Changes in daily measures of blood pressure and heart rate improve weight-based detection of heart failure deterioration in patients on telemonitoring

Joshi, R.; Cuba Gyllensten, I.G.L.

Published in:
IEEE Journal of Biomedical and Health Informatics

DOI:
10.1109/JBHI.2018.2856916

E-pub ahead of print: 18/07/2018

Document Version
Accepted manuscript including changes made at the peer-review stage

Please check the document version of this publication:

• A submitted manuscript is the author's version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

Citation for published version (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal?

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 05. Nov. 2018
Changes in Daily Measures of Blood Pressure and Heart Rate Improve Weight-based Detection of Heart Failure Deterioration in Patients on Telemonitoring

Rohan Joshi and Illapha Cuba Gyllensten

Abstract—Blood pressure (BP) and heart rate (HR) are often captured in conjunction with weight in telemonitoring systems, but the additional prognostic potential of daily measurements of BP and HR in providing information on upcoming hospitalizations for worsening heart failure (HFH) have not been explored thoroughly. We retrospectively analyzed 267 daily home-telemonitored heart failure (HF) subjects. We extracted those episodes of HFHs that had sufficient data entries in the days leading up to hospitalization and tested the prognostic potential of 48 trend features based on weight, systolic BP, diastolic BP, pulse pressure (PP), and HR with a Naïve Bayesian model. The single best-performing trend feature—with a cross-validated estimate of 0.64 for the area under the curve (AUC) with a standard deviation (SD) of 0.01—is based on a 2-day weight trend. The best multivariate feature set (cross-validated AUC=0.70, SD=0.01) comprises of 2-day trend features based on weight, systolic BP, pulse pressure, and HR. There were large variations in the weight trends preceding hospitalizations and weight-change alone had a modest predictive ability. Readily interpretable features capturing trends in BP and HR provided additional prognostic information and can be used for improving classification.

Index Terms—Telemedicine, Blood pressure, Heart rate, Heart Failure, Decompensation.

I. INTRODUCTION

The path of deterioration from “stable” Heart Failure (HF) to acute decompensated HF is often preceded by a period of worsening hemodynamics, spanning several days or even weeks before manifesting itself acutely. Therefore, theoretically, a window of opportunity presents itself during which decompensation might be reflected in physiological signals and thus could be acted upon. Home-telemonitoring could be used as a tool to allow monitoring of the vital parameters of patients and provide education and follow-up for them. Together with promoting medication adherence and optimized self-care behaviors, this could not only improve patient outcomes but also reduce costs and the burden on the healthcare systems by keeping the patients stable.

Despite the many purported advantages of such an integrated system, the evidence for such implementation with non-invasive measures of weight, blood pressure, and heart rate is conflicting. Although most meta-analyses of HF telemonitoring trials have concluded that in comparison with usual care, telemonitoring reduces all-cause hospitalizations, hospitalizations for worsening heart failure (HFH), and all-cause mortality, different conclusions have been drawn from recent larger trials. A cohesive interpretation of the findings is complicated by inter-trial differences among the interventions [1]–[9]. Better selection of telemonitoring service model, integration of multi-disciplinary team and improving the use and interpretation of the data captured by the system should improve this.

In addition to observations by medical personnel, telemonitoring also provides the possibility of automatically analyzing the physiological signals measured at the patient’s home in order to predict an increased risk of decompensation. In particular, some episodes of decompensated heart failure are preceded by an increase in body weight—as a result of water retention—and this finding has been exploited as part of developing predictive models that “recognize” deterioration [10]–[14]. While the prognostic potential of an increase in weight before heart failure has been established, its sensitivity is poor [11], [13]–[15]. Existing studies that have used weight-based features have employed a range of analyses and different techniques to evaluate algorithms, often using a small number of hospitalized patients. Therefore, comparisons between various studies are difficult and, as a result, many issues remain unresolved. Improved detection of deterioration could help telemonitoring nurses to more effectively direct their attention to patients at higher risk and provide better care through telemonitoring if properly integrated [16].

This study aimed to determine if there was any additional value of including trends in blood pressure (BP) and heart rate (HR),
which are often captured in conjunction with weight, into the
detection of deterioration.

II. METHODS

A. Patient Population

The data used for this study was retroactively collected from the
Trans-European Network-Home-Care Management System
(TEN-HMS) [17] and from an ongoing telemonitoring service in
Hull, UK (Motiva-Hull) in which patients with a recent
admission for worsening HF could be enrolled. In both studies
the same system was used to log weight and blood pressure.
Amongst other criteria, at the time of enrolment, the TEN-HMS
study excluded subjects with a left ventricular ejection fraction
(LVEF) greater than 40%. With the intention of homogenizing
the two databases, we excluded those patients in the Motiva-
Hull database who, upon enrollment, presented with an LVEF
greater than 40% and those with simultaneously missing
levels were perhaps due to greater water retention despite
hospitalizations were also included.

With regard to the number of HFHs, in the combined dataset
of 267 patients, 111 presented with 256 HFHs during the period
of monitoring (until death/censoring). The remaining 156
subjects did not suffer from any episode of HFH (or die from
HF) during the period of monitoring. No significant differences
(p > 0.05) were present in the baseline characteristics between
these two groups with the exceptions of NT-proBNP and serum sodium values, which were higher and lower respectively in the
111 subjects who got hospitalized. The lower serum sodium
levels were perhaps due to greater water retention despite
titrated diuretic therapy [21].

B. Selecting Cases of HFHs and Controls

Since not all episodes of HFHs have sufficient
telemonitoring data in the days preceding hospitalization, not
all cases of HFHs can be used for analyses. For analysis, data
from the 30-day window preceding an HFH is vital, since this
is the period during which physiological changes are expected
to manifest [10], [20]. Therefore we retained only those cases
of HFHs that had at least 50% data entries (empirical choice) in
the 30-day and the eight-day windows preceding hospitalization. While data sufficiency in the days preceding

All subjects were required to use a telemonitoring device
installed in their homes for recording their weight, BP, and HR
every day, at a fixed time. Each instance of measurement
resulted in a single value for weight, systolic BP, diastolic BP,
HFH is important, it is also necessary to ensure that the data is solely reflective of the impending HFH and free from the influence of any past “events” experienced by the patient. To avoid such “contamination,” only those instances of HFHs were retained where the 75 days preceding HFH were free of “events,” defined as enrollment into the study or discharge from an earlier HFH. The mean values of the physiological signals between the 30th day and the 45th-day preceding HFH served as a “stable baseline” reference for every subject.

To determine whether the physiological variations before HFH are distinctive, they need to be contrasted against naturally occurring changes in physiology, hereafter referred to as control cases. To obtain the time-series corresponding to control cases, we used those subjects who did not experience an episode of HF leading to hospitalization (or death) during the period of monitoring (until censoring/death). Data sufficiency criteria, identical to those used for the cases of HFHs were applied at five random points (control “events”) on the time-series data of these subjects. Here, too, a “control-case” had to be at least 75 days away from enrollment into the study. These data, corresponding to both the cases of HFHs and the controls, were used to generate and test features.

Previously researched algorithms that have attempted to capture weight-gain preceding HF have had limited sensitivity. We sought to explain these results by quantifying the episodes of HF that do exhibit weight-gain preceding hospitalization. Weight-trends were stratified as “up trend,” “down trend” or “no trend,” based on whether the mean weight in the ten days preceding HFH increased by, decreased by, or stayed within one kilogram of the stable baseline value.

While the data preceding HFH is relevant for analyzing the predictive potential of the different physiological signals, the post-discharge data, reflective of treatment, can hold useful information on the return to baseline. To observe this effect, we applied certain data sufficiency criteria to the stream of data following discharge from HFH. These criteria required at least 50% data in the ten-day and the 30-day window post-discharge, along with an absence of “events” in the 60 days post-discharge. Here, “events” refers to HF leading to hospitalization or death. Fig. 1 shows the selection criteria that were applied to the cases of HFHs and controls.

### C. Feature Generation

A battery of 48 features was developed for evaluating the prognostic potential of time-series data of weight (16 features), systolic BP (eight features), diastolic BP (eight features), PP (eight features), and HR (eight features) for identifying periods of deteriorating HF. All features were developed from three “families of features.”

The first feature-family was the rule-of-thumb algorithm that was motivated by the guideline thresholds of two kg weight gain in two to three days or two percent weight gain in two days [11]. Such features were intended to capture acute discriminative changes in weight. Six weight-based features corresponding to one-, two-, and three-day differences in absolute and percentage weight change were employed. Since evidence of an increase in weight starting up to a month before HF deterioration was available [10], we also developed features corresponding to the differences (absolute and percentage) between the mean weights measured over a week, a fortnight, and a month with respect to stable baseline values giving rise to six features.

For the systolic BP, diastolic BP, PP, and HR, four features apiece were developed to capture the absolute differences between the mean values measured over the past week, the penultimate week (this was an empirical choice motivated by observing a bump in the population-level systolic BP trend in the week penultimate to HFH), the past fortnight, and the past month with respect to stable baseline values.

The second feature-family used the moving average convergence divergence (MACD) algorithm, i.e., the difference between the mean of short-term data (short-term window size (STWS)) and a long-term average (long-term window size (LTWS)). The use of this feature-family, which was motivated by the work of Zhang et al. [15], was employed for weight, BP (systolic, diastolic and pulse pressure), and HR using LTWSs of 20 and 40 days respectively, along with an STWS of five days. In effect, the MACD algorithm acts like the combination of a differential operator and a low pass filter and is intended to capture fast changes w.r.t. to slower-changing trends.

The third feature-family was based on the standard score (SS), which was defined as:

$$SS(t) = \frac{x(t) - \mu}{\sigma}$$

where, $x(t)$ is the value of the time-series signal at time $t$ while “$\mu$” and “$\sigma$” are the mean and standard deviation, respectively, calculated from a predefined length of past values. SS-based features were employed for weight, BP (systolic, diastolic and pulse pressure), and HR using sample lengths of 20 and 40 days respectively. The SS-based features were designed to capture changes in a subject’s physiological signals that differ significantly from the naturally exhibited variation of an individual. Both, the MACD and the SS-based features used the original, uninterpolated data.

### D. Classification Evaluation and Statistics

While decompensation is believed to be reflected in physiological changes, the precise point in time when such deterioration peaks before requiring hospitalization are unknown. Therefore, the analysis was carried out using different lengths of data (two, five, eight, and twelve-day windows) from the subject’s past measurements. In this scenario, the entire “window of data” of a subject is considered to correspond to “HFH” or a “control,” as the case may be and is fed to the classifier. For example, a five-day window would yield 5 data points corresponding to HFH or of control. It should be noted that constructing all features required data from the past. So to calculate a feature value on the 12th day prior to HFH, data from before the 12th day is needed. This is the reason why we applied the criteria of 50% data availability in the 30-day period before HFH.

The Naïve Bayes classifier was used to distinguish daily data belonging to the period leading up to a HFH from control periods. The class priors were defined to be uniform while
continuous data were represented by estimating non-parametric densities using multiple kernels of the normal distribution. The predictive strength of the classifier was evaluated by calculating the area under the curve (AUC) from the concatenated left out folds of a fivefold cross-validation (CV) stratified by HFH and control data, carried out five times to provide an average AUC estimate and estimates of its dispersion (standard deviation).

While the AUC measures the predictive strength of the classifier, it does not explain how changes in the individual features affect the probability of decompensation. Thus, to make the classification process easier to interpret, the calculated probabilities of decompensation were extracted for different values of the individual features. The means and standard deviations of these values were then calculated over the different runs of the CV process for a model (chosen to serve as an example) and plotted individually for each feature.

Differences in the baseline characteristics between the cohorts were evaluated with Welch’s test for the continuous variables and with the chi-squared test for categorical variables, a nominal p-value of 0.05 was assumed throughout. All models were developed and evaluated in a MATLAB environment (MATLAB R2014 MathWorks).

III. RESULTS

A. Characteristics of the HFH and the Control Subjects

Upon implementing the selection criteria described in section 2.2, 50 instances of HFHs and 112 of controls were found suitable for quantitative analysis. These instances originated from 40 and 63 patients from the hospitalized and the control group respectively, baseline characteristics of which are shown in Table II. The large decrease in the number of HFHs available for analysis, from 256 to 50, is primarily attributable to most episodes of HFHs occurring within 75 days of discharge from a previous HFH or from enrolment into the study. This is not entirely surprising since the burden of disease in these patients was high.

B. Physiological Trends Preceding HFH and Following Discharge

While 50 episodes of HFHs had fulfilled the pre-admission criteria, only 44 episodes of HFHs remained after applying the post-discharge criteria. These were used for visualizing the influence of hospitalization on physiological signals. Out of the five physiological signals considered, only weight changed visibly in the days leading up to HFH and after discharge. A closer analysis of the weight trends, based on stratifying trends leading up to hospitalization (see Fig. 2) shows that 41% of decompensations do not exhibit any increase in weight before HFH. In fact, weight decreased in 20% of the cases in the weeks before hospitalization with some increase, possibly edematous, in the days immediately preceding HFH.

![Fig. 2. Stratification of the instances of HFH by weight (stable baseline removed) showing an uptrend, a downtrend, or a flat trend in the weeks preceding HFH. These trends depict the mean population weight, and the error bars represent the standard error of the mean. The legend indicates the number of subjects in each subgroup. The thick black line at "0" represents admission weight, and the error bars represent the standard deviation in the days immediately preceding HFH.](Image)

C. Univariate analysis

For univariate analysis, features originating from each of the five physiological signals were individually tested on the four different window lengths of data using the Naïve Bayes classifier. The best-performing feature from each physiological category could change for different window lengths. Table III details the list of features which were best-performing for at least one window length while Fig. 3 shows the corresponding AUCs.

Interestingly, while the performance of the best weight-based feature does not change much with longer windows, the best HR-based features become more discriminatory as window length increases, in fact outperforming the best weight-based feature for a window of eight days. The performance of the BP-based features is lower than that of weight and HR-based ones and tends to decrease with increasing window-lengths.
f-11 give the best result for an eight-day window while f-2, f-4, and f-11 perform best for a twelve-day window. Overall, the best performance across window size decreases with increasing the window length.

The changes in feature-values that determined classification for the top-performing features from the five-day window are shown in Fig. 5. For example, the graph corresponding to f-3 shows that the probability of decompensation increases with both increasing and decreasing values of mean weight of the past month with respect to baseline. A relatively stable weight, on the other hand, reduces the probability of decompensation (value below 0.5). Similarly, the graph corresponding to f-6 shows that lower values of systolic BP, after adjusting for the inherent variability in the systolic BP (which is what the SS measures), increase the risk of decompensation. The fact that the standard deviations are small for the classifier estimated probabilities in Fig. 5 suggests that the training was stable with little evidence of overfitting.

Since the Naïve Bayes classifier assumes independence of the features, the probability of decompensation for any subject using multivariate classification can be obtained by multiplying the probabilities of decompensation obtained from each feature-value followed by normalization. For example, if on a certain day the probabilities of decompensation corresponding to feature values of f-3, f-4, f-6 and f-11 are 0.70, 0.75, 0.60 and 0.50, then the probability of decompensation is 0.91 ( = 0.70 × 0.75 × 0.60 × 0.50 / (0.70 × 0.75 × 0.60 × 0.50 + (1-0.7) × (1-0.75) × (1-0.60) × (1-0.50)) ). Interpretation should take into account the high prior probability assumed by the model (0.50) that can be adjusted to baseline characteristics.

### Table III

<table>
<thead>
<tr>
<th>Feature no.</th>
<th>Feature definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>f-1</td>
<td>Change in mean weight of the past week with respect to (w.r.t.) baseline</td>
</tr>
<tr>
<td>f-2</td>
<td>Change in mean weight of the past fortnight w.r.t. baseline</td>
</tr>
<tr>
<td>f-3</td>
<td>Change in mean weight of the past month w.r.t. baseline</td>
</tr>
<tr>
<td>f-4</td>
<td>Weight-based MACD with LTWS = 40 days, STWS = 5 days</td>
</tr>
<tr>
<td>f-5</td>
<td>Systolic BP-based MACD with LTWS = 40 days, STWS = 5 days</td>
</tr>
<tr>
<td>f-6</td>
<td>Systolic BP-based SS with sample length = 20 days</td>
</tr>
<tr>
<td>f-7</td>
<td>Systolic BP-based SS with sample length = 40 days</td>
</tr>
<tr>
<td>f-8</td>
<td>Change in mean diastolic BP of the past week w.r.t. baseline</td>
</tr>
<tr>
<td>f-9</td>
<td>Diastolic BP-based SS with sample length = 20 days</td>
</tr>
<tr>
<td>f-10</td>
<td>Diastolic BP-based SS with sample length = 40 days</td>
</tr>
<tr>
<td>f-11</td>
<td>Change in mean heart rate of the past month w.r.t. baseline</td>
</tr>
<tr>
<td>f-12</td>
<td>Heart rate-based MACD with LTWS = 20 days, and STWS = 5 days</td>
</tr>
<tr>
<td>f-13</td>
<td>Heart rate-based MACD with LTWS = 40 days, and STWS = 5 days</td>
</tr>
<tr>
<td>f-14</td>
<td>Heart rate-based SS with sample length = 20 days</td>
</tr>
<tr>
<td>f-15</td>
<td>Change in mean pulse pressure of the past month w.r.t. baseline</td>
</tr>
<tr>
<td>f-16</td>
<td>Pulse pressure-based MACD with LTWS = 20 days and STWS = 5 days</td>
</tr>
<tr>
<td>f-17</td>
<td>Pulse pressure-based MACD with LTWS = 40 days and STWS = 5 days</td>
</tr>
<tr>
<td>f-18</td>
<td>Pulse pressure-based SS with sample length = 40 days</td>
</tr>
</tbody>
</table>

Features that contribute to the best-performing feature combinations in multivariate analysis, irrespective of window length, are highlighted in bold.
IV. DISCUSSION

Since HFHs are costly to the healthcare system and detrimental to the patient, predictive monitoring of the same is clinically important. While most studies have used weight-based features for classification, this study aimed to determine if there was any additional value of including BP and HR, parameters often captured in conjunction with weight, into the detection of deterioration.

The primary finding of this study is that medium to short-term changes (≤ 8 days) in HR and BP can add additional information to classifiers of decompensation risk in HF patients on telemonitoring. HR-based features were included in the best-performing classifiers for all time windows under consideration, and BP-based features were included for window lengths of two and five days. Weight change features were included in all time windows, and weight was overall the most predictive feature class in the univariate analysis. This finding is similar to that of Koulaouzidis et al. with weight-based features holding the most prognostic potential. However, as opposed to their findings we find that, both, the features based on HR and on systolic BP can improve upon the prognostic potential of weight-based features alone, although this difference might be attributable to differences in the study-populations [22]. In addition, the study of Koulaouzidis et al. studied 27 HFHs, as opposed to 50 in this study, and was not tailored to investigate the additional potential HR and BP.

Weight gain in the week preceding HF hospitalizations occurred in the majority of the visualized cases. However, 41% of the cases had a mean weight increase of less than one kilogram, and in half of these cases, the mean weight decreased by more than one kilogram. This data highlights the heterogeneity of acute heart failure with varying presentations and shows that weight gain, although an integral part of monitoring and self-care, is insufficient as a single telemonitoring target [23]. Furthermore, the multitude of weight patterns preceding HFH can explain findings in the literature of low selectivity of weight-gain algorithms in predicting worsening heart failure [11], [13]–[15].

Redistribution, rather than fluid retention, has been proposed as a mechanism for decompensation [24] that could explain the lack of substantial weight gain. We further found a subgroup with a decreasing weight trend preceding HFH. Although the reason for this weight change cannot be determined from our data, the post-discharge trend towards the baseline weight suggests that treatment following the HFHs reversed the weight loss. There are many processes that can lead to weight loss in HF [25], [26], some of which may be aggravated during the period preceding HFH, e.g., bowel wall edema and reduced intestinal blood flow causing loss of appetite and reduced nutrient uptake that could revert following treatment. If this is
an underlying factor segmental bioimpedance of the legs, abdomen, and thorax may provide a method to distinguish and classify the same [27]. In a large study of 6933 chronic HF patients, Pockock et al. found that loss of weight in the last 12 months was an important predictor of all-cause mortality [28]. Furthermore, a small reduction in weight from admission to discharge and a post-discharge trend of increasing weight has been linked to an increase 180-day post-discharge mortality [29]. Our finding supports the importance of weight loss for HFH and the link with trends at presentation and discharge, which could be followed with telemonitoring.

Although neither BP nor HR presented any clear trends on a population level in the days preceding hospitalizations, both added to the classification of decompensation cases. In general, the features that were predictive focused on longer time frames (weeks to months) than usually considered for weight, which might be because active telemonitoring is targeted to respond to acute increases in weight and that more subtle changes are overlooked. However, it could also indicate that sharp changes in physiological signals are an uncommon presentation of acute HF, at least in patients on telemonitoring. This analysis suggests that long-term changes are important in HF monitoring and, in particular, that HR-based features appeared to be discriminatory when a window of data of 5-12 days was considered.

The detailed analysis of the 5-day risk classifier in Fig. 5 shows that both significant increases and decreases in weight increase the risk of decompensation. Also, a falling systolic BP, after adjusting for the inherent variability in the BP, (f-6) without any change in HR add to the risk of decompensation. Possibly, this points to the body’s inability to adapt the HR despite a falling BP and the associated reduction in cardiac output. Unmeasured confounders are another possibility.

The strength of this study comes from analyzing well-characterized hospitalizations for acute heart failure, using multi-resolution analysis based on non-invasive measurements of HR, BP, and weight at home. Further, the results of the study indicate an added benefit of incorporating measures of HR and BP, parameters not well studied in literature, to weight-based measures for improving classification. Based on our findings of weight gain prior to HFH, future studies may benefit from developing weight-based features that capture an uptick in weight in an otherwise downward trend.

However, there are limitations in the data used for analysis. First, only a minority of HFHs passed the data sufficiency criteria and could be analyzed. This is largely because repeat HFHs are common and occur in close proximity to one another and many within 30 days following discharge. In conjunction to algorithms, such as the one presented herein which requires, for instance, 75 event-free days prior to deployment for predictive monitoring, telemonitoring systems will need algorithms that are tailored to predict the risk of HFH in the days immediately following discharge from HFH or enrolment in to the study. This can serve as a direction for future research. Second, there exist some underlying differences between the patient populations of the two databases with regard to body mass index (p-value=0.003), LVEF (p-value = 0.0003), and serum potassium (p-value = 0.05). In the Motiva-Hull subjects, the mean values of all three characteristics are higher by 8.6%, 18.5%, and 3% respectively. Also, there are differences in the baseline medications between the subjects of the two databases, especially in the use of beta-blockers and vasodilators. Some of these differences were likely because the TEN-HMS database dates to the period from 2000-2002 while the Motiva-Hull database was recorded during 2008-2010. Finally, active telemonitoring acts as a confounding factor for feature/algorithm development since medical personnel may dynamically change the pharmaceutical regimen of the patient and provide lifestyle/diet advice by observing the subject’s vital parameters. It is likely that the discriminatory potential of features is higher when the subject-characterization is more homogenous (for instance from the same period and from the same dataset) and if they are not actively telemonitored.

V. CONCLUSION

Home telemonitoring is posited to become an integral part of all chronic disease management, including HF. Algorithms that can indicate risk of decompensation by automatically analyzing daily measurements of vital signs can augment these systems and may help clinicians keep the patients stable. So far, conventional noninvasive systems have focused on weight increase alone. We have shown that although the majority of patients gained weight, some are hospitalized for worsening HF with only a small or no increase in weight. Including weight decrease and analyzing measurements of HR and BP, which are often captured in conjunction with weight, can give additional information of decompensation risk. Algorithms, such as the Naive Bayesian approach used here, could provide a visualization of how each variable contributed to the risk estimation, thereby aiding clinicians reviewing the alert.

REFERENCES


