

Editorial

David G. Grenache*

Variable accuracy of home pregnancy tests: truth in advertising?

DOI 10.1515/cclm-2014-1033

Pregnancy testing dates back to the 14th century BC when Egyptians described a process for identifying a woman who might be pregnant: she could urinate on seeds of wheat and barley over several days. If the barley grew she was pregnant with a male child while the growth of the wheat indicated a female gestation. Growth of neither was interpreted as the absence of pregnancy [1]. Over the last 33 centuries, many methods of identifying the pregnant patient were considered but it was not until the early 20th century AD that assays for detecting and/or measuring human chorionic gonadotropin (hCG) were developed [2].

Presently there are numerous brands of home pregnancy tests (HPTs) available to consumers worldwide with more than 60 brands available in the US alone [2]. These devices are intended to be used with a random urine sample for the qualitative detection of hCG as an indicator of possible pregnancy. Unfortunately, all devices are not created equal and their analytical and clinical performance vary considerably [3, 4]. The analytical sensitivity of any qualitative hCG test is a principle determinant of the device's ability to detect pregnancy. The more analytically sensitive the device, the more likely it would produce a positive result in early pregnancy.

Interestingly, there is lack of consensus regarding the optimal analytical sensitivity of qualitative hCG devices. Some have argued that the detection limit should be no less than 25 IU/L to avoid detecting 'biochemical pregnancies' (i.e., very early miscarriage) or the confusion that results from the detection of hCG that may occur post-menopause [5]. Others contend that a detection limit of 5 IU/L is needed to reduce the potential for false-negative results in early pregnancy due to dilute urine samples [6].

The claimed analytical sensitivities of qualitative urine hCG tests ranges from 6.3 to 50 IU/L with most falling between 20 and 25 IU/L [2]. However, claims are not always consistent with actual performance. Several studies have documented that devices are either more sensitive [4] or less sensitive [7] than claimed. Reasons for

such differences include the molecular heterogeneity of hCG and its variants, the antibody pairs used by device manufacturers, and, perhaps most importantly, the quantitative method used to assign an hCG value to a testing material.

As highlighted by Johnson et al. in this issue of the journal, there is a need for internationally standardized methods for expressing the performance of HPTs [8]. Many countries have no national standards that address this issue and among those that do, such as the US, such guidance may be outdated and inadequate [9]. In light of this, Johnson and colleagues evaluated the analytical sensitivity relative to the claims of eight HPTs manufactured by seven different companies located in six countries. The claimed detection limits of the devices ranged from 12.5 to 25 IU/L.

The devices were evaluated with the use of three urine standards containing hCG derived from pregnancy urine at concentrations of 0, 15, and 25 IU/L as determined by the AutoDELFI system calibrated against the 4th international standard for hCG. This system only recognizes dimeric hCG variants (intact hCG and nicked hCG). While these variants were present in the urine standards tested, so were variants that are not detected by the assay such as the free β subunit, the nicked free β subunit, and the β core fragment of hCG. Each device was tested and interpreted 24 times with the three standards by technicians. Additionally, 72 female volunteers representing typical users of HPTs also interpreted each test result within the time specified by the device's instructions for use.

All devices performed best when interpreted by a technician rather than a volunteer. The technician identified no false-positive results when the 0 IU/L standard was tested while volunteers called 1% of these test results positive. Much more alarmingly, when the standards that contained hCG were tested and the results evaluated against the claimed detection limits, 40.5% and 63.6% of results were interpreted as negative by the technician and volunteers, respectively! Notably, the distribution of these false-negative results was heavily weighted towards specific device brands. When interpreted by both the technician

and volunteers, the same two brands produced <50% agreement between test results and claims. Conversely, three brands produced >95% agreement between test results and claims when interpreted by a technician with two of these brands achieving the same performance when interpreted by volunteers.

The study design also called for a repeat interpretation of each device evaluation by the technician immediately after the volunteer interpretation of the same device. When evaluated by standards at or above the claimed detection limit, some devices that were originally interpreted as negative were interpreted as positive suggesting that the recommended read times provided for some devices may be inadequate.

Following their interpretations of a device brand, the volunteers, who were blinded to the expected results, completed a survey in which they scored the brand on their trust and satisfaction of the test, how clear the results were to interpret, and their assessment of the test's accuracy. Two devices were highly rated in all categories by >50% of the volunteers. The device with >90% of volunteers providing high ratings in all categories is formatted to display the words 'pregnant' or 'not pregnant' in its interpretation window rather than rely on the appearance of a colored line or symbol to indicate a positive result, a feature that was clearly preferred by the volunteers. Of great concern was the other device that received high ratings from 60% to 70% of the volunteers. When tested with the 15 IU/L standard, this device, with a claimed detection limit of 10 IU/L, was interpreted as negative in 96% of evaluations. When challenged with the 25 IU/L standard, the false-negative rate marginally decreased to 86%. The availability of such poorly performing HPTs that appeal to consumers clearly highlights the need for guidelines and/or regulations for manufacturers that produce and market these tests.

While any erroneous hCG test result creates the potential for harm to the mother or fetus, false-negative results are of particular concern. When they occur in HPTs, a pregnant individual may not modify adverse social behaviors or may take medications that could harm the fetus. Similarly, potentially dangerous medical interventions may cause harm if false-negative hCG test results occur in devices used at the point-of-care in healthcare settings. Several qualitative point-of-care hCG devices have been shown to be susceptible to false-negative results due to an analytical sensitivity that is insufficient for detecting early pregnancy [7] or susceptibility to inhibition by the β core fragment of hCG [10].

An important limitation of the study was the quantitative hCG method used to determine the hCG concentrations of the urine standards. As the method used is not able

to detect non-dimeric hCG variants, it is very likely that the actual hCG concentration of the standards was higher than stated. The impact that this variable has on the data reported is impossible to know. One could hypothesize that there would be more disagreements between device performance and claims, yet without knowing the analytical specificity of the HPTs evaluated and the method each manufacturer used to set the performance claims of their device, this is merely speculative.

It is hoped that the work of Johnson et al. [8] gains the attention of HPT manufacturers. Consumers of these devices often lack an appreciation of the strengths and limitations of qualitative hCG screening tests and will rely heavily on marketing claims to guide their purchasing decisions. Perhaps more importantly, manufacturer claims and product appearance are also likely to influence how an individual perceives the reliability and accuracy of the test. HPTs that fail to meet their performance claims yet are regarded as trustworthy by consumers present a considerable risk to patient safety. Clear and consistent regulatory requirements are needed to guide HPT manufacturers and ensure consumer confidence and safety.

Author contributions: All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission.

Financial support: None declared.

Employment or leadership: None declared.

Honorarium: None declared.

Competing interests: The funding organization(s) played no role in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the report for publication.

References

1. National Institutes of Health. A thin blue line. Available from: <http://history.nih.gov/exhibits/thinblueline/index.html>. Accessed 19 October, 2014.
2. Braunstein GD. The long gestation of the modern home pregnancy test. *Clin Chem* 2014;60:18–21.
3. Tomlinson C, Marshall J, Ellis JE. Comparison of accuracy and certainty of results of six home pregnancy tests available over-the-counter. *Curr Med Res Opin* 2008;24:1645–9.
4. Cervinski MA, Lockwood CM, Ferguson AM, Odem RR, Stenman UH, Alfthan H, et al. Qualitative point-of-care and over-the-counter urine hCG devices differentially detect the hCG variants of early pregnancy. *Clin Chim Acta* 2009;406:81–5.
5. Stenman U-H, Alfthan H. Optimal sensitivity for pregnancy tests. *IVD Technol* 2003;9:18–9.

6. Terwijn M, van Schie A, Blankenstein MA, Heijboer AC. Pregnancy detection by quantitative urine hCG analysis: the need for a lower cut-off. *Clin Chim Acta* 2013;424C:174.
7. Greene DN, Schmidt RL, Kamer SM, Grenache DG, Hoke C, Lorey TS. Limitations in qualitative point of care hCG tests for detecting early pregnancy. *Clin Chim Acta* 2012;415:317–21.
8. Johnson S, Cushion M, Bond S, Godbert S, Pike J. Comparison of analytical sensitivity and women's interpretation of home pregnancy tests. *Clin Chem Lab Med* 2015;53:391–402.
9. FDA. Guidance for over-the-counter (OTC) human chorionic gonadotropin (hCG) 510(k)s. FDA Guidance Document. 2000;1–12.
10. Nerenz RD, Song H, Gronowski AM. Screening method to evaluate point-of-care human chorionic gonadotropin (hCG) devices for susceptibility to the hook effect by hCG β core fragment: evaluation of 11 devices. *Clin Chem* 2014;60:667–74.

***Corresponding author: David G. Grenache**, Department of Pathology, University of Utah School of Medicine, 15 North Medical Drive, Salt Lake City, UT, USA, Phone: +1 801 583 2787, E-mail: david.grenache@path.utah.edu