

THE STRENGTHS AND WEAKNESSES OF HEART RATE VARIABILITY AS A RECOVERY METRIC AND WHY WHOOP RECOVERY INCLUDES RESTING HEART RATE AND SLEEP PERFORMANCE

CHRISTOPHER ALLEN, EMILY BRESLOW, JOHN CAPODILUPO
DEPARTMENT OF PHYSIOLOGY AND ANALYTICS
WHOOP, INC.

Copyright © 2016 WHOOP, inc. All rights reserved.

The information contained in this document is the proprietary and exclusive property of WHOOP, inc. Except as otherwise indicated. No part of this document, in whole or in part, may be reproduced, stored, transmitted, or used for design purposes without the prior written permission of WHOOP.

The information in this document is provided for informational purposes only. WHOOP expressly disclaims all warranties, express or implied, including without limitation those of merchantability, fitness for a particular purpose and noninfringement.

Table of Contents

Abstract	2
HRV for Training Prescription	3
Incorporating HRV Prescription into WHOOP	3
HRV Measurements: SD1 vs. RMSSD	4
Findings: Recovery by HRV Prescription	5
Implications	6
Including Resting Heart Rate	7
Including Sleep Performance	8
Conclusion	13
References	14



Abstract

Over the past ten years, scientific research has investigated the use of Heart Rate Variability (HRV) as a tool to optimize training in athletes. This research has involved several studies that illustrate the relationship between pre-training HRV and adaptation.^{1,2,3,4,5} Specifically, the research has suggested that HRV can be used as an indicator for how physiologically ready the body is to reap the benefits from training at a given point in time, and that it can be used to monitor athletes during training blocks.

The WHOOP Data Science team has built on the last decade of medical and academic HRV research to develop a physiological monitoring platform capable of delivering athletic-performance optimizing analytics. The driving philosophy at WHOOP is that with continuous, actionable data, our athletes can make better informed and scientifically grounded training and recovery decisions to give them a competitive advantage over athletes training with the more common periodized training approach or simply training by feel. Our proprietary Recovery algorithm is one of the core metrics of the performance optimization system, as it indicates to the user his or her physiological readiness to optimally respond to exercise.

HRV is one of the main inputs into the WHOOP Recovery algorithm. This paper demonstrates how our use of this metric leverages the scientific research on HRV. It then illustrates how the inclusion of resting heart rate (RHR) and measures of the quality and duration of a user's sleep in our measure of Recovery allows for more nuanced and physiologically-relevant training recommendations.

¹ Kiviniemi et al., 2007

² Kiviniemi et al., 2010

³ Vesterinen et al., 2013

⁴ Plews et al., 2013

⁵ Plews et al., 2012



HRV for Training Prescription

In a 2007 study, Kiviniemi et al. explored the utility of daily exercise prescription based purely on HRV in recreational men. They observed that men who adapted their program based on HRV had significantly larger improvements in cardiovascular fitness than did men who followed a pre-defined training program with high-intensity and moderate-intensity days scheduled in advance.⁶ This result was particularly interesting because the men training via the HRV-based prescription model did not train any more than the men on the predetermined training schedule. This means that for the same amount of exercise, the HRV-based group experienced greater returns on investment.

The researchers extended this study in 2010 by testing exercise prescription on the basis of HRV in both men and women. In this study, 24 men and 24 women were split into a control group, a pre-defined training group (defined as two moderate-intensity workouts and three high-intensity workouts per week, in whichever order was most convenient to simulate real life), and a group that trained based on morning HRV levels for 8 weeks of training. The researchers also had 12 additional women train using an HRV-based training program tailored towards women, hypothesizing that differences in hormone production, thermoregulation, and hemoglobin concentration in women may cause differences in response to exercise. The subjects in the HRV-based training group recorded their morning HRV using a HR monitor (Polar Electro Oy) after awakening. Then, the researchers declared the morning's HRV as "increased" (INC), "decreased" (DEC), or "unchanged" (UNCH) relative to the HRV values from the previous 10 days. The HRV-based training group containing both men and women performed moderate exercise if HRV was decreased, and high-intensity otherwise. The HRV-based group geared towards women performed high intensity exercise only when HRV was increased. VO₂ max and maximal workload were measured by a bicycle ergometer test both before and after the training period to quantify fitness improvements.

The researchers found that for men, the improvement in maximal workload was significantly higher ($p < 0.05$) in the HRV-trained group compared to the pre-defined training group, but not for women. The researchers also observed that the women in the HRV-based group tailored towards women displayed a significantly larger improvement in fitness, which they attributed to the lower frequency of high-intensity exercise.⁷

Incorporating HRV Prescription into WHOOP

During the development of the WHOOP Recovery algorithm, no attempts were made to optimize the output to agree with that of the Kiviniemi et al. models. However, our retrospective analysis, shown in detail below, reveals a high degree of consistency between the recommendations WHOOP would make based on the results of its Recovery algorithm and the recommendations Kiviniemi et al. set out to test in their proposed training model.

⁶ Kiviniemi, et al., 2010

⁷ Kiviniemi et al., 2010

To demonstrate this, HRV-based recommendations from Kiviniemi et al.'s 2010 study were compared to Recovery reported by WHOOP. R-R intervals used to calculate Recovery, as well as Recovery itself, were collected for 61 WHOOP users for each night's sleep between August 2nd, 2016 and October 16th, 2016. These users, who represented collegiate, professional, and recreational athletes of both genders, were selected at random where the only inclusion criterion was high compliance (at least 70 days of data) during this 75-day period.

With these R-R intervals, daily HRV was calculated and a training prescription was generated according to the WHOOP Recovery algorithm and the Kiviniemi method. Then, the training recommendation by Kiviniemi et al. was compared to that of WHOOP.

Compared to Kiviniemi et al., the WHOOP Recovery scale is highly granular, using all integer values between 0 and 100. In order to map these to the INC, UNCH, and DEC designations of Kiviniemi et al., we divide our Recovery scale into three equally-sized bins such that values from 67% to 100% Recovery mapped to INC SD1, 34-66% mapped to UNCH, and 0-33% Recovery mapped to a DEC recommendation.

Because the distributions of Recovery for INC SD1 and DEC SD1 were non-Gaussian, the Kruskal-Wallis test was used to determine whether there was a statistically significant difference in Recovery for the three classifications of SD1. The Mann-Whitney U-test was utilized as a post hoc test.

HRV Measurements: SD1 vs. RMSSD

HRV is a measure of the irregularity of the heart's rhythm over time. The utility of HRV in many physiological applications, including those related to the training of athletes, has been widely explored.⁸ There are many accepted methods to measure HRV⁹, and WHOOP and Kiviniemi et al. use different calculations.

Kiviniemi et al. used a metric called Standard Deviation 1 (SD1) as their measure of HRV, while WHOOP uses the root mean square of successive differences (RMSSD). **Equation 1** and **Equation 2** are the formulas for calculating SD1 and RMSSD.

Equation 1

$$SD1 = \sqrt{\frac{\sum_{i=1}^{N-1} (RR_i - RR_{i+1})^2}{\sqrt{2}(N-1)}}$$

⁸ Breslow, 2016

⁹ Task Force of the European Society of Cardiology, 1996

Equation 2

$$RMSSD = \sqrt{\frac{\sum_{i=1}^{N-1} (RR_i - RR_{i+1})^2}{N - 1}}$$

A close examination of the formulas shows that the equations for SD1 and RMSSD are simply scaled versions of each other. Other than the inclusion of $\sqrt{2}$ in the denominator (emphasized by red text) in **Equation 1**, SD1 and RMSSD both process beat-to-beat (R-R interval) variability the same way. One can easily convert between them via multiplication or division by $\frac{1}{\sqrt{2}}$. Therefore, the different choices of HRV metrics by WHOOP and Kiviniemi et al. have no bearing on the interpretation of the results given that analysis in both cases is based on the trends in the data, making them independent of absolute scale.

WHOOP selected RMSSD as its HRV measurement because it is a popular metric in studies examining HRV in athletes, and it has a reputation as one of the most reliable HRV indices.¹⁰

Findings: Recovery by HRV Prescription

Table 1 below shows how often WHOOP Recovery is consistent with the recommendations by Kiviniemi et al., and **Figure 1** displays the distributions of Recovery for each of the three classifications by Kiviniemi et al.

Kiviniemi et al.'s SD1 Classification	Recovery	Percent Agreement
INC SD1	Green (> 66)	66.2%
UNCH SD1	Yellow (> 33 and <= 67)	66.8%
DEC SD1	Red (<=33)	46.0%

Table 1. A breakdown of how often each classification of SD1 lines up with green, yellow, and red Recovery.

¹⁰ Al Haddad et al, 2011.

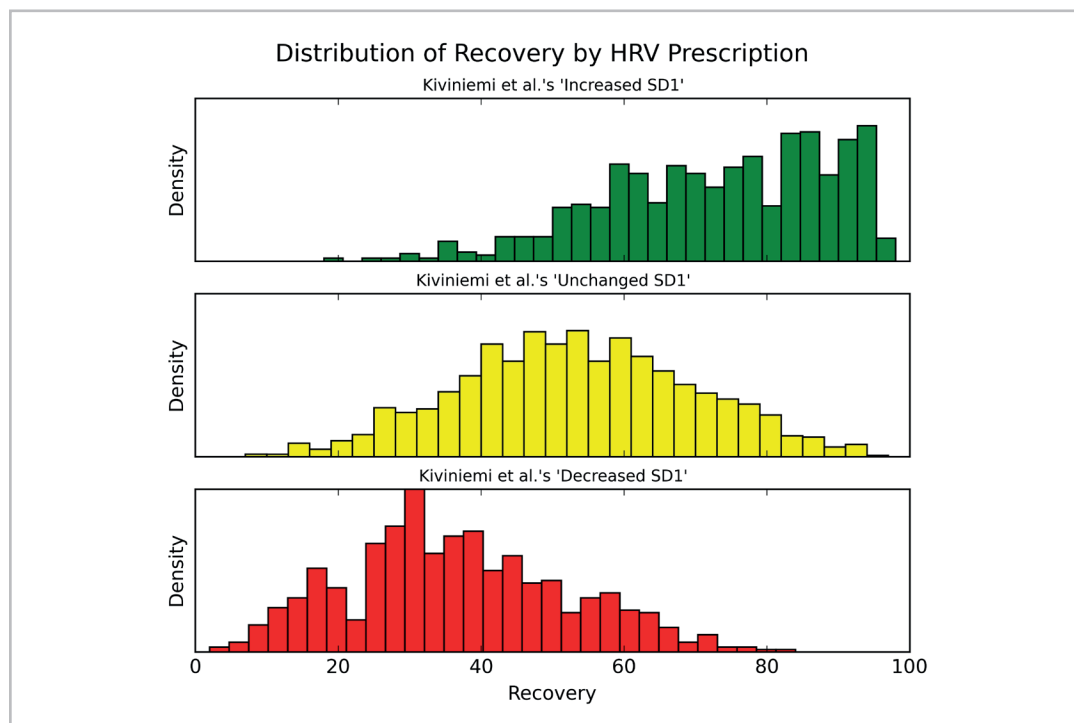


Figure 1. The distributions of Recovery for each of Kiviniemi et al.'s three SD1 classifications. Due to sample size differences, the three histograms are all normalized for improved clarity.

Recovery was significantly higher for INC SD1 compared to UNCH SD1 ($p < 0.0001$) and was also significantly lower for DEC SD1 compared to UNCH SD1 ($p < 0.0001$).

Implications

These results indicate profound agreement between Kiviniemi et al.'s “HRV-only” model and the WHOOP multi-factor Recovery model. Agreement is greatest when recommending INC or UNCH training (green and yellow in the figure above). In cases when the Kiviniemi et al. model would recommend DEC training, WHOOP agrees only 46% of the time (red in the figure above) and is within 10% of a DEC recommendation 68% of the time. This suggests that the Kiviniemi approach is more conservative than that of WHOOP, and would therefore likely result in an overall lower prescribed training load.

For most users most of the time, the physiological markers of training readiness are fully captured in HRV. However, there exists edge cases in which an HRV-only based recommendation is wildly misleading, several of which are detailed further on in this publication. The Recovery algorithm, by considering additional inputs, is able to identify these exceptions, and therefore make better-informed conclusions about physiological state. Specifically, in the Recovery algorithm, weight is given to HRV, sleep and resting heart rate (RHR), while Kiviniemi et al. strictly based their training recommendations on HRV only.

We also note that there is a difference in the time period from which HRV is calculated. While academic studies such as Kiviniemi et al.'s study record R-R intervals to derive HRV right after awakening, WHOOP uses R-R intervals collected during sleep, where there is the smallest chance of exogenous factors influencing the recording. This would affect the amount of noise in the recording but would not fundamentally change the way the results are interpreted.

The following sections use WHOOP data to demonstrate the importance of including RHR and sleep in a training-prescription model.

Including Resting Heart Rate

RHR represents the number of times a heart beats during 60 seconds while at rest. It is widely understood that a lower RHR is associated with a higher level cardiovascular fitness. Many factors determine HR and RHR of an individual, however it is widely accepted that “stronger” hearts can pump more blood per contraction (Cardiac Output) to working muscles and organs. To this end, RHR was often used to monitor training status before new research provided a focus on HRV as a training tool.¹¹

Despite this shift in focus, there is some recent evidence that RHR provides important insights alongside HRV. In a 2013 review, Plews et al. addressed unequivocal results in research on the associations between HRV and training adaptation. Specifically, Plews et al. detailed limitations with using HRV alone to monitor exercise performance. The assumption that HRV alone can project an athlete's physiological readiness to perform relies on a direct linear relationship between HRV and R-R interval length, and therefore a direct relationship between HRV and fitness. In most early studies investigating HRV and performance, including the 2010 study by Kiviniemi et al., the subjects observed were only recreationally fit. Plews et al. were the first to note that this assumption does not generalize to elite athletes with a history of a high training load; for elite athletes with lower resting heart rates, decreases in HRV have been observed.¹² This has been attributed to two processes.

The first is a decrease in parasympathetic activity that has been observed to occur with reductions in training load (i.e. tapering). This may also be connected to increased sympathetic activity that comes from pre-competition stress.^{13,14} The second process, known as “parasympathetic saturation”, reflects the saturation of acetylcholine receptors at the myocyte level. A heightened vagal tone then causes longer parasympathetic control and therefore decreases HRV.¹⁵ Generally speaking, the relationship between RHR and HRV is quadratic rather than linear.

¹¹ Achten and Jeukendrup, 2003

¹² Plews et al., 2013

¹³ Mateo et al., 2012

¹⁴ Morales et al., 2013

¹⁵ Malik et al., 1993

Replicating the analysis from Plews et. al., **Figure 2** shows a plot of the natural log of RMSSD (as suggested by Plews et al.) versus the average R-R interval, for a WHOOP user that is a 3-time Olympian. The graph shows that the relationship between RHR and HRV seems to increase linearly until longer average R-R intervals (i.e. lower RHR), after which the relationship levels off.

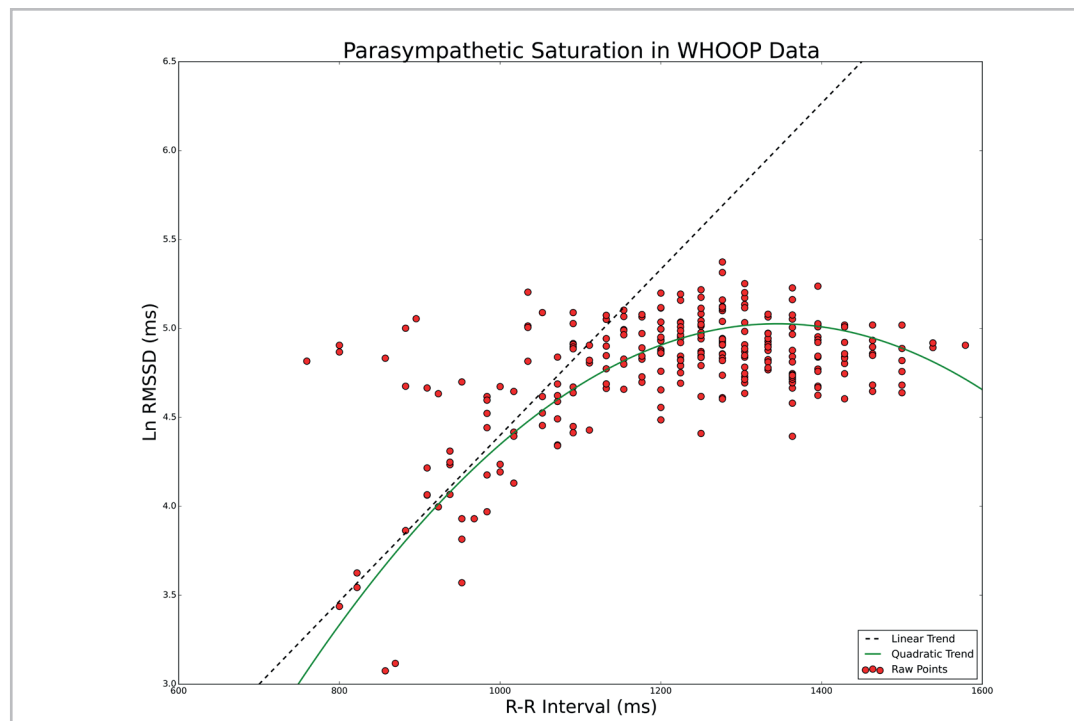


Figure 2. Parallel to analysis from Plews et al.: A demonstration of parasympathetic saturation through a plot of R-R interval length and Ln RMSSD. As R-R interval length increases, Ln RMSSD increases linearly and then saturates. The black dotted line represents a linear trend, while the green solid lines represents a quadratic trend.

Because elite athletes tend to have low RHR and go through high training loads, they are prone to parasympathetic saturation. An elite athlete who then trains based on strictly HRV, such as in the protocol tested by Kiviniemi et al. in their 2010 study, would be given incorrect training recommendations. When the athlete tapers before competition and is rested, HRV-based training would suggest that the athlete is not physiologically ready to perform due to decreased HRV from parasympathetic saturation. RHR clearly provides insight into an athlete's fitness level and can be used to differentiate maladaptation to training with parasympathetic saturation, and is therefore considered essential by WHOOP to include in Recovery.

Including Sleep Performance

The relationship between sleep and exercise performance has been widely explored in research for decades. There have been numerous studies demonstrating the negative effects of sleep deprivation on both physical and cognitive human functioning, including but not

limited to reaction times, puzzle solving, perception of effort, and exercise treadmill performance.^{16,17,18} There has also been research indicating that various forms of athletic performance improve with extra sleep.^{19,20,21}

Figure 3 and **Figure 4** demonstrate with WHOOP data the importance of sleep in athletes that cannot necessarily be reflected by HRV only. **Figure 5** shows a plot of reported energy levels, where a greater value represents greater energy and a lower value represents less energy, as a function of sleep duration for WHOOP users from September 1st, 2016 to November 2nd, 2016. **Figure 4** shows a plot of reported energy levels as a function of RMSSD.

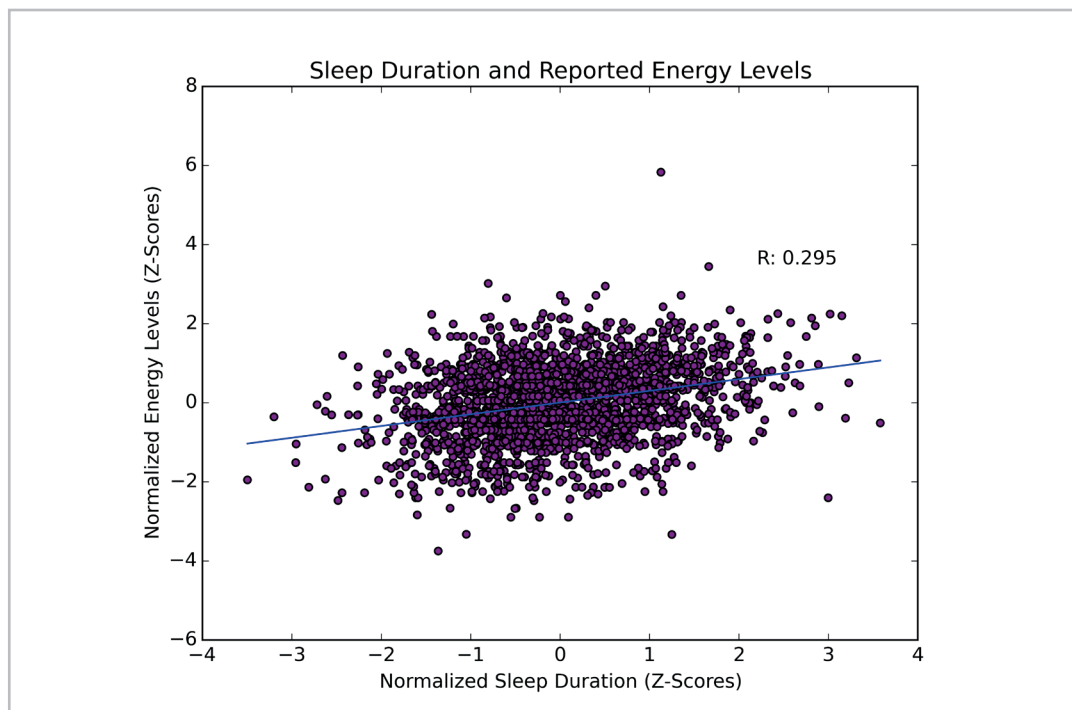


Figure 3. A scatterplot of sleep duration and reported energy levels for WHOOP users between 9.1.2016 and 11.2.2016. Both variables are standardized into z-scores because users may tend to skew themselves in either direction when reporting energy levels and because average sleep duration may vary depending on the user.

These visualizations demonstrate that there is almost no correlation between HRV and reported energy level ($R = 0.077$), while there is a slightly stronger positive relationship between sleep duration and reported energy level ($R = 0.295$). While the association between sleep and reported energy level is not very strong, it should be acknowledged that many factors could play a role in influencing reported energy levels beyond sleep duration and

¹⁶ Pilcher and Huffcutt, 1996

¹⁷ Walker and Stickgold, 2005

¹⁸ Oliver et al., 2009

¹⁹ Mah et al., 2011

²⁰ Schwartz and Simon, 2015

²¹ Venter, 2012

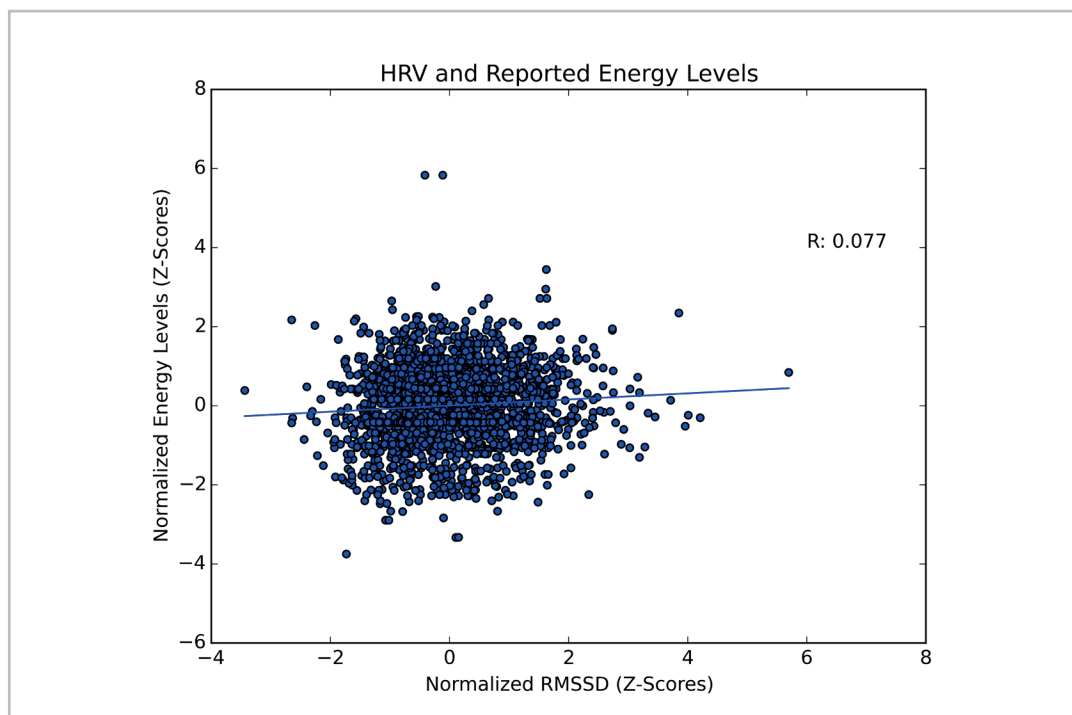


Figure 4. A scatterplot of RMSSD and reported energy levels for WHOOP users between 9.1.2016 and 11.2.2016. As in Figure 5, both variables are standardized into z-scores. RMSSD is standardized due to each individual having their own individual HRV range.

concerns with survey bias, including but not limited to the sleep stage right before awakening²² and sleep debt accumulated from sleep deprivation or restriction.

WHOOP aims to address this issue by the use of their proprietary Sleep Performance metric as an input to the Recovery algorithm. Sleep Performance is calculated by dividing total sleep time (a measure of sleep duration which does not give credit for periods of time during which the user was briefly awake) divided by a dynamic and personalized calculation of a user's sleep need, which takes into account a learned physiological baseline, strain over the course of the day, naps during the day, and sleep debt accumulated from past nights. Because Sleep Performance takes all of these into account, WHOOP is able to provide a more complete understanding of restedness than could simple calculations of time in bed or time asleep alone.

Figure 5 shows the distributions of Sleep Performance for each of WHOOP's four possible energy level survey responses. Using the non-parametric Mann-Whitney U-test, Sleep Performance was significantly different ($p < 0.05$ for energy level reported as "energized" versus "rested", $p < 0.0001$ for all other pairwise tests) for every reported level of energy.

²² Cavallero et al., 2003

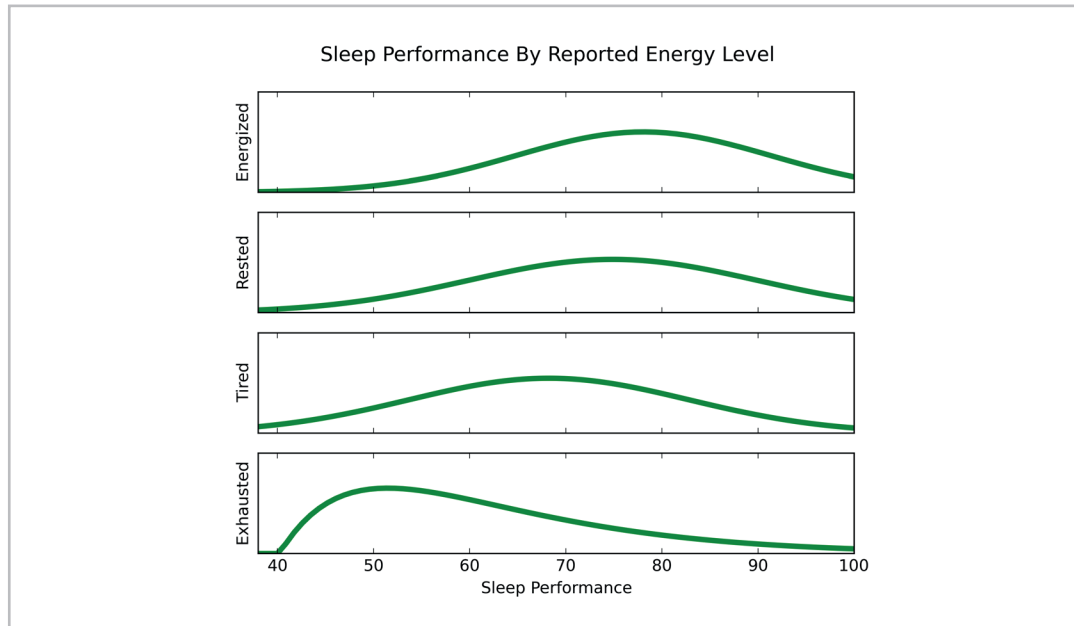


Figure 5. The distribution of Sleep Performance of WHOOP users from 9.1.2016 to 11.2.2016 for each of the four possible reported energy levels.

Figure 6 shows the distributions of RMSSD (normalized into z-scores) for each of WHOOP's four possible energy level survey responses. RMSSD does not change for each of the four energy levels.

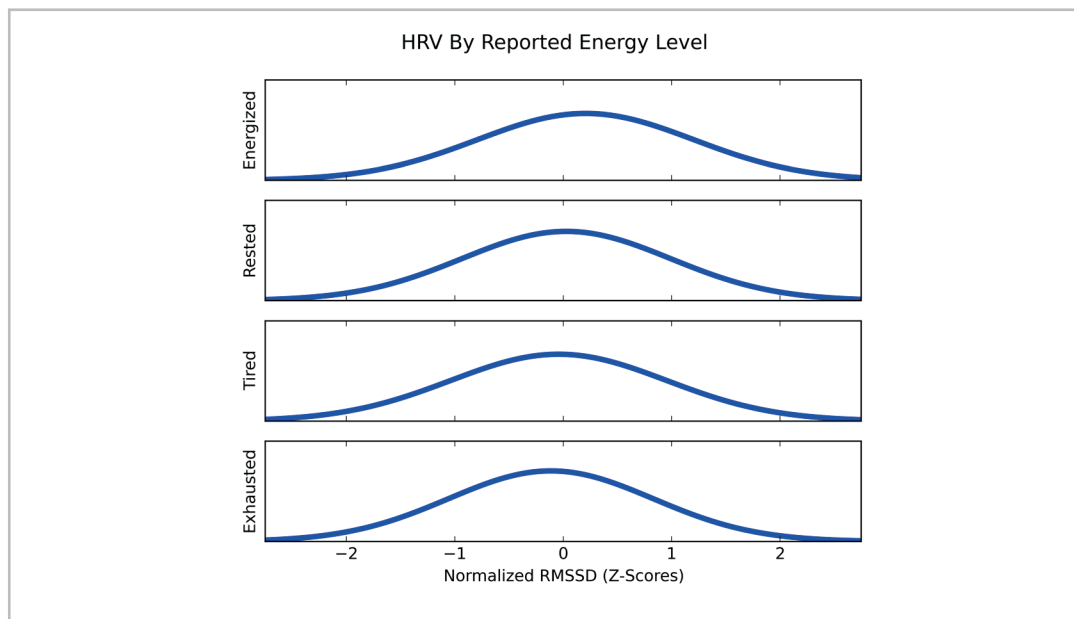


Figure 6. The distribution of RMSSD of WHOOP users from 9.1.2016 to 11.2.2016 for each of the four possible reported energy levels. As in Figure 6, RMSSD is standardized into z-scores.

Together, **Figures 3-6** demonstrate that Sleep Performance provides meaningful insight into assessing the body's ability to take on strain. However, it should be acknowledged that sleep alone does not provide a complete picture of recovery. For example, when feeling sick, one may sleep more, but may still not be physiologically ready to perform. WHOOP data shows that the average user who self-reports waking up feeling sick attained a slightly higher than average Sleep Performance (75.6% vs. 73.0%). This increase alone might, in isolation, be interpreted as a positive indicator of Recovery, but Recovery in the 4,125 instances of user-reported illness was actually lower than the overall WHOOP average Recovery (56.4% vs. 58.2%). In these cases, the apparent lack of agreement is being driven by decreasing HRV and increasing RHR.



Conclusion

This report demonstrates that the WHOOP physiological monitoring platform makes recommendations consistent with academic HRV research. It then reviews the strengths and weaknesses of an HRV-only training prescription model, ultimately concluding that the inclusion of RHR and sleep allows for a physiologically nuanced and better-informed training recommendation.

Most physiological research, including Kiviniemi et al.'s study, analyzes recreationally fit subjects. This publication illustrates several situations in which conclusions drawn from these types of studies do not generalize to physiologically unique populations, such as elite athletes. WHOOP uses its proprietary, continuous monitoring technology to do its physiology research on the elite athletes for whom the product is built. Doing our research directly on these elite athletes enables us to observe these edge cases which Kiviniemi et al.'s model would have missed, and therefore build a robust model that reflects a deeper understanding of the elite athlete population. Doing our own research in this less-studied population in addition to learning from the extant medical and academic literature allows us to improve the understanding of the human body and unlock peak performance.

References

- Achten, Juul, and Asker E. Jeukendrup. "Heart rate monitoring." *Sports medicine* 33, no. 7 (2003): 517-538.
- Al Haddad, H., P. B. Laursen, D. Chollet, S. Ahmaidi, and M. Buchheit. "Reliability of resting and postexercise heart rate measures." *International journal of sports medicine* 32, no. 08 (2011): 598-605.
- Breslow, Emily. WHOOP, Inc. Department of Physiology and Analytics. *Heart Rate Variability – A Coach's Review of the Uses and Value of HRV Data in Athletes* (Boston, MA: 2016).
- Cavallero, Corrado, and Francesco Versace. "Stage at awakening, sleep inertia and performance." *Sleep Res Online* 5 (2003): 89-97.
- Kiviniemi, Antti M., Arto J. Hautala, Hannu Kinnunen, and Mikko P. Tulppo. "Endurance training guided individually by daily heart rate variability measurements." *European journal of applied physiology* 101, no. 6 (2007): 743-751.
- Kiviniemi, Antti M., Arto J. Hautala, Hannu Kinnunen, Juuso Nissilä, Paula Virtanen, Jaana Karjalainen, and Mikko P. Tulppo. "Daily exercise prescription on the basis of HR variability among men and women." *Med Sci Sports Exerc* 42, no. 7 (2010): 1355-1363.
- Mah, Cheri D., Kenneth E. Mah, Eric J. Kezirian, and William C. Dement. "The effects of sleep extension on the athletic performance of collegiate basketball players." *Sleep* 34, no. 7 (2011): 943-950.
- Malik, Marek, and A. John Camm. "Components of heart rate variability—what they really mean and what we really measure." *The American journal of cardiology* 72, no. 11 (1993): 821-822.
- Mateo, M., Blasco-Lafarga, C., Martinez-Navarro, I., Guzman, J. F., and Zabala, M. (2012). Heart rate variability and pre-competitive anxiety in BMX discipline. *Eur. J. Appl. Physiol.* 112, 113–123. doi: 10.1007/s00421-011-1962-8
- Morales, J., V. Garcia, X. García-Massó, P. Salvá, and R. Escobar. "The use of heart rate variability in assessing precompetitive stress in high-standard judo athletes." *International journal of sports medicine* 34, no. 02 (2013): 144-151.
- Oliver, Samuel J., Ricardo JS Costa, Stewart J. Laing, James LJ Bilzon, and Neil P. Walsh. "One night of sleep deprivation decreases treadmill endurance performance." *European journal of applied physiology* 107, no. 2 (2009): 155-161.
- Pilcher, June J., and Allen J. Huffcutt. "Effects of sleep deprivation on performance: a meta-analysis." *Sleep: Journal of Sleep Research & Sleep Medicine* (1996).
- Plews, Daniel J., Paul B. Laursen, Andrew E. Kilding, and Martin Buchheit. "Heart rate variability in



elite triathletes, is variation in variability the key to effective training? A case comparison." *European journal of applied physiology* 112, no. 11 (2012): 3729-3741.

Plews, D. J., Laursen, P. B., Stanley, J., Kilding, A. E., & Buchheit, M. (2013). Training adaptation and heart rate variability in elite endurance athletes: opening the door to effective monitoring. *Sports medicine*, 43(9), 773-781.

Schwartz, Jennifer, and Richard D. Simon. "Sleep extension improves serving accuracy: A study with college varsity tennis players." *Physiology & behavior* 151 (2015): 541-544.

Task Force of the European Society of Cardiology. "Heart rate variability standards of measurement, physiological interpretation, and clinical use." *Eur Heart J* 17 (1996): 354-381.

Venter, Rachel E. "Role of sleep in performance and recovery of athletes: a review article." *South African Journal for Research in Sport, Physical Education and Recreation* 34, no. 1 (2012): 167-184.

Vesterinen, V., K. Häkkinen, E. Hynynen, J. Mikkola, L. Hokka, and A. Nummela. "Heart rate variability in prediction of individual adaptation to endurance training in recreational endurance runners." *Scandinavian journal of medicine & science in sports* 23, no. 2 (2013): 171-180.

Walker, Matthew P., and Robert Stickgold. "It's practice, with sleep, that makes perfect: implications of sleep-dependent learning and plasticity for skill performance." *Clinics in sports medicine* 24, no. 2 (2005): 301-317.

