Published in final edited form as:

Stat Med. 2013 November 20; 32(26): 4624–4638. doi:10.1002/sim.5882.

A comparison of methods for analyzing time series of pulsatile hormone data

N. E. Carlson^{*,†}, K. W. Horton, and G. K. Grunwald

Department of Biostatistics and Informatics, University of Colorado Denver, 13001 E. 17th PL, MS B119, Aurora, CO, 80047, U.S.A

Abstract

Many endocrine systems are regulated by pulsatile hormones – hormones that are secreted intermittently in boluses rather than continuously over time. To study pulsatile secretion, blood is drawn every few minutes for an extended period. The result is a time series of hormone concentrations for each individual. The goal is to estimate pulsatile hormone secretion features such as frequency, location, duration, and amount of pulsatile and non-pulsatile secretion and compare these features between groups. Various statistical approaches to analyzing these data have been proposed, but validation has generally focused on one hormone. Thus, we lack a broad understanding of each method's performance. By using simulated data with features seen in reproductive and stress hormones, we investigated the performance of three recently developed statistical approaches for analyzing pulsatile hormone data and compared them to a frequently used deconvolution approach. We found that methods incorporating a changing baseline modeled both constant and changing baseline shapes well; however, the added model flexibility resulted in a slight increase in bias in other model parameters. When pulses were well defined and baseline constant, Bayesian approaches performed similar to the existing deconvolution method. The increase in computation time of Bayesian approaches offered improved estimation and more accurate quantification of estimation variation in situations where pulse locations were not clearly identifiable. Within the class of deconvolution models for fitting pulsatile hormone data, the Bayesian approach with a changing baseline offered adequate results over the widest range of data.

Keywords

simulation; deconvolution; birth-and-death MCMC; spline; model validation; Bayesian

1. Introduction

The pattern in which hormones are released into the blood is critically important for regulation of endocrine systems [1–6]. The modes of release vary from continuous release to intermittent pulsatile release. Many of the hormones responsible for regulating systems such as stress and reproduction are secreted intermittently as pulses. Thus, investigators are

^{*}Correspondence to: N. E. Carlson, Department of Biostatistics and Informatics, University of Colorado Denver, 13001 E. 17th PL, MS B119, Aurora, CO, 80047, U.S.A.

[†]nichole.carlson@ucdenver.edu

interested in estimating the features that define pulsatile release and comparing them between groups of subjects to understand both the physiology and pathophysiology of a hormone. To compare pulsatile features, researchers collect samples of blood every few minutes (5–15 min) for a period of 6–24 h and assay each sample for various hormone concentrations [7–16]. This results in time series of various hormone concentrations for each individual in the study. In standard analyses, each hormone series for each subject is fitted separately using one of a variety of statistical models to obtain estimates of the number and locations of the pulse secretion, along with the amount of hormone secreted in each pulse, the non-pulsatile secretion (baseline) and the hormone elimination (e.g., half-life). In a second stage of analysis, the estimated pulse features are compared between groups using t -tests or other appropriate regression type analyses.

In the past decade, numerous statistical models and estimation algorithms were developed for quantifying pulsatile secretion features in time series of hormone data [17–22]. These approaches resulted in two major contributions. In one set of research, the pulsatile hormone models were expanded to allow pulsing to occur on top of a slowly changing baseline component. This allowed for modeling circadian patterns in the non-pulsatile components of the secretion. In a second set of research, joint estimation of the pulse locations and other model parameters was developed using Bayesian estimation techniques. In non-Bayesian approaches, estimation of the model parameters is performed conditional on the estimated number and location of the pulses. Pulses number and locations are estimated in separate ad hoc steps external to parameter estimation. Thus, bias in the estimated number or location in the first stage of analysis often translates into bias in the second stage of analysis.

Development of each model and methodology have generally focused on fitting one hormone and comparisons between the various approaches has been extremely limited [23, 24]. Thus, it remains unclear whether one method is more generally applicable compared to other methods. The goal of this work was to (1) understand whether the models that allow for circadian patterns would also effectively fit series that do not have a circadian pattern, making them applicable to a wider range of hormones; and (2) understand the strengths and limitations of the Bayesian approach compared with the other non-Bayesian approaches.

The current investigation was carried out using simulated data. The simulated data were created to cover issues arising across reproductive and stress hormones including luteinizing hormone in the follicular and luteal phases, adrenocorticotrophic hormone, and cortisol (Figure 1), for which pulsatile secretion is widely studied. The values chosen for the simulation were a hybrid of parameter ranges for luteinizing hormone, adrenocorticotrophic hormone, and cortisol and were derived from data we have analyzed in reproductive aged women. However, the principles underlying the simulation are likely applicable to other hormones and patient populations.

Using these data, we compared and contrasted four methods for fitting pulsatile hormone data (Table I). AUTODECONV [26, 27] is potentially the most widely used method in the clinical literature and served as a gold standard. BayesDeconv [18] is the Bayesian analog to this approach and jointly estimates pulse locations and model parameters. These two methods assume that parameters do not systematically change over the observation period.

The other two methods studied allow for a changing baseline component using splines [19, 22].

The performance of these algorithms was investigated as the signal-to-noise ratio decreased and when there were circadian (time changing) patterns in either pulse mass, baseline or both. The effect of the signal-to-noise ratio on estimation was investigated by changing model error (noise) and changing the average size of the pulses (signal). The performance of the models under different time changing patterns was investigated by simulating data with a non-changing baseline and pulse mass components and by simulating weak and strong circadian changes over the period of observation.

The model estimation was compared by summarizing the false positive and false negative rates of pulse detection, that is, how well did the methods work at identifying the correct number and location of the latent pulse locations, as well as the frequentist properties of bias and coverage of the other parameters defining the models (decay, pulse mass, and baseline).

2. Methods of characterizing pulsatile hormone data

2.1. The deconvolution model of hormone concentration and parameters of interest

The deconvolution model is explained in detail here because it forms the model frame work for all methods evaluated. The deconvolution model is based on the convolution integral, which is a mathematical construct that represents how independent processes act together to yield an observed outcome. In the case of pulsatile hormone data, the two independent processes are hormone secretion and elimination. Thus, the true hormone concentration at time t for a single subject, is the integral up to time t of the product of the pulsatile secretion, $[S(z)]$, and hormone elimination, $[E(z)]$ plus non-pulsatile, baseline concentration, $[B(t)]$. The subject subscript is suppressed for simplicity. The true hormone concentration for a single subject is then linked with the observed hormone concentration, $y(t)$, through one of the following two statistical models:

$$y(t) = B(t) + \int_0^t S(z)E(t-z)dz + \varepsilon(t) \quad (1)$$

$$\log[y(t)] = \log[B(t) + \int_0^t S(z)E(t-z)dz] + \varepsilon'(t), \quad (2)$$

where $\varepsilon(t)$ and $\varepsilon'(t)$ represent technical and biological error on the original and log scales, respectively. The log transformation is often used to accommodate the fact that hormone concentration is positive. These error choices also imply that model (1) has an additive error structure and model (2) has a multiplicative error structure, on the data's original scale.

The pulsatile secretion $[S(z)]$ is further defined as the sum of the single pulse secretion events, that is, $S(z) = \sum_{k=1}^{N_p} s_k(z)$ where N_p are the number of pulsatile secretion events in the period of observation and $s_k(z)$ is the k th pulsatile secretion event. Each pulsatile secretion event is defined by three quantities: (1) the amount of hormone secreted in the pulse (called mass); (2) duration; and (3) location. For three of the models investigated, a single secretion event is modeled by a Gaussian curve,

$s_k(z) = \frac{\mu_{p,k}}{\sqrt{2\pi\sigma_{p,k}^2}} \exp \left\{ \frac{-1}{2\sigma_{p,k}^2} (z - \tau_{p,k})^2 \right\}$, with $\mu_{p,k}$ defining the mass of the pulse, $\sigma_{p,k}^2$ the duration, and $\tau_{p,k}$ the location. The non-pulsatile secretion $[B(t)]$ is either constant, represented by one parameter B or modeled semi-parametrically by a spline. The elimination is modeled by a single exponential decay, $E(t-z) = \exp \left[\frac{-\log(2)}{h} (t-z) \right]$ where h is the half-life in minutes. The secretion and elimination functions differ slightly for the PULSE model and the differences are defined later on.

In the following sections, we further define how the pulse secretion features are modeled and give a brief outline of each of the four estimation algorithms.

2.2. Methods with a constant baseline concentration

2.2.1. AUTODECONV [27]—The AUTODECONV method is a modification of the original estimation algorithm proposed by Veldhuis and Johnson (1992) [26]. The model was developed on the observed scale following Equation (1) previously. *Model specifics:*

All pulse specific parameters ($\mu_{p,k}$, $\sigma_{p,k}^2$ and $\tau_{p,k}$) are treated as fixed effects. The pulse duration is assumed common across all pulses to reduce the number of parameters in the model. The baseline secretion is constant in the AUTODECONV model. *Estimation:* AUTODECONV uses a three-step process: (1) Given the initial parameter values and an initial set of pulse locations, the model parameters are estimated using weighted nonlinear least squares. (2) Pulses are then added to the model one at a time at the observation time that results in the largest change in the residual error (called variance-of-fit). After each pulse is added the model parameters are reestimated. Pulses are added until adding a pulse does not significantly reduce the variance-of-fit. (3) Pulses are then removed one at a time and the parameters reestimated. The pulse is permanently removed when the variance-of-fit for the reduced model is not significantly reduced compared to the model with the pulse included. The algorithm iterates between the insertion and deletion steps until no new pulses are added after one complete cycle of the insertion and deletion steps.

2.2.2. Bayesian Deconvolution [18] (BayesDeconv)—BayesDeconv is the Bayesian analog to the AUTODECONV model; however, the observed hormone concentrations are modeled on the log scale (Equation 2 previously). *Model specifics:* The pulse masses and widths are assumed to be random effects with means m_m and d_m , respectively, and variances s_m^2 and s_d^2 , respectively (Table I). As in AUTODECONV, the baseline is assumed constant in the model. To complete the Bayesian specification vague priors are placed on m_m , d_m , s_m^2 , s_d^2 , B , and h . In addition, vague priors are placed on the number and locations of the pulses. The exact prior distributions can be found in [18]. The summary measures of interest in the Bayesian formulation of the deconvolution model are the means of the posterior distributions of the mean pulse mass, m_m , the mean pulse duration, d_m , the baseline concentration, B , and the half-life, h . Unlike the non-Bayesian approaches, the result is a distribution of the number of pulses (N_p) and not a single model with a single number of pulses. In addition, the posterior distributions of the other parameters are averaged over all models, which may have differing numbers and locations of the pulses. Thus, the summary measures contain the estimation variation in the number and location of the pulses, unlike in

the non-Bayesian approaches. *Estimation:* The parameters are estimated using a birth-and-death MCMC algorithm [28]. The algorithm simultaneously estimates the pulse number and locations along with the other model parameters by iterating between two major steps: (1) conditioned on the current values of all parameters, the pulse number and other pulse parameters are simulated using a birth-and-death process; and (2) conditioned on the current pulse number and values of the pulse parameters, a traditional MCMC algorithm is used to simulate all model parameters. The full algorithm iterates between steps 1 and 2. Convergence of this algorithm is established in [28]. More specifically the algorithm is as follows:

Given the current state of the parameters (baseline, half-life, model error variance, pulse number, location, mass, and width, and mean mass and width) at iteration r , simulate the $(r + 1)$ st draw according to the following steps:

1. For each subject, sample the pulse number and pulse specific parameters $(\mu_{p,k}, \sigma_{p,k}^2, \text{ and } \tau_{p,k})$ by running a virtual birth–death process given the current values of the other parameters defining the model.
2. Conditioned on the number of pulses identified, use a Metropolis–Hastings within Gibbs hybrid [29] sampler to update the baseline, half-life, model error, mean pulse mass and width, and variance of the pulse masses and widths.
3. Use a Metropolis–Hastings within Gibbs hybrid sampler to update the pulse specific parameters.

Each simulation was run for 150,000 iterations. The first 20,000 iterations of each chain were treated as burn-in. Every 50th iteration was saved and used to summarize the posterior distributions. The variances in the proposal distributions for the Metropolis–Hastings steps were chosen to obtain acceptance rates between 25% and 50%. Convergence and mixing were visually assessed using serial plots of the draws. When starting values were chosen in biologically plausible ranges, the algorithm had adequate performance.

2.3. Methods with changing basal concentration

2.3.1. PULSE [22]—The secretion and elimination functions in this deconvolution model follow development in O’Sullivan and O’Sullivan [30] and result in the following model of observed concentration, which has an error structure as in Equation (1) previously:

$$y(t) = B(t) + \sum_{k=1}^{N_p} \mu_{p,k} p(t - \tau_k; \delta, h) + \varepsilon(t),$$

where

$$p(t - \tau_k; \delta, h) = \begin{cases} \exp\{\delta(t - \tau)\}, & t < \tau, \tau \in \{\tau_k\} \\ \exp\left\{-\frac{\log(2)}{h}(t - \tau)\right\}, & t \geq \tau, \tau \in \{\tau_k\}. \end{cases} \quad (3)$$

The pulse mass is defined by $\mu_{p,k}$, as in the other models, δ is the infusion rate and similar in interpretation to pulse duration in the other models, and h is the half-life in minutes as in the other models. As in the Bayesian models, the pulse masses are modeled as random effects with mean m_m and variance s_m^2 . The infusion rate is fixed and assumed common across the pulses. The baseline function is allowed to change over time and modeled by a smoothing spline. *Estimation:* The estimation algorithm is a three step process. First, parameters are initialized, then pulse detection occurs, and finally, given a pulse number and set of pulse locations, the final estimates of the parameters are obtained. Parameter estimation is performed using the semiparametric nonlinear mixed effects algorithm developed in [31]. Briefly, parameters in the spline and hormone concentration model are estimated by minimizing a double-penalized log-likelihood in one step and maximizing the profile-likelihood in a second step. In the pulse detection step, this algorithm is applied to a series of models with a different number of pulses in each model. The sequence of models is chosen by removing the pulses with the smallest mass in sequence. Then the final model is chosen using a modified AIC selection criterion where the modification is in the measure of model complexity to account for the pulses and the spline. We used the following model options at the various stages: (1) we initialized pulse locations using the CLUSTER option and used a value of $\alpha = 0.1$ because it appropriately erred on the side of false positives as needed; (2) we used the shift and spline options to initialize the baseline; (3) we assumed fixed effects for the parameters when obtaining initial values for the parameters; and (4) we assumed a fixed effect for the infusion and decay rates and a mixed effect for the pulse masses. For some series, the final estimation algorithm did not converge when using the previous options. In these cases, adjusting the α level up or down slightly led to a different number of initial pulses and also alleviated the convergence issues.

2.3.2. Bayesian deconvolution with a changing baseline [19]

(BayesDeconvChbl)—BayesDeconvChbl extends the Bayesian deconvolution algorithm by modeling the baseline $[B(t)]$ as a natural cubic b-spline [32]. The error structure follows Equation (2) previously. All other features of the model are as in the Bayesian deconvolution model. The estimation algorithm is expanded to include two birth-and-death processes. The first simulates the number and locations of the pulses and the second simulates the number and locations of knots in the b-spline. More specifically the algorithm is as follows:

Given the current state of the parameters (baseline, half-life, model error variance, pulse number, location, mass, and width, and mean mass and width) at iteration r , simulate the $(r + 1)$ st draw according to the following steps:

1. For each subject, sample the pulse number and pulse specific parameters ($\mu_{p,k}$, $\sigma_{p,k}^2$, and $\tau_{p,k}$) by running a virtual birth–death process given the current values of the other parameters defining the model.
2. Conditioned on the number of pulses identified, sample the number of knots in the b-spline by running a virtual birth–death process given the current values of the other model parameters.

3. Conditioned on the number of knots in the b-spline, use a Metropolis–Hastings within Gibbs hybrid sampler to update the b-spline coefficients.
4. Conditioned on the number of pulses and the current value of the b-spline, use a Metropolis–Hastings within Gibbs hybrid sampler to update the half-life, model error variance, mean pulse mass and width and variance of the pulse masses and widths.
5. Use a Metropolis–Hastings within Gibbs sampler to update the pulse specific parameters.

The implementation of the algorithm is the same as for the Bayesian deconvolution approach.

3. Simulated data

Table I lists the compared methods and Table II the 10 parameter combinations used to simulate the hormone series. Figure 2 provides an example of simulated time series for the different simulation scenarios. Our interest was in elucidating what method, if any, had superior performance in terms of pulse detection and parameter estimation as one or more of the following features existed in the data: (1) the signal-to-noise ratio increased (or decreased) by a change in the technical and biological error; (2) the signal-to-noise ratio decreased (or increased) by a change in the average pulse mass; or (3) baseline, pulse mass, or both changed systematically over the period of observation. To investigate these properties we simulated 10 sets of 500 series. The true hormone concentration was simulated using the deconvolution model of hormone concentration. The observed hormone concentration was simulated by log transforming the true hormone concentration, adding Gaussian model error, and then back transformed to the observed hormone concentration scale (Equation (2) in Section 2.1). The parameters that varied in the model are shown in Table II. For each series, we assumed a sampling interval of 10 min for a period of 24 h to correspond to the study design for many pulsatile hormone studies. This resulted in 144 observed hormone concentration values for each simulated series. We first simulated the pulse locations using a renewal process, that is, the time between pulses was randomly drawn from a gamma distribution with a mean of 120 min and a variance of 600 min². A minimum interpulse interval of 40 minutes was imposed to prevent two pulses from occurring too close in time. The half-life was set to 45 min for all scenarios. Given the number and location of the pulses for a series, the pulse masses and widths were treated as random effects and each drawn separately from truncated *t*-distributions with 4 degrees of freedom (df) constrained to be positive. To draw from the *t*-distribution, we used the conditional gamma formulation [33], [*i.e.*, if $x \mid \kappa \sim N(\mu, \kappa^{-1} \Sigma)$ and $\kappa \sim \text{Gamma}(r/2, r/2)$, then the marginal distribution of x is a *t*-distribution with r degrees of freedom]. When the mean of the masses did not change over the observation period the mean of the random effects distribution ranged from 0.75 to 6.5 concentration units with the reference mean being 3.5 units (Table II). The mean of the normal distribution for the pulse widths was 5 min (or 1/2 a sampling unit) and the variance was 1 min. When the baseline concentration was constant over the period of observation, the value was 2.6 concentration units. When incorporating a circadian pattern in the baseline or mean of the pulse mass distribution, we

used a cosinor model (i.e., $m_m = a + b\cos(2\pi t/1440) + c\sin(2\pi t/1440)$, where t is time in minutes). The parameters used in the cosinor can be found in Tables I and II. We simulated two cosinor functions for both baseline and mass. For each, one reflected a strong circadian pattern and one reflected a weak circadian pattern. The model errors were independently drawn from a Gaussian distribution with variances corresponding to constant coefficient of variation on the observed scale ranging from 5% to 20%. The actual variance values are shown in Tables I and II.

3.1. Methods of comparison

The methods were compared on the basis of the following metrics:

1. False positive rate (FP): This is 100 times the number of estimated pulses that are not within 20 min of a true pulse location divided by the number of pulses that were estimated. For the Bayesian approaches, the false positives were identified for each iteration and the percentage calculated by summing the false positives across all iterations and dividing by the total number of pulses estimated across all iterations.
2. False negative rate (FN): This is 100 times the number of true pulses that have no estimated pulse within 20 min divided by the number of pulses that were simulated. For the Bayesian approaches, the false negatives were identified for each iteration and the percentage calculated by summing across all iterations and dividing by the total number of simulated pulses times the number of iterations.
3. Average bias: This is the average of the bias (estimated value minus the true parameter value) for each of the 500 simulated data sets. For the Bayesian approaches the bias is the difference between the posterior mean and the true parameter value. To summarize the models and methods with a changing baseline, the average baseline was computed for the simulated and data fits. The bias is then the estimated average baseline minus the average true baseline.
4. 95% coverage: The percentage of 95% CI that contain the true parameter value. For the Bayesian approaches the equal tails 95% credible interval was used as an analog to a frequentist 95% CI. AUTODECONV does not provide variability for the estimates and thus is missing this summary measure. Coverage was not computed for the average baseline because it is only available for one algorithm (BayesDeconv).
5. Graph of baseline curves: The average estimated baseline concentration at each observation time point was calculated and graphed to assess whether the shape of the estimated baseline was similar to the simulated true baseline curves.

4. Results

4.1. Influence of signal-to-noise ratio

4.1.1. Pulse detection—For the reference and high signal-to-noise models, the FP and FN rates were low (~ 3% FP and 3–10% FN) for three of the methods (AUTODECONV and Bayesian approaches; top panel of Figure 3). This equates to less than one misidentified

pulse on an individual over a 24-h period for the simulated pulse frequency. A fivefold increase in the model error resulted in only a slight increase in the FP rates (6–12% FP); however, the FN rate more than doubled from 16 to 30%. Reducing the average pulse mass by 75% (from 3.5 to 0.75 units/pulse on average) resulted in a substantial increase in the both the FP and FN rates (27–57% FP and 42–59% FN). The PULSE algorithm had the highest false positive and negative rates for all sizes of model error. These rates did not improve when using an alternative pulse detection algorithm. Bayesian deconvolution (BayesDeconv) generally had the lowest false positive and negative rates. Allowing for flexibility in the baseline in the Bayesian approach resulted in similar, if not lower, FP rates for all signal-to-noise models, but resulted in higher FN rates. Unlike the other approaches, the Bayesian approaches estimate the number of pulses, thus, calculating the coverage of the number of pulses is possible. The 95% coverage probability was > 99% for both Bayesian methods. However, the width of the 95% credible intervals of the number of pulses becomes quite wide as the signal-to-noise ratio decreased. The 95% CI widths ranged from 6.5 to 10.5 pulses for low signal-to-noise models versus CI widths ranging from 1.5 to 3 pulses for high signal-to-noise models. The change in the CI widths likely reflected the lack of identifiability of the exact number of pulses in the models as the signal-to-noise ratio decreased.

4.1.2. Baseline and half-life parameter estimation—For the reference and high signal-to-noise models, bias in the baseline is lower for methods that assume a constant baseline compared with methods that assume a changing baseline (~6% vs. 20%, respectively; second panel of Figure 3). Increasing the model error resulted in a larger positive bias in the baseline estimates. The pattern between the constant and changing baseline models remained similar. Decreasing the pulse mass resulted in less bias in the baseline estimates compared with increasing the model error. The methods with a changing baseline fit a relatively constant baseline (Figure 4). Given the strong correlation between estimates of baseline and half-life, the bias in the half-life mimicked that of the baseline. The coverage of the 95% CI for half-life was greater than 90% for all but one Bayesian model (Table III). The coverage of the PULSE model was substantially smaller (< 60%) than coverage of the Bayesian approach.

4.1.3. Average pulse mass parameter estimation—As the signal-to-noise ratio decreased, the pulse mass tended to have a larger positive bias (bottom panel Figure 3). This corresponded to the increase in the FN rate. Small pulses were not detected, and thus, the average pulse mass was estimated from the larger pulses of the set of true pulses. When the average pulse mass was quite large, the bias in the pulse mass was negative. This potentially resulted from one pulse being modeled as two smaller pulses. Given the restriction on closeness of pulses in the BayesDeconv algorithm, the pulse mass has a low bias for this measure when the signal-to-noise ratio is strong. As the signal-to-noise ratio decreased the bias increased as it did for the other methods. Except when the average pulse mass was quite small, the coverage of the 95% CI was high for all methods and all signal-to-noise conditions.

4.2. Accommodating circadian rhythms

4.2.1. Pulse detection—The Bayesian method with a changing baseline (BayesDeconvChbl) had the lowest FP and FN rates for all the models with circadian patterns (top right panel Figure 3). Having a weak circadian pattern in the pulse mass did not influence the FP and FN rates for any of the methods. However, a strong circadian pattern in the pulse mass increased the FN rates for all the methods and increased the FP rate for the Bayesian methods without a changing baseline (BayesDeconv). There was fairly small influence of a circadian pattern in the pulse mass for AUTODECONV compared with the other methods. This possibly implies that the random effects distribution is not able to capture the large amount of variation over the day when there is a circadian pattern.

Ignoring the circadian baseline in model fitting when it exists in the data had a much more detrimental effect on the FP rates compared to ignoring a circadian pattern in the pulse mass. The FP rate was high ($> 20\%$) for all methods for both strong and weak circadian patterns in the baseline. The FN rate was not influenced by misspecifying the baseline. This indicates that pulses are being added to the model in places that they do not exist to increase the hormone concentration level to fit the observed hormone concentration. When appropriately accounting for the circadian baseline, the Bayesian approach had much lower FP and FN rates compared with PULSE that had FP and FN rates of approximately 30%.

4.2.2. Baseline and half-life parameter estimation—Ignoring a circadian pattern in the pulse mass when fitting the data did not result in any change in the bias of the half-life (Figure 3 third panel). There was a slight increase in the bias of the baseline (Figure 3 second panel). This pattern is consistent across all four methods. All the methods fitted a flat baseline (Figure 4). Coverage was high ($\sim 90\%$) for the half-life for the Bayesian approaches compared with PULSE.

Fitting a model assuming a constant baseline when the baseline is actually circadian resulted in extremely large positive bias in the half-life and negative bias in the baseline. AUTODECONV fits the baseline to the minimum hormone concentration, while the Bayesian approach smoothed through the data more and added more false pulses (Figure 4). This resulted in much less bias in both the half-life and baseline estimates compared to AUTODECONV. When fitting a circadian baseline, both the Bayesian and PULSE models had similar bias and fit the shape of the baseline equally well, but the Bayesian approach had much higher coverage. There was a slightly higher negative bias in the half-life for PULSE, but this can likely be explained by the difference in the statistical and simulation models for this approach.

4.2.3. Average pulse mass parameter estimation—With the exception of AUTODECONV, the methods tended to underestimate the pulse mass for all the circadian pattern models. AUTODECONV generally had a large positive bias in the pulse mass estimates. Bias in the average pulse mass tended to be more influenced by ignoring circadian patterns in the baseline than ignoring a pattern in the pulse mass (Figure 3 bottom panel). In general, the Bayesian approach with a changing baseline (BayesDeconvChbl) had the lowest

bias of the parameter estimates. Also, generally, coverage of the 95% CI for the average pulse mass was high for all methods.

5. Discussion

AUTODECONV and its Bayesian analog produced similar results when the data had reasonable signal-to-noise ratios. However, when the signal was not strong, the Bayesian approach offered additional ways to quantify the lack of identifiability of the model. Although the FP and FN rates and biases in the parameters increased with a decreasing signal-to-noise for all approaches, the coverage of the parameters was still good for the Bayesian fits. This potentially shows the MCMC algorithm's ability to identify lack of separation of the parameters compared with the weighted least squares minimization approach employed in AUTODECONV.

Allowing for flexibility in the baseline when fitting introduced model identifiability issues between baseline and half-life regardless of whether the baseline was in truth changing or constant. This bias resulted in increased FP and FN rates for the changing baseline models. Thus, it may be useful to assume a constant baseline when fitting when it is believed to be true. However, the bias introduced when using a more flexible approach was small. Thus, if the goal is to use one statistical package most of the time or if the form of the baseline is not known, the Bayesian method with a changing baseline may be preferable to the other approaches as it introduced only a small amount of bias in most cases.

Ignoring patterns in the pulse mass when fitting did not influence the results. The bias in pulse mass was minimized given the modeling flexibility of the random effects model for the individual pulse masses. However, when a circadian baseline exists and is ignored in fitting, all parameter estimates are highly biased and the amount of pulsatile secretion incorrectly estimated. When ignoring a circadian baseline, the methods fit to the minimum concentration in the circadian pattern. In regions where the circadian baseline has a higher concentration, extra pulses are added to the model to increase the fitted hormone concentration. Thus, ignoring the circadian baseline often results in an over estimate of the amount of pulsatile hormone secretion (and an increased FP rate) and an underestimate of the basal secretion. When the circadian patterns are similar between groups of subjects, these bias may wash out when analyzing the difference between the groups; however, if the circadian patterns differ between groups, incorrect conclusions regarding group differences will likely result.

The PULSE algorithm was expected to perform slightly worse compared with the other methods because the simulation and statistical models differed slightly. However, the FP and FN rates were much higher than the other approaches even in cases where performance would be expected to be good. These situations also showed bias in the estimates of the other parameters. We conclude that PULSE has fitting issues that appear to go beyond just model misspecification. One reason may be that the algorithm does not iterate between adding and removing pulses. Thus, estimating the correct number of pulses depends upon over-identifying pulses in the first step, which may not be simple to do using initial estimates of the other pulse parameters.

The additional computation time and complexity in model interpretation of the Bayesian approach adds little in a well-defined series; however, the Bayesian approach has several strengths as noise and circadian patterns are added to the system. First, uncertainty in the number of pulses is captured in the posterior distribution of the number of pulses. Second, the parameter estimates are averaged over all possible models of pulse number and location. Thus, a single model of the number of pulses is not needed in situations where the true number of pulses is difficult to identify. Third, the improved coverage of the parameters in the Bayesian approach probably more accurately reflects the ill-posed nature of pulsatile hormone models. It is disappointing that the original deconvolution model (AUTODECONV) has removed estimates of variability in their estimates from the software. This likely reflects the difficulty of obtaining appropriate estimates of variance in a complex nonlinear model; however, it is a significant limitation of the approach.

With the exception of the PULSE model, the baseline was the main hormone feature that differed between the simulation and the model fitting. In addition, the deconvolution model used for simulation and model fitting is likely also oversimplified in its model of the pulse shape and the elimination function. Use of the Gaussian pulse shape and restriction to a single exponential decay for the elimination function have generally been used to aid in model fitting and parameter identifiability. In theory, each of these methods can be extended to incorporate more accurate (and complex) biology. However, improvements in estimation may be marginal given the sparse sampling intervals (every 10–15 min) relative to biology. We would expect the biases owing to model simplification to be similar across the three deconvolution models. Thus, these findings should generalize to more complex underlying deconvolution models.

The most important contribution of the deconvolution model to the modeling of pulsatile hormone levels is incorporating the simultaneous action of the secretion and elimination. Given the sampling interval relative to the timing of the biology it is likely less important to get the exact shape of the secretion and elimination functions correct. This distinction is exhibited in the findings for the PULSE algorithm, which assumes that only one of secretion or elimination acts at a particular time. This resulted in substantial bias in the estimates of elimination. Using the Gaussian shape minimizes computational complexity, but does likely introduce some bias in the estimation of duration and half-life in analysis of real data. Half-life would likely be underestimated and duration overestimated in real data analysis compared with an analysis with a skewed pulse shape, which is more biologically plausible. This simplification needs to be balanced with the computational burden that would be introduced in approximating a more complex integral. Using only a single exponential decay instead of a bi-exponential decay likely results in an overestimate of the decay. However, it is unclear that data are collected frequently enough to achieve adequate estimation of the fast decay in the biexponential. Mauger, et al. (1995) [23] showed that at least four data points are needed on a half-life to achieve good estimation. Furthermore, because the final inference is between group comparisons, as long as the misspecification is relatively similar between groups, the final group-to-group comparisons will remain valid.

Many of the models used random effects for the pulse mass and width, but some assume a constant pulse mass and width across the series. The following considerations are likely

important for achieving good estimation properties. The pulse mass should be allowed to vary between pulses (either through a random effect or brute force estimation of each pulse size) for adequate pulse detection. If the pulse masses are assumed constant, the false negative rate may increase and the average pulse mass may be biased high because small pulses may be difficult to detect in the presence of large pulses. Fixing the pulse width across the series likely introduces fewer problems with bias and pulse identification because width is the least identifiable in estimation. There are often only one or two observations defining the up side of a pulse and the down side is confounded with elimination. Thus, if further model simplification is necessary for fitting, pulse width could be modeled as constant across pulses without much sacrifice in estimation performance.

We focused on implementing the algorithms as provided. Modifications to MCMC fitting algorithms may result in further advantages for the Bayesian approaches. The false positive rates may decrease and bias in the pulse mass may reduce if the point process underlying the birth–death MCMC algorithm were changed from a modified Poisson process to a repulsive process such as a Strauss [34] or hard core process [35], because pulses will be even more likely to be modeled by a single Gaussian rather than several Gaussians with smaller masses close together in time. In addition, the biases in the baseline may be reduced by restricting the potential number of knots or the set of potential locations of the knots in the birth–death MCMC algorithm for identifying knot locations. Further improvements in the fit of the baseline might also be achieved by moving from a standard [36] to a natural b-spline [32].

On the basis of our investigation, we conclude that the most general model for fitting pulsatile hormone data is the Bayesian approach with the changing baseline. The additional model flexibility in the baseline introduces minimal issues with estimation while maintaining the strengths that come with joint estimation of locations and other model parameters. However, implementing the Bayesian approaches does come with more up front time for guaranteeing appropriate specification of priors and tuning of the MCMC algorithm. Therefore, it will be necessary to develop tutorials and menu driven software interfaces for the Bayesian approaches to offer clinicians more guidance on how to implement the Bayesian estimation algorithms.

Acknowledgments

The authors would like to thank Thomas Glorioso, a graduate student at the University of Colorado Denver, for his assistance in creating the graphs in this manuscript. We also thank the referees for their thoughtful comments that helped to improve the article.

References

1. Sherwood, L. Fundamentals of Human Physiology: A Human Perspective. 3. Brooks-Cole; Belmont, CA: 2005.
2. Greenspan, FS.; Gardner, DG. Basic and Clinical Endocrinology. 7. McGraw-Hill; New York: 2004.
3. Marshall JC, Dalkin AC, Haisenleder DJ, Griffin ML, Kelch RP. GnRH pulses—the regulators of human reproduction. Transactions of the American Clinical and Climatological Association. 1993; 104:31–46. [PubMed: 1343446]

4. Carroll BJ, Curtis GC, Mendels J. Neuroendocrine regulation in depression I. Limbic system-adrenocortical dysfunction. *Archives of General Psychiatry*. 1976; 33:1039–1044. [PubMed: 962488]
5. Carroll BJ, Curtis GC, Mendels J. Neuroendocrine regulation in depression II. Discrimination of depressed and nondepressed patients. *Archives of General Psychiatry*. 1976; 33:1051–1058. [PubMed: 962489]
6. Schally AV, Arimura A, Kastin AJ, Matsuo H, Baba Y, Redding TW, Nair RM, Debeljuk L, White WF. Gonadotropin-releasing hormone: one polypeptide regulated secretion of luteinizing and follicle-stimulating hormones. *Science*. 1971; 173:1036. [PubMed: 4938639]
7. Clarke IJ, Cummins JT. The temporal relationship between gonadotropin releasing hormone (GnRH) and luteinizing hormone (LH) secretion in ovariectomized ewes. *Endocrinology*. 1982; 111:1737–1749. [PubMed: 6751801]
8. Vermeulen A, Kaufman JM, Deslypere JP, Thomas G. Attenuated luteinizing hormone (lh) pulse amplitude but normal lh pulse frequency, and its relation to plasma androgens in hypogonadism of obese men. *Journal of Clinical Endocrinology and Metabolism*. 1993; 76:1140–6. [PubMed: 8496304]
9. Abelson JL, Curtis GC. Hypothalamic-pituitary-adrenal axis activity in panic disorder: 24-hour secretion of corticotropin and cortisol. *Archives of General Psychiatry*. 1996; 53:323–331. [PubMed: 8634010]
10. Young EA, Carlson NE, Brown MB. Twenty-four hour ACTH and cortisol pulsatility in depressed women. *Neuropsychopharmacology*. 2001; 25:267–276. [PubMed: 11425510]
11. Young EA, Abelson J, Lightman SL. Cortisol pulsatility and its role in stress regulation and health. *Frontiers in Neuroendocrinology*. 2004; 25:69–76. [PubMed: 15571755]
12. Breen KM, Oakley AE, Pytiak AV, Tilbrook AJ, Wagenmaker ER, Karsch FJ. Does cortisol acting via the type II glucocorticoid receptor mediate suppression of pulsatile luteinizing hormone secretion in response to psychosocial stress? *Endocrinology*. 2007; 148:1882–1890. [PubMed: 17204556]
13. Jain A, Polotsky AJ, Rochester D, et al. Pulsatile luteinizing hormone amplitude and progesterone metabolite excretion are reduced in obese women. *Journal of Clinical Endocrinology and Metabolism*. 2007; 92:2468–73. [PubMed: 17440019]
14. Wagenmaker ER, Breen KM, Oakley AE, Pierce BN, Tilbrook AJ, Turner AI, Karsch FJ. Cortisol interferes with the estradiol-induced surge of luteinizing hormone in the ewe. *Biology of Reproduction*. 2009; 80:458–463. [PubMed: 19056703]
15. Kunugi H, Hori H, Adachi N, Numakawa T. Interface between hypothalamic–pituitary–adrenal axis and brain-derived neurotrophic factor in depression. *Psychiatry and Clinical Neurosciences*. 2010; 64:447–459. [PubMed: 20923424]
16. Walker JJ, Terry JR, Lightman SL. Origin of ultradian pulsatility in the hypothalamic–pituitary–adrenal axis. *Proceedings of the Royal Society B*. 2010; 277:1627–1633. [PubMed: 20129987]
17. Guo W, Wang Y, Brown MB. A signal extraction approach to modeling hormone time series with pulses and a changing baseline. *Journal of the American Statistical Association*. 1999; 94:746–756.
18. Johnson TD. Bayesian deconvolution analysis of pulsatile hormone concentration profiles. *Biometrics*. 2003; 59:650–660. [PubMed: 14601766]
19. Johnson TD. Analysis of pulsatile hormone concentration profiles with nonconstant basal concentration: a Bayesian approach. *Biometrics*. 2007; 63:1207–1217. [PubMed: 18078483]
20. Keenan DM, Evans WS, Veldhuis JD. Control of lh secretory-burst frequency and interpulse-interval regularity in women. *American Journal Physiology, Endocrinology, and Metabolism*. 2003; 285:E938–E948.
21. Keenan DM, Roelfsema F, Veldhuis JD. Endogenous ACTH concentration-dependent driver of pulsatile cortisol secretion in the human. *American Journal Physiology*. 2008; 287:E652–E661.
22. Liu A, Wang Y. Modeling of hormone secretion-penetrating mechanisms with splines: a pseudo-likelihood approach. *Biometrics*. 2006; 63:201–208. [PubMed: 17447946]
23. Mauger DT, Brown MB, Kushler RH. A comparison of methods that characterize pulses in a time series. *Statistics in Medicine*. 1995; 14:311–325. [PubMed: 7724916]

24. Johnson, TD.; Wang, Y. Recent advances in the analysis of episodic hormone data. In: Biswas, A.; Datta, S.; Fine, J.; Segal, M., editors. *Statistical Advances in the Biomedical Sciences*. John Wiley & Sons; Hoboken, NJ: 2008. p. 527-546.
25. Young EA, Midgley AR, Carlson NE, Brown MB. Alteration in the hypothalamic–pituitary–ovarian axis in depressed women. *Archives of General Psychiatry*. 2000; 57:1157–1162. [PubMed: 11115329]
26. Veldhuis JD, Johnson ML. Deconvolution analysis of hormone data. *Methods in Enzymology*. 1992; 210:539–575. [PubMed: 1584051]
27. Johnson ML, Pipes L, Veldhuis PP, Farhy LS, Nass R, Thorner MO, Evans WS. AutoDeconv: a robust numerical method for the quantification of pulsatile events. *Methods in Enzymology*. 2009; 454:367–404. [PubMed: 19216935]
28. Stephens M. Bayesian analysis of mixture models with an unknown number of components-an alternative to reversible jump methods. *The Annals of Statistics*. 2000; 28:40–74.
29. Tierney L. Markov chains for exploring posterior distributions. *The Annals of Statistics*. 1994; 22:1701–1762.
30. O’Sullivan FO, O’Sullivan J. Deconvolution of episodic hormone data: an analysis of the role of season on the onset of puberty in cows. *Biometrics*. 1988; 44:339–353. [PubMed: 3390503]
31. Ke C, Wang Y. Smoothing spline nonlinear nonparametric regression models. *Journal of the American Statistical Association*. 2004; 99:1166–1175.
32. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning*. 2. Springer-Verlag; New York: 2008.
33. Kleinman KP, Ibrahim JG. A semiparametric Bayesian approach to the random effects model. *Biometrics*. 1998; 54 :921–938. [PubMed: 9750242]
34. Strauss D. A model for clustering. *Biometrika*. 1975; 63:467–475.
35. Møller, J.; Waagepetersen, RP. *Statistical Inference and Simulation for Spatial Point Processes*. Chapman and Hall/CRC; London: 2004.
36. de Boor, C. *A practical Guide to Splines*. Springer-Verlag; New York: 1978.

Figure 1.

Example experimental pulsatile hormone data. (A) Adrenocorticotrophic hormone (ACTH) sampled every 10 min [10] for 24 h; (B) cortisol sampled every 10 min for 24 h [10]; (C) luteinizing hormone (LH) sampled every 10 min for 12 h during the luteal phase of the menstrual cycle [25]; and (D) LH sampled every 10 min for 12 h during the follicular phase of the menstrual cycle [25].

Figure 2.

An example simulated series for 9 of the 10 models. Parameters defining these models can be found in Table II. Data were generated at 10-min intervals for a 24-h period. Each series has 144 simulated concentration values.

Figure 3.

False positive and negative rates (top panel), average bias in average baseline (2nd panel), average bias in half-life (3rd panel) and average bias in average pulse mass (bottom panel). Column 1 shows the influence of changes in the signal-to-noise ratio and column 2 show the influence of circadian changes in the baseline and pulse mass. Standard error bars are not provided because they are visually negligible for most models. Each bar is the average from fits of 500 simulated series.

Figure 4.

Average fitted baseline concentration curve versus the true baseline concentration curve. The range of the y-axis is a smaller range than the range of the observed hormone concentration levels to emphasize differences in the baseline estimates. The range of the hormone concentration levels in the simulation are typically 0 to 15 concentration units.

Table 1

Features of the methods and simulation.

Quantity	Parameter	AUTODECONV [26, 27]	BayesDeconv [18]	PULSE [22]	BayesDeconvChbl [19]	Simulations
Baseline	$B(t)$	Constant	Constant	Spline	Spline	Constant: 2.6 Weak circadian: $1.5 + 1.0 \cos x + 1.0 \sin x$ Strong circadian: $3 + 2.0 \cos x + 2.0 \sin x = 2\pi/1440$
Pulse mass	$\mu_{p,k}$	Fixed effect	Random effect ^a $\mu_{p,k} \sim N^+(m_m, s_m^2)^c$	Random effect $\mu_{p,k} \sim N(m_m, s_m^2)$	Random effect ^a $\mu_{p,k} \sim N^+(m_m, s_m^2)$	Random effect Ref: $t_4^+(3.5, 1)b$ Small mass: $t_4^+(0.75, 0.25)$ Large mass: $t_4^+(6.5, 5)$ Weak circadian: $5.0 + 1.5 \cos x + 1.5 \sin x$ Strong circadian: $4.0 + 2.0 \cos x + 2.0 \sin x = 2\pi/1440$
Pulse duration	$\sigma_{p,k}^2$	Fixed effect σ_p^2	Random effect $\sigma_{p,k}^2 \sim N^+(m_d, s_d^2)$	Fixed effect ^d δ	Random effect $\sigma_{p,k}^2 \sim N^+(m_d, s_d^2)$	Random effect $t_4^+(5, 1)b$
Pulse location	τ_k	Fixed effect	Random effect $\tau_k \sim 3\text{rd order stat. of } (3N_p + 2) \text{ i.i.d. } Unif(0, T) \text{ r.v.}$	Fixed effect	Random effect $\tau_k \sim 3\text{rd order stat. of } (3N_p + 2) \text{ i.i.d. } Unif(0, T) \text{ r.v.}$	Renewal Gamma with mean = 120 min, var = 600 min ²
Elimination (half-life)	h	Constant	Constant	Constant	Constant	$hl = 45 \text{ min}$
Model error variance	σ_e^2	Constant	Constant	Constant	Constant	Ref: 0.02 Low: 0.005 High: 0.10

^aIn the Bayesian framework, these are the priors on the parameters.

^b $t_4^+(m, v) = t$ -distribution with 4 degrees of freedom; truncated below 0; m =mean, v =var.

^c $N^+(m, s_m^2) =$ Gaussian distribution truncated below 0; m_m =mean, s_m^2 =var.

^dThe PULSE model defines duration with an infusion rate.

Larger values for the infusion rate correlate with larger values for $\sigma_{p,k}^2$ in the other models.

Table II

Parameter combinations in the 10 simulation models.

Model	Model Error	Baseline	Average Pulse Size
Reference	0.02	2.6	3.5
Changes in signal-to-noise ratio			
Low Error	0.005	2.6	3.5
High Error	0.10	2.6	3.5
Large Pulses	0.02	2.6	6.5
Small Pulses	0.02	2.6	0.75
Variation in circadian patterns			
Weak circadian baseline	0.02	$B(t) = 1.5 + 1.0 \cos x + 1.0 \sin x$	3.5
Strong circadian baseline	0.02	$B(t) = 3.0 + 2.0 \cos x + 2.0 \sin x$	3.5
Weak circadian pulse mass	0.02	2.6	$\mu_{p,k}(x) = 5 + 1.5 \cos x + 1.5 \sin x$
Strong circadian pulse mass	0.02	2.6	$\mu_{p,k}(x) = 4 + 2.0 \cos x + 2.0 \sin x$
Both circadian	0.02	$B(t) = 3.0 + 2.0 \cos x + 2.0 \sin x$	$\mu_{p,k}(x) = 4 + 2.0 \cos x + 2.0 \sin x$

$x = 2\pi t/1440$.

Table III

Coverage of 95% Confidence (credible) intervals for all models and methods.

Parameter	Algorithm	Reference	Effect of changes in signal-to-noise ratio				Effect of circadian patterns					
			Model Error		Pulse Size		Circadian Pulse Mass		Circadian Baseline		Both Circadian	
			Low	High	Large	Small	Strong	Weak	Strong	Weak		
Half-life	AUTODECONV	–	–	–	–	–	–	–	–	–	–	–
	BayesDeconv	86	93	69	92	46	11	36	85	92	6	6
	PULSE	39	62	10	40	24	43	47	20	25	15	15
	BayesDeconvChbl	86	93	68	94	62	64	64	84	91	79	79
Average Pulse Mass	AUTODECONV	–	–	–	–	–	–	–	–	–	–	–
	BayesDeconv	98	99	97	92	99	99	78	99	98	80	80
	PULSE	92	87	96	93	18	92	86	97	90	97	97
	BayesDeconvChbl	100	99	100	97	63	100	99	100	100	99	99