Journal of Human Performance in Extreme Environments

Volume 18 | Issue 1 Article 1

Rim-to-Rim Wearables at the Canyon for Health (R2R WATCH): Physiological, Cognitive, and Biological Markers of Performance Decline in an Extreme Environment

Kristin Divis
Sandia National Laboratories, kmdivis@sandia.gov

Robert Abbott
Sandia National Laboratories, rgabbot@sandia.gov

Cathy Branda
Sandia National Laboratories, cbranda@sandia.gov

See next page for additional authors

Recommended Citation

Divis, Kristin; Abbott, Robert; Branda, Cathy; Emmanuel-Aviña, Glory; Femling, Jon; Huerta, J. Gabriel; Jelinková, Lucie; Jennings, J. Kyle; Pearce, Emily A.; Ries, Daniel; Sanchez, Danielle; and Silva, Austin (2023) "Rim-to-Rim Wearables at the Canyon for Health (R2R WATCH): Physiological, Cognitive, and Biological Markers of Performance Decline in an Extreme Environment," *Journal of Human Performance in Extreme Environments*: Vol. 18 : Iss. 1, Article 1.
DOI: 10.7771/2327-2937.1134
Available at: https://docs.lib.purdue.edu/jhpee/vol18/iss1/1

Rim-to-Rim Wearables at the Canyon for Health (R2R WATCH): Physiological, Cognitive, and Biological Markers of Performance Decline in an Extreme Environment

Cover Page Footnote

Acknowledgements We would like to thank the volunteers from the University of New Mexico, the Grand Canyon National Park rangers, and the Sandia researchers who braved the elements and odd hours to help collect data for the R2R WATCH project. This project was funded by the Defense Threat Reduction Agency (DTRA), Project CB10359. Sandia National Laboratories is a multi-mission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. SAND2020-0499 J. This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

Authors

Kristin Divis, Robert Abbott, Cathy Branda, Glory Emmanuel-Aviña, Jon Femling, J. Gabriel Huerta, Lucie Jelinková, J. Kyle Jennings, Emily A. Pearce, Daniel Ries, Danielle Sanchez, and Austin Silva

Rim-to-Rim Wearables at the Canyon for Health (R2R WATCH): Physiological, Cognitive, and Biological Markers of Performance Decline in an Extreme Environment

Kristin Divis[1], Robert Abbott[1], Cathy Branda[1], Glory Emmanuel-Aviña[1], Jon Femling[2],
J. Gabriel Huerta[1], Lucie Jelinková[2], J. Kyle Jennings[1], Emily A. Pearce[2], Daniel Ries[1],
Danielle Sanchez[1], and Austin Silva[1]

[1]*Sandia National Laboratories*
[2]*University of New Mexico*

Abstract

Success in extreme environments comes with a cost of subtle performance decrements that if not mitigated properly can lead to life-threatening consequences. Identification and prediction of performance decline could alleviate deleterious consequences and enhance success in challenging and high-risk operations. The Rim-to-Rim Wearables at the Canyon for Health (R2R WATCH) project was designed to examine the cognitive, physiological, and biological markers of performance decline in the extreme environment of the Grand Canyon Rim-to-Rim (R2R) hike. The study utilized commercial off-the-shelf cognitive and physiological monitoring techniques, along with subjective self-assessments and hematologic measurements to determine subject performance and changes across the hike. The multiyear effort collected these multiple data streams in parallel on a large sample of participants hiking the R2R, leading to a rich and complex data set. This article describes the methodology and its evolution as devices and measurements were assessed after each data collection event. It also highlights a subset of the patterns of results found across the data streams. Subsequent work will draw on this data set to focus on building more sophisticated, predictive statistical models and dive deeper into specific analyses (such as the physiological and biological profiles of hikers who were left behind by their hiking partners).

Keywords: physiological markers, cognitive markers, biological markers, human performance, Grand Canyon

---

1 Introduction

In the face of a spectrum of environmental stressors, people strive to maintain high performance across a wide range of tasks. Athletes push themselves to ever better performances, military personnel and wildland firefighters strive to complete their missions safely, and park rangers work to keep visitors safe no matter the environmental conditions. Technology has come a long way in helping us to shape our environment, but human nature indelibly finds ways to push our boundaries and see just what we can do despite environmental constraints.

Nature enthusiasts attempting to traverse the Grand Canyon from the South Rim to the North Rim (or vice versa) in a single day often have this pioneering mindset. The Rim-to-Rim (R2R) hike has become increasingly popular, with more than 1,000 hikers observed attempting to complete it during the opening and closing weekends of the 2015 season (Pearce et al., 2019). Hiking R2R provides a range of environmental stressors, making it an excellent candidate for studying human performance in an extreme environment.

Here, we first review the physiological, cognitive, and biological consequences of a variety of stressors found in extreme environments and then describe in detail the study of R2R hikers at the Grand Canyon.

1.1 Environmental Temperature

High environmental temperatures are one of the leading causes of environmental fatalities in the United States. The primary mechanism for dealing with thermal stress by humans is through evaporative cooling mediated primarily through sweating. This core mechanism helps to explain the multifactorial compensatory changes that occur with increased thermal load, including increased cardiac output, blood flow to the skin, and sweat (Sawka & Wenger, 1988). These compensatory

http://dx.doi.org/10.7771/2327-2937.1134

efforts may stress the cardiovascular system (Rodahl, 2002); dehydration, heat cramps, heat syncope, heat exhaustion, heat stroke, and/or electrolyte imbalance can result as the body tries to prevent a rise in core body temperature (Lipman et al., 2014). When these compensatory systems are inadequate, body temperature begins to rise outside of tolerable norms resulting in cell damage, increased gastrointestinal permeability, inflammation, organ dysfunction, and potentially death (Epstein & Yanovich, 2019). Many of these declines can be detected hematologically before catastrophe occurs, including cell damage and dehydration (discussed later).

Even small rises in environmental temperature can disrupt cognitive and physiological performance (for reviews, see Hancock & Vasmatzidis, 2003; Staal, 2004); skills such as time estimation, reaction time, tracking, and cognitive ability show a decline at environmental temperatures above about 30℃ (Hancock, 1982). Maruff and colleagues (2006) found decrements in psychomotor speed, attention, and higher-order executive functions in a male subject as he progressed through a 17-day hike across the Simpson Desert in Australia, where the environmental temperature fluctuated from 13 to 42℃. A meta-analysis of the field (Hancock et al., 2007) revealed that psychomotor and perceptual abilities appear to be the most susceptible to heat stress, with cognitive abilities tending to be more robust; these reductions in performance likely occur due to the need to divert resources to cope with the thermal environmental pressures.

High environmental temperatures pose a significant risk to Grand Canyon hikers. From 2004 to 2009, emergency medical services in the Grand Canyon National Park responded to 474 cases of heat-related illnesses—six of which resulted in fatalities (Noe et al., 2013).

While heat remains one of the top risks to hikers crossing the Grand Canyon, exposure to cold temperatures can also lead to deleterious effects on physical and cognitive performance. In cold environments, short-term exposure (less than 120 minutes) in temperatures less than 10℃ decreased performance on reasoning, learning, and memory tasks (Pilcher et al., 2002). Exposure to cold air has also been shown to reduce performance on serial choice reaction time tasks (Ellis, 1982; Ellis et al., 1985). Navy special operations forces training in a cold environment suffered from reduced hand strength and fine motor skills (Hyde et al., 1997). Even mild hypothermia results in compensatory changes that range from changes in subjective measures of discomfort, to physiologic changes in cardiovascular function, to diuresis and electrolyte changes (Brown et al., 2012).

1.2 Dehydration

The detrimental effects of dehydration on performance have been well studied in the scientific community in both controlled laboratory and field settings (for reviews, see Adan, 2012; Murray, 2007; Popkin et al., 2010; Sawka et al., 1983). The negative impact of dehydration increases as severity of dehydration increases, with symptoms first appearing at as little as 1–2% bodyweight reduction due to dehydration and becoming worse as dehydration increases (Adan, 2012; Cheuvront et al., 2003; Murray, 2007; Sawka et al., 2015). The adverse effects of dehydration are exacerbated by extended periods of sustained activity, increased heat load, increased intensity, lack of heat acclimatization, and/or high altitudes (Cheuvront et al., 2005; Murray, 2007; Sawka et al., 1983, 2015).

Research on ultramarathoners reveals that dehydration is common in endurance athletes (Holtzhausen & Noakes, 1995), and mild states of dehydration are linked to greater rates of success in endurance competition (Knechtle et al., 2010, 2012; Landman et al., 2012). This apparent contradiction is likely related to how the measurements were performed and a limited understanding of the precise body compartments from which fluid loss has occurred. In the case of simple dehydration, volume loss seen primarily as a decrease in circulatory volume is clearly detrimental. However, in the case of endurance athletes, where glycogen metabolism results in a release of water into the circulatory system, total body dehydration is apparent with maintenance of circulatory plasma volume (Hoffman et al., 2019). In fact, in the setting of endurance events maintenance of body weight with a perceived goal of euhydration may come at the cost of profound electrolyte disturbance and performance decrement. If the expected losses during extended activities are exceeded, dehydration can lead to a vicious interplay of reduced sweating, increased core temperature, increased heart rate and cardiovascular strain, and changes in both the metabolic and central nervous systems (Murray, 2007; Popkin et al., 2010; Sawka et al., 1983; Strydom & Holdsworth, 1968). As the body compensates for dehydration, there is a contraction of circulatory volume that can be measured by hemoconcentration; as dehydration persists and blood flow to vital organs such as the kidneys becomes restricted, an increase in hematologic markers such as creatinine and blood urea nitrogen is easily measured. As a person experiences increased thermal stress, they begin by sweating efficiently, reabsorbing sodium and chloride. As thermal stress continues the sweating mechanism becomes less efficient with both loss of fluids and increased loss of sodium and chloride. This inefficiency can then lead to electrolyte disturbances that can be compounded by inappropriate oral replacement (Baker, 2017).

Dehydrated individuals also tend to give higher subjective effort and fatigue ratings and perform worse on attention, psychomotor, and short-term memory tasks (Adan, 2012; Popkin et al., 2010). Adan (2012) noted in her review of the literature on cognition that dehydration does not appear to have a strong influence on executive function, long-term memory, or working memory capabilities; however, the inconsistencies in methodologies between studies and individual differences make the intersection between dehydration and cognition an area in need of further refinement.

1.3 Altitude

Unacclimated individuals operating in moderate (e.g., 1,500 to 3,500 m) to high (e.g., 3,500 to 5,300 m) altitude environments may experience altitude-related symptoms (see Cudaback, 1984; Keupper & Classen, 2002). Symptoms can range from increased breathing and heart rates (starting at as little as 1,000 m; Bärtsch & Saltin, 2008; Higgins et al., 2010) to acute mountain sickness or even life-threatening pulmonary or cerebral edema (Cudabeck, 1984). Air in higher-altitude environments has a lower partial pressure of oxygen, which leads to reduced oxygen throughout the body (hypoxia). Symptoms of acute mountain sickness due to hypoxia include headaches, insomnia, lack of appetite, vomiting, shortness of breath, dizziness, and/or lack of muscle control; in extreme cases, seizures and pulmonary and/or cerebral edema can also occur (see Banderet & Burse, 1991; Cudabeck, 1984; Maggiorini, 1990; Virués-Ortega et al., 2004). Several hematologic changes have been described as the body adapts to altitude. These include changes in partial pressures of gases, increases in hematocrit, and a shift in the balance of carbon dioxide and bicarbonate in order to maintain homeostatic pH (Zouboules et. al, 2018). There is an acute adaptive change when a new altitude has been reached that then, with changes in renal compensatory function, adapt further in order to maintain homeostasis.

Altitude-related symptoms are more likely to occur at higher elevations, and some symptoms tend only to occur in high-altitude environments (see Cudabeck, 1984). They are also more likely to occur with fast-paced ascents; however, symptoms can vary based on time since ascent, prior acclimation, and other individual differences (Cudabeck, 1984). Significant performance decrements have been found even when the afflicted individual does not report feeling symptomatic (Cudabeck, 1984; Stück et al., 2005), a finding particularly concerning for situations involving high-consequence decision making (see also Wickens et al., 2015).

Reviews of the interplay between cognition and altitude reveal mixed effects driven by variants in experimental design and individual differences (see Bahrke & Shukitt-Hale, 1993; Lowe et al., 2007; Virués-Ortega et al., 2004). Reaction time in complex tasks is one of the most accepted and sensitive measures, with multiple studies revealing poorer performance as altitude increases (see Virués-Ortega et al., 2004). Memory and learning—particularly encoding—appear to also be sensitive to changes in altitude (Kramer et al., 1993; Virués-Ortega et al., 2004). While decrements in psychomotor speed and precision have been found at high elevations, this may be due to the corresponding increases in fatigue rather than altitude itself; increased elevation has also been linked to visual hallucinations and decreases in language and metacognitive abilities (see Virués-Ortega et al., 2004).

1.4 Fatigue

Human performance tends to drop as one becomes fatigued (for a review, see Staal, 2004). Definitions of fatigue have both acute and chronic implications. There is no single definition or metric of fatigue, but several domains have been defined, including mental, central, and peripheral (Finsterer & Mahjoub, 2014). Muscle function as measured by metrics such as peak power, maximum repetitions, and maximum contraction has been studied in controlled settings (Areta & Hopkins, 2018). Fatigue from a biochemical standpoint can be measured by the consumption of energy-producing resources. Skeletal muscle consumption of glycogen during exercise has been shown to be related to nutrition, exercise intensity, and extent of exercise (Halperin et al., 2015). Unfortunately, muscle glycogen concentration is impractical to measure in the field, but there has been some work evaluating systemic markers of fatigue with the goal of understanding the nonlocal effects of fatigue. These markers include changes in blood lactate, pH, and potassium (Areta & Hopkins, 2018).

Two causes of fatigue relevant to cognitive performance during the R2R hike are lack of sleep and extended periods of activity. In a meta-analysis of the literature on changes in performance due to sleep deprivation, Pilcher and Huffcutt (1996) found deficits in mood, cognition, and psychomotor abilities (in descending order of magnitude); overall, participants who were sleep-deprived functioned like the bottom 10% of participants who were not. Subsequent research also highlighted that sleep is particularly important to higher-level thinking such as executive function and working memory, along with decision making in unexpected or atypical situations (Durmer & Dinges, 2005; Harrison & Horne, 2000). The effects of exercise-induced fatigue on cognitive performance are more complex and understudied when it comes to long-duration exercise activities. The majority of studies have looked at the effects of exercise-induced fatigue over relatively short durations (e.g., an hour or less). The results are mixed—relatively short bouts of exercise can actually *improve* performance across a range of tasks but there have also been noted reductions in performance metrics such as response time and accuracy along with memory, with some researchers hypothesizing that the performance pattern is similar to the classic Yerkes–Dodson curve of activation (see Moore et al., 2012; Tomporowski, 2003; Yerkes & Dodson, 1908). Studies looking at extended endurance races (which often combine exercise-induced fatigue with sleep deprivation and extreme environmental temperatures) have revealed that while deficits in cognitive performance are usually found, they do not necessarily decline in a steady or linear fashion, making it challenging to pinpoint the influence of underlying causal factors such as time of day, length of activity, proximity to finish, and environmental temperature fluctuation (e.g., Doppelmayr et al., 2005; Hurdiel et al., 2015, 2018).

1.5 R2R WATCH

1.5.1 Environment

The Grand Canyon R2R is a grueling hike spanning 24.2 miles, starting at 7,260 feet in elevation at the South Rim and descending 4,780 feet in elevation down the Kaibab Trail to the Colorado River; it then crosses the bottom of the canyon and ascends 5,760 feet in elevation to North Rim (at 8,240 ft in elevation; Rim to Rim, 2012; also see Figure 1). During the hike, there are few opportunities for shade or shelter along the trail, and water sources are unreliable. The gradient of the trails is consistently steep with variable footing, including loose gravel and steep railroad tie steps. Hikers face extreme temperature differentials, with an average temperature difference of 30 degrees Fahrenheit from the rim to the canyon bottom. Summer temperatures at the canyon bottom may exceed 120 degrees Fahrenheit in the shade, with significantly hotter temperatures felt on the trail surface and when exposed to the sun. Timing is important when it comes to the R2R, and savvy hikers often hike in the cooler months of spring or fall and start during the predawn dark of early morning to avoid being in the bottom of the canyon during the hottest part of the day (mean start time during opening and closing weekends in May and October of the 2015 season was 05:30 a.m.; Pearce et al., 2019).

Common R2R mistakes include starting too late in the day or not bringing adequate supplies (e.g., assuming water will be available along the trail). Even prepared hikers can run into trouble on the R2R. Over the last decade, the Grand Canyon National Park's emergency medical services staff responded to an average of approximately 300 search and rescue incidents and over 1,000 emergency medical services incidents each year throughout all of the Grand Canyon National Park (National Park Service, 2019), with anecdotal reports of increased requests for emergency medical services along the R2R trail itself (Pearce et al., 2019). Hikers suffer from a variety of ailments (e.g., hyponatremia, injury, etc.) and the park has a dedicated preventive search and rescue team patrolling the trails for struggling hikers. However, R2R hiking has become increasingly popular, with multiple visitors attempting to complete it in less than 24 hours.

1.5.2 Partnership and Context

The R2R WATCH study was funded by the Defense Threat Reduction Agency (Project CB10359) and led by Sandia National Laboratories in partnership with the University of New Mexico (UNM) School of Medicine's Department of Emergency Medicine and the support of Grand Canyon National Park. It built upon early work

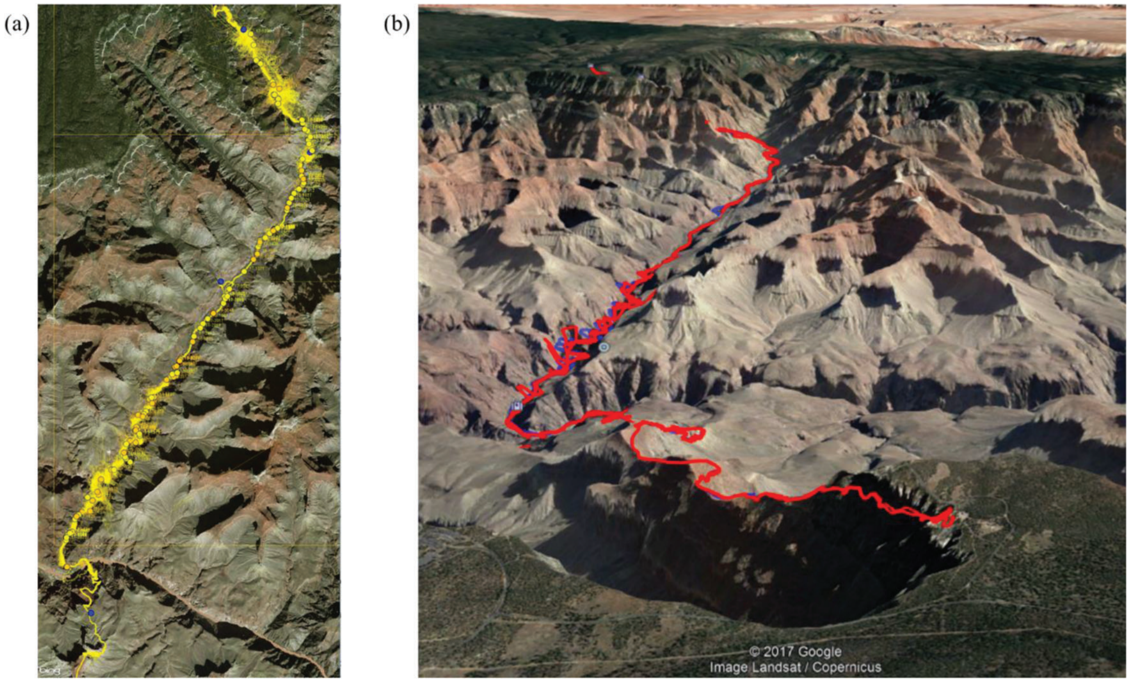

Figure 1. Visualizations of Grand Canyon R2R hike from South Kaibab trailhead to North Kaibab trailhead. Viewed from (a) a satellite image with the yellow lines and circles coming from GPS tracks of R2R WATCH participants and (b) a satellite image from the south rim, with the red line showing the GPS track from a single participant. Note that the GPS tracks are not always aligned with the trail due to error in the GPS measure from poor satellite reception in the canyon (especially in steep sections).

conducted by the UNM team examining the influence of physiological and nutritional components on R2R hikers (for further details, see Jelínková et al., 2017; Pearce et al., 2019).

R2R WATCH collected data on hikers at the Grand Canyon four times over the course of three years, with the first serving as a pilot study (Divis et al., 2018; Emmanuel-Aviña et al., 2017). This article focuses on the data from the three studies after the pilot study.[1]

2 Methodology

2.1 Overall Methodology

We collected data from R2R hikers at the Grand Canyon once a year over the course of three years (not including the pilot study; see Divis et al., 2018; Emmanuel-Aviña et al., 2017). Complementary to the goal of identifying markers of performance decline, we also evaluated the robustness and appropriateness of various devices and techniques in this environment. While the broad methodology remained the same across all three studies, the details of the experimental design were adjusted between studies as needed. Techniques and devices that were strong performers from the beginning were replicated across the studies; underperforming ones were replaced or eliminated.

Each study collected physiological, environmental, cognitive, blood, and survey data on R2R hikers. The physiological and environmental data were collected via wearable devices (e.g., smart watches) that passively recorded throughout the hike. Basic measurements such as weight and peripheral capillary oxygen saturation (SpO_2) were also taken at the start and finish of the hike. The cognitive data came from a short set of tasks completed on a mobile device periodically throughout the hike. Hikers also completed surveys at the start, midpoint, and end of the hike (a subset also completed a follow-up survey a few weeks after the hike). Blood samples were drawn before and after the hike. See the Appendix for details on which metrics were included in each study.

The following sections cover the methodology, including any changes, for all three studies.

2.2 Study 1: Details

Study 1 focused on testing a breadth of device packages and device configurations following results from our pilot study (see Divis et al., 2018; Emmanuel-Aviña et al., 2017). We enrolled participants both in advance and at the trailhead, achieving a relatively large sample size in this study. Our goal was to determine which design elements, devices, and package configurations provided data that best allowed us to answer our research questions.

2.2.1 Participants

Participation in the R2R WATCH study was voluntary. Grand Canyon R2R hikers who planned on completing the Grand Canyon R2R hike in a single day during Study 1 data collection (two days in May during the opening weekend of the 2017 season) were invited to participate in the study. Both military and civilian hikers participated. Prior to data collection, military participants were invited to participate in the study as part of a training exercise (they were compensated for their travel costs per government rates). Civilian participants were asked if they would like to participate in the study after arriving at the trailhead and indicating they were planning on doing the R2R in a single day. Researchers halted enrollment in the study by 9:00 a.m. each day or when all packages of wearable devices had been used, whichever came first.

Twenty-five military hikers (76% male) and 48 civilian hikers (46% male) volunteered to participate in the study, for a total of 73 participants.

2.2.2 Design and Materials

Wearable Devices. Participants were outfitted with a package of wearable devices. We included both "basic" (lower cost, lighter weight, lower functionality) packages and "advanced" (higher cost, higher weight, higher functionality) packages in this study. All packages included a mobile device with a set of cognitive assessments, a fitness watch, and an environment sensor. The advanced packages added additional devices and capabilities such as electrocardiograph (ECG) heart rate monitoring (rather than just optical), more accurate elevation monitoring via a barometric altimeter (rather than GPS, which performs poorly in steep sections of canyons), skin temperature monitoring, and foot-based cadence monitoring (superior to wrist-based cadence). These devices were chosen following bench testing and the pilot study run the prior closing weekend at the Grand Canyon (Divis et al., 2018; Emmanuel-Aviña et al., 2017). We intentionally included a variety of devices in Study 1 in order to test the viability of a range of commercial off-the-shelf (COTS) devices within the Grand Canyon R2R environment. See Table 1 for all package options, devices, and capabilities.

Cognitive Tasks. The cognitive assessments were administered via a customized version of Digital Artefact's BrainBaseline application.[2] They included two Likert-scale subjective fatigue questions (one each for mental and physical fatigue), a visual short-term memory task (VSTM; Cowan, 2001; Luck & Vogel, 1997), and a Go/No-Go task (Conners & Sitarenios, 2011). Criteria for choosing tasks included:

---

[1] All R2R WATCH human research activities were conducted with a research permit from the National Park Service and reviewed and approved by the Sandia National Laboratories Human Subjects Board and the University of New Mexico Health Sciences Center Human Research Protections Office.

Table 1
Wearable device package options in Study 1.

		Package:			
Device	**Metrics**	**Basic-1**	**Basic-2**	**Advanced-1**	**Advanced-2**
Apple iPod touch (6th generation)	BrainBaseline cognitive assessments	✓	✓	✓	✓
SensorPush	Thermometer; hygrometer	✓	✓	✓	✓
Garmin vívoactive HR	Wrist-based optical heart rate monitor; accelerometer; altimeter; GPS	✓	—	(✓)	(✓)
Fitbit Charge HR	Wrist-based optical heart rate monitor; accelerometer; altimeter	—	✓	—	—
Garmin fēnix 3 HR	Wrist-based optical heart rate monitor; accelerometer; barometric altimeter; GPS	—	—	✓	—
Suunto Spartan Ultra	Accelerometer; barometric altimeter; GPS	—	—	—	✓
Empatica E4	Wrist-based optical heart rate monitor; sleep monitor; galvanic skin response monitor; thermometer (underside of watch)	—	—	(✓)	(✓)
Garmin eTrex 10	GPS	(✓)	—	—	—
Wahoo TICKRx	ECG heart rate monitor; accelerometer	—	—	✓	—
Suunto Smart Sensor	ECG heart rate monitor	—	—	—	✓
LifeBEAM SmartHat	Forehead-based optical heart rate monitor	—	—	✓	✓
Garmin Foot Pod	Accelerometer	—	—	✓	—
Polar Stride Sensor Bluetooth Smart	Accelerometer	—	—	—	✓
Garmin tempe	Thermometer (under chest strap)	—	—	✓	—
Myontec Mbody Shorts	Quadricep and hamstring muscle group monitoring	—	—	(✓)	—

Note: Checkmarks without parentheses indicate the device was included in all packages of that type. Dashes indicate the device was not included in that package. Checkmarks with parentheses indicate the device was included in a subset of the packages of that type. For example, for watch devices an Advanced-1 package might only include a Garmin fēnix 3 HR watch or both a Garmin fēnix 3 HR watch and Garmin vívoactive HR watch (worn on opposite wrists).

potential connection to fatigue, ability to be successfully self-administered, minimal administration time, reduced practice effects, touchscreen compatibility, and ability to fit on small screen. The VSTM task consisted of 34 trials (50% match); the Go/No-Go task consisted of 50 trials (80% "go") with a delay of 500 to 1850 ms.[3] Performance feedback (including the participant's performance on previous sessions) was given after completing each task. Order of trial type was randomized within each session of a task. The fatigue questions were always administered first; VSTM and Go/No-Go task order was counterbalanced between participants.

Surveys. Participants were given surveys at the start, midpoint, and end of the hike. The start and end surveys also included space for researchers to fill in measurements such as pack weight, heart rate, and SpO_2. Participants were also given a questionnaire at the end of the post-hike survey with three personality assessments: the Big Five Inventory 10-item version (BFI-10; Rammstedt & John, 2007), the Recklessness component of International Personality Item Pool (IPIP; Goldberg, 1999; Goldberg et al., 2006) Temperament and Character Inventory (TCI), and the Citizenship/Teamwork component of Peterson and Seligman's Revised IPIP-Values in Action scale (IPIP-VIA; Peterson & Seligman, 2004). See the Appendix for further information on survey questions.

Blood. Whole blood was collected by venipuncture before and after the R2R. At each timepoint, one aliquot was placed in a serum separator tube, allowed to clot for approximately thirty minutes, and spun down for fifteen minutes. An additional aliquot was placed in a plasma tube with an anticoagulant. Serum and plasma samples were kept refrigerated for up to three days and frozen at −80℃ upon returning from the Grand Canyon. Serum samples were analyzed by TriCore Reference Laboratories for a basic metabolic panel and creatine kinase.

2.2.3 Procedure

Because military participants were assigned advanced packages (see Table 1) with devices that were more challenging to put on at the trailhead, researchers met them the afternoon prior to the hike and gave the hikers the subset of wearable devices that they would need to wear to the trailhead (chest strap, Mbody shorts, and Polar foot pod); military participants also gave pre-hike blood samples at this time (rather than at the trailhead). Civilian hikers were asked at the trailhead if they were planning on hiking the R2R in a single day and, if so, whether they would like to participate in our study. All other procedures were identical for military and civilian hikers.

---

[2] See https://www.brainbaseline.com for additional details.

[3] The pilot study (Divis et al., 2018) revealed that the set of pilot cognitive tasks was too long and compliance was lower than preferred. We sought ways to shorten the task to alleviate those issues. Analysis of the pilot data revealed the Flanker task provided no significant additional information and the trial numbers of the VSTM and Go/No-Go tasks could be cut off in half. We implemented those changes for Study 1 of the current paper, along with feedback (as a motivator to continue doing the task throughout the hike). This methodology was further refined in subsequent studies.

Hikers who agreed to participate in the study were met at the South Kaibab trailhead at the Grand Canyon. Participants gave verbal consent to the study and completed the pre-hike survey (including measurements such as SpO_2). They also gave a pre-hike blood sample and were assigned a wearable device package. Military participants received advanced packages and civilian participants received lightweight, basic packages. All devices were started and placed on participants or their backpacks. All devices (except the cognitive tasks) passively recorded data, with no input needed from the participants. Hikers completed one session of the cognitive assessments at the trailhead, with a researcher present to answer questions. Participants were instructed to complete the cognitive tasks approximately every three hours when they were at a good stopping point along the trail (the mobile device chimed when three hours had passed since the last session). After finishing the first session of the cognitive tasks, hikers' devices were given a final check, and participants were directed to the trail to begin their hike.

Researchers were stationed at Phantom Ranch at the bottom of the canyon to field any questions or concerns participants had in the middle of the hike and to administer the mid-hike survey. Upon completing the hike, participants were met and congratulated at the North Kaibab trailhead. Researchers offered the hikers a cold beverage and place to rest prior to finishing the remaining items in the study. The wearable devices were taken off the participants and stopped, and post-hike blood samples were drawn. Participants also completed a final session of the cognitive assessments and filled out the post-hike survey, including the personality questionnaire.

2.3 Study 2: Details

Study 2 was designed to be more focused and streamlined, narrowing in on the best devices and methodologies based on our findings from both Study 1 and the pilot study. Instead of four separate wearable device package configurations, we focused on a single high-quality package. We also targeted pre-enrolling all participants the night prior to the hike rather than catching civilians at the trailhead. Pre-enrollment the night prior (1) helps to minimize the time needed for the study early in the morning when participants are hoping to quickly get started on their hike and (2) allows for multiple training sessions on the cognitive assessments to reduce practice effects. We also updated the cognitive task timing to be location-based rather than timing-based to increase consistency in completing the tasks.

2.3.1 Participants

Once again, both military and civilian hikers planning to complete the R2R in a single day were invited to voluntarily participate in the study (one day in October during the 2017 season closing weekend). The study was announced via military list serves, community pages, and R2R social media outlets. Hikers already planning on completing the R2R were encouraged to contact the R2R WATCH research team if they were interested in joining the study. In contrast to Study 1, both military and civilian hikers were allowed to sign up in advance (pre-enroll) rather than at the trailhead. Due to last-minute deployments in our military population and civilian cancelations, we also opportunistically enrolled an additional eight civilian participants at the trailhead (as in Study 1) once all the pre-enrolled participants had started their hike.

Twelve military hikers (92% male) and 15 civilian hikers (53% male) participated in the study, for a total of 27 participants.

2.3.2 Design and Materials

In Study 2, all participants were outfitted with the same types of devices: a mobile device with the cognitive assessments, an environmental sensor, a fitness watch, a chest strap and body temperature sensor, and a foot pod. See Table 2 for further details.

As in Study 1, participants used the BrainBaseline application on the mobile device for the cognitive assessments. Study 2 updated the application to include three training sessions, no post-task feedback, and new session timing. Instead of completing the cognitive assessment at the trailheads and every three hours as in Study 1, participants in Study 2 completed the cognitive tasks at the trailheads and approximately every five miles during the hike (their watches buzzed as a reminder). They also worked through the cognitive tasks three times the day prior to arriving at the trailhead (training sessions), and limits were placed on how long a given trial could last (e.g., 10 seconds, which is well beyond a reasonable time to complete the trial). Every other aspect of the cognitive assessment was identical to Study 1. These updates were intended to further reduce practice effects, increase compliance (i.e., completing the cognitive tasks at the correct time), prevent abnormally long sessions, and increase consistency between participants (by focusing on spatial location rather than timing).[4] Only minor modifications were made to the survey questions, with the exception of adding an online follow-up survey. See the Appendix.

---

[4] Study 1 and the earlier pilot study (Divis et al., 2018) revealed substantial variability between participants in number of cognitive assessment sessions completed when sessions were based on time rather than location. Moving to a location-based cue helped to alleviate those differences between participants. Additionally, including training sessions substantially reduced the practice effects previously seen.

Table 2
Wearable devices in Study 2.

Device	Metrics
Apple iPod touch (6th generation)	BrainBaseline cognitive assessments
SensorPush	Thermometer; hygrometer
Garmin fēnix 3 or 5 HR	Wrist-based optical heart rate monitor; accelerometer; barometric altimeter; GPS
Wahoo TICKR or TICKRx	ECG heart rate monitor; accelerometer
Garmin Foot Pod	Accelerometer
Garmin tempe	Thermometer (under chest strap)

2.3.3 Procedure

Researchers met the pre-enrolled study participants the afternoon prior to the hike.[5] Any pre-hike survey questions that could be completed in advance (e.g., demographics) were filled out and the pre-hike blood samples were collected. Researchers also gave the participants the chest strap, foot pod, and mobile device. During this meeting, researchers explained the cognitive assessments and had each participant work through them twice while the researcher was present to answer any questions. Participants were instructed to complete the cognitive tasks one more time before going to bed (for a total of three training sessions on the cognitive tasks) and to wear the other devices to the trailhead the next morning.

The procedure at the South Kaibab trailhead for Study 2 for pre-enrolled participants was similar to that for the military participants in Study 1 for the remaining metrics. Participants checked in, remaining pre-hike survey questions and measurements (e.g., SpO_2) were collected, the remaining wearable devices were placed on the participant (e.g., watch), all devices were started, and participants completed a session of the cognitive assessments before beginning their hike.

Participants were instructed to complete the cognitive assessments approximately every five miles along the trail (their watches were set to buzz at five-mile increments based on GPS distance). The remainder of the procedure was identical to that of Study 1, except participants could also access a follow-up survey online a few weeks after the hike.

2.4 Study 3: Details

Study 3 closely replicated the methodology for the pre-enrolled population from Study 2. The changes that were made allowed more variables to be measured (rather than taking away or significantly changing previously measured factors). During this data collection, we had access to an on-site blood gas analyzer (allowing us to measure previously unavailable analytes). The chest strap was also updated to one which provided more types of data, and an additional cognitive assessment was included.

Study 3 allowed for not only a replication of previous findings on a new population on a different day but also provided the possibility of combining data across the related collections for a more powerful look at underlying effects.

2.4.1 Participants

Once again, military list serves, community pages, and R2R social media outlets were used to invite both military and civilian hikers planning to complete the R2R in a single day to voluntarily participate in the study (data collection was in September 2018). All participants were pre-enrolled.

Sixteen military hikers (75% male) and 22 civilian hikers (77% male) participated in the study, for a total of 38 participants.

2.4.2 Design and Materials

Study 3 was similar to Study 2 in terms of design and materials with the following exceptions.

The Equivital EQ02 device was the primary heart rate measurement system. It could collect full ECG traces across the hike, allowing for higher resolution of heart signals. When the device was not able to be used (often due to fitting issues of the device or lack of available fabric sensor holsters in a suitable size), the Wahoo TICKRx was used instead (four participants). See Table 3.

The experimental design for the cognitive assessments was identical to that of Study 2 except for the addition of a third assessment. The Balloon Analogue Risk Task (BART) was added to the end of each set of tasks. The BART is a classic risk-taking task that looks at the balance between risky behavior and reward (Lejuez et al., 2002; Wallsten et al., 2005; White et al., 2008). Every time participants completed the cognitive assessments, they did so in the following order: (1) mental and physical fatigue questions, (2) Go/No-Go and VSTM tasks (order counterbalanced across participants), and (3) BART. The BART

---

[5] Eight of the civilian participants were opportunistically enrolled at the trailhead (similar to civilians in Study 1) due to last-minute cancelations in our pre-enrolled populations (leaving extra wearable packages available after the pre-enrolled participants had begun the hike). They completed the pre-hike survey and carried the wearable devices. However, they did not provide blood samples and did not complete all the training sessions for the cognitive assessment. While aspects of the wearable and survey data should be comparable to the rest of the population, the cognitive data should be interpreted with caution.

Table 3
Wearable devices in Study 3.

Device	Metrics
Apple iPod touch (6th generation)	BrainBaseline cognitive assessments
SensorPush	Thermometer; hygrometer
Garmin fēnix 3 or 5 HR	Wrist-based optical heart rate monitor; accelerometer; barometric altimeter; GPS
Equivital EQ02	High fidelity ECG heart rate monitor; R-R interval; respiratory rate; thermometer; accelerometer
Wahoo TICKRx	ECG heart rate monitor; accelerometer
Garmin foot pod	Accelerometer
Garmin tempe	Thermometer (under chest strap)

was added to the end of the tasks to more easily allow for comparison of the fatigue questions, Go/No-Go task, and VSTM task to prior data collections.

Access to a blood gas analyzer on site during Study 3 allowed us to examine more analytes than were available to us in the prior two studies. Blood was collected via venipuncture in heparinized tubes and processed immediately with a Nova Biomedical Prime Plus Blood Gas Analyzer. All values, except for creatine kinase, were quantified on site using the Nova analyzer. Creatine kinase was measured by TriCore Reference Laboratories using stored serum collected at the same time as the initial blood sample was collected (as in Studies 1 and 2). See the Appendix for analytes collected.

2.4.3 Procedure

The general procedure in Study 3 for participants was similar to that of pre-enrolled participants in Study 2, except the personality questionnaire was moved from the post-task survey to the pre-task survey at pre-enrollment. This was done to save time at the trailhead and ensure our participants were able to catch the shuttle back to the south rim.

3 Results

The results reported here are intended to give a high-level snapshot of the effects found in the R2R WATCH data. These analyses intentionally focused on the subsets of the data that give us the cleanest view (as noted in each section). For example, the VSTM cognitive task data are drawn from participants with full training sets in Studies 2 and 3, excluding data from participants who did not have full training sets. Follow-on work using this data set will dive into deeper, more specialized analyses (see the Discussion and Future Directions section). Here, we describe (a) participant characteristics, (b) changes in subjective mental and physical fatigue ratings, (c) environmental temperature, (d) changes in heart rate and how they tie to a model of fitness level, (e) changes in VSTM and Go/No-Go scores, (f) changes in blood analytes, and (g) correlations among the data streams.

3.1 Participant Characteristics

Table 4 describes characteristics of our 138 participants across the three studies. While participant characteristics were not identical across the studies (e.g., participants tended to come from a higher elevation in Study 3 and home elevations had high variability), some consistent patterns emerge. Most notable is this population's generally high level of fitness and experience, as indicated by their recent longest distance and prior experience with the R2R hike. The average longest distance our participants walked, ran, or hiked in the last six months was 16.6 miles, and over 30% of them had previously completed the R2R hike. On average, it took participants over ten hours to hike the 24.2-mile R2R Kaibab trail.

3.2 Fatigue Ratings

We collected multiple subjective fatigue ratings from our participants. The cognitive assessment mobile phone app asked participants to rate their mental and physical fatigue every time they stopped to complete the cognitive assessments. The start, middle, and end surveys asked participants to give their current level of overall fatigue and the most fatigued they felt during different sections of the hike. Here, we focus on the mental and physical fatigue ratings given on the mobile device before the hike, during the hike (averaged across all mid-hike sessions), and after the hike. These analyses look at data from 124 participants across the three studies (after excluding participants whose mobile devices failed or who were not compliant with instructions for using them).

Analyses of changes in fatigue ratings across the hike were fit via the *lme4* package in R software (Bates et al., 2015) with Sattherwaite approximation to degrees of freedom (see Luke, 2017). Mixed-effect models allow one to account for variance due to factors such as participant and trial type, essentially "soaking up" some of the unexplained variance seen in more traditional ANOVA models. Here, we only included additional random effects if they significantly improve the fit of the model. All statistical tests were held at the $\alpha = 0.05$ level.

Table 4
Participant characteristics across the three studies in R2R WATCH.

	Study 1	**Study 2**	**Study 3**
Count	73 hikers	27 hikers	38 hikers
Gender	56% male	70% male	76% male
Age	40.3 years *(stdev = 9.1)*	37.1 years *(stdev = 8.5)*	37.1 years *(stdev = 9.4)*
Elevation (residence)[†]	1603.5 ft *(stdev = 2018.6.6)*	1548.1 ft *(stdev = 2193.7)*	2948.8 ft *(stdev = 2320.8)*
Weight*	169.7 lb *(stdev = 31.0)*	174.8 lb *(stdev = 25.8)*	178.6 lb *(stdev = 30.1)*
SpO_2*	94.7% *(stdev = 5.9)*	95.6% *(stdev = 3.5)*	94.7% *(stdev = 2.8)*
Heart rate*	75.3 bpm *(stdev = 13.6)*	82.9 bpm *(stdev = 15.0)*	81.3 bpm *(stdev = 13.2)*
Body fat*	—	20.3% *(stdev = 4.6)*	21.0% *(stdev = 5.5)*
BMI*	—	24.6 *(stdev = 3.2)*	25.5 *(stdev = 3.0)*
Sleep prior night	5.3 hrs *(stdev = 1.5)*	5.4 hrs *(stdev = 1.1)*	5.0 hrs *(stdev = 1.5)*
Longest distance	17.2 mi *(stdev = 11.3)*	16.1 mi *(stdev = 5.2)*	13.4 mi *(stdev = 9.7)*
Prev. completed R2R	39.7% yes	29.6% yes	13.2% yes
Total hike time	10.4 hrs *(stdev = 2.8)*	9.6 hrs *(stdev = 2.1)*	11.1 hrs *(stdev = 2.7)*

*Prior to beginning the hike.
[†]Elevation along the R2R trail ranges from 2400 to 8200 ft.

A mixed-effects model predicting mental fatigue rating from the fixed effects of hike location (pre-hike, mid-hike, or post-hike) and military membership (military or civilian) along with the random effects of participant and session revealed a significant increase in mental fatigue from the start of the hike to the mid-hike sessions ($t(464.8) = 9.740$, $p < 0.001$) and from the mid-hike sessions to the end of the hike ($t(525.1) = 12.827$, $p < 0.001$).

A similar mixed-effects model predicting physical fatigue ratings from the fixed effects of hike location and military membership along with the random effect of participant revealed the same pattern of results: participants indicated increased levels of physical fatigue from the start to middle of the hike ($t(528.6) = 14.610$, $p < 0.001$) and from the middle to end of the hike ($t(529.4) = 14.622$, $p < 0.001$). See Figure 2.

Unsurprisingly, participants showed a consistent increase in fatigue ratings—both mental and physical—across the hike. This pattern was reflected in the other measures of fatigue as well.

3.3 Environmental Temperature

Figure 3 shows the median ambient temperature recorded for participants at each hour of the day from 5 a.m. to 8 p.m. on each date of data collection. Environmental temperature data were recorded by wearable sensors zip-tied to the top of participants' packs. This time range was chosen because it covers a wide swath of our hikers' time spent along the trail and includes both the coolest and warmest parts of the day. The simultaneous temperature readings at each hour varied because hikers started at different times and proceeded at different rates, and elevation and shade vary by location. The median temperatures observed varied between approximately 5℃ (41℉) and 35℃ (95℉), with individual participants experiencing extremes from −0.5℃ (31℉) up to 47℃ (116℉).

3.4 Physiological (Heart Rate)

3.4.1 Overall Heart Rate Patterns

Data on participants' heart rates during the hike can give (albeit imperfect) insight into how hard the participants were working during the hike. The analyses below look at heart rate data from 68 participants across the three studies. Of these 68 participants with reliable heart rate data (determined by complete data without extreme values or implausible patterns when visually inspected), 58 were from chest strap ECG devices and 10 were from a wrist-based optical sensor (Garmin fēnix).[6]

Figure 4 gives two examples of heart rate data from participants. The heart rate on the left comes from a fast hiker while the heart rate on the right comes from an average-paced hiker. Identifying times when participants took breaks is relatively easy, but there are not many obvious trends when simply looking at the raw heart rate data. In general, heart rates rose as the hike progressed, especially during the uphill portion, but they often decreased slightly in the final quarter of the hike. Further inspection using functional data analysis could identify interesting trends or relationships in these patterns.

Figure 5 shows the mean (left) and standard deviation (right) of heart rates from the 68 participants at locations along the trail. As expected, mean heart rates increased as the participants completed more of the hike, with a significant increase in the uphill section. The black cross indicates the location of the maximum mean heart rate (standard deviation of heart rate on the right), indicating the highest mean heart rate occurred about halfway back up the

---

[6] ECG heart rate monitoring devices consistently produced cleaner, more reliable data than their optical heart rate counterparts, which is one of the reasons Studies 2 and 3 increased the use of ECG heart rate monitors even though it led to fewer participants willing to wear the device and enroll in the study (for a discussion, see Emmanuel-Aviña et al., 2018).

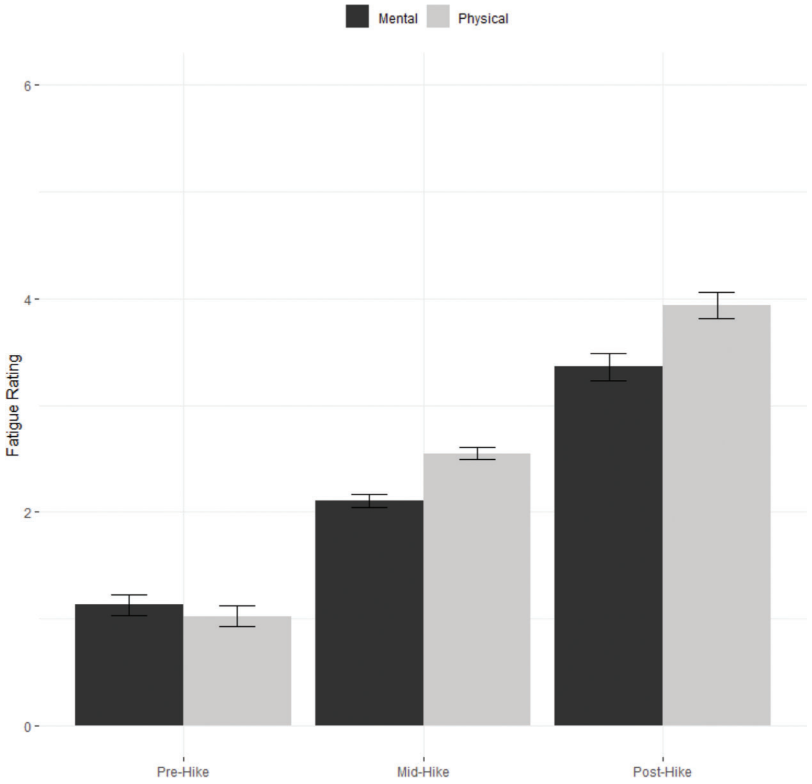

Figure 2. Subjective mental and physical fatigue ratings given on the mobile device right before starting the hike (pre-hike), averaged across the ratings during the hike (mid-hike), and after completing the hike (post-hike). Error bars represent standard error of the mean.

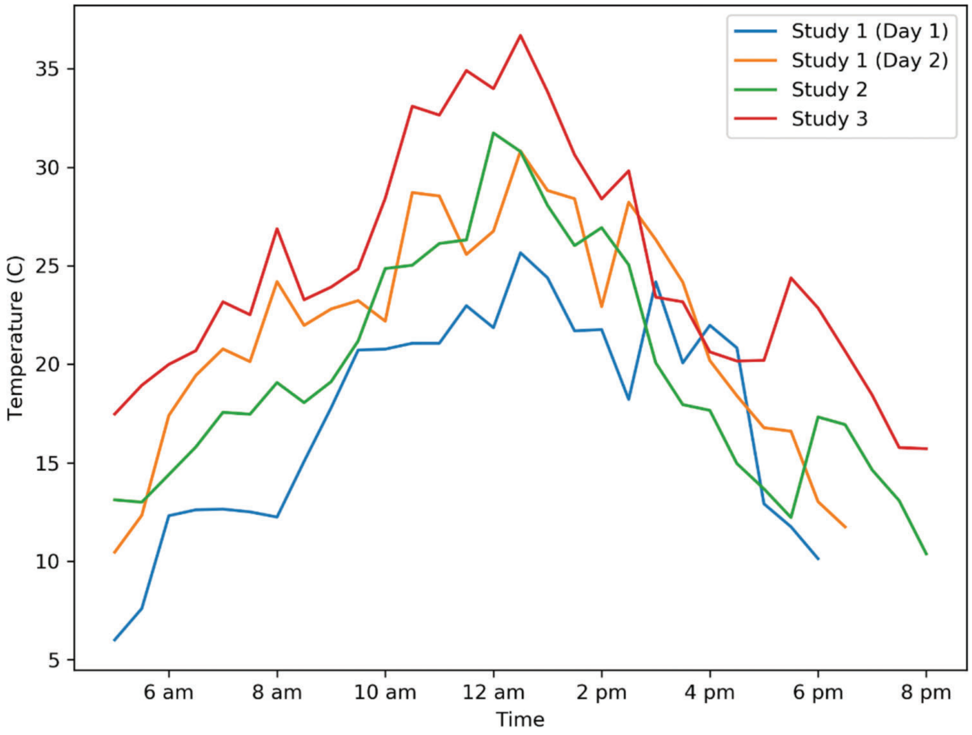

Figure 3. Median temperature as a function of time of day across all three studies.

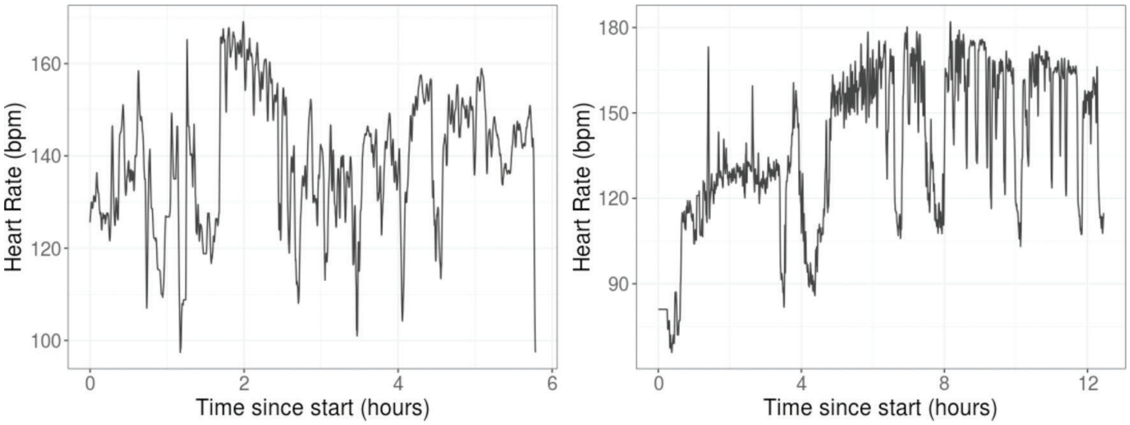

Figure 4. Example heart rate patterns across the hike for two participants.

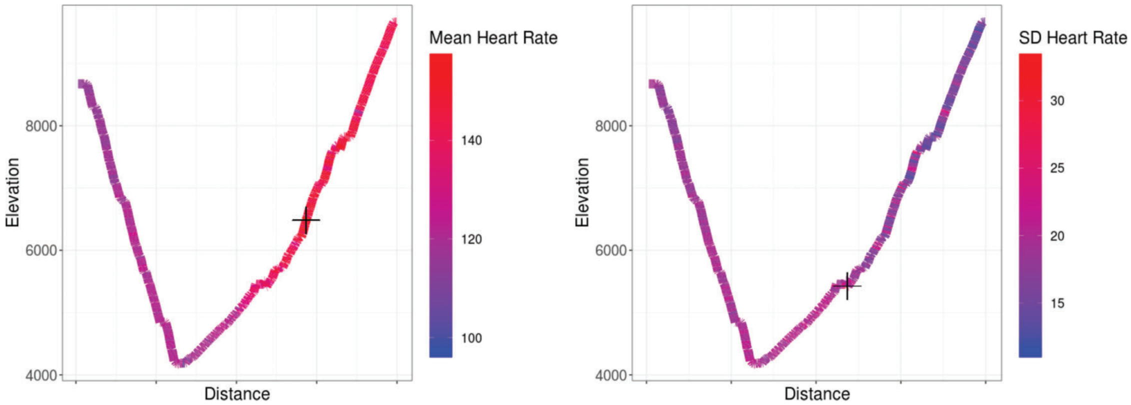

Figure 5. Mean heart rate (right) and standard deviation of heart rate (left) across the hike as a function of elevation and distance.

canyon. One possible explanation for this is that participants are beginning to feel the effects of physical fatigue; they can no longer keep their pace and intensity, and thus heart rate is high. Variability in heart rate across participants, as shown in Figure 5 (right), decreases toward the end of the hike. The peak variation is about a quarter of the way back up the canyon, after which participants' heart rates begin to converge. There are locations where the standard deviation is particularly low, which may reflect a popular resting point for many hikers toward the end of the trail.

3.4.2 Fitness Scores

The physiological heart rate data from the chest straps, along with trail position can be used to come up with an overall fitness score. Here, we describe the building of these fitness scores based on data across all three studies. Fitness scores can be derived by the stress of changes in heart rate (HR) of participants that arise from changes in "work" measured by recent physical activity as participants progress along the hike. A linear regression model was used to build the fitness scores that considered *heart rate (HR)* as the response variable. The predictor variables were *average cadence (CAD)* and *changes in elevation* over the last five minutes. The changes in elevation were determined through two variables that measured the effort of each participant as s/he went *up (ElvUp)* into and *down (ElvDown)* out of the canyon during the hike, separating the effects in elevation by positive and negative changes. The regression model takes the form:

$$HR(t) = \beta_0 + \beta_1 CAD(t) + \beta_2 ElvUp(t) + \beta_3 ElvDown(t) + \varepsilon$$

where ε is the model error and β_0, β_1, β_2, and β_3, are the effects associated to each predictor variable, respectively. The fitness score is defined as the effect of elevation going up into the hike $\boldsymbol{\beta_2}$ (higher scores are indicative of poorer fitness). The time series of heart rates for each participant were screened to assess quality of the data, and those with anomalous data (e.g., impossibly high heart rate based) were discarded. This left a total of 68 participants across all three studies. Heart rate was measured via ECG devices (e.g., Wahoo TICKR, Suunto SmartSensor, Equivital) or optical heart rate monitors (e.g., Garmin fēnix). Where possible, the ECG heart rate data were used; if they were

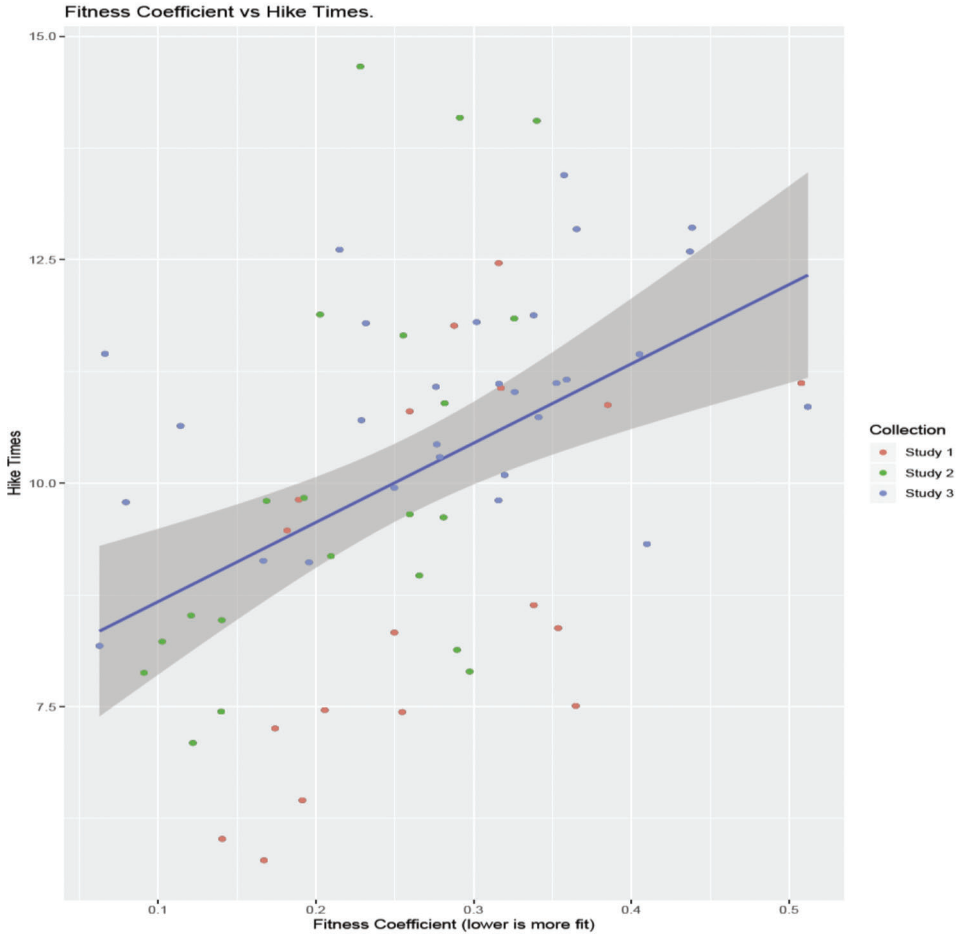

Figure 6. Fitness scores versus hike times for all data collections.

not available, fēnix optical heart rate data were used for that participant instead. Optical heart rate data from other devices (e.g., Garmin vívoactive) were not used due to consistently insufficient data quality from those devices.

The linear regression above was fit to each participant. We also considered total *hike time* (in hours) for each participant. Figure 6 presents a scatterplot for the fitness score (coefficient β_2) versus the corresponding *hike time* for each participant. Notice that there is a positive linear relationship between fitness scores and hike times, indicating that as the fitness score increases, the hike times are larger. The Pearson correlation between fitness scores and hike time is $r^2 = 0.46$.[7]

Notably, there are still data points that are scattered away from the estimated regression line between hike time and fitness score. This shows that we have individuals that could be considered “fit” in terms of their fitness scores but may still have long hike times (perhaps due to taking it easy or hiking with a less fit partner). We also see individuals that may not be very “fit” according to this assessment of fitness, but nonetheless have fast hike times. These variants highlight that while the fitness scores may capture a chunk of the variance in our population, they are by no means a perfect measure.

We also assessed the predictive performance of the regression of the *HR* equation above by comparing the observed time series of heart rates with the predictive or fitted values of the regression model. Figure 7 provides a sample of these comparisons for four participants from Study 3. The actual time series of heart rates are shown in red while the fitted values appear in blue. Overall, the fitted values capture the general behavior of heart rate as time progressed across the hike; this was also observed for all other Study 3 participants. However, some portions of the model show potential misfit, especially toward the start of the hike. Looking at Figure 7, this is especially noticeable for participant 18-1001. Some moderate lack of fit is also observed for fitted values of the regression model for

---

[7] We also considered the relationship between hike time and fitness coefficient as given by β_2 where the predictor variables of average cadence (CAD) and change in elevation were measured over a 1-minute span (instead of a 5-minute span). In this case, the Pearson correlation coefficient was $r^2 = 0.49$ across all three studies. The calculation of fitness score showed some slight variations between the 5-minute versus 1-minute averages for cadence and elevation, but the overall trend remained.

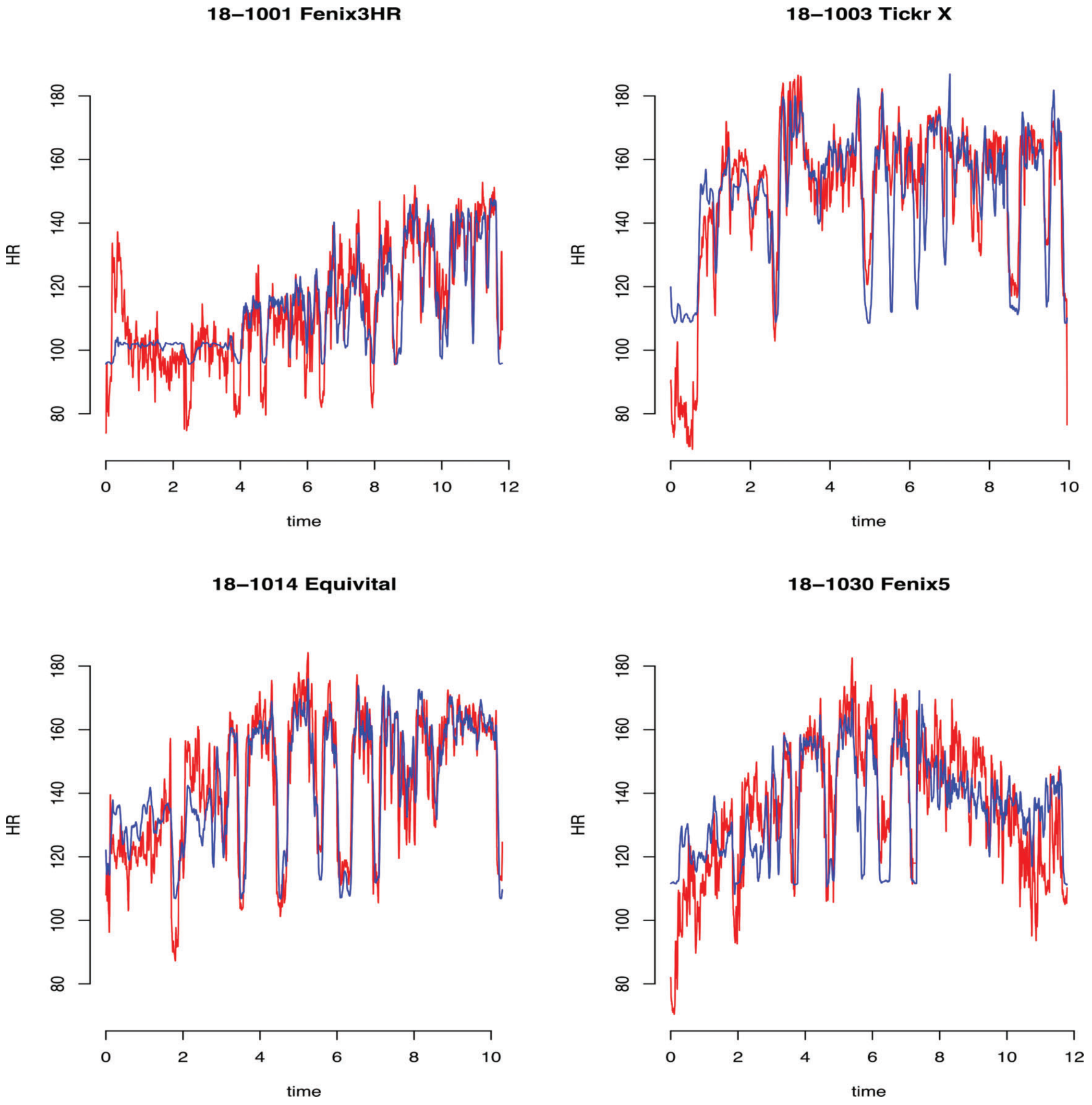

Figure 7. Heart rates (in red) for four Study 3 participants as a function of time. Predictive values of regression model (in blue).

participant 18-1003 at the start of the hike and during the middle of the hike. Residual analysis of the regression models for heart rate confirms these findings. These deviations could be partially explained by the chest strap not being damp enough to provide good connectivity (and therefore good data) at the beginning of the hike; the model likely does not sufficiently capture the "warming up" phase. In addition, traditional regression models often do not account for all the variability that is present in the data. The next section offers a possible extension of the model to help address this issue.

Model Variant. Time variations of fitness may occur across the hike, but the above regression approach treats the regression coefficients as static in time. The deviations seen in Figure 7 suggest that a more flexible regression approach could offer better predictive performance. Moving to statistical or machine learning models (dynamic models) with time-varying coefficients could produce a time-varying representation of fitness across the hike. A model to consider in this case is a dynamic regression of the form:

$$HR(t) = \beta_0(t) + \beta_1(t) CAD(t) + \beta_2(t) ElvUp(t) + \beta_3(t)\, ElvDown(t) + \varepsilon$$

where the model coefficients $\beta_i(t)$; $i = 0, 1, 2, 3$ are allowed to vary smoothly in time. This model permits a time series representation of "fitness" per participant that would estimate the variations of fitness and performance across the hike (e.g., better capturing warming-up periods or hard sections of the hike). This model also has a multivariate representation that allows for estimate commonality and data correlations among participants. Future work could be aimed at calculating point estimates and uncertainty bands for a time-varying fitness coefficient.

3.5 Cognitive Assessments

We also examined changes in performance on the VSTM and Go/No-Go cognitive tasks across the hike. Due to the changes in protocol, only data from Study 2 and Study 3 were used in these analyses. Participants who were noted as noncompliant with the instructions (e.g., reported to a researcher s/he stopped doing the cognitive tasks, completed too few training sessions, or had evidence of "button mashing" such as chance performance) or had device failures (e.g., iPod died prior to the end of the hike) were excluded from the analyses. This left data from 53 participants for the VSTM task and 52 participants for the Go/No-Go task.

One method of examining performance decline in the cognitive assessments over the course of the hike is to compare the pre-hike scores (beginning of the hike at the South Kaibab trailhead) to the post-hike scores (end of the hike at the North Kaibab trailhead). This approach takes full advantage of all three training sessions during pre-enrollment the day before the hike, allowing participants to become more familiar with the tasks and better mitigate practice effects. However, a potential confound is that most participants start the hike very early in the morning where factors such as grogginess and the cold weather could artificially lower performance on the cognitive tasks. For example, response time in the Go/No-Go task at the start of the hike was correlated with temperature and time (with faster response times for those starting later in the morning, which was also when temperatures were warmer).

In order to mitigate the potential early morning effects of the pre-hike sessions, we also examined the change in performance from the *final training session* to the end of the hike. The final training session has the advantage of being in a temperature-controlled environment the evening prior to the hike. It has the disadvantages of only allowing for two prior training sessions to overcome practice effects and less consistency in environment between participants (e.g., some might have completed it at 5 p.m. in the hotel lobby while others completed it at 10 p.m. in their hotel rooms).

The analyses reported below use both the final training session (the evening before the hike at the hotel) and the pre-hike session (immediately prior to the hike at the South Kaibab trailhead) as baselines relative to the post-hike session (immediately after finishing the hike at the North Kaibab trailhead). While not strictly necessary, converging evidence from both approaches provides stronger support for conclusions drawn about performance decline over the course of the hike.

Analyses of performance differences between the start of the hike and end of the hike and between the final training session and the end of the hike were run using mixed-effect models (as with the fatigue ratings above).

VSTM Task. The VSTM task is primarily an accuracy task, so the critical metric of interest is accuracy on a trial-by-trial basis. Mixed-effect models predicting trial accuracy in the VSTM task from the fixed effects of hike location (pre-hike versus post-hike or final training versus post-hike) and military membership (military versus civilian) along with the random effects of participant and trial type (match versus no match) revealed a significant reduction in performance from both the start of the hike to the end of the hike ($Z = 3.05$, $p = 0.002$) and from the final training session to the end of the hike ($Z = 4.55$, $p < 0.001$). Performance appeared to drop by the end of the hike, regardless of whether one baselined to the final training session or the start of the hike. See Figure 8.

Go/No-Go Task. The Go/No-Go task is primarily a response time task. Accuracy tends to be near ceiling on this task, making error analysis difficult without very large sample sizes. The primary metric of interest for the

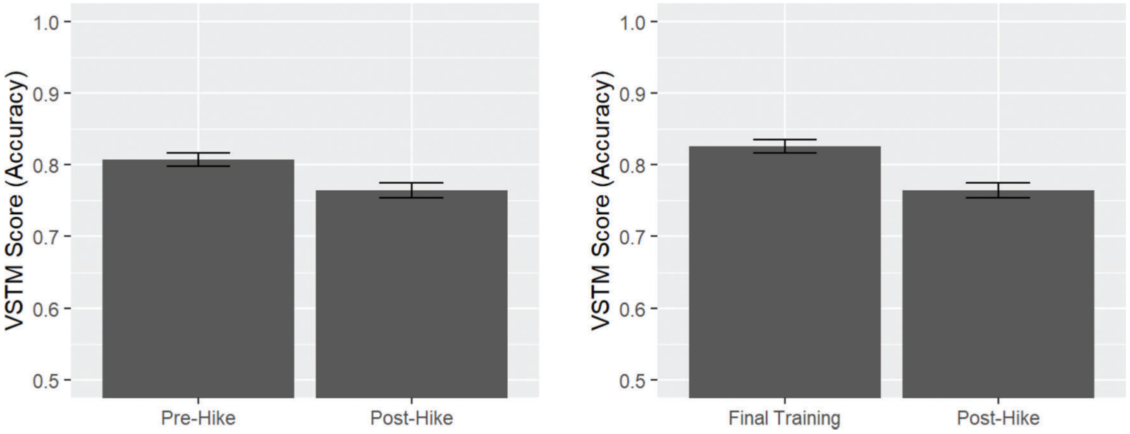

Figure 8. Accuracy on the VSTM task for Studies 2 and 3. Error bars represent standard error of the mean.

Go/No-Go task is response time on accurate "go" trials. Mixed-effects models predicting trial response time in the Go/No-Go task for accurate "go" trials from the fixed effects of hike location and military membership and the random effects of participant and interstimulus interval (ten levels ranging from 500 to 1850 ms) were fit with Sattherwaite approximation to degrees of freedom, comparing both pre-hike versus post-hike performance and final training versus post-hike performance. The models revealed a marginal but statistically significant *reduction* in response time (i.e., participants got faster) from the final training session to the end of the hike ($t(4019) = 2.03$, $p = 0.043$) but no significant change in response time between the start and end of the hike. See Figure 9.

Overall, we saw a drop in VSTM accuracy across the hike but did not find a similar reduction in Go/No-Go response time performance. This mixed pattern of results is not entirely surprisingly considering that prior studies looking at the cognitive effects of similar environmental stressors have also revealed mixed results on cognitive measures (e.g., Adan, 2012; Hancock et al., 2007; Maruff et al., 2006; Pilcher & Huffcutt, 1996; Virués-Ortega et al., 2004).

3.6 Blood

The results here focus on the available blood data from participants in all three studies. The onsite blood gas analyzer added during Study 3 allowed us to collect additional analytes of interest that were not available without the onsite analyzer during Studies 1 and 2. We narrowed in on the following analytes of interest: pH, pCO_2, lactate, pO_2, sodium, potassium, ionized calcium, ionized magnesium, creatine kinase, hematocrit (HCT), blood urea nitrogen (BUN), and creatinine. These analytes were chosen because they all represent direct measurements and are consistent with the most common clinical laboratory tests that are run. They are of high value and highly informative (with the exception of the ionized magnesium which is more experimental). Sodium, potassium, creatine kinase (CK), creatinine, and BUN were measured in all three studies; the other analytes were only measured in Study 3. These laboratory tests represent the base investigations most clinicians would do in evaluating patients with the addition of some blood gas evaluation commonly performed on more critically ill patients.

3.6.1 Overall Changes

Statistical analyses were done using a Student's T-test of paired samples. Figure 10 shows the 12 analytes of interest with annotations. The left axis is for the pre- and post-hike measurements, which are linked by a line in the before–after plot. The right axis is for the delta (finish value minus start value), represented with a violin plot with all points also shown. The grey bracket on each subplot represents the clinically normal range for the pre- and post-data as determined by our emergency medicine professional. If there are sex-specific normal ranges, the outer range was used for simplicity (e.g., male CK normal range is 37–242, female CK normal range is 28–203, graph tick marks were put at 28 and 242).

Additionally, the red "AKI" notation on the creatinine graph (lower right) shows the delta value denoting acute kidney injury. The pCO_2 normal range is set for normal individuals acclimated to their current altitude. Normal hematocrit levels for women go down to 36, which is below the range shown on the graph (normal low range for men is 42). Indicated performance decrements such as dehydration are relative to this population. Outside of this specific clinical scenario there is a broader differential of etiologies. For example, an elevated BUN and creatinine in the setting of chronic diabetes and hypertension may be a sign of chronic kidney damage rather than acute dehydration.

Several changes are noted in this panel of 12 analytes that fall under four broad areas of performance degradation: respiratory effort, muscle cellular damage, electrolyte changes, and dehydration.

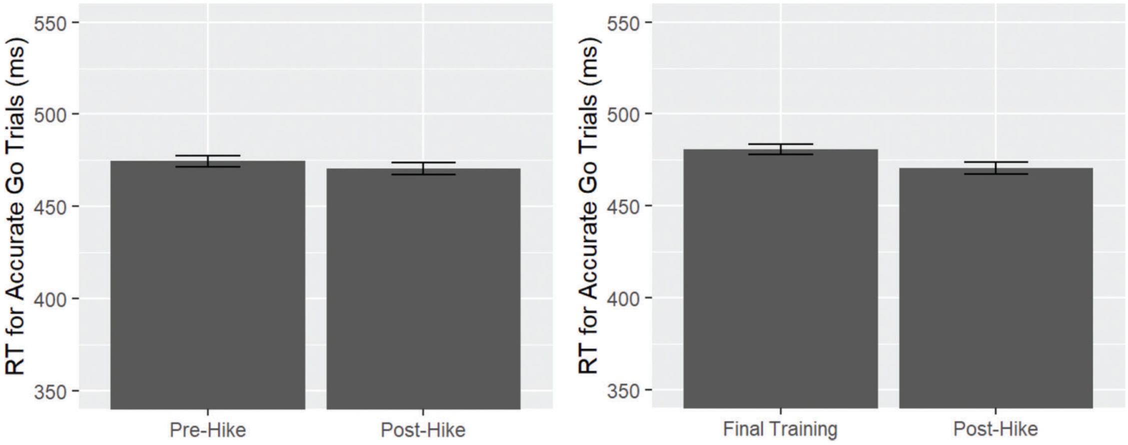

Figure 9. Response time (RT) for accurate "go" trials for the Go/No-Go Task for Studies 2 and 3. Error bars represent standard error of the mean.

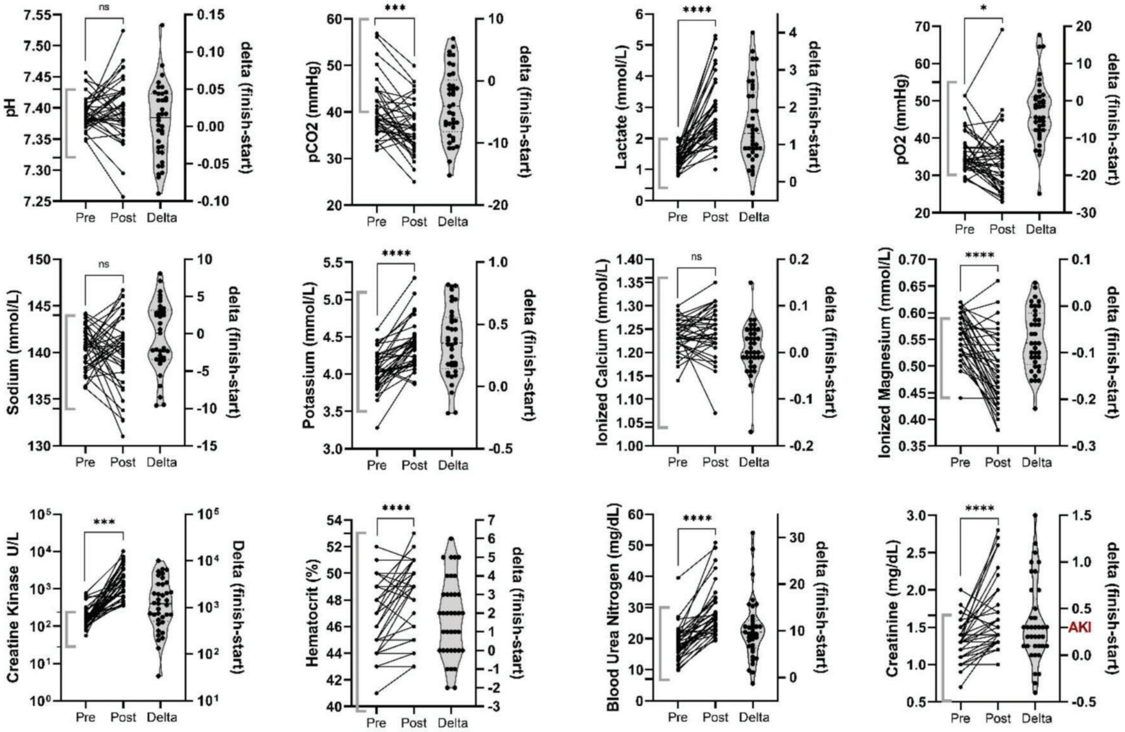

Figure 10. Blood analytes showing pre-and post-hike blood measurements with the delta values on the right. Grey brackets represent the normal clinical ranges. The asterisks indicate level of significance (* = $p<0.5$, ** = $p<0.1$, *** = $p<0.01$, **** = $p<0.001$).

Respiratory Effort. The first area of performance measurement is the combination of pH, pCO_2, and lactate. Together these metrics give insights into the effort and reserve of the participant. High pH coupled with low pCO_2 is a sign of increased respiratory effort such as someone sprinting at the end or breathing hard to try to relieve an oxygen debt. A normal pH coupled with a low pCO_2 and an elevated lactate is indicative of an individual who put themselves into a metabolic debt but was able to compensate. An individual with a pH below normal range, a low pCO_2, and an elevated lactate is an example of a participant who lost the ability to compensate and experienced a performance degradation.

Muscle Cellular Damage. CK is an excellent measure of muscle cellular damage by mechanical injury. It is an intracellular protein that leaks out during damage and can lead to kidney dysfunction. As an objective measure of damage, in this population it shows the most consistent elevation (as well as a several logarithmic increase in some participants). We would expect that anyone participating in this task would see an increase in CK, but that the highest elevations are consistent with performance degradation.

Electrolyte Changes. The third performance metric is related to electrolyte changes. Significant changes in electrolytes, especially sodium and potassium, have been shown to have devastating consequences. In healthy participants hiking in the heat, exercise-associated hyponatremia is a known problem resulting in vomiting, seizures, and potentially death. Some of these changes are behavior-based—poor nutritional strategies are known contributors to this problem. In this population, half of the participants had an increase in sodium while half decreased. It is not clear that this is all nutritionally based, suggesting a possible alternative explanation such as differences in arginine vasopressin responses.

Dehydration. The last performance metric we can see in these data is the effect of hydration. Although related to electrolyte changes, there is a separate metric of adequate fluid intake. Elevations in hematocrit, blood urea nitrogen, and creatinine all signal inappropriate hydration that may lead to performance degradation.

3.6.2 Elevation Acclimatization

We also analyzed the effect of elevation acclimation by focusing on physiologic changes over the hike in individual metrics (with some interpretation given to the bigger picture provided by looking at consistent changes over several analytes). As mentioned previously, this is consistent with how a diagnostician would evaluate performance over time, especially given that performance degradation is multifaceted and in many cases is amenable to timely intervention if appropriately identified.

Because we were able to have a blood gas analyzer on site during Study 3, we were able to evaluate some metrics that can only be measured soon after sample acquisition. Among these are lactate, pH, and pCO_2. These analytes are more susceptible to significant changes in a short period of time. The interaction of pH and pCO_2 is especially interesting because it offers insights into adaptation as well as stress. Clinically, pCO_2, total CO_2, and HCO_3^- are considered together with pH because they are intimately related and in constant flux in an effort to keep blood chemistry buffered. The pertinent equations for this balance are as follows: $CO_2 + H_2O <\!\!-\!\!> H_2CO_3 <\!\!-\!\!> H^+ + HCO_3^-$ for the relationship between CO_2 and bicarbonate. It is important to realize that CO_2 concentrations are modulated by respiratory effort. Heavy breathing will offload more CO_2, driving the equation above to the left. The relationship between CO_2 and pH is as follows: $pH = 6.1 + \log HCO_3^-/(0.03 \times pCO_2)$. This relationship is crucial because it reveals two mechanisms by which we can modulate our pH: the retention or loss of HCO_3^- or CO_2. The nature of that relationship also provides insight into the adaption of an individual. This can be seen in Figure 11. Participants who came from a "low" home altitude (less than 1,000 m above sea level) had a different pH to pCO_2 relationship pattern than those from a "high" home altitude. This is consistent with the adaptive changes in HCO_3^- metabolism and is related to how acetazolamide (a medication used for altitude acclimatization) works in humans. These changes persisted throughout the hike and were further exaggerated at the finish measurements. This is likely multifactorial, with a higher finish than starting elevation and additional metabolic perturbations such as lactate elevation.

These correlations are crucial for several reasons. They show that there is an effect of acclimatization that can be easily measured. It also demonstrates that there are persistent changes in some participants that place them in concerning ranges (pH less than 7.35). Additionally, it is also important from the standpoint that pCO_2 (as estimated by end-tidal CO_2) is an easily field-measurable variable by means of handheld instruments that could be deployed. Lastly, it is important to note that the relationship between pCO_2 and pH is modified by acclimatization and home altitude. These factors would need to be considered when using CO_2 as a metric in future analyses.

3.7 Combining Data Streams

We also explored the correlations between the data streams. While these analyses were exploratory in nature, they highlight potential areas of future work to confirm patterns noted here.

We examined correlations in the changes (deltas) of our metrics across the hike. For example, was an increase in fatigue also associated with a decrease in performance on the cognitive assessments? Do more easily collected measures such as the VSTM task reflect important biological changes in one's blood chemistry?

The correlation plot (created using Student's T-test on complete cases from Study 3) in Figure 12 highlights patterns of interest. Correlations outlined in black are significant at an $\alpha = 0.05$ level; those in gray are significant at an $\alpha = 0.10$ level. The variables of interest were:

- Changes in HCT in the blood, which is a measure of the percentage of red blood cells and is one of the best measures of dehydration.
- Changes in pCO_2 in the blood, which can be used to understand respiration rate. Decreases in pCO_2 can be caused by hypoxia and hyperventilation.
- Changes in pH levels in the blood, which can result from respiration and lactate changes.
- Changes in mental and physical fatigue ratings, which are indicative of how tired participants feel throughout the hike.
- Changes in VSTM accuracy, which is a measure of visual working memory capacity.
- Changes in Go/No-Go response time, which is a measure of inhibition and response time.
- Hours of sleep the night prior, which is known to influence both physical and cognitive performance.

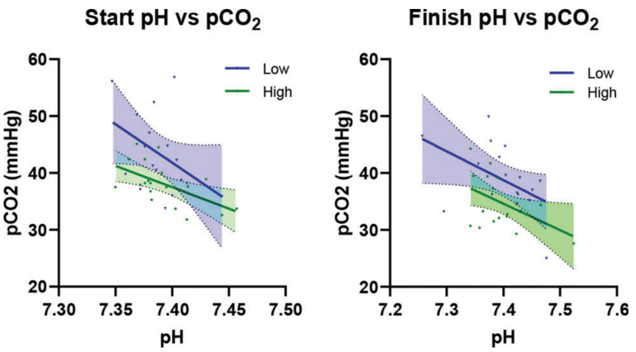

Figure 11. Relationship between pH and pCO2 in subjects coming from a "low" home altitude (less than 1000 meters above sea level) compared to a "high" home altitude (greater than 1000 m).

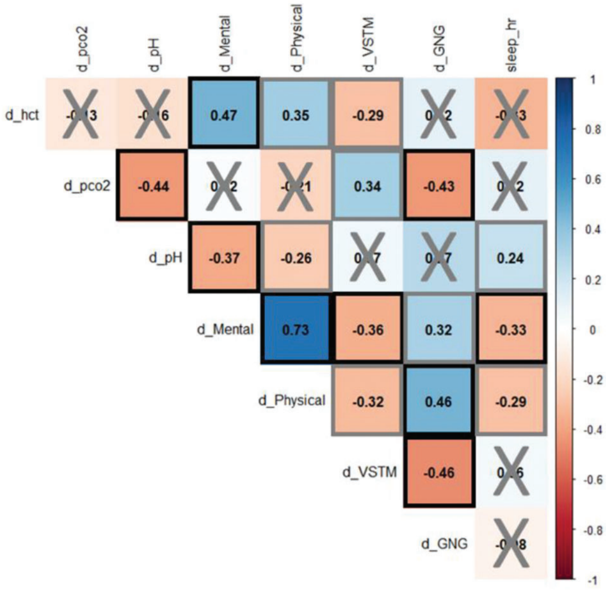

Figure 12. Correlation plot for changes in HCT, pCO_2, pH, mental fatigue, physical fatigue, VSTM accuracy, and Go/No-Go response time from the start to end of the hike, along with hours of sleep the night prior for Study 3. Correlations with a black box are statistically significant at an $\alpha = 0.05$ level; those with a gray box are significant at an $\alpha = 0.10$ level. Those with an "X" are not statistically significant.

Increases in hematocrit levels were associated with increased fatigue ratings ($r = 0.47$ and $r = 0.35$) and reduced accuracy on the VSTM task ($r = -0.29$); it was not associated with Go/No-Go performance or sleep the night prior. This indicates that participants who were more dehydrated by the end of the hike (as indicated by hematocrit levels) also felt more fatigued and performed more poorly on the VSTM task by the end of the hike.

Decreases in pCO_2 were associated with increased pH levels ($r = -0.44$), reduced accuracy on the VSTM task ($r = 0.34$), and slower response time on the Go/No-Go task ($r = -0.43$) across the hike; however, changes in pCO_2 were not significantly correlated with changes in fatigue ratings. Interestingly, those who were working harder at breathing by the end of the hike (as indicated by the pCO_2 levels) did not report a similar increase in feelings of fatigue. However, changes in pCO_2 were still associated with reduced performance on the cognitive tasks across the hike.

Decreased pH levels were associated with increased fatigue ratings across the hike ($r = -0.37$ and $r = -0.26$) and less sleep the night prior ($r = 0.24$), along with the previously noted increase in pCO_2; however, they were not significantly correlated with changes in performance on the cognitive tasks. While changes in pH levels were associated with amount of sleep the night prior and changes in fatigue ratings, they were not reflected in either of the cognitive assessments.

Increases in either mental or physical fatigue across the hike were correlated with reduced performance on both cognitive tasks (reduced accuracy in VSTM and slower response times in Go/No-Go; $0.32 \leq |r| \geqslant 0.46$). They were also associated with fewer hours of sleep the night prior ($r = -0.33$ and $r = -0.29$), along with the previously noted increase in hematocrit and reduction in pH levels. Hikers who felt more fatigued by the end of the hike also showed cognitive decrements, and those who got less sleep the night prior also had a greater increase in fatigue across the hike.

Finally, participants who showed reduced performance in one cognitive task (e.g., reduced VSTM accuracy) across the hike were likely to also show reduced performance on the other cognitive task (e.g., slower Go/No-Go response time; $r = -0.46$). Changes in mental and physical fatigue ratings were also highly correlated ($r = 0.73$).

4 Discussion and Future Directions

The R2R WATCH project measured physiological, cognitive, and biological markers of performance decline in tandem in an extreme environment. The series of three studies on hikers completing the Grand Canyon R2R yielded both a rich data set and insight into which methodologies and devices performed best in this environment. This article highlighted a subset of high-level results revealed by initial analyses on the data set, but more remains to be gleaned from this data set. It has also spurred insights for future directions in the field.

4.1 R2R WATCH: Summary of Current Findings and Future Analyses

Our participants showed a general increase in both mental and physical fatigue over the course of the 24.2-mile hike, which on average took over 10 hours to complete. Hikers were exposed to both cool (around freezing) and hot (up to 116℉) environmental temperatures, along with a range of moderately high elevations (starting at 7,260 feet and ending at 8,240 feet, with about a mile up and down in between). We had participants with home elevations at sea level and those with home elevations more consistent with the heights of the canyon rims—and home elevation appeared to make a difference in the changes in pH and pCO_2 levels in their blood, with those more acclimated to the higher elevations doing better.

Overall, heart rate tended to increase across the hike, but variations in that pattern hint toward more complex underlying variables. For example, heart rate tended to drop toward the end of the hike. This could indicate that several metrics of fatigue had been reached. Stored energy utilization (i.e., glycogen stores) would have been depleted at this point. An individual's ability to compensate with nutritional intake as well as metabolic adaptations would also be important considerations at this point. Waste products, including lactate and urea, would have accumulated alongside potassium and creatinine kinase released during muscle damage. This combination leads to a decrease in performance and an inability to reach maximal muscular output (Areta & Hopkins, 2018).

Our participants were also not required to maintain a consistent level of effort across the hike (either within themselves across the hike or relative to other hikers). That means that some might have pushed themselves harder than others and individual effort level likely varied throughout the hike. Many of our participants hiked with others. A particularly fit hiker might have taken a relatively easy pace to stay with their group while another pushed harder so as not to hold the rest of the group back. We also observed interesting helping (or dropping behaviors)—we had reports of hikers slowing to help a complete stranger who was struggling to make it out of the canyon (and of groups splitting apart as individuals wanted to move faster). We are currently investigating the patterns in the performance data associated with differing group dynamics.

The relatively simple combination of heart rate, percentage up/down in the canyon, and cadence used to create a fitness score did a decent job of predicting the success variable of hike time. However, areas where the prediction was misaligned indicate that a more nuanced model that better factors in the effects of time and/or additional variables such as home elevation might lead to even better results. In general, while we have unveiled many of high-level effects, the heart rate data still leave more to be gleaned. Of particular interest is analyzing the higher-resolution heart rate variability data collected by the more sophisticated chest strap added in Study 3; fine-grained HRV has been shown to be an excellent indicator of cardiac performance and physical stress (see Acharya et al., 2006; Dong, 2016; Makivić et al., 2013).

Throughout the hike, participants completed cognitive tasks including a Go/No-Go task (measuring processing speed and inhibition) and VSTM task (measuring visual short-term memory). The VSTM task was linked to the strongest effects. VSTM accuracy performance dropped across the hike and was correlated with both an increase in fatigue and changes in hematocrit and pCO_2 levels indicative of dehydration and increased respiratory effort. The Go/No-Go task did not show a general decrease in performance across the hike; however, participants who had slower response times by the end of the hike also tended to report higher levels of fatigue and increased respiratory effort (as indicated by changes in pCO_2 levels). Here, we examined changes in performance from the start to the end of the hike; however, future work could also factor in performance from the sessions during the hike. While participants did not necessarily complete those sessions at the same times or locations (e.g., due to GPS error), they could be tied to percentage down/up the canyon or proximity to known challenging areas (e.g., Supai Tunnel).

Changes between the pre-hike and post-hike blood samples revealed indicators of respiratory effort, muscle cellular damage, electrolyte changes, and dehydration, all of which are known to be tied to overall performance. Furthermore, a subset of the blood analytes was correlated with measures from the other data streams. Increased dehydration (as indicated by elevated levels of hematocrit) was moderately correlated with both increased fatigue and reduced performance on the VSTM task. Changes in respiratory effort (as measured by pCO_2) were linked to cognitive performance changes (though not fatigue ratings); in contrast, elevated lactate levels (as measured by pH) were associated with increased fatigue across the hike but not the cognitive tasks. Unfortunately, the fidelity of measurement of some of these metrics was not yet at the level needed to directly link to the cognitive testing. Future studies with newer available technologies, such as point-of-care end-tidal CO_2 measurements and microneedle-type sensors, could better delineate the waxing and waning effects of fatigue.

There remain additional variables in the R2R WATCH data set that remain relatively underutilized. Not reported here but of interest to the team are the personality questionnaires and additional survey questions (e.g., nutritional intake), which are potential mediators of activity level and performance. Additional sensor data (e.g., body temperature and/or respiration rate) and individual session data from the cognitive tasks (along with BART) are also candidates for future analyses. One of the most promising directions to take the current data set is building more sophisticated prediction models that pull together all the data streams to more easily predict performance decline.

4.2 Challenges and Opportunities

Multiple challenges arise when working in a field study environment like the Grand Canyon R2R. One of the goals of the R2R WATCH project was to put to the test different techniques and technologies in this environment.

After initial benchmarking to test a wider range of devices, throughout the lifecycle of the project we sent over 900 devices across the canyon coming from 17 different device types. A set of key patterns emerged (see also Emmanuel-Aviña et al., 2018). Two of the most important factors for our COTS wearable devices were reliability of device and quality of data. Battery life was crucial—particularly since the hardest sections of the hike were at the end. Numerous candidate devices were discarded prior to the benchmark stage due to relatively short battery lives. Even so, those that were incorporated into the study did not always perform as advertised at the canyon. Despite having extraneous functions turned off and being fully charged before starting the hike, the device batteries sometimes failed before participants finished. Other devices dropped data for periods of time or just stopped recording all together, despite remaining battery power. More concerning were devices that appeared to automatically fill in or smooth missing sections of the data (this might be acceptable for less stringent applications such as general activity level tracking but is problematic for more fine-grained analyses). ECG heart rate sensors (e.g., chest straps) provided higher-quality data than their counterpart optical heart rate sensors (often found in watches or hats)—but the chest straps were also less convenient and comfortable for participants to wear. Chest straps also tended to perform more poorly at the start of the activity. Electrode connectivity is important; if participants had not started sweating or did not use gel (or it was dried out), the device tended not to perform as well at the beginning. We also found it was critical to include a device with a barometric altimeter in order to accurately calculate elevation and distance along the trail. Devices with only GPS sensors tended to perform poorly in steep sections of the canyon, artificially adding to distance metrics.[8] This erroneous GPS behavior is particularly concerning for hikers checking their GPS distance to see how close they are to the finish. Mistakenly thinking they are close to the end of the hike could lead to both frustration and dangerous decisions regarding regulation of remaining resources (e.g., finishing off remaining food or drinking water). While we were able to narrow in on the devices that worked best in this environment, no COTS device consistently worked exactly as advertised (i.e., across all steps from collecting data to quality of data to offloading data). Those with well-established wearable sensors (e.g., ECG) tended to work better than those with "cutting edge," new sensors (such as those monitoring muscle activation).

Measuring cognitive performance in a realistic field setting also has a wide swath of challenges and factors that influence performance. Individual differences—many of them challenging to measure in this setting—can have large effects on cognitive performance. Cognitive performance also generally cannot be measured in the same fine-grained, passive way as physiological measures such as heart rate sensors (i.e., participants need to stop to directly interact with the cognitive task). Additionally, cognitive performance cannot be assumed to change in a linear fashion in response to activity in extreme environments. Relatively short bursts of activity or of moderate duration have been shown to lead to an initial *increase* in performance, and it has been posited that the compensatory effects in response to minor changes indicate that more intense or longer-duration activities are needed to see a consistent decline in cognitive performance (e.g., Chang et al., 2014; Davranche et al., 2015; Rattray & Smee, 2016; Tempest et al., 2017). Control groups are also of particular interest for cognitive measures in field studies—Kramer and colleagues (1993) measured cognitive performance at high altitude and found that while the hikers' performance remained stable, the control group showed an improvement (indicating that the increased elevation likely negatively affected the hikers). Consistency and compliance with the tasks, along with mitigating apparent practice effects were all considerations that needed to be accounted for. These factors combine to make cognition particularly challenging to measure and compare across field studies in extreme environments. However, improvements in methodologies and technologies that would allow for more passive and frequent collection of cognitive measurements, along with appropriate control groups, could strengthen the field's understanding of changes in cognitive abilities in response to performing under the conditions of extreme environments.

The environment itself provided challenges to the research team as well as the hikers. During the cold, early morning hours at the start of the hike, electronics (e.g., bioimpedance device, centrifuge) needed to be kept warm to keep the batteries functioning properly. All measures needed to be self-contained since the researchers could only interact with participants at three points along the trail (and signs are not allowed by the park service). Furthermore, participants were essentially released into the "wild" of the Grand Canyon—once they made it past the Phantom Ranch checkpoint, the research team had little information on where participants were or how they were doing until they

---

[8] See the top half of Figure 1(a) for an example of this behavior. Zoomed-in views of particularly poor GPS tracks make it appear as though the participants were bungee jumping in the steep sections of the canyon.

finally made it out of the canyon.[9] Remote environments with distributed research teams also make it more challenging to communicate critical information about the study and/or participants. All of these factors need to be taken into account and mitigated where able (both in design and in analyses for unanticipated challenges). Many of the changes in methodology across the studies reflect these opportunities for improvement.

While some of the particulars outlined here are unique to the R2R WATCH studies, most moderately complex field studies—especially those with repeated variables or in remote settings—will require a similar level of flexibility and tight understanding of what the resultant data can and cannot indicate based on the nuances of data collection.

4.3 Future Studies

The relatively unconstrained environment of R2R hikers at the Grand Canyon provided a wealth of information on biological, cognitive, and physiological markers of performance; however, future studies could implement additional changes to better measure variables of interest.

Measuring performance during a more extreme task would likely lead to more plentiful detrimental health events in participants, allowing for predictive models to more easily extract the early makers indicative of those later health events and resultant performance decline. While promising from a model-building perspective, it is nontrivial to balance exposing participants to a more extreme task in a natural environment and appropriately prioritizing their safety.

While R2R WATCH went for breadth across a range of data streams in a natural setting, future follow-up studies could narrow in on specific analyses of interest (e.g., focus on group dynamics) or directly test measures in a more controlled setting (e.g., hold participants to a particular pace and/or regulate breaks). It could also be beneficial to study the same participants in different environments. For example, collecting data in a relatively unconstrained setting (as in the present studies) but then following up the next day with additional tests in a more highly controlled setting. This approach could give insight into the individual differences between the participants (e.g., VO_2 max), building up personal profiles and leading to better interpretation of the data from the more naturalistic setting.

Future studies could also pull in additional metrics. For example, continuous interstitial fluid collection (e.g., Miller et al., 2018). It is possible to extract analytes with microneedles that would be minimally invasive to the participants while allowing for real-time, on-board measurements to better understand subtle and sudden changes (rather than just before and after the event). Extended cognitive measurements (not constrained to be fast, self-administered, and run on a small mobile device) would also lead to better insights into changes in cognitive performance. Direct measurement of water and caloric intake along with sleep (rather than self-reported survey questions) would also allow for a tighter understanding of how these factors interact with cognitive, physiological, and biological measures during the task.

Physiologic measurement platforms are continuing to evolve at a rapid pace, with each year seeing commercial devices able to measure important changes with greater and greater fidelity. Devices have been deployed capable of continuous ECG monitoring with the potential to compare not only R to R intervals but also to do more sophisticated evaluations of ECG segment changes. Additionally, point-of-care testing devices (e.g., pocket-sized end-tidal CO_2 monitors) and newer watch-based monitors with SpO_2 sensors have improved. Together, these improvements will allow for a more refined examination of changes in physiology and the determinants that influence them. This will allow for even better enhancement of human performance in challenging environments.

There is also rising interest to develop tools enabling both accurate prediction and augmentation of human performance. While there are many routes for generating such tools, the converging advances in genomic technologies, human genome annotation, and methodologies for assessing physiological and cognitive states in real time during extreme performance tasks have made it possible to envision the elucidation of how deployment of the genome impacts future performance, as well as how pharmacology and pharmacogenetics can be used to augment performance.

4.4 Conclusions

The R2R WATCH project was the first of its kind to collect physiological, cognitive, and biological markers of performance in parallel on a large sample size in a natural extreme environment. The performance data on R2R hikers at the Grand Canyon revealed consistent patterns of performance decline across the data streams, highlighting the strengths and weaknesses of a range of COTS devices and methodologies while paving the way for future analyses and studies.

Acknowledgements

We would like to thank the volunteers from the University of New Mexico, the Grand Canyon National Park rangers, and the Sandia researchers who braved the elements and odd hours to help collect data for the R2R WATCH project. This project was funded by the Defense

---

[9] As an example, in one study we had a pair of hikers who were long overdue to the North Rim for the last shuttle back to the South Rim. It turns out they were so exhausted from the hike that they chose to stop for the night to sleep in one of the bathrooms along the trail. Members of the research team went on a late-night hike, found them, and were able to assist them out of the canyon.

Threat Reduction Agency (DTRA), Project CB10359. Sandia National Laboratories is a multi-mission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. SAND2020-0499 J. This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

Reference

Acharya, U. R., Joseph, K. P., Kannathal, N., Lim, C. M., & Suri, J. S. (2006). Heart rate variability: A review. *Medical and Biological Engineering and Computing*, *44*(12), 1031–1051. https://doi.org/10.1007/s11517-006-0119-0

Adan, A. (2012). Cognitive performance and dehydration. *Journal of the American College of Nutrition*, *31*(2), 71–78. https://doi.org/10.1080/07315724.2012.10720011

Areta, J. L., & Hopkins, W. G. (2018). Skeletal muscle glycogen content at rest and during endurance exercise in humans: A meta-analysis. *Sports Medicine*, *48*(9), 2091–2102. https://doi.org/10.1007/s40279-018-0941-1

Bahrke, M. S., & Shukitt-Hale, B. (1993). Effects of altitude on mood, behaviour and cognitive functioning. *Sports Medicine*, *16*(2), 97–125. https://doi.org/10.2165/00007256-199316020-00003

Baker, L. B. (2017). Sweating rate and sweat sodium concentration in athletes: A review of methodology and intra/interindividual variability. *Sports Medicine*, *47*(1), 111–128. https://doi.org/10.1007/s40279-017-0691-5

Banderet, L. E., & Burse, R. L. (1991). Effects of high terrestrial altitude on military performance. In R. Gal & A. D. Mangelsdorff (Eds.), *Handbook of military psychology* (pp. 233–254). John Wiley & Sons.

Bärtsch, P. & Saltin, B. (2008). General introduction to altitude adaptation and mountain sickness. *Scandinavian Journal of Medicine & Science in Sports*, *18*(S1), 1–10. https://doi.org/10.1111/j.1600-0838.2008.00827.x

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using *lme4*. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Brown, D. J., Brugger, H., Boyd, J., & Paal, P. (2012). Accidental hypothermia. *New England Journal of Medicine*, *367*(20), 1930–1938. https://doi.org/10.1056/NEJMra1114208

Chang, Y. K., Chi, L., Etnier, J. L., Wang, C. C., Chu, C. H., & Zhou, C. (2014). Effect of acute aerobic exercise on cognitive performance: Role of cardiovascular fitness. *Psychology of Sport and Exercise*, *15*(5), 464–470. https://doi.org/10.1016/j.psychsport.2014.04.007

Cheuvront, S. N., Carter, R. III, Castellani, J. W., & Sawka, M. N. (2005). Hypohydration impairs endurance exercise performance in temperate but not cold air. *Journal of Applied Physiology*, *99*(5), 1972–1976. https://doi.org/10.1152/japplphysiol.00329.2005

Cheuvront, S. N., Carter, R. III, & Sawka, M. N. (2003). Fluid balance and endurance exercise performance. *Current Sports Medicine Reports*, *2*(4), 202–208. https://doi.org/10.1249/00149619-200308000-00006

Conners, C. K., & Sitarenios, G. (2011). Conners' Continuous Performance Test (CPT). In: Kreutzer, J. S., DeLuca, J., Caplan, B. (Eds.), *Encyclopedia of Clinical Neuropsychology*. Springer, New York, NY. https://doi.org/10.1007/978-0-387-79948-3_1535.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*, 87–185. https://doi.org/10.1017/S0140525X01003922

Cudaback, D. D. (1984). 4-Km altitude effects on performance and health. *Publications of the Astronomical Society of the Pacific*, *96*(580), 463–477. https://doi.org/10.1086/131365

Davranche, K., Brisswalter, J., & Radel, R. (2015). Where are the limits of the effects of exercise intensity on cognitive control? *Journal of Sport and Health Science*, *4*(1), 56–63. https://doi.org/10.1016/j.jshs.2014.08.004

Divis, K. M., Anderson-Bergman, C., Abbott, R. G., Newton, V. E., & Emmanuel-Aviña, G. (2018). Physiological and cognitive factors related to human performance during the Grand Canyon Rim-to-Rim hike. *Journal of Human Performance in Extreme Environments*, *14*(1), Article 5. https://doi.org/10.7771/2327-2937.1095

Dong, J. G. (2016). The role of heart rate variability in sports physiology. *Experimental and Therapeutic Medicine*, *11*(5), 1531–1536. https://doi.org/10.3892/etm.2016.3104

Doppelmayr, M. M., Finkernagel, H., & Doppelmayr, H. I. (2005). Changes in cognitive performance during a 216 kilometer, extreme endurance footrace: A descriptive and prospective study. *Perceptual and Motor Skills*, *100*(2), 473–487. https://doi.org/10.2466/pms.100.2.473-487

Durmer, J. S. & Dinges, D. F. (2005). Neurocognitive consequences of sleep deprivation. *Seminars in Neurology*, *25*, 117–129. https://doi.org/10.1055/s-2005-867080

Ellis, H. D. (1982). The effects of cold on the performance of serial choice reaction time and various discrete tasks. *Human Factors*, *24*(5), 589–598. https://doi.org/10.1177/001872088202400509

Ellis, H. D., Wilcock, S. E., & Zaman, S. A. (1985). Cold and performance: The effects of information load, analgesics, and the rate of cooling. *Aviation, Space, and Environmental Medicine*, *56*(3), 233–237.

Emmanuel-Aviña, G., Abbott, R. G., Anderson-Bergman, C., Branda, C., Divis, K. M., Jelínková, L., Newton, V., Pearce, E., & Femling, J. (2017). Rim-to-Rim Wearables at the Canyon for Health (R2R WATCH): Experimental design and methodology. *Augmented Cognition, Neurocognition, & Machine Learning*, 263–274. https://doi.org/10.1007/978-3-319-58628-1_21

Emmanuel-Aviña, G., Divis, K. M., & Abbott, R. G. (2018). Physiological state in extreme environments. *Pervasive & Mobile Computing*, *50*, 137–147.

Epstein, Y., & Yanovich, R. (2019). Heatstroke. *New England Journal of Medicine*, *380*(25), 2449–2459. https://doi.org/10.1056/NEJMra1810762

Finsterer, J., & Mahjoub, S. Z. (2014). Fatigue in healthy and diseased individuals. *American Journal of Hospice and Palliative Medicine*, *31*(5), 562–575. https://doi.org/10.1177/1049909113494748

Goldberg, L. R. (1999). A broad-bandwidth, public domain, personality inventory measuring lower-level facets of several five-factor models. In I. Mervielde, I. Deary, F. De Fruyt, & F. Ostendorf (Eds.), *Personality psychology in Europe* (Vol. 7, pp. 7–28). Tilburg University Press.

Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. C. (2006). The International Personality Item Pool and the future of public-domain personality measures. *Journal of Research in Personality*, *40*, 84–96. https://doi.org/10.1016/j.jrp.2005.08.007

Halperin, I., Chapman, D. W., & Behm, D. G. (2015). Non-local muscle fatigue: Effects and possible mechanisms. *European Journal of Applied Physiology*, *115*(10), 2031–2048. https://doi.org/10.1007/s00421-015-3249-y

Hancock, P. A. (1982). Task categorization and the limits of human performance in extreme heat. *Aviation, Space, & Environmental Medicine*, *53*(8), 778–784.

Hancock, P. A., Ross, J. M., & Szalma, J. L. (2007). A meta-analysis of performance response under thermal stressors. *Human Factors*, *49*(5), 851–877. https://doi.org/10.1518/001872007X230226

Hancock, P. A., & Vasmatzidis, I. (2003). Effects of heat stress on cognitive performance: The current state of knowledge. *International Journal of Hyperthermia*, *19*(3), 355–372. https://doi.org/10.1080/0265673021000054630

Harrison, Y., & Horne, J. A. (2000). The impact of sleep deprivation on decision making: A review. *Journal of Experimental Psychology: Applied*, *6*(3), 236. https://doi.org/10.1037//1076-898x.6.3.236

Higgins, J. P., Tuttle, T., & Higgins, J. A. (2010). Altitude and the heart: Is going high safe for your cardiac patient? *American Heart Journal*, *159*(1), 25–32. https://doi.org/10.1016/j.ahj.2009.10.028

Hoffman, M. D., Stellingwerff, T., & Costa, R. J. (2019). Considerations for ultra-endurance activities: Part 2—Hydration. *Research in Sports Medicine*, *27*(2), 182–194. https://doi.org/10.1080/15438627.2018.1502189

Holtzhausen, L. M., & Noakes, T. D. (1995). The prevalence and significance of post-exercise (postural) hypotension in ultramarathon runners. *Medicine & Science in Sports & Exercise*, *27*(12), 1595–1601.

Hurdiel, R., Pezé, T., Daugherty, J., Girard, J., Poussel, M., Poletti, L., Basset, P., & Theunynck, D. (2015). Combined effects of sleep deprivation and strenuous exercise on cognitive performances during The North Face® Ultra Trail du Mont Blanc® (UTMB®). *Journal of Sports Sciences*, *33*(7), 670–674. https://doi.org/10.1080/02640414.2014.960883

Hurdiel, R., Riedy, S. M., Millet, G. P., Mauvieux, B., Pezé, T., Elsworth-Edelsten, C., Martin, D., Zunquin, G., & Dupont, G. (2018). Cognitive performance and self-reported sleepiness are modulated by time-of-day during a mountain ultramarathon. *Research in Sports Medicine*, *26*(4), 482–489. https://doi.org/10.1080/15438627.2018.1492401

Hyde, D., Thomas, J. R., Schrot, J., & Taylor, W. F. (1997). *Quantification of special operations mission-related performance: Operational evaluation of physical measures*. Naval Medical Research Institute.

Jelínková, L., Pearce, E., Bossart, C., Garcia, R., & Femling, J. (2017). Rim-to-Rim Wearables at the Canyon for Health (R2R WATCH): Correlation of clinical markers of stress with physiological COTS data. *Augmented Cognition, Neurocognition, & Machine Learning*, 312–322. https://doi.org/10.1007/978-3-319-58628-1_25

Knechtle, B., Knechtle, P., Wirth, A., Alexander Rüst, C., & Rosemann, T. (2012). A faster running speed is associated with a greater body weight loss in 100-km ultra-marathoners. *Journal of Sports Sciences*, *30*(11), 1131–1140. https://doi.org/10.1080/02640414.2012.692479

Knechtle, B., Rosemann, T., Knechtle, P., & Lepers, R. (2010). Predictor variables for a 100-km race time in male ultra-marathoners. *Perceptual & Motor Skills*, *111*(3), 681–693. https://doi.org/10.2466/05.25.PMS.111.6.681-693

Kramer, A. F., Coyne, J. T., & Strayer, D. L. (1993). Cognitive function at high altitude. *Human Factors*, *35*(2), 329–344. https://doi.org/10.1177/001872089303500208

Kuepper, T., & Classen, J. (2002). Single grand mal seizures provoked by altitude? *Journal of Travel Medicine*, *9*(2), 94–96. https://doi.org/10.2310/7060.2002.21498

Landman, Z. C., Landman, G. O., & Fatehi, P. (2012). Physiologic alterations and predictors of performance in a 160-km ultramarathon. *Clinical Journal of Sport Medicine*, *22*(2), 146–151. https://doi.org/10.1097/JSM.0b013e318243ffdc

Lejuez, C. W., Read, J. P., Kahler, C. W., Richards, J. B., Ramsey, S. E., Stuart, G. L., Strong, D. R., & Brown, R. A. (2002). Evaluation of a behavioral measure of risk taking: the Balloon Analogue Risk Task (BART). *Journal of Experimental Psychology: Applied*, *8*, 75–84. https://doi.org/10.1037//1076-898x.8.2.75

Lipman, G. S., Eifling, K. P., Ellis, M. A., Gaudio, F. G., Otten, E. M., & Grissom, C. K. (2014). Wilderness Medical Society practice guidelines for the prevention and treatment of heat-related illness: 2014 update. *Wilderness & Environmental Medicine*, *25*(4), S55–S65. https://doi.org/10.1016/j.wem.2018.10.004

Lowe, M., Harris, W., Kane, R. L., Banderet, L., Levinson, D., & Reeves, D. (2007). Neuropsychological assessment in extreme environments. *Archives of Clinical Neuropsychology*, *22S*, S89–S99. https://doi.org/10.1016/j.acn.2006.10.010

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*(6657), 279–281. https://doi.org/10.1038/36846

Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods*, *49*(4), 1494–1502. https://doi.org/10.3758/s13428-016-0809-y

Maggiorini, M., Buhler, B., Walter, M., & Oelz, O. (1990). Prevalence of acute mountain sickness in the Swiss Alps. *BMJ*, *301*(6756), 853–855. https://doi.org/10.1136/bmj.301.6756.853

Makivić, B., Nikić Djordjević, M., & Willis, M. S. (2013). Heart rate variability (HRV) as a tool for diagnostic and monitoring performance in sport and physical activities. *Journal of Exercise Physiology Online*, *16*(3).

Maruff, P., Snyder, P., McStephen, M., Collie, A., & Darby, D. (2006). Cognitive deterioration associated with an expedition in an extreme desert environment. *British Journal of Sports Medicine*, *40*(6), 556–560. https://doi.org/10.1136/bjsm.2004.016204

Miller, P. R., Taylor, R. M., Tran, B. Q., Boyd, G., Glaros, T., Chavez, V. H., Krishnakumar, R., Sinha, A., Poorey, K., Williams, K. P., Branda, S. S., Baca, J. T., & Polsky, R. (2018). Extraction and biomolecular analysis of dermal interstitial fluid collected with hollow microneedles. *Communications Biology*, *1*(1), 173. https://doi.org/10.1038/s42003-018-0170-z

Moore, R. D., Romine, M. W., O'Connor, P. J., & Tomporowski, P. D. (2012). The influence of exercise-induced fatigue on cognitive function. *Journal of Sports Sciences*, *30*(9), 841–850. https://doi.org/10.1080/02640414.2012.675083

Murray, B. (2007). Hydration and physical performance. *Journal of the American College of Nutrition*, *26*(5), 542S–548S. https://doi.org/10.1080/07315724.2007.10719656

National Park Service. US Department of the Interior. (2019). *Grand Canyon 2018 annual report branch of emergency services*. Retrieved from https://www.nps.gov/grca/learn/management/upload/2018_BES_Annual_Report_508_Compliant.pdf

Noe, C. R., Choudhary, E., Cheng-Dobson, L. J., Wolkin, A. F., & Newman, C. S. (2013). Exertional heat-related illnesses at the Grand Canyon National Park, 2004–2009. *Wilderness & Environmental Medicine*, *24*(4), 422–428. https://doi.org/10.1016/j.wem.2013.06.008

Pearce, E. A., Jelínková, L., Fullerton, L., Malcom, C. J., Heinrich, H. L., Norwil, E. J., Plate, A., Hall, M., Bossart, C. S., Schaller, B. S., Harrell, A. J., & Femling, J. K. (2019). Observational study of Grand Canyon rim-to-rim day hikers: Determining behavior patterns to aid in preventive search and rescue efforts. *Wilderness & Environmental Medicine*, *30*(1), 4–11. https://doi.org/10.1016/j.wem.2018.08.001

Peterson, C., & Seligman, M. E. P. (2004). *Character strengths and virtues: A handbook and classification*. American Psychological Association.

Pilcher, J. J., & Huffcutt, A. I. (1996). Effects of sleep deprivation on performance: a meta-analysis. *Sleep*, *19*(4), 318–326. https://doi.org/10.1093/sleep/19.4.318

Pilcher, J. J., Nadler, E., & Busch, C. (2002). Effects of hot and cold temperature exposure on performance: A meta-analytic review. *Ergonomics*, *45*(10), 682–698. https://doi.org/10.1080/00140130210158419

Popkin, B. M., D'Anci, K. E., & Rosenberg, I. H. (2010). Water, hydration, and health. *Nutrition Reviews*, *68*(8), 439–458. https://doi.org/10.1111/j.1753-4887.2010.00304.x

Rammstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality*, *41*, 203–212. https://doi.org/10.1016/j.jrp.2006.02.001

Rattray, B., & Smee, D. J. (2016). The effect of high and low exercise intensity periods on a simple memory recognition test. *Journal of Sport and Health Science*, *5*(3), 342–348. https://doi.org/10.1016/j.jshs.2015.01.005

Rim to Rim. (2012). *Hiking map: North Kaibab, South Kaibab, and Bright Angel trails*. Retrieved from https://www.rimtorim.org/hike-it/

Rodahl, K. (2003). Occupational health conditions in extreme environments. *Annals of Occupational Hygiene*, *47*(3), 241–252. https://doi.org/10.1093/annhyg/meg033

Sawka, M. N., Cheuvront, S. N., & Kenefick, R. W. (2015). Hypohydration and human performance: Impact of environment and physiological mechanisms. *Sports Medicine*, *45*(Suppl. 1), 51–60. https://doi.org/10.1007/s40279-015-0395-7

Sawka, M. N., Toner, M. M., Francesconi, R. P., & Pandolf, K. B. (1983). Hypohydration and exercise: Effects of heat acclimation, gender, and environment. *Journal of Applied Physiology*, *55*(4), 1147–1153. https://doi.org/10.1152/jappl.1983.55.4.1147

Sawka, M. N., & Wenger, C. B. (1988). Physiological responses to acute exercise-heat stress (No. USARIEM-M-14/88). *Army Research Institute of Environmental Medicine.*

Staal, M. A. (2004). Stress, cognition, and human performance: A literature review and conceptual framework (NASA Technical Memorandum 212824). NASA Ames Research Center.

Strydom, N. B., & Holdsworth, L. D. (1968). The effects of different levels of water deficit on physiological responses during heat stress. *European Journal of Applied Physiology*, *26*(2), 95–102. https://doi.org/10.1007/BF00699507

Stück, M., Balzer, H. U., Hecht, K., & Schröder, H. (2005). Psychological and psychophysiological effects of a high-mountain expedition to Tibet. *Journal of Human Performance in Extreme Environments*, *18*(1), 11–20. https://doi.org/10.7771/2327-2937.1042

Tempest, G. D., Davranche, K., Brisswalter, J., Perrey, S., & Radel, R. (2017). The differential effects of prolonged exercise upon executive function and cerebral oxygenation. *Brain & Cognition*, *113*, 133–141. https://doi.org/10.1016/j.bandc.2017.02.001

Tomporowski, P. D. (2003). Effects of acute bouts of exercise on cognition. *Acta Psychologica*, *112*(3), 297–324. https://doi.org/10.1016/S0001-6918(02)00134-8

Virués-Ortega, J., Buela-Casal, G., Garrido, E., & Alcázar, B. (2004). Neuropsychological functioning associated with high-altitude exposure. *Neuropsychology Review*, *14*(4), 197–224. https://doi.org/10.1007/s11065-004-8159-4

Wallsten, T. S., Pleskac, T. J., & Lejuez, C. W. (2005). Modeling behavior in a clinically diagnostic sequential risk-taking task. *Psychological Review*, *112*, 862–880. https://doi.org/10.1037/0033-295X.112.4.862

White, T. L., Lejuez, C. W., & de Wit, H. (2008). Test-retest characteristics of the Balloon Analogue Risk Task (BART). *Experimental & Clinical Psychopharmacology*, *16*, 565–570. https://doi.org/10.1037/a0014083

Wickens, C. D., Keller, J. W., & Shaw, C. (2015). Human factors in high-altitude mountaineering. *Journal of Human Performance in Extreme Environments*, *12*(1), Article 1. https://doi.org/10.7771/2327-2937.1065

Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology & Psychology*, *18*(5), 459–482. https://doi.org/10.1002/cne.920180503

Zouboules, S. M., Lafave, H. C., O'Halloran, K. D., Brutsaert, T. D., Nysten, H. E., Nysten, C. E., Steinback, C. D., Sherpa, M. T., & Day, T. A. (2018). Renal reactivity: Acid-base compensation during incremental ascent to high altitude. *Journal of Physiology*, *596*(24), 6191–6203. https://doi.org/10.1113/JP276973

Appendix: Measures Across Studies

This appendix highlights the details of and changes in methodology across the studies. Tables A1, A2, A3, A4, and A5 indicate which items were included in each study, along with key updates (such as timing of the cognitive tasks).

Table A1
Enrollment, survey, and personality inclusion across the R2R WATCH studies.

	Enrolment			**Survey**				**Personality**	
Data collection	**Trailhead**	**Pre-enrolled**	**Total**	**Start**	**Mid**	**Finish**	**Follow-Up**	**Collected**	**Location**
Study 1	✓	✓	73	✓	✓	✓	—	✓	North Rim
Study 2	✓	✓	27	✓	✓	✓	✓	✓	North Rim
Study 3	—	✓	38	✓	✓	✓	✓	✓	Pre-enrollment

Note. Check marks (✓) indicate that item was included in the study; dashes (—) indicate that it was not included.

Table A2
Cognitive task inclusion and design across the R2R WATCH studies.

	Cognitive Tasks						
Data collection	**Fatigue**	**Go/No-Go**	**VSTM**	**BART**	**Training (pre-enrollment)**	**Feedback**	**Timing**
Study 1	✓	✓	✓	—	—	✓	3 hours
Study 2	✓	✓	✓	—	Partial (Pre-Enrolled Only)	—	5 miles
Study 3	✓	✓	✓	✓	✓	—	5 miles

Note. Check marks (✓) indicate that item was included in the study; dashes (—) indicate that it was not included.

Table A3
Device inclusion across R2R WATCH studies.

	Data collection		
Device	**Study 1**	**Study 2**	**Study 3**
Apple iPod Touch 6	✓	✓	✓
SensorPush Smart Sensor	✓	✓	✓
Fitbit Charge HR	✓	—	—
Garmin vívoactive HR	✓	—	—
Garmin fēnix 3 HR	✓	✓	✓
Garmin fēnix 5 HR	—	✓	✓
Suunto Spartan Ultra	✓	—	—
Garmin tempe	✓	✓	✓
Garmin eTrex 10	✓	—	—
Suunto Smart Sensor	✓	—	—
Wahoo TICKRx	✓	✓	✓
LifeBEAM Smart Hat	✓	—	—
Myontec Mbody Shorts	✓	—	—
Garmin foot pod	✓	✓	✓
Polar Bluetooth Smart Stride	✓	—	—
Empatica E4	✓	—	—
Equivital	—	—	✓

Note. Check marks (✓) indicate that a device was sent out during that study. Dashes (—) indicate that it was not included in that study. Note that a given participant may not have worn all the devices that were sent out during that study.

Table A4
Blood analytes included across the R2R WATCH studies.

		Blood measures			
Analysis	**Category**	**Measure**	**Study 1**	**Study 2**	**Study 3**
Tricore Laboratory	Basic metabolic panel	Anion gap	✓	✓	—
		Blood urea nitrogen (BUN)	✓	✓	
		Calcium	✓	✓	—
		Chloride	✓	✓	—
		Carbon dioxide (CO_2)	✓	✓	—
		Creatinine	✓	✓	—
		Est. glomerular filtration rate (eGFR)	✓	✓	—
		Glucose	✓	✓	—
		Potassium	✓	✓	—
		Sodium	✓	✓	—
	Creatine kinase (CK)	Creatine kinase (CK)	✓	✓	✓
Blood gas analyzer	Direct measurements	pH	—	—	✓
		Partial pressure of carbon dioxide (PCO_2)	—	—	✓
		Partial pressure of oxygen (PO_2)	—	—	✓
		Oxygen saturation ($SO_2\%$)	—	—	✓
		Hematocrit	—	—	✓
		Sodium (Na^+)	—	—	✓
		Potassium (K^+)	—	—	✓
		Chloride (Cl^-)	—	—	✓
		Total carbon dioxide (TCO_2)	—	—	✓
		Calcium (Ca^{++})	—	—	✓
		Magnesium (Mg^{++})	—	—	✓
		Glucose	—	—	✓
		Lactate	—	—	✓
		UTea (BUN)	—	—	✓
		Creatinine	—	—	✓
	CO-oximetry tests	Deoxyhaemoglobin (HHb)	—	—	✓
		Oxyhemoglobin (O2Hb)	—	—	✓
		Methemoglobin (MetHb)	—	—	✓
		Carboxyhemoglobin (COHb)	—	—	✓
		Total hemoglobin (tHB)	—	—	✓
		Oxygen saturation ($SO_2\%$)	—	—	✓
		Total bilirubin (tBil)	—	—	✓
		Fetal Hemoglobin (HbF)	—	—	✓

Note. Check marks (✓) indicate that item was included in the study; dashes (—) indicate that it was not included.

Table A5
Survey questions included in (a) pre-hike, (b) mid-hike, (c) post-hike, and (d) follow-up surveys.

(a) Start (South Kaibab)		**Study 1**	**Study 2**	**Study 3**
Demographics	Gender	✓	✓	✓
	Age	✓	✓	✓
	Zip	✓	✓	✓
	When arrived at Grand Canyon	—	—	✓
	Longest distance (6 months)	✓	✓	✓
Vitals	Weight	✓	✓	✓
	Pack weight	✓	✓	✓
	HR (standing)	✓	✓	✓
	SpO_2	✓	✓	✓
	Body fat (%)	—	✓	✓
	BMI	*	✓	✓
	Height	✓	—	—
R2R/Fatigue	R2R prior?	✓	✓	✓
	R2R in single day?	✓	✓	✓
	How long expect R2R to take?	✓	✓	✓
	Preparedness (1-10)	✓	✓	✓
	Fatigue now (1-10)	✓	✓	✓
Exercise	Times exercise/week (6 months)	✓	✓	✓
	Exercise alone or others	✓	✓	✓
	Time of day exercise	✓	✓	✓
Food/Fluids	Alcohol since 4 p.m. yesterday?	✓	✓	✓
	Caffeine since 4 p.m. yesterday?	✓	✓	✓
Sleep	Sleep last night (hours)	✓	✓	✓
	Sleep two nights ago (hours)	—	—	✓
	Sleep quality last night (1-10)	✓	✓	✓
	Sleep quality two nights ago (1-10)	—	—	✓
	Trouble sleeping—problems?	✓	✓	✓
	Trouble sleeping—reasons?	✓	✓	✓
	Typical hours sleep (6 months)	✓	✓	✓
	Needed hours of sleep	✓	✓	✓
	Morning or night person	✓	✓	✓
(b) Mid-hike (Phantom Ranch)		**Study 1**	**Study 2**	**Study 3**
	Fluid since last checkpoint?	✓	✓	✓
	Preparedness (1-10)	✓	✓	✓
	Fatigue now (1-10)	✓	✓	✓
	Most fatigued going downhill (1-10)	✓	✓	✓
(c) Finish (North Kaibab)		**Study 1**	**Study 2**	**Study 3**
Vitals	Weight	✓	✓	✓
	Pack weight	✓	✓	✓
	HR (standing)	✓	✓	✓
	SpO_2	✓	✓	✓
	Body fat (%)	—	✓	✓
	BMI	*	✓	✓
	Height	✓	—	—
R2R/Fatigue	Preparedness (1-10)	✓	✓	✓
	Fatigue when finished (1-10)	✓	✓	✓
	Most fatigued (1-10)	✓	✓	✓
Food/Fluids	Fluid since Phantom?	✓	✓	✓
	Pain meds?	✓	✓	✓
	Blood pressure meds or water pills?	✓	✓	✓
	Caffeine during hike?	✓	✓	✓
	Sodium/salt during hike?	✓	✓	✓
Sleep	Sleep affect performance?	✓	✓	✓
Groups	Hike with others? Do it again? Speed?	✓	✓	✓
(d) Follow-up (online)		**Study 1**	**Study 2**	**Study 3**
	Preparedness (1-10)	—	✓	✓
	Physical/mental week later	—	✓	✓
	Injuries, symptoms, medical treatment	—	✓	✓

Note. Check marks (✓) indicate questions of that kind were included in that study; dashes (—) indicated they were not.
*BMI was not explicitly measured but could be calculated from the other metrics.