Discrete Heart Rate Values or Continuous Streams? Representation, Variability, and Meaningful Use of Vital Sign Data

Keith Feldman, PhD, Annie J. Rohan, PhD, RN, FAAN, FAANP, Nitesh V. Chawla, PhD

Documentation and review of patient heart rate are a fundamental process across a myriad of clinical settings. While historically recorded manually, bedside monitors now provide for the automated collection of such data. Despite the availability of continuous streaming data, patients' charts continue to reflect only a subset of this information as snapshots recorded throughout a hospitalization. Over the past decade, prominent works have explored the implications of such practices and established fundamental differences in the alignment of discrete charted vitals and steaming data captured by monitoring systems. Limited work has examined the temporal properties of these differences, how they manifest, and their relation to clinical applications. The work presented in this article addresses this disparity, providing evidence that differences between charting techniques extend to measures of variability. Our results demonstrate how variability manifests with respect to temporal elements of charting timing and how it can facilitate personalized care by contextualizing deviations in magnitude. This work also highlights the utility of variability metrics with relation to clinical measures including associations to severity scores and a case study utilizing complex variability metrics derived from the complete set of monitor data.

KEY WORDS: Heart rate, Meaningful use, MIMIC database, Monitoring, SOFA scores

IMPLICATIONS FOR CLINICAL PRACTICE

- The method of heart rate documentation (provider-selected "charted" values vs computer-selected values) should be considered when interpreting heart rate data in the ICU.
- Heart rate documentation practices that require the bedside provider to "select a representative value" may diminish representation of heart rate variability.
- Changes in heart rate SD may be a more sensitive indicator of patient deterioration than changes in discrete heart rate values.
- Heart rate interpretation should be based upon data that represent as accurate a picture of heart rate variability as possible to optimize detection of patient deterioration.

The observation of hospitalized patients represents an integral component of their overall care. Although clinicians may formalize these observations through a variety of subjective descriptions and objective assessments, one of the most prominent remains the collection of vital sign data.[1] The documentation and review of patient vital signs are a fundamental process across a myriad of clinical settings, aiding in the discovery of symptoms reflective of patient deterioration, as well as identifying patients at risk of adverse events.[2–5] The information gathered by monitoring vital signs has been long utilized for numerous other clinical tasks: ranging from high-level objectives such as setting and monitoring treatment goals, or improving physician-patient communication, to specific calculations of well-established early warning scoring systems and severity score measures.[6–10] As personalized and precision healthcare initiatives continue to rise, there exists a strong belief that vital sign measures present an opportunity not only to assess conditions, but also to improve patient care.

While historically collected and recorded in a manual fashion, the advent of bedside monitors now provides for the automated collection of vital sign data, in particular that of heart rate, where measures of stability/variability have repeatedly been to be valuable metrics in assessing patient outcomes.[11–13] Continuous monitoring devices and telemetry systems now link directly to a patient's electronic medical record and can autopopulate this document. Yet, despite the continuous

Author Affiliations: Department of Computer Science and Engineering and iCeNSA, University of Notre Dame, IN (Drs Feldman and Chawla); SUNY Downstate Health Sciences University, College of Nursing, Brooklyn, NY (Dr Rohan).

This work was supported in part by the National Science Foundation (grant IIS-1447795).

The authors have disclosed that they have no significant relationships with, or financial interest in, any commercial companies pertaining to this article.

The data utilized for this study were drawn from one of the largest publicly available collections of clinical data for ICU patients, the MIMIC (Multiparameter Intelligent Monitoring in Intensive Care) database. Data sets that are made available to the public and do not require special permissions to access the data are considered publicly available data sets. Analysis studies of publicly available data sets are exempt from requiring ethical approval.

Corresponding author: Annie J. Rohan, PhD, RN, FAAN, FAANP, SUNY Downstate Health Sciences University, College of Nursing, 450 Clarkson Ave, Brooklyn, NY 11203 (annie.j.rohan@gmail.com).

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site (www.cinjournal.com).

DOI: 10.1097/CIN.0000000000000728

stream of data made available through these monitors; a patient's chart often reflects only a subset of this information with a series of heart rate snapshots recorded throughout a hospitalization. Over the past decade, this disparity has gained attention, with prominent works establishing differences in the magnitude and alignment between the discrete set of charted vitals and the full steam of data captured by monitoring systems.[14] However, limited work explores the implications of such differences.

The work presented in this article addresses exactly this. Utilizing a large repository set of secondary-use clinical data extracted from electronic medical record records from acute care patients on telemetry, this study sets out to demonstrate that measures of alignment in magnitude represent only one aspect in the broader discussion around utilizing snapshots of a patient's data, specifically illustrating how the attribute of variability differs between two common charting sources, manually charted values and those automatically recorded from beside monitors. In doing so, we present a quantitative framework to compare measures of alignment and variability between patients' data throughout admission to the ICU, highlighting implications of identified changes with respect to common clinical metrics such as severity scores, and ultimately explore the utility of utilizing the complete set of waveform data to differentiate between mortality outcomes of patients in the ICU.

DATA

The data utilized for this study were drawn from one of the largest publicly available collections of clinical data for patients cared for in the ICU, the MIMIC (Multiparameter Intelligent Monitoring in Intensive Care) database.[15] The database was populated over an 11-year period (2001–2012) with patient records from the Beth Israel Deaconess Medical Center, in Boston, MA. The center comprised both a 620-bed tertiary academic medical center and a 77-bed level I trauma center. For this work, we utilized a specialized subset of the database comprised of patients whose clinical records have been matched with physiologic and vital sign data (known as waveforms) from a patient's bedside monitor. These monitors continually record several physiologic signals, such as patient's arterial blood pressure, respiration rate, oxygen saturation, and heart rate. In total, the matched records provided by the MIMIC database contained 5266 waveforms, representing 2809 unique patients. Overall, heart rate was well represented, with 4365 of the 5266 total waveforms producing at least one overlapping time period for evaluation.

It is important to note that, resulting from the natural course of care, patients may be disconnected from a monitor for an extended period, for example, to undergo procedures. As a result, gaps longer than 1 hour were split into two records, allowing multiple monitor-recorded segments to be associated with a patient over the course of their admission. Recognizing we cannot control for changes in patient's condition over this time, all analyses in this work treat each waveform as a unique instance.

A fundamental premise of this work relies on the contrast between monitor-recorded and charted data, so it is important to distinguish the processes by which each of these data types is collected. While monitor data are collected as extracted directly from a patient's waveform, the process of recording values in a patient's chart is somewhat more complex. Discussions with the MIMIC team yielded a workflow that, while heart rate data recorded into a patient's chart are often collected by electronic monitoring systems, they require a nurse to select the specific measurement from the series or overwrite the value with manually collected data. Nurses may also insert additional discrete values to the set of monitor data. For clarity, these discreetly curated heart rate values (whether validated, overwritten, or inserted) will be henceforth referred to as *charted* values, whereas data extracted directly from the waveform will be referred to as automatically *recorded* collected data.

PRELIMINARY STEPS

Creating an Analytical Framework

The ability to execute the analyses presented throughout this work required more than an availability of data. It required a structured framework to extract, match, and compare charted heart rates values with those collected by monitoring systems. However, this undertaking represented a nontrivial exercise, as charted values can occur at nonstandardized time points, and reoccur at nonstandard frequencies (ie, 15, 30, 45, 60+ minutes apart). As a result, to accomplish this task, we created a representation deemed an interval that allowed for standardized comparisons and analysis between overlapping charted heart rate values and monitor data collected over the same time period.

Interval Creation

To begin, raw heart rate values were extracted as a time series from the monitor data of a patient, functionality provided by the PhysioBank Waveform Database package.[16] Next, the values and times of charted heart rate data were overlaid across the waveform data. While it is possible patient vitals are recorded prior to being connected to a bedside monitor, we focus on the timeframe for which they overlap and thus can be directly compared.

Intervals were then created in an iterative fashion beginning with the earliest charted value. For each iteration, the subset of waveform data occurring after the current charted value up to (and including) the next charted heart rate was extracted. The current charting value was then updated to represent the end point of the prior interval, and the process

is repeated until the last charted value in the overlapping segment is reached. To prevent biasing the data, any automatically recorded data recorded prior to the first charted value are discarded. Without a prior charted value to act as a reference point, the interval would not truly capture data between clinical evaluations and may provide a biased measure of alignment. Thus, the first interval is extracted between the first and second charted values.

Feature Extraction

Next, each interval was utilized to create a feature vector used for analysis. First, the charted heart rate associated with the end time point in the interval pair was added to a vector. Next, each interval was linked to an extensive set of features. These included a set of clinical features, such as the patient's length of stay in the ICU; demographic features, such as the patient's sex and age; and derived features, such as the elapsed time since the previous charted value (representing the overall length of time captured by the interval). Finally, the feature vector was associated with summary statistics calculated across the subset of waveform data captured by the interval. These included the mean, 25th and 75th percentile heart rate values, and the corresponding SD.

Unfortunately, automatically recorded data are inherently noisy, containing artifacts such as intermittent spikes and flatlines that can be caused by natural phenomena, such as hiccups or a lead falling off.[17,18] Yet, their presence has the potential to artificially skew the data distribution and bias analysis. As a result, each subset of waveform data underwent outlier thresholding utilizing the median absolute deviation (MAD).[19] Outlier detection was done on an interval-by-interval basis, so as not to bias data for patients whose heart rate changed drastically from their admission to discharge, a distinct possibility for patients admitted to ICUs. For all analyses, outliers are identified as those values at a threshold of ±3 MAD.

Interval Example

A demonstration of interval creation is shown in Figure 1 for a patient with six charted heart rate values overlapping the timeframe where their heart rate was recorded by a bedside monitor, recorded hourly from 11:00 AM to 4:00 PM. Figure 1A presents the raw waveform data, while Figure 1B

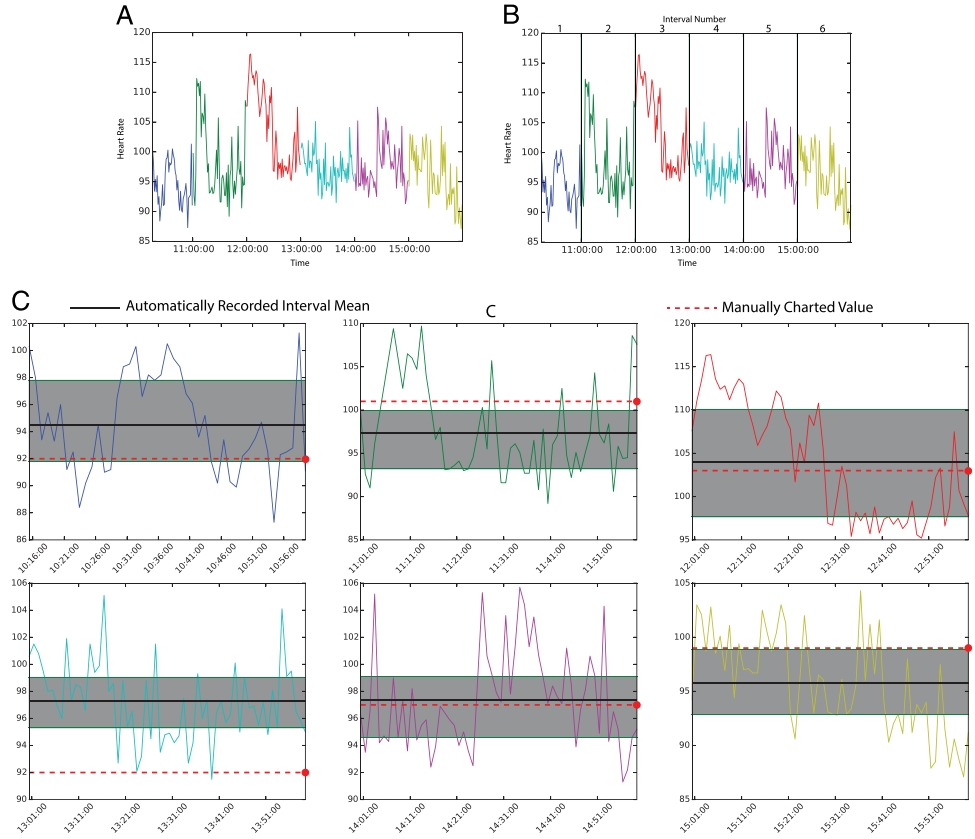

FIGURE 1. Overview of interval creation process. A, Baseline data: example of an automatically recorded heart-rate waveform for a patient's hospitalization. B, Segmentation of interval windows: each line segment represents the period of monitor data between instances of heart rate values recorded in patient charts. C, Interval details: subplots representing derived summary statistics for each interval extracted from the complete waveform.

highlights the six unique interval periods between each of the six charted heart rates. Finally, Figure 1C presents a visual representation of a feature vector corresponding to an extracted interval. Each subplot presents a snapshot of the automatically recorded data contained within each interval. The point on the far right denotes the charted value associated with the time interval. Although the charted value is collected at the end of each interval, a line representing this value was extended back across the subplot for the reader's convenience. Overlaid on each are summary statistics including the mean waveform heart rate and a shaded area representing 1 SD above and below the derived mean.

Data Cleaning

Notably, data collected in EHRs (and subsequently the MIMIC database) are not collected for analytical purposes, but as a means to facilitate existing clinical purposes of both care and billing. As a result, there exists a substantial amount of noise and potential for bias associated with the data when focused on a specific research question. Thus, to improve the reliability of the resulting analyses, once the interval feature vectors were extracted, an extensive set of data-cleaning steps was performed.

Broadly, this cleaning represented four distinct steps. First, to account for potential variability in telemetry usage and practices between different care units, any patient whose admission spanned more than one ICU was removed.[20] Second, to account for potential confounding effects of variability-reducing factors, such as implanted pacemakers, patients whose automatically recorded data produced an interval with 0 variations were removed. Third, to ensure sufficient data per patient for analysis, patients with the lowest percentage of overlapping charted and waveform data were removed; capturing those at or below the 25th percentile (less than 17 overlapping values). Fourth, as the waveform and charted data were matched by an automated process, we removed patients whose two data sources differed significantly over the course of an admission, notably those whose charted values were outside 3 SDs of their automatically recorded data across more than 50% of their matched values. Additional details around this criterion can be found in the e-supplement. It is important to note that although intended to reduce noise and incorrect data, removing patients well outside the norm may itself introduce bias into this analysis. However, should the underlying data prove correct, those patients misaligned at such a high degree for the majority of time likely represent a fundamentally different population outside the scope of this study. As such, the limitation of this work to focus on typical patients (as empirically defined by our data) presents a more rigorous and uniform study population on which to draw conclusions.

A comprehensive review of the design decisions for each step can be found in Supplemental Digital Content 1 (http://links.lww.com/CIN/A87), and for completeness, the corresponding number of patients, waveform segments, and extracted intervals remaining after each of the cleaning steps can be found in Table 1.

Data Partitioning

Prior to addressing the specifics of analysis, it is important to remember the data are drawn across a real-world distribution of ICU patients. As a result, we utilize a set of partitioning criteria to help control for confounding attributes. In particular we create 96 unique partitions that capture variance across age, sex, length of stay (as a proxy for severity), and the frequency between charting intervals (as the number of data captured can influence the summary statistics). Justification for partition criteria is provided in Supplemental Digital Content 1 (http://links.lww.com/CIN/A87).

METHODS

The analyses presented in this article examine three aspects of variability: (1) identification of variability as a factor differentiating manually charted and automatically recorded heart rate data; (2) study of how variability can contextualize established differences in magnitude; and (3) case studies exploring the potential clinical application using such knowledge. Details of each are provided in the sections below.

Identifying and Exploring Variability as a Differentiating Factor

The initial analyses formalize the existence of statistical differences in the variability of heart rate data captured by charted values and those extracted from a monitor waveform. We then explored the way such differences may arise from a bias of charted data to reflect specific temporal periods of a waveform.

Table 1. Cohort Selection: Overview of the Resulting Data for Each Data-Cleaning Step (n = 5266 Total Waveforms)

Filtering Criteria	Total Sample Size (n)		
	No. of Unique Patients	No. of Waveforms	No. of Intervals
Multiple units	2727	4267	251 196
Variability	2308	3440	179 091
Overlap	1947	2605	171 911
Alignment	1906	2164	168 187

Measures of Variability and Central Tendency

We began with a direct analysis quantifying alignment between the heart rate values recorded by each method utilizing two measures of variability (SD, range). Further, in line with previous work, we included two measures of central tendency (mean, median). To do so, within a single partition, all intervals extracted from the same waveform were grouped together, providing internal consistency within the charting of an individual. Next, each of the four metrics was computed across the charted and automatically recorded for each individual.

For the charted values, heart rates associated with each interval were used directly, while for monitor data, the mean waveform heart rate in each interval was employed. Together, these values act as repeated measures of individuals within the partition, and for each metric, a paired t test was performed across to identify partitions where the patient's charted and monitor data significantly differed.

Temporal Considerations

Having identified differences in variability, we next highlight a way in which they manifest, examining how intermittent values of monitor data relate to the value recorded in a patient's chart. To accomplish this task, intervals were regenerated at far more granular level. Rather than extracting the waveform data between the charted end points as a single continuous stream, data were broken into segments, each representing 10% of the time captured by the interval.

Utilizing these modified interval segments, we then identified which percentile the mean heart rate was closest to the charted value recorded at the end of the full interval. Ultimately, a χ^2 goodness-of-fit test was performed to assess whether the distribution of closest segments was equal among the percentiles, or there existed a temporal bias to charted values.

Variability as a means to Contextualize Measures of Distance and Alignment

With the understanding that variability represents another differentiating characteristic between charted and monitor-recorded data, we next sought to illustrate how measures of variability can provide deeper insights into the differences in magnitude documented by prior comparisons of the two data sources. The investigation of such centered around two primary measures, distance, and alignment.

Distance

To explore variability with respect to deviations magnitude, we first computed perhaps the most prominent comparative measure between the two data sources: directly calculating the distance, measured in beats per minute (BPM), between the charted heart rate and the mean heart rate extracted from the monitor data. The mean distance was calculated for each interval within a distinct partition, and ultimately a set of summary statistics was calculated as an average across all 96 partitions. Further, as charted values can be either above or below the monitor average, this results in a centering of the summary statistics around an average value of 0. To more accurately capture distance, we replicate the analysis using the absolute difference between the two values.

Of note, by computing data on an interval-by-interval basis, a potential source of bias remained, as the heart rates of a patient may be an inherent more/less difficult to chart based on factors of their condition. Thus, depending on the number of intervals extracted, their data may skew the results for difficult-to-chart patients. As such, we repeated the analysis, averaging the data for each patient within a partition prior to calculating the partition average and subsequently the overall summary statistics.

Alignment

Although distance metrics provide an objective measure of coherence between charted and monitor vitals, we next demonstrate how the use of variability can provide improved context to the magnitude of these deviations. Utilizing the SD of the interval waveform data, we were able compute the percentage of intervals for which the charted values fall within the bounds of 1, 2, and 3 SDs above or below the waveform mean, helping to qualify the magnitude of the misalignment with respect to each patient's own interinterval variability.

Case Studies of Variability in Common Clinical Applications

Thus far, each analysis has focused on establishing variability as a differentiating factor between the two data sources. However, such differences do not necessitate a relevance to clinical practice. In this final section, we provide two clear examples of such: first, comparing the association of variability extracted from charted data and monitor data to a patient's maximal Sequential Organ Failure Assessment (SOFA) score; and second, providing a case study demonstrating how monitor data collected over the course of the patient stay can be used to derive a metric to differentiate mortality outcomes of patients.

Correlational Comparison

Within our first clinical analysis, we aimed to answer the question: Does the variability captured by charted data provide the same analytic capability as that captured by monitor? To address this, we performed two correlation analyses, between a patient's maximal SOFA score and (1) SD of heart rates extracted from monitor data and (2) SD of the charted HR values. Given the acute care setting of the patient population and the likely upward bias of patient severity scores,

each analysis was performed utilizing the nonparametric Spearman correlation. The SOFA score was selected as there exists previous literature associating its magnitude to changes in heart rate variability.[21,22]

It should be noted that as this analysis utilizes a single external data element (maximal SOFA score) for each patient, we must modify our partitioning criteria to ensure that the independence of samples is maintained. As a single patient can be split across multiple partitions, the frequency criteria were removed, and the analysis was performed across a reduced set of partitions (24 vs the full 96).

The Value of Variability in Waveform Data

Continuing to evaluate the clinical implications of variability, our final analysis provides a case study highlighting the potential in fully utilizing measures of variability available from the complete set of waveform data, in particular demonstrating how granular measures of variability change can differentiate the mortality of patients admitted to the ICU.

In a similar fashion to the temporal analysis, we begin by segmenting the waveform data into 10 percentiles. However, as we are no longer comparing charted and automatically collected waveform data, this analysis could utilize the total length of time captured within each waveform. Yet, the raw heart rate data collected from monitor systems cannot be used directly to compare individuals, as it has been well established that demographics such as age and gender have an effect on heart rate variability.[23,24] To account for this, we moved to a derived measure of change. To do so, the SDs of waveform data within each percentile were computed. Then, the difference in consecutive percentiles was computed to create a new temporal series, henceforth referred to as a "first-order change." An example of the process can be found in Figure 2A.

Unfortunately, as a result of directionality, as patients who are successfully discharged can appear quite different under this metric based on their admission condition. For example, a patient may be admitted to an ICU with highly unstable vitals, which are stabilized to normal levels over the course of their admission prior to a successful discharge, while another patient may be admitted in a comatose state with extremely low viability in their vitals, where over the course of their stay, viability steadily increased to normal levels prior to a successful discharge.

To capture this level of detail, we move to a difference-in-differences approach. In a similar fashion, the difference was computed between the computed differences (first-order differences) in the SD of consecutive percentiles, henceforth referred to as a "second-order" difference. The process is shown in Figure 2B. Moving to the difference-in-differences approach provides a derived measure for the consistency of increasing or decreasing variability, a particularly valuable clinical attribute when evaluating a patient's condition. Through this metric, a flatline indicates a steady increase or decrease in variability between consecutive percentiles of the patient's stay, where a value of 0 represents a change in variability of equal value over each percentile. Thus, we postulated successfully discharged patients will result in overall lower values.

To statistically test for such a difference, we integrated examined across the series of second-order differences, comparing if the value for patients who were successfully discharged differs from that of patients who die during their hospitalization with a Kruskal-Wallis test (a nonparametric analysis of variance). It should be noted the final percentile (90%–100%) was excluded from the calculated integral to protect against the potential bias introduced by end-of-life data artifacts.

RESULTS

Results corresponding to each analysis are provided within the respective subsections, while a broader dialogue of the implications of these results can be found in the discussion section to follow.

Identifying and Exploring Variability as a Differentiating Factor

Measures of Variability and Central Tendency

Quantifying the manifestation of deviations between the two charting methods, the aggregates (count and percentage) of the statistically significant ($P \leq .05$) results from the paired t test performed between the respective aggregate measures between charted and automatically recorded data for each of the 96 distinct partitions were as follows: We note only a small percentage of partitions indicated differences between measures of central tendency, with n = 26 (27.08%) and n = 17 (17.71%) reaching significance for the mean and

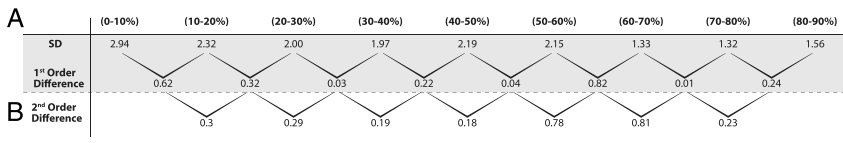

FIGURE 2. Demonstration of a first- and second-order difference for SD values of one interval. A, First-order differences computed between consecutive raw SD values of each percentile. B, Second-order differences computed between consecutive pairs of first-order differences.

Table 2. Closest Percentile: Number of Intervals for Which the Respective Percentile of Monitor-Recorded Data Was Closest to the Associated Charted Value

Percentile	0%–10%	10%–20%	20%–30%	30%–40%	40%–50%	50%–60%	60%–70%	70%–80%	80%–90%	90%–100%
n	16 488	11 473	10 871	10 883	11 366	12 187	13 403	15 944	19 686	33 599

median, respectively. However, a majority of partitions differed with respect to variability measures; n = 63 (65.63%) and n = 67 (69.79%) for SD and range, respectively.

Temporal Considerations

Looking next, we attempt to identify characteristics of intervals that may result in the misalignment of charted and automatically collected data. Table 2 details the percentile segment for which the associated mean heart rate was closest to the charted value recorded at the end of each interval. The resulting χ^2 goodness of fit resulted in $P < .05$, indicating that the temporal position of the closest percentile was not equally distributed across the interval segments.

Variability as a Means to Contextualize Measures of Distance and Alignment

Distance

Table 3 provides an overview of the raw and absolute differences in heart rate values in BPM. For clarity, raw differences capture both the positive (charted values greater than the mean heart rate calculated from the associated waveform data) and negative (charted values less than the mean heart rate calculated from associated interval of waveform data), whereas the absolute difference represents the absolute distance between the two measures without regard to directionality.

Alignment

Expanding the analysis of distance to include measures of variability, Table 4 provides a measure of alignment indicating the average percentage of charted data that fell within 1, 2, and 3 SDs of the waveform mean.

As noted in Methods, the measure of charting alignment shown in Table 4 is in fact an average over the partitions. To ensure that these results are representative of the data as a whole, we have provided Supplemental Digital Content 1 (http://links.lww.com/CIN/A87), which presents the mean alignment for each partition with SD represented as error bars. The stability of alignment measures at the 3 SD thresholds across all 96 partitions clearly demonstrates the partitions' ability to provide a high level of reliability across various confounding factors.

Case Studies of Variability in Common Clinical Applications

Correlational Comparison

Turning next to the correlation with the SOFA severity score, we find that the measures of variability collected from automatically collected data may capture information not available from the series of charted heart rates. Noting variability extracted from waveform heart rate provides a statistically significant correlation across 16 of 24 partitions, averaging a correlation value of −0.284. Such consistent correlation is expected as there are known associations between such variability and SOFA scores. However, the correlations to the SD computed across charted heart rate values were statistically correlated to far fewer (six) partitions, with an average correlation value of 0.22.

The Value of Variability in Waveform Data

Moving to our final analysis, an example of the raw standard division and first- and second-order differences for three sample partitions can be seen in Figure 3. (The complete set of plots for all partitions can be found in Supplemental Digital Content 1, http://links.lww.com/CIN/A87.) With respect to the derived metric, we find integral values for the series of second-order differences for patients who were successfully discharged was 8.83, whereas patients who die averaged 11.14. This difference was found to be statistically significant at $P < .05$.

DISCUSSION

The advent of continuous bedside monitoring systems has fundamentally changed the manner in which heart rate data are collected and recorded. In an effort to seamlessly transition new technology into existing clinical processes, recorded data

Table 3. Charting Difference: Difference Between Monitor-Recorded Data and Charted Heart Rate in BPM

		Mean	SD	Min	25%	50%	75%	Max
Interval	Raw difference	0.663	0.948	−1.638	0.213	0.412	0.938	4.914
	Absolute difference	4.725	0.886	2.795	4.150	4.498	5.216	7.400
Patient grouped	Raw difference	0.669	1.022	−1.654	0.210	0.480	1.064	4.078
	Absolute difference	4.697	0.824	2.767	4.128	4.563	5.020	7.757

Table 4. Charting Alignment: Estimation of Distance Between Monitor-Recorded Data and Charted Values When Considering Variability of Each Interval, Computed as a Percentage of Intervals In-Range

		Mean	SD	Min	25%	50%	75%	Max
Interval	1 SD	44.209	7.295	31.579	40.264	43.432	46.582	77.778
	2 SD	72.083	7.922	56.471	66.651	72.133	76.853	95.833
	3 SD	83.554	7.237	67.161	79.696	84.905	87.759	100.000
Patient grouped	1 SD	42.010	2.684	34.554	40.440	41.720	43.880	49.271
	2 SD	69.647	3.291	59.919	67.005	69.665	72.359	77.302
	3 SD	81.648	2.740	75.480	79.636	81.591	83.322	89.065

have remained largely unchanged in appearance, and patient charts continue to reflect a series of discrete values. Utilizing of structured internal framework to align and compare values of these charted (validated, overwritten, or inserted) heart rate data with waveform streams, the analyses presented in this article establish quantifiable differences in variability between the two sources. Our work then illustrates how such knowledge can contextualize magnitude differences seen in prior literature and highlights novel associations between variability and clinical measures of patient severity and mortality. A discussion of the results for each analysis, as outlined in their respective Methods sections, can be found in the corresponding sections to follow.

Identifying and Exploring Variability as a Differentiating Factor

Measures of Variability and Central Tendency

We began with a comparative analysis of central tendency (mean, median), capturing the alignment of the average

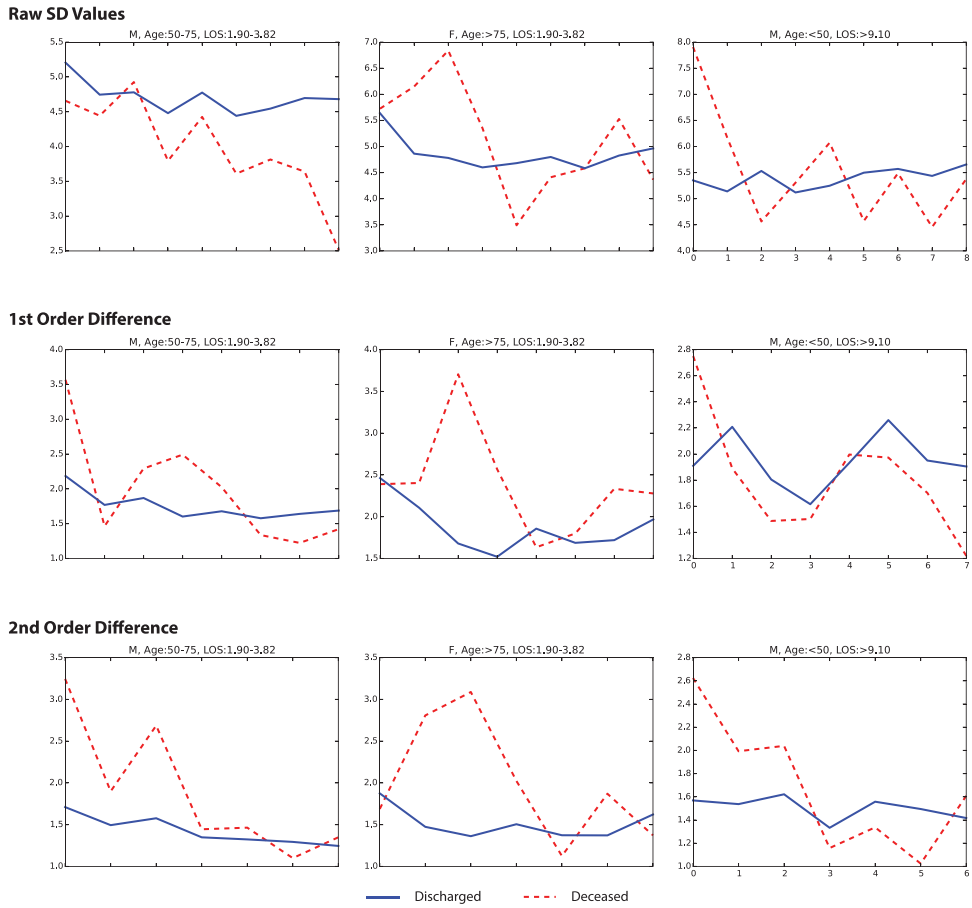

FIGURE 3. Deriving metrics from waveform data: average raw SD, first- and second-order differences of deceased versus discharged patients in three highlighted partitions. Each x tick represents 10% of the waveform data, from 0% to 90%.

automatically recorded data and a patient's charted heart rate within each partition. In line with prior works, our results reaffirm that there exists a misalignment in magnitude between the two data sources.[24–28] However, we note that the differences in many partitions did not reach significance. Conversely, however, with respect to measures of variability (SD and range), we find a much more widespread occurrence of differences across the values of the waveform data and the temporal series of charted values. In total, over 50% partitions were found to have statistically significant deviations across both metrics.

We believe that these results capture a phenomenon known as *smoothing*. That is, the charted data were significantly less likely to reflect values outside of normal physiologic parameters compared to automatically recorded data. This expression of such smoothing is consistent with several prior works. In their work, Reich et al[29] examined discrepancies between handwritten (manual) and computerized (automatic) anesthesia records, noting that extreme values were rarely noted in handwritten records. More recently, Sapo and colleagues[14] examined discrepancies between manually and automatically charted heart rate values and the phenomenon of *smoothing* in a neurocritical care. Despite overall agreement in mean heart rate values between manual and automatic values, researchers identified that manual data had fewer extreme values.

While there are several possible explanations for these findings, prior work suggests it may be an artifact that practitioners, within the broad parameters of physiologic truths, avoid or prevent the documentation of abnormal values. Work by Taenzer and colleagues[25] describe the existence of consistent inflation (closer to normal physiologic parameters) of manually recorded oxygen saturation values when compared to values recorded by monitor systems.

Regardless of explanation, the effect of smoothing is nonetheless a serious one in terms of capturing variability in heart rate values for critical care patients.[30–33]

Temporal Considerations

To better understand how the identified differences manifest within the temporal patterns of automatically recorded and charted values, we next identified the percentile of each waveform in which the mean heart rate was closest to the charted value for a given interval. A χ^2 test indicated that the associations were not equal across the waveform percentiles, and at an observational level, it appears that there may be a temporal bias. As the mean value of the 90th to 100th percentile was closest to the manually charted value in almost twice as many cases when compared to nine earlier percentiles (Table 2). In turn, this further suggests charted values may not be wholly reflective of the data captured between charting instances, but rather reflective of a smaller subset of patients' vital sign data at a time point close to when the observation is taken.

Variability as a Means to Contextualize Measures of Distance and Alignment

Distance and Alignment

Having established differences between charting methods, our next analysis provided a clear example of how such information could better quantify their degree of alignment, or lack of it. Drawing on the raw/absolute measures of distance from Table 3, we find that charted and automatically recorded heart rate values differ by only a few BPMs. However, these values do not tell the whole story. Utilizing the metrics of variability, we note for a large percentage of intervals that this distance represents of 2 to 3 SDs away from the mean of the waveform data for the same period (Table 4). To frame this observation from a clinical perspective, although these variations in BPM may not seem drastic, they represent deviations far outside a patient's normal heart rate for the period in question.

Case Studies of Variability in Common Clinical Applications

Differences in variability do not merely represent an abstract concept in the comparison between two measurement techniques. Rather, these differences have implications to existing clinical measures. Accordingly, the final component of this article focuses on the examination of how the characteristics of different data collection methods can impact the ability to accurately assess the patient condition.

Correlational Comparison

Focusing first on a common risk score, we investigate the correlation between a patient's maximal SOFA score and the variability (as measured by SD) in heart rate across both charting methods. In doing so, our results make clear measures of variability drawn from automatically collected data to provide a stronger and far more consistent correlation effect across the set of patients partitions when compared to the variability that can be extracted from the series of charted data alone. This result supports the premise that information available within the subset of charted heart rates may at its core lack either the granularity or another more complex property necessary to capture increasingly nuanced aspects of clinical phenomena.

It is interesting to note that the sign of the average correlation value between the maximal SOFA scores for the charted values was the inverse of that of the waveform SD. However, such a phenomenon can be explained with an understanding of what the data represent. While the SD of the waveform data represents the average variability for each

interval, the measure of variability available from charted data is averaged between intervals. Thus, while an increase of intrainterval variation may have a negative relationship to SOFA scores, interinterval variation may indicate a rise in SOFA value. These differences again support the idea that the data captured by each charting method may be different at a fundamental level.

The Value of Variability in Waveform

Our last analysis provides a case study of the clinical value in utilizing the intermittent measures of variability available through the complete set of waveform data. This example focuses on a core clinical application in the recognition of patient deterioration and death.

Three studies have previously found heart rate variability to be consistently and reproducibly altered in certain disease states, and the degree of alteration to be prognostic of illness severity.[34–36] As such, this case study allowed for the creation of a complex metric that could statistically differentiate mortality outcomes in the ICU built off the characteristics of variability uncovered in the earlier analyses. Specifically, findings that automatically recorded data presented an improved ability to capture granular SD as well as knowledge-charted data may capture a biased subset of data across percentiles of a charting interval.

Of note, although we did not create an optimal predictor of mortality (a model that would require consideration of factors such as admission condition, procedures, medications, etc), this case study highlights the importance of understanding that each of the processes by which heart rates are recorded offers its own strengths and limitations that are inherently present in the data they generate.

Finally, it is important to highlight that measures of variability directly extracted from waveform data were not enough to differentiate mortality patterns among patient groups. Rather, due to the varied reasons patients are admitted to the ICU, differentiation between outcomes was obtainable only when variability was calculated as a difference-in-differences approach between the percentiles of a hospitalization. The need to extend the raw variability measures serves as an important reminder that, regardless of technology, analytics aimed to provide information for clinicians requires not only the collection of frequent and detailed heart rate data, but also insight into the clinical scenarios for which these values are being collected and applied.

CONCLUSION

The collection, recording, and monitoring of heart rate data are unlikely to change as an integral component of the hospital care process. As technology advances, an opportunity exists to advance the conversation around meaningful use of technology and its role in data-driven clinical practice.

This article highlights an example of exactly this, demonstrating how advancements in telemetry offer a means to better understand the data we collect as part of routine patient monitoring in acute settings and the implications of utilizing discrete snapshots of a patient's heart rate data. With a systemic evaluation, this work offers compelling evidence that differences between discretely charted data and continuous automatically recorded waveforms extend beyond accepted measures of alignment and magnitude, to measures of variability. In doing so, we highlight how this temporal variability can impact clinical processes, with respect to the association with established risk scores, as well as the broader goal to derive complex automated metrics in predictive scenarios.

The results presented here illustrate a clear need to provide context around the data we collect, as magnitude alone fails to capture valuable information between charting events as even in the variability between discrete events, it is clear that some degree of variation remains masked. However, the speed of comprehending a singular number rather than a complete multihour waveform is valuable. Thus, research such as this, identifying differentiating factors between methodologies, allows the broader community a chance to consider future work drawing on the strengths of each to provide a more comprehensive view of a patient. Perhaps then simple additional recording of SD at the time of heart rate charting—a value that can be selected for display on most monitors—can improve the representation of heart rate variability between chartings.

More broadly, this work lays the foundation on which to focus on the notion of context from which data are collected and utilized. These additional factors can provide insights into the existence of underlying biases. In the long term, a richer composite can enhance our overall understanding of the relationship between documented heart rates and a patient's condition for practice or research, aiding in the growing mission to understand how the multitude of data becoming available today can provide improvements for patient care tomorrow.

References

1. Mok WQ, Wang W, Liaw SY. Vital signs monitoring to detect patient deterioration: An integrative literature review. *International Journal of Nursing practice*. 2015;21(suppl 2): 91–98.
2. Belle A, Ansari S, Spadafore M, et al. Signal processing approach for detection of hemodynamic instability before decompensation. *PLoS One*. 2016;11: e0148544.
3. Liu NT, Holcomb JB, Wade CE, Darrah MI, Salinas J. Utility of vital signs, heart rate variability and complexity, and machine learning for identifying the need for lifesaving interventions in trauma patients. *Shock*. 2014;42: 108–114.
4. Fairchild KD. Predictive monitoring for early detection of sepsis in neonatal ICU patients. *Current Opinion in Pediatrics*. 2013;25(2): 172–179.
5. Günther A, Salzmann I, Nowack S, et al. Heart rate variability—a potential early marker of sub-acute post-stroke infections. *Acta Neurol Scand*. 2012;126: 189–196.

6. Gavish B, Bursztyn M. Blood pressure and heart period variability ratios derived from 24-h ambulatory measurements are predictors of all-cause mortality. *Journal of Hypertension*. 2015;33(3): 491–498; discussion 498.
7. Pei Z, Shi M, Guo J, Shen B. Heart rate variability based prediction of personalized drug therapeutic response: the present status and perspectives. *Current Topics in Medicinal Chemistry*. 2020;20(18): 1640–1650.
8. Williams B, Alberti G, Ball C, et al. *National Early Warning Score (NEWS): Standardising The assessment of Acute-Illness Severity in the NHS*. London: The Royal College of Physicians; 2012.
9. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. *Critical Care Medicine*. 1985;13: 818–829.
10. Le Gall JR, Lemeshow S, Saulnier F. A new Simplified Acute Physiology Score (SAPS II) based on a European/North American multicenter study. *Journal of the American Medical Association*. 1993;270: 2957–2963.
11. Mezentseva LV, Pertsov SS, Kopilov FY, Lastovetsky AG. Mathematical analysis of the stability of heart-rate dynamics in postinfarction patients. *Biophysics*. 2017;62: 499–502.
12. Shabestari AA, Dalirooyfard M, Mazloomzadeh S. Electrocardiographic correted QT dispersion value as a predictor for estimation of neontatal mortality in preterm infants. *Acta Inform Med*. 2019;227(3): 158–161.
13. Ball J, Carrington MJ, Thompson DR, Horowitz JD, Stewart S, Atrial Fibrillation SpEcific managemenT studY (SAFETY) Investigators. Post-discharge electrocardiogram Holter monitoring in recently hospitalised individuals with chronic atrial fibrillation to enhance therapeutic monitoring and identify potentially predictive phenotypes. *European Journal of Cardiovascular Nursing*. 2015;14: 384–394.
14. Sapo M, Wu S, Asgari S, et al. A comparison of vital signs charted by nurses with automated acquired values using waveform quality indices. *Journal of Clinical Monitoring and Computing*. 2009;23: 263–271.
15. Saeed M, Villarroel M, Reisner AT, et al. Multiparameter Intelligent Monitoring in Intensive Care II (MIMIC-II): a public-access intensive care unit database. *Critical Care Medicine*. 2011;39: 952–960.
16. Goldberger AL, Amaral LA, Glass L, et al. Physiobank, physiotoolkit, and physionet components of a new research resource for complex physiologic signals. *Circulation*. 2000;101: e215–e220.
17. Clifford GD, Azuaje F, McSharry P. ECG statistics, noise, artifacts, and missing data. *Advanced Methods and Tools for ECG Data Analysis*. 2006;6: 18.
18. Clifford GD, Behar J, Li Q, Rezek I. Signal quality indices and data fusion for determining clinical acceptability of electrocardiograms. *Physiological Measurement*. 2012;33: 1419–1433.
19. Iglewicz B, Hoaglin DC. *How to detect and handle outliers (Vol. 16)*. Milwaukee, WI: ASQ Quality Press; 1993.
20. Flanders K, Hudson Z. Appropriate use of telemetry in the acute care setting. *Nursing Management*. 2020;51(5): 44–51.
21. Trimmel K, Sacha J, Huikuri HV. Heart rate variability: clinical applications and interaction between HRV and heart rate. *Frontiers Media SA*. 2015.
22. Barnaby D, Ferrick K, Kaplan DT, Shah S, Bijur P, Gallagher EJ. Heart rate variability in emergency department patients with sepsis. *Academic Emergency Medicine*. 2002;9: 661–670.
23. Stein PK, Carney RM, Freedland KE, et al. Severe depression is associated with markedly reduced heart rate variability in patients with stable coronary heart disease. *Journal of Psychosomatic Research*. 2000;48(4–5): 493–500.
24. Stein PK, Kleigr RE, Rottman JN. Differing effects of age on heart rate variability in men and women. *The American Journal of Cardiology*. 1997;80(3): 302–305.
25. Taenzer AH, Pyke J, Herrick MD, Dodds TM, McGrath SP. A comparison of oxygen saturation data in inpatients with low oxygen saturation using automated continuous monitoring and intermittent manual data charting. *Anesthesia & Analgesia*. 2014;118: 326–331.
26. Rinta-Koski OP, Hollmén J, Leskinen M, Andersson S. Variation in oxygen saturation measurements in very low birth weight infants. *Proceedings of the 8th ACM International Conference on PErvasive Technologies Related to Assistive Environments*. 2015;29.
27. Maslove DM, Dubin JA, Shrivats A, Lee J. Errors, omissions, and outliers in hourly vital signs measurements in intensive care. *Critical Care Medicine*. 2016;44: e1021–e1030.
28. Smith I, Mackay J, Fahrid N, Krucheck D. Respiratory rate measurement: a comparison of methods. *British Journal of Healthcare Assistants*. 2011;5: 18–23.
29. Reich DL, Wood RK Jr., Mattar R, et al. Arterial blood pressure and heart rate discrepancies between handwritten and computerized anesthesia records. *Anesthesia & Analgesia*. 2000;91: 612–616.
30. Cygankiewicz I, Zareba W. Heart rate variability. *Handbook of Clinical Neurology*. 2013;117: 379–393.
31. Huikuri HV, Mäkikallio T, Airaksinen KE, Mitrani R, Castellanos A, Myerburg RJ. Measurement of heart rate variability: a clinical tool or a research toy? *Journal of the American College of Cardiology*. 1999;34: 1878–1883.
32. Monteiro F, Meloni F, Baranauskas JA, Macedo AA. Prediction of mortality in intensive care units: a multivariate feature selection. *Journal of Biomedical Informatics*. 2020;23: 107.
33. Helfand M, Christensen V, Anderson J. Technology assessment: early sense for monitoring vital signs in hospitalized patients. In: *VA Evidence Synthesis Program Evidence Briefs*. Washington, DC: Department of Veteran Affairs (US); 2016.
34. Zhang Y, Zhou B, Qiu J, Zhang L, Zou Z. Heart rate variabilty changes in patients with panic disorder. *Journal of Affective Disorders*. 2020;267: 267–306.
35. Lombardi C, Peveri G, Cani D, et al. In-hospital and long-term mortality for acute heart failure: analysis at the time of admission to the emergency department. *ESC Heart Failure*. 2020;7(5): 2650–2661.
36. Johnston BW, Barrett-Jolley R, Krige A, Welters ID. Heart rate variability: measurement and emerging use in critical care medicine. *Journal of the Intensive Care Society*. 2020;21(2): 148–157.