then be pooled. The major drawbacks of this method are that it is difficult to evaluate what absolute hormone levels are related to disease, and to assess dose–response relationships. Another method used is to pool the risk estimates for a doubling (or tripling, etc.) of sex hormone concentrations within each study and then pool the risk estimates. However, the only way to assess how the absolute levels of a hormone compare across studies is to reassess a subset of samples from each laboratory used in each of the different studies (i.e. a calibration study) using a gold-standard assay.

Another approach to evaluating hormone levels is to assess genetic variation (e.g. single nucleotide polymorphisms (SNPs)), in the gene(s) associated with the hormone. Although estimates vary by hormone, a substantial component of the between-person variation observed in circulating hormone levels is genetically determined (e.g. ~40% for IGF-I (97)). For example, to assess the role of the steroid vitamin D in osteoporosis, several studies have evaluated SNPs in the vitamin D receptor (98). Furthermore, many studies have assessed variation in the sex hormone metabolizing pathway in relation to cancer risk (99–101). Advantages to this approach include: retrospective case-control designs can be used without concern of the disease altering circulating hormone levels, genetic variation may provide information on the tissue hormonal environment, and genetic assays tend to be robust and with little to no variation in measures across studies. Disadvantages include not knowing the function (if any) of the SNPs measured, and small effect sizes. Ultimately, clearer answers likely will be obtained with approaches that evaluate multiple SNPs in a gene or across multiple genes in a pathway, but this requires extremely large sample sizes and complex statistical tools that are still in development. Although promising, this approach to evaluate the hormonal milieu has yet to provide substantial insight into the hormone-disease relationship (102). Yet with increasing sample sizes and method development, the potential of this approach should be realized soon.

**Within-person stability over time**

A particularly important source of measurement error in hormone studies is random variation in biomarker levels within an individual over time. Thus one measurement of the biomarker, as is common in many epidemiologic studies, may not accurately reflect an individual’s long-term exposure. Measurement error correction, or inclusion of multiple samples per participant, can ameliorate the attenuating effects of biomarkers with a high intra-individual variability over time (103–105).

The intraclass correlation coefficient (ICC) can be used to measure the stability or reliability within individuals over time or across different assay platforms (103,106-108). It is the ratio of the between-person variance with the total variance (between-plus within-person variance), and ranges from 0 to 1.0 (80,109). The ICC is distinct from a Pearson or Spearman correlation coefficient in that a common mean is assumed between repeated measures. The ICC can be assessed on the natural log-transformed or untransformed scale, although if the data are skewed it is best to log-transform. An advantage of the ICC is that the impact of the within-assay variability is considered relative to the total variation. For example, a somewhat high assay coefficient of variation (CV) may not be acceptable if there is very limited between-person variation (resulting in a low ICC), as the additional laboratory variability could overshadow true differences between individuals.

Overall, most sex steroid hormones are reasonably stable within postmenopausal women over a 1–3 year period, with intraclass correlations ranging from 0.5 to 0.9
In a setting such as this, where measurement error is higher (since the correlation of hormone levels within woman over time is lower), statistical methods that account for this error in the calculation of relative risks should be used (see Chapter 8), or collecting multiple samples per subject considered.

Future directions and challenges

Overall, the use of hormone measurements is becoming more common in epidemiologic studies. Several factors must be considered in the study design, sample collection, assay choice, and statistical analysis. Of greatest importance is a suitable choice of study population and sample type(s) (collected at an appropriate time), as well as proper storage facilities. Any non-standard methods should be pilot-tested before conducting the formal study. Pilot studies should also be conducted when considering the use of a new assay or laboratory. Choice of assay type can have a large impact on measurement error, and, ultimately, the interpretation of results. Although not discussed in this chapter, assessment of laboratory precision and reproducibility on an on-going basis is extremely important (36). Additionally, studies assessing hormone stability within an individual over time are important to conduct, particularly if the hormone is the exposure of interest. It is important to ultimately address all these issues to obtain results that are both reliable and valid. A better understanding of the role of hormones in human disease will benefit immensely from well-conducted epidemiologic studies.

References


for biospecimens: defining the sample et al.-10-0058 PMID:20233852


Collection, processing, and storage of biological samples in epidemiologic studies: sex hormones, carotenoids, inflammatory markers, and proteomics as examples. Cancer Epidemiol Biomarkers Prev, 15:1578–1581. doi:10.1158/1055-9965.EPI-06-0629 PMID:16985015


Relationship of plasma adiponectin with sex hormone and insulin-like growth factor levels. Obesity (Silver Spring), 15:2217–2224. doi:10.1038/oby.2007.263 PMID:17890489

Serum adiponectin levels are inversely associated with overall and central fat distribution but are not directly regulated by acute fasting or leptin administration in humans: cross-sectional and intervention studies. J Clin Endocrinol Metab, 88:4823–4831. doi:10.1210/jc.2003-030214 PMID:14557461


Two urinary catecholamine measurement indices for applied stress research: effects of time and temperature until freezing. Hum Factors, 45:563–574. doi:10.1518/hfes.45.4.5278 PMID:15055454


Two urinary catecholamine measurement indices for applied stress research: effects of time and temperature until freezing. Hum Factors, 45:563–574. doi:10.1518/hfes.45.4.5278 PMID:15055454


