

Appraising Endocrine Pulse Signals at Low Circulating Hormone Concentrations: Use of Regional Coefficients of Variation in the Experimental Series To Analyze Pulsatile Luteinizing Hormone Release¹

JOHANNES D. VELDUIS, JEFFREY WEISS, NELLY MAURAS, ALAN D. ROGOL,
WILLIAM S. EVANS, AND MICHAEL L. JOHNSON

Departments of Internal Medicine [J.M. W.S.E.], Physiology [J.W.], Pediatrics [N.M., A.D.R.], and Pharmacology [A.D.R., N.L.J.], University of Virginia School of Medicine, Charlottesville, Virginia 22908

ABSTRACT. We have developed a regional dual-threshold method for detecting endocrine pulses at variably low hormone concentrations such as those that occur in the pediatric age group. This algorithm uses the local intra-series coefficient of variation to test for a significant increase and decrease (peak or pulse) in the experimental series. The method is unique in combining the following considerations: 1) use of internal measurement error present in the actual sample replicates (rather than that estimated from external standards); 2) an ability to accommodate a wide range of nonuniformity in intraassay coefficients of variation; 3) utilization of both upstroke and downstroke threshold criteria; and 4) adaptability to specify a physiologically relevant proximity of the postpeak nadir to the center of the peak. We have validated the behavior of this method in relation to false-positive rates for various program parameters and have illustrated the application of this new regional dual-threshold method to luteinizing hormone pulse detection in pubertal boys who have relative hypogonadotropism, which renders reliable luteinizing hormone pulse detection particularly difficult. (*Pediatr Res* 20: 632-637, 1986)

Abbreviation

LH, luteinizing hormone

The accurate detection of endocrine pulse signals represents an important objective of physiological investigations, since the nature of the hormonal signal is a critical determinant of target organ function. The visual identification of presumptive endocrine pulses fails to provide an adequate objective and communicable technique for reproducible pulse detection. Thus, several objective computerized methods have been put forward to iden-

tify distinct but episodic hormone fluctuations objectively (1). These methods exhibit individual strengths and weaknesses, which render them more or less suitable for specific analytical tasks.

A special problem in endocrine pulse detection arises when hormone concentrations approximate the sensitivity of the assay. At hormone concentrations that approach the lower limit of detectability, greater degrees of measurement error are introduced into the assay estimates. This is particularly likely to occur in pathophysiological states associated with reduced circulating hormone concentrations, or in the pediatric age group prior to pubertal activation of endocrine axes. Unfortunately, available pulse detection methods do not make explicit use of actual measurement errors inherent within the experimental series, and therefore do not account for either variable or high degrees of assay imprecision that may be present under these conditions.

In an effort to approach the problem of accurate pulse identification in clinical conditions associated with a wide range of hormone measurement errors, we have developed and validated a simple modification of the threshold method of Santen and Bardin (2). Rather than using the original fixed threshold value of 20% to identify significant hormone increases (peaks), we have chosen to relate the threshold to the intraassay coefficient of variation present in the replicates within each region of the data series being analyzed for pulses. This use of a "regional" coefficient of variation permits an analysis of hormonal data in which coefficients of variation may be nonuniform either in relation to concentrations of hormone measured or portions of the assay (e.g. relative position or order of the sample in the assay). In addition, we have required that a peak be characterized by a threshold degree of both increase and decrease; hence, the term dual-threshold.

In the present report, we have validated the performance of this regional, dual-threshold method for assay-related false-positive error rates (for various pertinent combinations of threshold criteria) and illustrated its applicability to studies of pulsatile LH release in pubertal boys.

METHODS

The original program of Santen and Bardin (2) defines a pulse as an increase in hormone concentrations of at least 20% (approximately 3-fold their average intraassay coefficient of variation) above preceding nadir, when this increase is followed immediately (in the next point) by any degree of decrease (2). We modified this program in the following manner. 1) The

Received November 12, 1985; accepted February 19, 1986.

Correspondence Dr. Johannes D. Veldhuis, P.O. Box 202, University of Virginia School of Medicine, Charlottesville, VA 22908.

This work was supported in part by NICHD RCDA 1 K04 HD00634, NIDA Grant DA 03315, University of Virginia Computer Services Grant (JDV), and NIH Biomedical Research Support Award S 07 RR05431 (ADR, JDV and WSE); by NIH Grants AM22125, AM30302, GM28928 (MLJ), and HD 00439 (WSE); by USPHS General Clinical Research Grant RR-847; by a Diabetes Research and Training Grant 5P60 AM 22125-05; and by a DuPont Fellowship (JW).

¹ Copies of this regional dual-threshold program will be made available to qualified investigators upon request.

program uses replicate (*e.g.* triplicate) values associated with each sample in the experimental series to calculate individual coefficients of variation for all data points. 2) The operator specifies a threshold for the upstroke (degree of ascent) of the putative peak, which is expressed as a designated multiple of the "regional coefficient of variation" (defined further below); 3) The investigator defines an additional threshold for the downstroke, which is expressed as a specified multiple of the "regional" coefficient of variation; 4) The downstroke is calculated from the putative peak value, and must be observed within a designated number of points (period of time) following that peak value; and 5) The investigator denotes the regional width or "window" within which the mean coefficient of variation is calculated from the corresponding replicates (*vide infra*). Thus, in utilizing the revised program, the investigator must specify: a particular threshold multiple of the regional coefficient of variation to be used to search for a significant upstroke; a particular threshold multiple to be used for a significant downstroke; the number of points after the peak to be searched to identify this significant downstroke; and the regional "window" wherein a local mean coefficient of variation is calculated. The upstroke and downstroke thresholds are expressed as corresponding selected multiples of the regional coefficient of variation. At the beginning and end of the series, either a significant increase or a significant decrease is accepted as a peak since both upstroke and downstroke may not be complete at the "edges" of the data series.

The performance of the revised pulse detection method was evaluated in relation to false-positive error rates for a range of intraseries coefficients of thresholds and a range of pulse detection parameters. The false-positive rates for various program parameters were defined as the numbers of "pulses" detected in 990 simulated LH values (each represented in triplicate). In these simulated series, the Z score distribution of measurement error within triplicates conformed to a Gaussian pattern, while the mean intrareplicate coefficients of variation were constrained to 4, 8, 12, 20, or 32%, as described earlier (3). Note that an intrareplicate coefficient of variation of X% corresponds to an intraseries coefficient of variation of X/\sqrt{n} %, where n is the number of replicates contributing to each sample mean. These series were created to mimic accurately the performance of the LH radioimmunoassay, which had been evaluated using 565 sets of triplicates derived from normal male serum pools (3). False-positive pulses were defined as "pulses" detected in these randomly varying data series.

To test for program parameters that resulted in stable and low false-positive error rates, the following program characteristics were each assessed for 990 simulated data points. 1) The threshold for a significant upstroke was defined as one of several possible multiples of the regional coefficient of variation. The particular multiples tested included 2, 3, 4, 6, and 8. 2) The threshold multiple for a significant downstroke, expressed in relation to the regional coefficient of variation, was tested for multiples of 1, 2, 3, or 4. 3) The number of points searched after the peak to locate a significant downstroke was tested at values of 1, 5, or 13. 4) The width of the "window," from which the regional mean coefficient of variation was calculated, was taken as $N = 1, 3, 10, 30, \text{ or } 100$. For a region of width N , the number of points used to calculate an average regional coefficient of variation was $2N + 1$ (N points prior to the possible peak, N points following the possible peak, and the peak itself). Any individual coefficient of variation of $\leq 1\%$ was set to 1% in order to eliminate spurious effects of identical assay replicates with a "zero" coefficient of variation. 5) Each of the above conditions was tested in relation to intrareplicate coefficients of variation of 4, 8, 12, 20, or 32%.

Overall, therefore, each of 1500 possible combinations of pulse-detection criteria was tested against 990 simulated data points, each of which was represented in triplicate.

The physiological applicability of the revised program was assessed on LH values derived from serum collected at 20- and

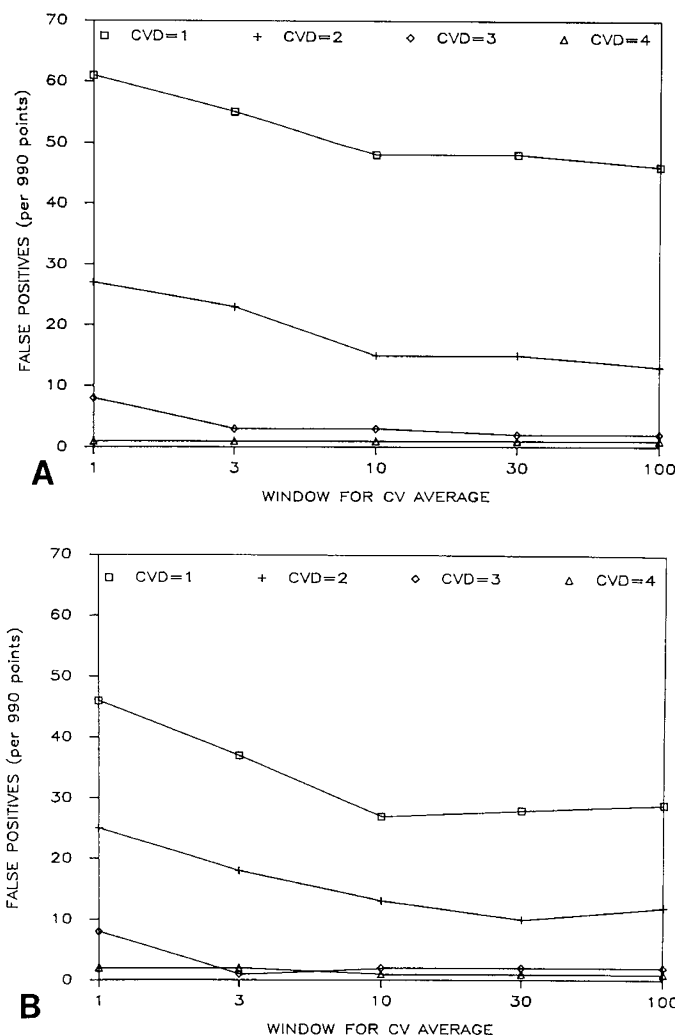


Fig. 1. Influence of the size of the "window" for the calculated regional coefficient of variation on false-positive rate. Simulated LH series were created with a Gaussian distribution of measurement error having with an intrareplicate mean coefficient of variation of 20% (A) or 8% (B). The regional dual-threshold program was set to require an upstroke minimum of three times the regional coefficient of variation, which was calculated for the indicated window (N), varying as 1, 3, 10, 30, or 100 points before and after the putative peak. The families of curves correspond to downstroke thresholds (CVD) of 1, 2, 3, or 4-fold the regional coefficient of variation. The false-positive rate is expressed as the number of spurious "pulses" detected per 990 mean values in the simulated series.

5-min intervals for 24 h in four pubertal boys (Tanner stage III-IV) and eight normal men, respectively. The LH measurements in these sera were performed by radioimmunoassay as described previously (4). Each sample was assayed in triplicate, and each subject's samples were assayed within the same run.

RESULTS

As shown in Figure 1, the width of the "window" for the regional coefficient of variation estimate significantly influenced the false-positive rate. In Figure 1, the width of the regional coefficient of variation window was made to vary from 1 to 100 points. The family of four curves represents false-positive rates for downstroke thresholds of 1, 2, 3, or 4 times the regional coefficients of variation, when the upstroke threshold was held constant at a multiple of $3 \times$ the regional coefficient of variation. In A, the simulated series exhibited a mean intrareplicate coef-

efficient of variation of 20%, and in *B*, 8%. In both cases, the program was permitted to search one point after the putative peak to identify a significant downstroke. Notably, the false-positive rate declined as the width of the window for the regional coefficient of variation increased. In addition, the false-positive rate declined as the downstroke threshold was made more restrictive. Moreover, the false-positive rate was lower when the intrareplicate coefficient of variation was 8% compared to 20%.

As the upstroke criterion was increased from a multiple of 3 to a multiple of 4, the false-positive rate decreased, and then declined further as the regional window width was increased. The false-positive rate was relatively stable for window widths of ≥ 10 –30 points (not shown).

A higher intrareplicate coefficient of variation of 20% compared to 8% in the simulated data series resulted in a higher false-positive rate for each window width and for each downstroke threshold criterion, as anticipated. Thus, the false-positive

rate can be made to decrease by expanding the regional window from which the mean coefficient of variation is calculated, and/or by increasing the upstroke and downstroke threshold criteria for the definition of a significant pulse.

The false-positive rate of this revised threshold method was also tested in relation to a range of coefficients of variation in the simulated data series. These coefficients of variation were made to vary from 4 to 32%, as shown in Figure 2: for an upstroke threshold multiple of 4 (*A*) or an upstroke threshold multiple of 3 (*B*). When the window for the regional coefficient of variation was held constant at 10, and 5 points after the peak were searched for a significant downstroke, we observed a significant impact (2- to 3-fold effect) on the false-positive rate of varying intraseries coefficients of variation from 4–32%. The adverse impact of increasing intraseries coefficients of variation was considerably less when upstroke and downstroke multiples were both in the range of 3 to 4, which yielded false-positive rates of less than 1% (fewer than 10 false-positive “pulses” per 990 points examined). On the other hand, less restrictive criteria resulted in significantly increased false-positive rates, which reached levels as high as 67 false-positive pulses per 990 points (approximately 7%). Accordingly, these figures provide the investigator with parameter choices that will maintain minimum estimated false-positive rates in selected ranges.

The false-positive rate was also critically influenced by the number of points searched after a peak. When the postpeak nadir was required to occur within 1, 5, or 13 points after the presumptive peak, the false-positive rate increased as the number of points tested after a possible peak increased (not shown). This increase in false-positive rate was attenuated by increasing the upstroke threshold for a fixed downstroke criterion, regional window width, and intrareplicate coefficient of variation.

In Table 1, we have summarized various applicable conditions which yield a false-positive rate of $\leq 1\%$ for intrareplicate coefficients of variation that span the range 4 to 32%. These program parameters will yield false-positive rates of $\leq 1\%$, despite coefficients of variation that vary from 4–32%. If desired, less stringent conditions can be inferred from Figures 1 and 2.

Application of the revised threshold method to LH pulsations in normal men. Based on the foregoing validation of false-positive rates, we chose the following program parameters: an upstroke threshold multiple of 4 times the regional coefficient of variation; a downstroke threshold multiple of 3 times the regional coefficient of variation (assuming some asymmetry in peak shape, with less rapid downstroke than upstroke); the requirement that a decrease after a putative peak be located within 10 points of

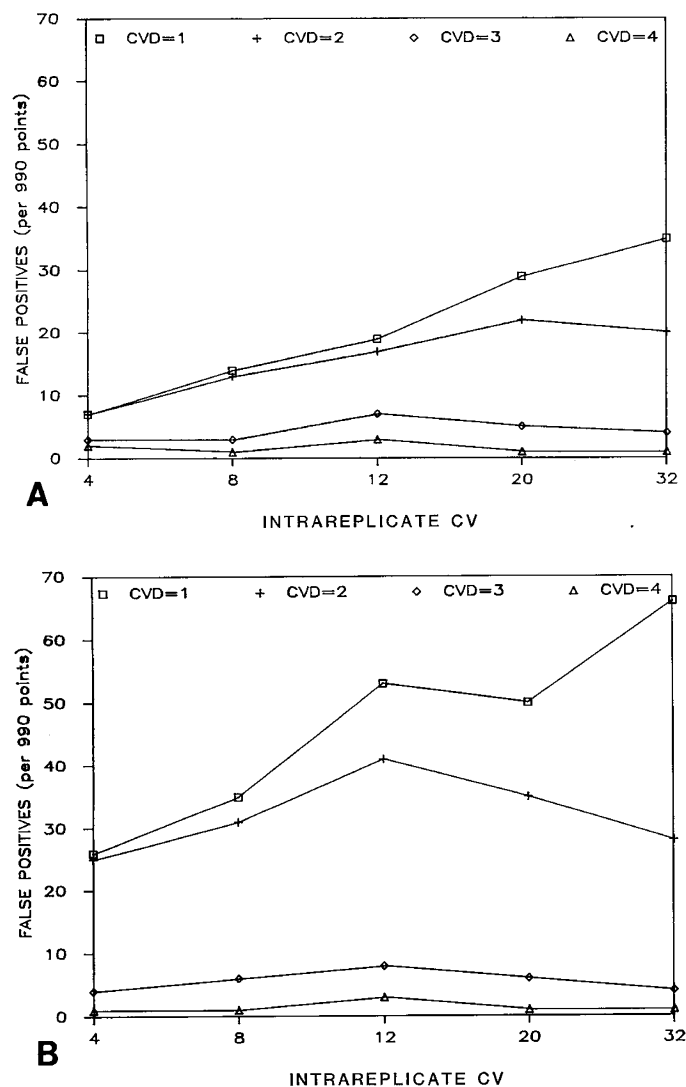


Fig. 2. Influence of a range of intrareplicate coefficients of variation on the false-positive rate for the regional dual-threshold method of pulse detection. Simulated LH series were generated as described in “Methods,” with various intrareplicate coefficients of variation (CV) that ranged from 4–32% (horizontal axis). The upstroke thresholds were defined as increases of 4-fold (*A*) or 3-fold (*B*) the corresponding regional CV. The families of curves indicate the range of downstroke criteria imposed, which varied as 1, 2, 3, or 4-fold the regional CV for a fixed window width of $N = 10$ and postpeak nadir proximity of 5. The false-positive rate is defined as given in the legend of Figure 1.

Table 1. Summary of program parameters yielding $< 1\%$ false-positive rates for intraseries coefficients of variations ranging from 4–32%

Threshold for upstroke of peak (as regional CV multiple)	Threshold for downstroke of peak (as regional CV multiple)	No. of points searched after peak*	Window for Regional CV estimate†
3	3	1	1–100
3	3	5	10–100
4	2	1	10–100
4	3	1	1–100
4	3	5	3–100
4	3	13	10–100
4	4	1	1–100
4	4	5	1–100
4	4	13	1–100

* No. of points following the putative peak that are searched for a significant downstroke.

† No. of points preceding and following the test peak from which a regional coefficient of variation is derived (see “Methods”).

that peak (since this number of points yields relatively stable false-positive rates); and a regional window width of $N = 10$ for the calculation of a local coefficient of variation against which the threshold multiples are applied. These conditions were selected because they yield a false-positive rate of 0.25–0.75% across a wide range of intraseries coefficients of variation that span 4% at the lower extreme and 32% at the upper extreme. Using these program parameters, the number of LH pulses detected in eight normal men sampled at 5-min intervals for 24 h averaged 14.8 ± 1.5 . These estimates contrast with a mean LH pulse frequency of 28.1 ± 2.7 per 24 h for the original method of Santen and Bardin (2) using a fixed threshold criterion of 20% ($p < 0.004$ versus the revised regional threshold method by Mann-Whitney test).

The foregoing observations suggest that when the original method of Santen and Bardin (2) using a fixed 20% threshold is applied to regions of the experimental series exhibiting relatively increased coefficients of variation (whether or not associated with a lower dose of the hormone), higher false-positive error rates would result. This prediction was confirmed by the high rate of erroneous identification of “pulses” within simulated series exhibiting a range (4–36%) of intraassay (intrareplicate) coefficients of variation (Fig. 3). In contrast, the regional threshold method maintained a stable false-positive rate of approximately 0.4% for the same data series.

Differences between the regional threshold and fixed threshold method: analyses of pulsatile LH secretion for a mock series and in pubertal boys. As illustrated by the mock series in Figure 4, the fixed threshold method is largely insensitive to varying degrees of assay imprecision. In this illustrative series, all the

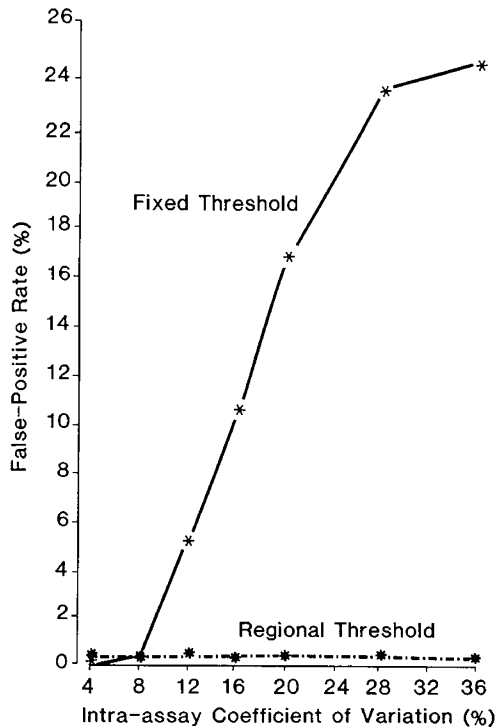


Fig. 3. Comparison of the regional threshold algorithm with the original fixed threshold method of Santen and Bardin (2) over a wide range of intraassay coefficients of variation. Simulated LH series were constructed with intratriPLICATE coefficients of variation ranging from 4–36%. The regional threshold procedure used the following parameters: upstroke and downstroke thresholds of respectively 4- and 3-fold the regional intrareplicate coefficients of variation; a regional coefficient of variation window width of $N = 10$; and a proximity of ≤ 10 for the peak and postpeak nadir. The values shown represent the mean false-positive rates per 100 samples as estimated from a series of 900 values each generated in triplicate.

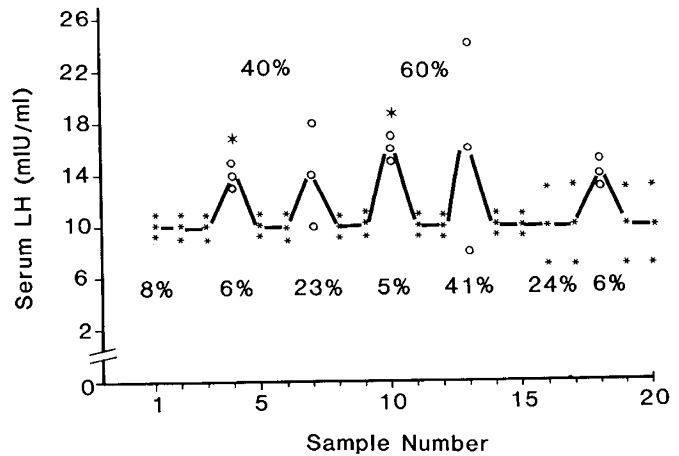


Fig. 4. Illustration of the performance of the dual-threshold regional coefficient of variation method, compared to the original fixed threshold method for pulse detection. Simulated serum LH concentrations (vertical axis) are shown in triplicate in relation to sample number (horizontal axis). The coefficients of variation within the triplicates are indicated by the values below the simulated LH pulse tracing. Above the pulse tracing, the values of 40 and 60% refer to the mean amplitude of the “peaks” above preceding nadirs. The fixed threshold method detected all mean increases (either 40 or 60%) as significant peaks. In contrast, the regional threshold method only detected the two peaks indicated by the asterisks, in which both the peak and nadir values reflected a higher degree of assay precision.

“peaks” were defined as significant using the original fixed threshold method of Santen and Bardin (2). Notably, although mean LH values did increase by $\geq 20\%$ in each peak, there were quite different coefficients of variation (shown by the values given below the data points) for replicates within the various peaks. The regional threshold method was sensitive to these variations and only detected two of the peaks (marked by the large asterisks above the tracing), in which both the peak and nadir values reflected a higher degree of assay precision. Conversely, when assay precision was poor in the interpulse nadir but good in the peak, the regional threshold method also responded appropriately (see end of data series). This is particularly important since interpulse hormone concentrations may be very low, especially in states of low mean hormone concentrations.

The performance of the regional threshold method is further illustrated on LH series from pubertal boys (Fig. 5). In these boys, serum LH concentrations measured in blood collected every 20 min for 24 h varied from low values (2–4 mIU/ml) in the daytime to normal adult levels at night. Pulses could be identified at a false-positive rate of $\leq 1\%$ in both daytime and nighttime samples (upstroke and downstroke thresholds of 4- and 3-fold respectively, nadir search range of 3 points after the peak, and regional window of $N = 10$). Note that our choice of a smaller number of points after the peak at 20-min than at 5-min sampling rates reflects the less intensive sampling scheduling (see “Discussion”). We have marked the various individual pulses detected by the regional dual-threshold method with arrows. Note the appearance of increased numbers of LH pulses detected at night, resembling the adult pattern, with markedly (approximately 3-fold) lower LH concentrations during the day in the same individuals.

DISCUSSION

Endocrine pulses are particularly difficult to identify accurately when hormone concentrations approach the lower limit of sensitivity of the assay. This problem becomes important in a variety of pathological and physiological conditions, e.g. in healthy children prior to puberty when gonadotropin concentrations are

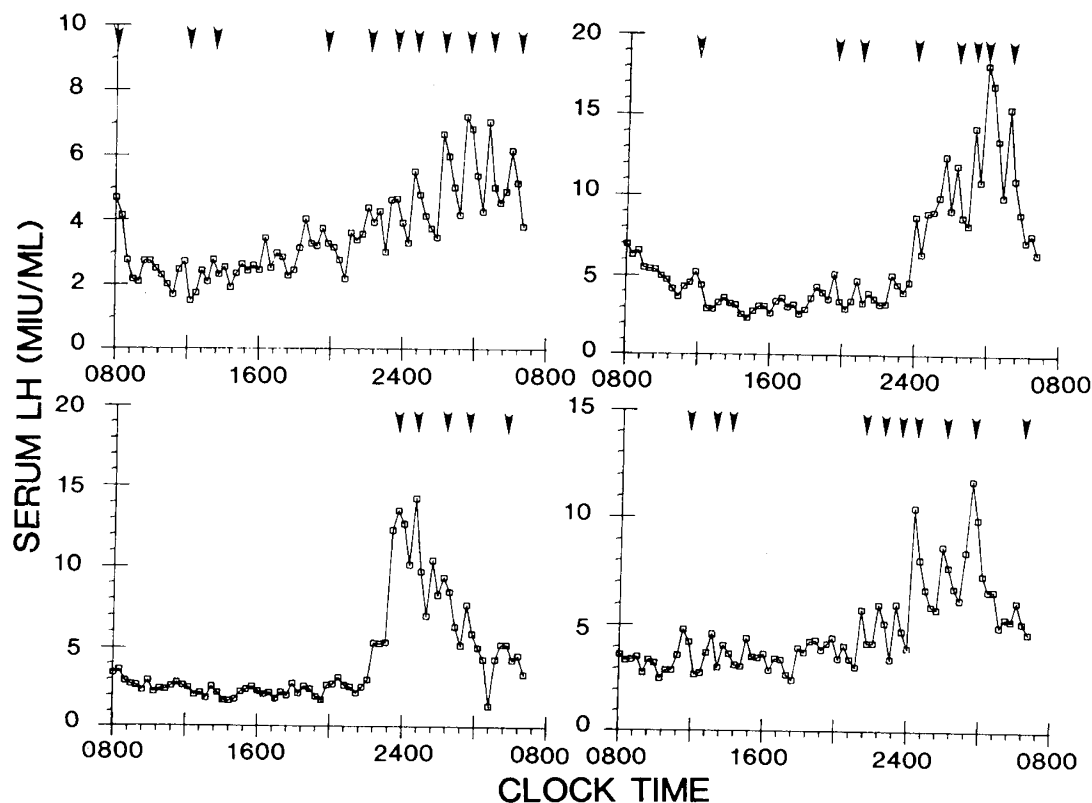


Fig. 5. Illustrative LH pulse profiles in four pubertal boys, in whom blood was sampled at 20-min intervals for 24 hr. Data are mean serum LH concentrations (mIU/ml) in relation to clock time. The arrows indicate significant LH pulses identified by the regional dual-threshold algorithm (see "Methods").

quite low, or in pathological states in adulthood when diminished hormone concentrations exist. However, important insights into pathophysiological mechanisms would still be gleaned from such statistically unfavorable data series, if parameters of endocrine pulse signaling are appraised reliably. Accordingly, we have developed and validated a dual-threshold pulse method, in which varying degrees of increased measurement error can be accommodated throughout the experimental series. As such, this method of estimating the frequency provides a useful addition to existing algorithms.

We have applied the concept of a "regional" coefficient of variation to address nonuniform measurement error in the experimental observations. This provides an expansion of the concept suggested by Ross *et al.* (5), in which a fixed multiple (3-fold) of the apparent intraassay coefficient of variation or paired *t* testing between quadruplicate assay estimates was used to find upstrokes. Our working algorithm relates the selected threshold criteria for both upstroke and downstroke to the pertinent "regional" coefficient of variation within the actual experimental series, and requires that a significant downstroke occur within a specified number of points of the presumptive peak. This procedure assures that regions of the experimental series that exhibit increased measurement error will not contribute a falsely increased number of "pulses." Our working algorithm differs from paired *t* testing for upstrokes alone (which requires many replicates in each sample) and from extrapolating measurement error from replicates within the standard curve or serum pools that are external to the actual experimental samples being tested for peaks. We have estimated false-positive rates for this new method directly using appropriate simulated data series that mimic the behavior of our LH immunoassay. This allows investigators to select program parameters that appropriately limit false-positive errors. Conversely, this algorithm should allow regions within the experimental series that exhibit a high degree of measurement precision to yield lower false-negative error rates. The latter

prediction cannot yet be tested, because there are no absolute and independent markers for "true" LH pulses.

Based on the foregoing results, we have applied this regional dual-threshold method to LH pulsations in normal men, in whom we found an LH pulse frequency of 14.8 ± 1.5 pulses/24 h. For this analysis, we estimated a false-positive error of 1% (threshold increases and decreases of respectively 4- and 3-fold the regional intrareplicate coefficients of variation, which were calculated from the peak and 10 flanking samples). In addition, we required that a significant decrease in LH concentrations occur within 10 samples (50 min) of the putative peak, which allows approximately three half-lives for initial LH disappearance at a 5-min sampling rate (6). In contrast to these results, the original procedure of Santen and Bardin (2) using a fixed 20% threshold yielded a pulse frequency estimate of $28.1 \pm 2.7/24$ h in the same individuals ($p < 0.004$ versus the regional threshold method). This difference presumably reflects in large part the adverse impact of nonuniform intraassay coefficients of variation on false-positive pulses within the LH series when a fixed threshold algorithm is used that is not adjusted for varying internal measurement error.

Our application of this method to pubertal boys illustrates a circumstance in which hormone concentrations vary from barely measurable to adult levels in the same individual. In this context, the range of measurement error is considerable. Moreover, in children, smaller quantities of blood can be withdrawn safely, so that less intensive sampling rates must be considered than in the adult. Accordingly, the program was made adaptable to permit the user to specify the number of samples (which would correspond to the time) after the peak within which the postpeak nadir must be observed. For example, at the 20-min sampling frequency illustrated here in pubertal boys, we required that the LH concentration fall significantly within three samples (60 min) of the maximum LH value contained in the peak. In addition, by accommodating the variable measurement error in daytime

and nighttime samples, the regional dual-threshold method permits one to maintain a constant false-positive rate of pulse detection despite several-fold variations in hormone concentrations. Thus, this algorithm should be particularly useful in accurately characterizing the neuroendocrine changes that accompany the transition from puberty to adulthood.

REFERENCES

1. Veldhuis JD, Clifton D, Crowley WF, Filicori M, Johnson ML, Maciel RJ, Merriam GR, Santoro MF, Steiner RA, Santen RJ 1986 Preferred attributes of objective pulse analysis methods. In: Crowley WF (ed) *Episodic Hormone Secretion* John Wiley and Sons New York (in press)
2. Santen RJ, Bardin CW 1973 Episodic luteinizing hormone secretion in man. Pulse analysis, clinical interpretation, and pathological mechanisms. *J Clin Invest* 52:2617-2628
3. Veldhuis JD, Rogol AD, Johnson ML 1984 Minimizing false-positive error rates in hormone pulse detection. *Am J Physiol* 248:E475-E481
4. Evans WS, Rogol AD, MacLeod RM, Thorner MO 1980 Dopaminergic mechanisms and luteinizing hormone secretion. I. Acute administration of the dopamine agonist bromocriptine does not inhibit luteinizing hormone release in hyperprolactinemic women. *J Clin Endocrinol Metab* 50:103-107
5. Ross JL, Barnes KM, Brody S, Merriam GR, Loriaux DL, Cutler GB 1984 A comparison of two methods for detecting hormone peaks: the effect of sampling interval on gonadotropin peak frequency. *J Clin Endocrinol Metab* 59:1159-1163
6. Veldhuis JD, Evans WS, Rogol AD, Drake CR, Thorner MO, Merriam GR, Johnson ML 1984 Performance of LH pulse detection algorithms at rapid rates of venous sampling in humans. *Am J Physiol* 247:554E-563E