A Sensitivity Analysis on an Economic Value Metric for Evaluating Exoskeleton Efficacy

Nundini D Rawal, Roberto Leo Medrano, Gray Cortright Thomas, Elliott J Rouse
University of Michigan Ann Arbor
Ann Arbor, USA
nundinir@umich.edu

***Abstract*—The minimum threshold of perceivable lower limb exoskeleton assistance is larger than the assistance provided by most state-of-the-art devices designed according to physiological metrics of success [1]. This discrepancy indicates that people are largely insensitive to metabolic benefits provided by these technologies, and the failure to perceive these benefits may impact adoption and use within society. This motivates the need for a metric that can capture exoskeleton performance along additional dimensions - ergonomics, comfort, and stability, to drive designs that users value in a holistic sense.**

In this paper, we perform a sensitivity analysis on a novel metric and associated methodology grounded in effort perception, to quantify exoskeleton success using the dollar ($) value of its assistance. Participants complete a strenuous uphill walking task during which they are periodically asked to assign a dollar value to the effort of walking (price-to-walk), while engaging in a Seller's Vickrey second-price auction. Over time, these dollar values (bids) create a specific fatigue profile which is used to determine the marginal value (MV) of exoskeleton use.

Price-to-walk ($) is guided by the pressures of the blinded Vickrey auction which features automated competing bidders modelled on human fatigue behavior (robobidders). In this study, we found that MV is insensitive to changing robobidder model parameters and was not significantly higher than our minimum detectable threshold of $82.098 or $24.91 log-dollars, necessary for a noticeable impact on participant price-to-walk.

***Index Terms*—Metrics, Vickrey Auction, Ankle-foot Exoskeletons, Lower-limb Wearable Devices, Effort Perception**

I. Introduction

A. High Level Overview

Powered lower limb exoskeletons reduce exertion and facilitate tasks by augmenting motion of specific joints in the leg. They have the potential to assist and even rehabilitate not only individuals with abnormal gait patterns and pathological conditions, but also healthy, able-bodied individuals [2] [3]. In the past decade, several studies have demonstrably broken the "metabolic barrier", showcasing improvements to walking and running economy that exceed the metabolic cost of wearing the device itself [4]. Despite the potential value of such assistive devices to consumers, they are still perceived as impractical for use beyond specialized applications that involve prolonged, often intense, physical labor. This perspective stems from a variety a reasons, such as lack of adaptability to variable environments and tasks critical to performing activities of daily living, weight of the device itself, safety and dependability, affordability and portability [5] [6] [7] [8] [9].

The engineering design process requires some metric(s) to guide the efficacy evaluation of the design at each iteration. For exoskeletons, such metrics are not just restricted to determining the integrity of the system and validating design choices, but also set a standard at which significant, valuable assistance is delivered to the user. The accuracy and sensitivity of these metrics are therefore correlated with differences or improvements in human performance. The chosen benchmarks must be able to capture user intent and be sensitive enough to recognize improvements in human performance, physiologically or otherwise.

There are several well-known metrics in wearable robotics that guide the design, efficacy and commercialization of assistive devices [10] [11] [12]. To name a few:

- Energy expenditure as indicated through metabolic cost
- Reductions in electromyographic (EMG) signal amplitude, indicating reduced muscle fiber activation
- Stability measures like gait symmetry, step speed and cadence
- Limb kinetics and kinematics with ground reaction forces and net joint torque reductions
- Functional or task dependent metrics related to performance time, quality of execution and speed
- Subjective measures of system usability, including pain or effort ranking through the use of the Likert Scale or Ratings of Perceived Exertion (RPE) scale

Apart from effort ranking, a majority of these metrics are collected directly from the human body and are privy to the nuances of human motion and intention, reflecting enhancements or changes in performance due to external inputs. Out of all of these metrics, metabolic cost is considered the 'gold standard' both in research settings and in the exoskeleton industry for detecting human performance improvements [13] [14]. Research has shown that metabolic cost stress markers, like blood lactate concentration and %VO2 max, are highly correlated with ratings of perceived exertion during exercise regimens [15] [16] [17].

The neurophysiology behind such metrics, however, is poorly understood as it relates to perception of effort compared to perception of exertion. Effort is defined as "the amount of mental or physical energy being given to a task" [7], while exertion is "associated with physical and physiological stress" [7] and is generally thought to be derived from afferent input [18]. There are 2 competing models to understand effort perception – the afferent feedback model and the corollary discharge model. The former contends that the recruitment of skeletal muscle fibers leads to the perception of effort, while the latter contends that the magnitude of communication between the motor and sensory neurons in the cerebral cortex influences perception of effort [18]. Some have proposed that perception of effort during exercise may be independent of afferent feedback altogether, suggesting that effort incorporates factors beyond exertion. Thus, related physiological metrics that depend on exertion may not be the most direct window into perception of effort.

Unlike standalone robotic systems, wearable systems must be able to account for the user, complicating the design and control since these devices must work in tandem with the human while recognizing the inherent limits to human motion and perception. There is an ongoing discussion about whether effective exoskeletons should only aid below the perceivable limit and mimic how seamlessly our natural limb operates or whether they should provide physical augmentation allowing users to perform beyond an average human capability. Medrano et al. recently found that users do not consistently perceive even large metabolic cost benefits provided by lower limb assistive systems compared to not wearing an exoskeleton during short term use. Specifically, the perceptual threshold of noticeable exoskeleton assistance was 22.7% metabolic improvement at 75% accuracy, while modern exoskeletons have consistently provided a 14% metabolic improvement compared to not wearing an exoskeleton. This discrepancy is troubling and has negative implications for device adoption, as research suggests that the design approach providing noticeable exoskeleton assistance (effort perception) is consistently valued more highly by users [1]. King et al. has found that there is a generalizable correlation between consciously perceiving usefulness of a technology and the intention to adopt it as well [19].

Current initiatives to incorporate individualized user experience directly into exoskeleton operation include Human-in-the-Loop optimization, which seeks to train a personalized predictive model to determine appropriate exoskeleton control parameters while accounting for measured human performance. While the metric of performance has generally been to minimize metabolic cost, user-preference as a metric has emerged as an additional optimization avenue. Various approaches have been taken to create frameworks to either learn entire preference landscapes or to optimize gait or comfort. Implementation of this preference-in-the-loop control in lower limb assistive devices has shown promising results [20] [21] [22]. For example, Tucker et. al demonstrated agreement between user-preference-optimized gait and quantitative gait measures of stability using the Atalante lower limb exoskeleton. Questions surrounding the variability in human self-perception, which such an optimization tactic is dependent on, have naturally arisen. Ingraham et al. identified repeatability of preference identification by engaging individuals in a paradigm that allowed for self-tuning of a timing based controller, ultimately determining that participants were able to reliably self-identify their preferences [23].

Altogether, designing exoskeletons and other lower limb augmentative systems around physiologic metrics such as metabolic cost may not capture criteria most valuable to users. The success of user-centered exoskeleton control and the potential for perception of effort to be measurable and reliable, suggests that a perception-based behavioral metric may be able to better capture noticeable improvements in exoskeleton performance, improving translation and adoption within society in the long run.

B. A Closer Look

1) An Additional Metric - Marginal Value: 'Marginal Value' (MV) evaluated through a Vickrey Auction paradigm is a novel metric under development that quantifies exoskeleton success using the perceived dollar value of its assistance. To obtain MV, individuals complete a fatiguing task to amplify their fatigue perception while being intermittently asked to report dollar amounts corresponding to their willingness to continue the task, termed Price-to-walk (PTW). It is expected that as individuals complete this fatiguing task, their PTW will also gradually increase, reflecting the increase in the task cost. This can be compiled as a unique fatigue profile. If an exoskeleton is providing noticeable assistance, the fatigue profile should be lower than the case where no exoskeleton is worn. MV, as measured by taking the difference between walking with an exoskeleton and some baseline state, promises to capture the exoskeleton value perceived by the user. In other words, MV is the area between two fatigue model curves, representing the value added to the user in terms of dollars when using the exoskeleton. An example of the resultant curves can be seen in **Figure 1**.

MV may allow us to evaluate exoskeletons in an intuitive manner, helping to gauge the monetary value individuals assign to assistive devices. It may have the potential to capture a more holistic representation of the exoskeleton interaction and use by considering nuanced ergonomic factors like comfort, stability, exertion, and speed that impact user experience. MV may therefore serve as an valuable additional metric for guiding exoskeleton design.

Given the potential for user-centric metrics in evaluating wearable devices, a methodology to capture honest user perception is necessary. For this reason, we turn to the Vickrey Auction paradigm common in behavioral economics.

2) An Introduction to Vickrey Auction: A Vickrey Auction is a type of strategic second-price blinded auction common in behavioral economics that is leveraged to elicit truthful valuations from participants regarding a certain item or, in our protocol, an activity. In particular, a Seller's Vickrey Auction was utilized in which the participant acted as a seller and sold their time to walk to the researcher while competing against other participants. Competition-induced pressures from engaging in the blinded auction ensure that truthful valuations are elicited from participants.

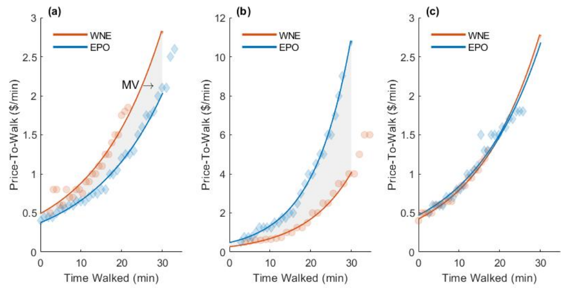

Fig. 1. A demonstration of price-to-walk curves: curves fitted through price-to-walk data for walking without the exoskeleton (WNE) and walking with exoskeleton powered on (EPO) represents the unique fatigue profiles for each activity for a specific subject across a standardized interval. Marginal value (MV) is the area between the two curves, representing the improvement from exoskeleton use. (a). A shallower EPO curve indicates positive assistance or value added through exoskeleton use as participants require less monetary incentive to continue uphill walking with the aid of the exoskeleton. (b). In this case, the exoskeleton is hindering the user as indicated by a steeper EPO curve as compared to the WNE curve (c). EPO and WNE curves that are nearly identical indicate that exoskeleton assistance is imperceptible.

The Seller's Vickrey Auction works by rewarding the individual with the lowest bid - the "winner" - with the payout of the second lowest bid. This is because the individual with the lowest bid requires the least incentive to continue walking. At any given point, the participant is only aware of their own bid, whether they have won or not, and the winning payout which is the second lowest bid. This system has inbuilt guardrails, where the sealed bids force participants to rely on their own internal value system, and the incentive structure limits inflated bidding ensuring that participants report their true value for a given task. Specifically, outrageously high bidding was discouraged as participants would not win any monetary incentive for doing so, while outrageously low bidding was discouraged as participants would have to sustain the discomfiting uphill walking task.

In our application, the bids or price-to-walk refers to the minimum amount of money in dollars it takes to encourage participants to start or sustain walking. As participants become more fatigued during walking, their price-to-walk will simultaneously increase as they require more monetary motivation to sustain the activity.

The Vickrey Auction paradigm has been primarily leveraged in the behavioral economics domain, with success in marketing applications for determining appropriate price points for certain goods and services. Variations of the auction have been used in electronic commerce and adopted for use in computational multi-agent systems. While there are limitations to the protocol, including vulnerability to bidder collusion and dishonesty i.e. strategic bidding when administered in-person, the blinded nature of the auction and the often computerized administration of the protocol can minimize these effects. When compared to an alternate economic auction - the random nth price auction - which was designed to reduce strategic bidding, the Vickrey Auction was shown to be acceptable and effective at eliciting honest valuations [24] [25] [26].

3) Automated Bidders: Logistically, the Seller's Vickrey Auction presents a challenge when coordinating the sealed auction, so robobidders that mimic competing human bidders were a practical "in-lab" solution. Robobidder behavior is governed by a first-order exponential model, which was determined from an initial pilot study that fitted resulting price-to-walk trends of real human participants to these functions. This model mimics the behavior of actual competing human participants, as the competing price-to-walk exponentially increases as participant fatigue increases. While a resting model following a two-minute walking interval was considered, it was not implemented as the price-to-walk did not sufficiently decrease in between walking cycles. One of the randomly selected robobidders out of the multiple that were implemented was selected to traverse the exponential fatigue model when the human participant lost a round of the Vickrey Auction, mimicking a real competing participant that has won and must walk the uphill incline for the next two minutes.

$$y = k \cdot e^{(b \cdot t_k)} + \sigma \tag{1}$$

In the equation above, y refers to the robot participant's bid, k is it's initial bid, b is the rate at which price-to-walk changes, t_k is the time at which the participant has walked, and σ is noise drawn from a zero mean gaussian distribution with a standard deviation of $0.01, meant to corrupt the behavior of the robobidders and replicate variability in real human bidding behavior.

Medrano et al.'s pilot experiments, with use of these robobidders, have shown that participant price-to-walk does indeed decrease as exoskeleton assistance increases, indicating that this metric can successfully detect a decrease in effort perception with exoskeleton aid. Specifically, this decrease in price-to-walk suggests that exoskeleton use adds a modest 4.2% value for the user [1].

4) Study Motivation: While the robobidders are a practical solution, they are also a major confound as the models governing their competitive behavior rely on parameters that may impact real participant behavior and subsequent price-to-walk trends. The number of competing robobidders may also impact how quickly participant bidding behavior converges. **Figure 2** depicts the various robobidder parameter combinations that were tested as part of this sensitivity analysis.

Since the number of robo-bidders, their growth rate and initial value (as represented by k and b in the governing **equation 1**) were kept constant throughout Medrano et. al's previous demonstration, the purpose of this study is to perform a sensitivity analysis to determine whether robo-bidder parameters have an impact on the resultant price-to-walk trends of actual participants it is competing against. Determining robobidder impact on value to walk will allow the previous results generated via Vickrey Auction to be accepted and will allow for the continued use of such robobidders.

II. Methods

1) Study Design: A fractional factorial experimental design was implemented to test the impact of varying robobidder

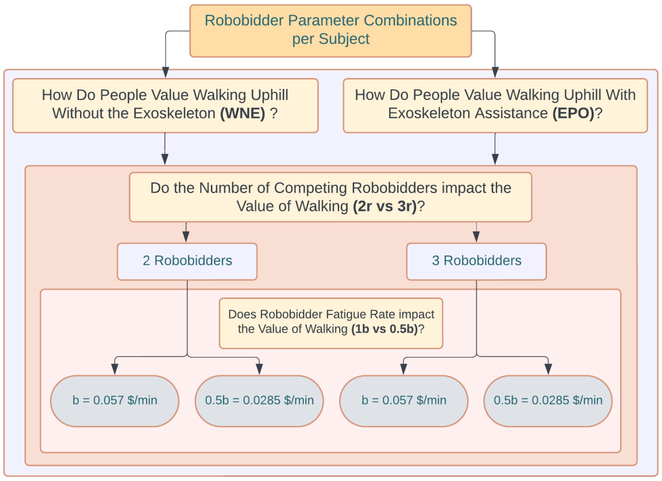

Fig. 2. Each of the following robobidder models were implemented to compete against participants - 2r0.5b, 2r1b, 3r0.5b, 3r1b. The value before the 'r' refers to the number of competing robobidders, while the value before the 'b' refers to a scaling of the nominal growth rate of $0.05735/minute established by Medrano et al.

model parameters on participant bidding behavior. Each participant completed four trials for two conditions – walking without the exoskeleton (WNE) and walking with the exoskeleton powered on (EPO). The robobidder growth rate (b) and number of robobidders were selected as model parameters to be varied. MV served as a dependent variable, while initial bidding value (k) was held constant. We hypothesized that participant bidding strategies would be influenced by the changing value system of robobidders, resulting in convergence behaviors. If MV significantly changed across these conditions, it would indicate that participant bidding behavior is sensitive to robobidder behavior. **Figure 2** above depicts the various robobidder parameter combinations that were tested as part of the sensitivity analysis. The nomenclature 2r0.5b and 3r0.5b refers to 2 and 3 competing robobidders respectively, that share a growth rate of 0.5b. 2r1b and 3r1b refers to the same robobidder count but with twice the growth rate. This experimental design set-up was used to reduce the number of experimental trials, but still yield significant results.

The growth rate and number of robobidders were chosen to be varied about the 2 robobidder, 1b growth rate of $0.05735/minute established by Medrano et al. Half the 1b growth rate and including an additional robobidder was deemed sufficient to test any differences in participant behavior as the perturbation was significant enough to be noticeable during pilot testing. For both WNE and EPO conditions, the initial bidding value k was kept fixed at $0.4395 as it would not foreseeably impact bidding behavior to the extent that the growth rate or number of bidders would. This is because k serves to vertically shift the expected exponential model of participant bidding behavior, and does not impact the underlying value captured when kept constant.

All participants were primed with knowledge of the inner workings of the Vickrey Auction protocol. To reinforce the protocol, an initial demonstration with hypothetical competitors was illustrated on paper, after which participants were asked to walk on the inclined treadmill and simulate a winning and losing round as part of the Vickery Auction. Participants were unaware of the presence of robobidders and were told that they were engaging with real competitors from different laboratory facilities as part of the blinded auction. Throughout, participants were discouraged from dishonest bidding and were instructed to bid values that fit with their internal value system while simultaneously accounting for the task at hand.

Participants walked uphill on a 10-degree incline at a speed of 1.25 m/s for a randomized duration between 50-70 minutes. Every two minutes within this larger interval, subjects reported their 'price-to-walk' and engaged in the Seller's Vickrey Auction Protocol, as described in **Section I-B2** above. Participants were queried about their price-to-walk value at the start of iterative 2-minute intervals and had 1 minute to respond with their bid. The win or lose status as well as the winning payout was announced 20 seconds before the end of the 2-minute interval. If the participant won the auction round, they would start walking or continue to walk for the next two minutes. If the participant lost, they would sit out for the next 2 minutes.

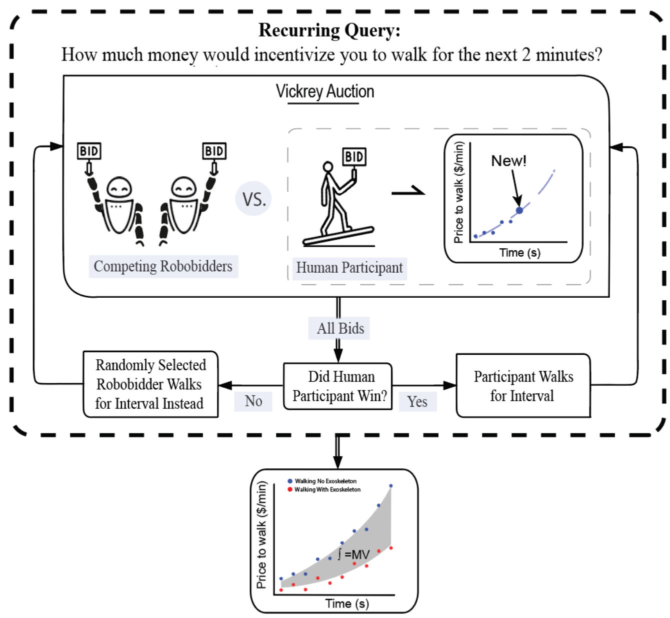

Fig. 3. The Vickrey Auction Paradigm: A human participant competes against automated bidders (robobidders) for a chance to complete a fatiguing uphill walking task and receive monetary compensation, by reporting an honest dollar value (termed price-to-walk) corresponding to their willingness to complete this fatiguing task. Participants are queried for this dollar value every 2-minutes and are required to walk if they win the round. If the human participant loses the round, they sit out for the next 2 minute interval while the winning robobidders is selected to "walk" and traverse an exponential fatigue model instead. Ultimately, the price-to-walk data points form a curve reflecting the task cost of the uphill walking activity, from which the marginal value (MV) metric can be extrapolated.

The commercially available Dephy ankle-foot exoskeleton, depicted in **Figure 4**, was used as the assistive platform of choice for the EPO condition due to its track record of introducing positive mechanical power and reducing metabolic cost of walking despite the added weight of the device itself, as well as being fully untethered and autonomous. The exoboot assists the wearer in plantarflexion by applying a specific torque profile initiated by the timing-based proprietary Dephy controller, which infers an individual's gait phase based on notable locomotion events and delivers the torque accordingly. This assistance works in parallel with the triceps surae muscles during the terminal stance phase of gait by leveraging an onboard brushless electric motor and a belt transmission to apply powered assistance in the saggital plane. Inversion and eversion motion of the foot is unpowered and unconstrained to allow for natural motion. Similarly, during swing phase the transmission introduces slack into the system to not impede the motion of the foot. The exoboot was selected due to it's usability, reliability and performance record as a state-of-the-art wearable device as it is able to inject an average of 13.4 ± 2.9 J during the full gait cycle.

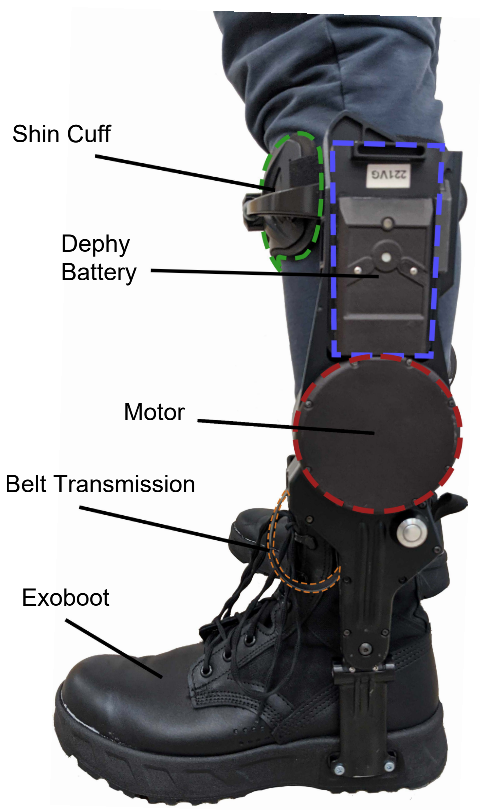

Fig. 4. The Dephy ankle-foot exoboot was used as the assistive platform of choice due to its reliable performance reflecting the cutting edge technologies in the lower limb exoskeleton space. The powered device augments plantarflexive movement using a belt transmission system that delivers torque according to a timing based proprietary controller.

In total, each individual completed eight trials, with an additional trial completed by participants that were randomly selected for a repeatability condition. These subjects were unaware of their participation in the repeatability trial and were duplicating the 3 robobidder, $1b$ condition. Participants were advised to walk no more than twice in one day, with a minimum duration of 1 hour between trials to prevent lingering fatigue effects from impacting their behavior in the subsequent trial. If participants self-reported extreme fatigue or were noticeably exhibiting signs of extreme fatigue, they were told to reschedule to another day. Participants that did not bid honestly were ineligible to continue participation. Despite the range of motivations for participation, the main criteria researchers used for continued participant eligibility was a willingness to walk a minimum of 10 minutes.

2) Determining Sample Size: Sample size to determine number of participants was estimated through a paired power analysis for Two One-Sided Tests (TOST) of Equivalence. The two-one-sided tests rely on equivalence bounds between which the difference between two groups is considered insignificant. Such tests are commonly used "for demonstrating the bioequivalence between two drug formulations in biopharmaceutical studies" and is relevant to our study as there is a possibility that MV is insensitive to changes in robobidder parameters. The equivalence boundary was determined based on intrasubject variability data obtained by Medrano et. al. who piloted the Vickrey Auction protocol. To account for intrasubject differences in average price-to-walk valuations, a least squares model was fit through compiled data accounting for the multiple averages of each subject. The resultant standard deviation of 3.38% accounted for inter-subject repeatability noise, and would form the basis of what minimum threshold is acceptable for recognizing a statistically significant difference in MV that would be indicative of robobidder influence. To capture overall variance of the of the repeatability trials in terms of MV, the difference between Gaussian's of WNE and EPO was found. Subtracting these distributions and accounting for a 95% confidence interval yielded an equivalence boundary of $\pm$ 9.38% MV, representing a likely range between which an insignificant difference in resultant MV between two and three robobidders or 'b' and '0.5b' is captured, and the groups are thus considered equivalent. This boundary can be generalized at the population level as intrasubject variability has been collapsed and accounted for within the single variance measure. This $\pm$ 9.38% MV boundary was subsequently taken as the standard deviation of differences within the power analysis computation to estimate an accurate sample size for the sensitivity analysis at hand.

A power analysis was conducted using R and the 'TOST' package. A significance level of 0.05 was selected while power was set at 80% and standard deviation of differences was chosen to be 4.78% MV as it represents overall variance of the dataset provided by Medrano et al. The resulting sample size was 3 pairs for a power of 80%, and 4 pairs for 95% power. Using a relatively conservative estimation, drawing upon clinical sample sizes within literature, and accounting for the minimum sample size recommended by the power analysis, a total of 14 subjects was determined to be appropriate to

TABLE I
PARTICIPANT CHARACTERISTICS

Participant	Gender	Weight (kg)	Age	Height (cm)
1	M	74.8	66	185.42
2	M	68	70	172.72
3	M	83.9	38	193.04
4	M	86.18	65	175.26
5	M	60.78	60	172.72
6	F	74.84	23	177.8
7	F	70.31	56	165.1
8	F	104.32	37	170.18
9	M	112.95	30	170.18
10	M	86.18	39	177.8
11	F	77.11	23	175.26
12	F	52.16	26	162.56
13	F	54.43	43	162.56
14	F	65.77	20	182.88

capture any statistically significant differences in MV due to robobidder behavior.

3) Participant Pool: Fourteen subjects, 7 female and 7 male (76.55 $\pm$ 16.69 kg, 68.71 $\pm$ 3.28in, 42.57 $\pm$ 17.03 yrs) between the ages of 18 and 80 were randomly selected and enrolled in the study. Subjects were pre-screened for any lower limb injuries, asymmetries, and chronic conditions which could preclude them from at least an hour of continuous walking, such as acute injury of the anterior cruciate ligament (ACL), stroke, asthma, spinal cord injury, vision impairment, and pregnancy at the time of enrollment. Subject selection was also age stratified to examine differences in bidding strategies dependent on this variable factor, which may have a significant impact on the value derived from exoskeleton use as the worth of compensation changes with age and socioeconomic factors simultaneously shift.

4) Data and Statistical Analysis: All data and statistical analyses was performed in MATLAB (MathWorks, Natick MA). Marginal value was determined for each participant by taking the area between the the two price-to-walk curves of WNE and EPO for the same robobidder condition. The area under each individual curve was referred to as the cumulative bid (CB) and indicates the value of the specific task in dollars. The following equations were used to compute MV:

$$MV = \overbrace{\int_{t_1}^{t_2} k_1 \cdot e^{(b_1 \cdot t)}\, dt}^{\text{exo condition 1}} - \overbrace{\int_{t_1}^{t_2} k_2 \cdot e^{(b_2 \cdot t)}\, dt}^{\text{exo condition 2}} \qquad (2)$$

The variables k_1 and k_2 refer to the initial bid, while b_1 and b_2 refer to the growth rates for each exoskeleton condition.

Prior to obtaining MV, raw bids for WNE and EPO were pre-processed. All bids before the first winning round were removed so as to not capture exploratory bidding behavior. A win rate was then computed and used to scale the total time of the trial to time walked in order to account for differences in walking period between multiple trials. The time walked also helps to account for the value of the completed task rather than accounting for the projected value of the walking task. Subsequently, a least squares fit through the expected exponentially trending data was taken to estimate the k and b values for each WNE and EPO curve, for each subject, and for each parameter combination. These k and b parameter values were then used to recreate the exponential model over a standardized 30 minute interval, which was determined to be the average total duration of walking that all subjects engaged in according to Medrano et al. Area below both curves was computed using the trapezoidal integration method and the difference in WNE and EPO areas yielded the value added in dollars by exoskeleton use, or MV. MV was converted to percent units by normalizing MV by the cumulative bid of the WNE curve. The area difference between the linear forms of the exponential models of WNE and EPO curves was also computed and denoted as 'log MV'. Goodness-of-fit between the linear form of the exponential model and the linear form of the raw data was evaluated by the coefficient of determination (R^2) metric.

An Anderson-Darling test, with α = 0.05, was performed on the MV's of the individual robobidder conditions to determine the goodness-of-fit of the data to a normal distribution. While the MV data collected for the 2r1b, 2r0.5b and 3r1b conditions failed to reject the null hypothesis that each data set was from a normal distribution, the 3r0.5b condition MV dataset rejected the null hypothesis, indicating that the dataset may not be normally distributed. The MV data for all conditions was then compiled and visualized as a histogram, determining that the dataset was closely approximated by a normal distribution. The presence of an outlier within the 3r0.5b condition was identified and may have led to the initial rejection of the null hypothesis.

To discern anticipated dependencies within the data, several linear mixed effects models (LMEM) were implemented to determine average effects and relationships on the population level between the various robobidder conditions and their impact on participant MV. To converge on an appropriate model, the Akaike information criterion (AIC) and the Bayesian information criterion (BIC), both of which comment on model fidelity and complexity, were iteratively minimized.

III. RESULTS

Repeat-ability of bidding behavior on different days reflected through MV was evaluated across 7 subjects when competing against 3 robobidders with a 1b growth rate. On average, participant repeater MV varied by a standard deviation of 30.37% or \$29.6184, which was used to determine the Minimum Detectable Change (MDC) or the minimum change in marginal value that is most likely not due to chance. MDC was found to be \$82.098 or 84.17% and was computed by taking the 95% confidence interval for the difference between the WNE and EPO gaussians. The MDC serves as a threshold below which it can be said that changes to the robobidder model have no significant impact on subject value for the exoskeleton. All repeater subjects appeared to have MV's below this threshold, indicating that their value was insensitive to changes in robobidder models. A 2-tailed paired t-test with a statistical significance threshold of 0.05 also revealed that participant MV was not significantly different between days, as observed in **Figure 5**.

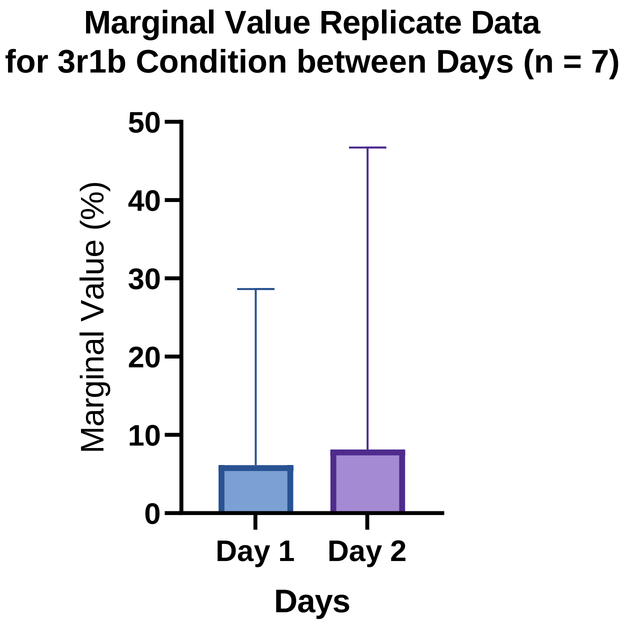

Fig. 5. Comparing MV between days for 7 randomly selected participants who competed against the 3r1b robobidder model.

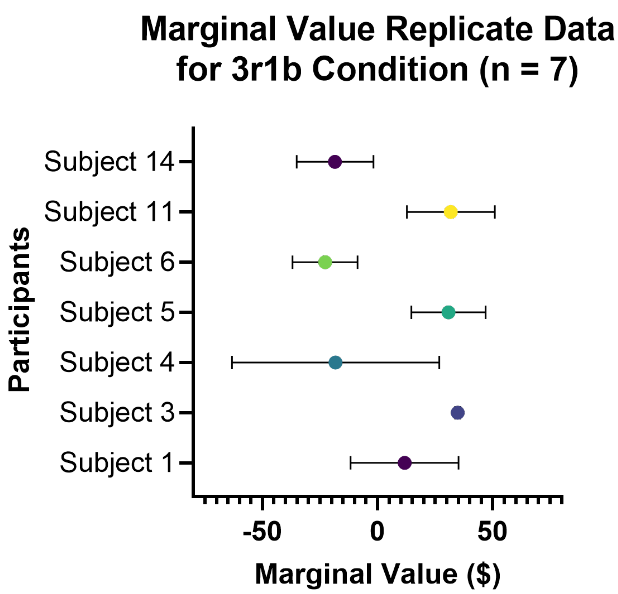

Fig. 6. Repeater data for 7 randomly selected participants who competed against the 3r1b robobidder model. On average, standard deviation was 30.3657%, indicating that perception of exoskeleton assistance or value was highly subjective and variable across subjects.

Participant MV was used as the measure of choice when performing a Linear Mixed Effects Model (LMEM) as opposed to log MV and cumulative bids, which were compiled for all subjects and are displayed in **Table II**.

TABLE II
MARGINAL VALUES.

Subject	MV (%)				
	2r0.5b	2r1b	3r0.5b	3r1b	Repeater
1	-67.27	2.72	-0.48	-4.78	28.38
2	-5.89	27.11	34.41	35.59	
3	1.95	97.62	44.63	13.62	34.11
4	13.22	-36.06	19.55	19.40	-50.06
5	25.35	24.41	18.10	-34.28	-12.79
6	-10.71	21.79	-3.41	-32.73	42.24
7	0.21	3.53	-9.48	28.69	
8	0.42	10.88	-102.40	29.91	
9	65.24	17.57	18.71	-63.81	
10	36.76	84.98	-282.52	-79.88	
11	37.76	-11.96	12.30	18.38	45.35
12	-103.21	-124.09	7.10	8.19	
13	-0.28	-16.46	37.22	25.24	
14	-13.48	9.05	25.54	-6.61	-30.25

A LMEM was implemented such that MV was designated as the response variable, while number of robobidders and robobidder growth rate was designed as fixed effects with potential interaction effects. Participants were designed as the random effect in order to account for their inherent randomness. The goal of this construction was to see whether there was a significant difference in participant MV due to varying robobidder parameters.

According to LMEM results, it was determined that there was a non-significant difference in MV when competing against 2 versus 3 robobidders, and between a $1b$ and $0.5b$ growth rate across both compiled robobidder number conditions within a detectable difference threshold of 84.17% or \$82.098. Further, there was no significant dependency between number of robobidders and its growth rate across all conditions. Intra and inter participant-level variation was found to be significant, indicating there is significant variation both within and between subjects bidding behavior. This corresponds with results generated by our repeater tests and results by Medrano et al, indicating that some participants may perceive a greater degree of locomotive assistance from exoskeleton use.

The 2 and 3 robobidder conditions were then isolated to ascertain within-condition robobidder growth rate effects on the response variable, MV. Participants were still designated as a random effect. Within each of the individual robobidder number conditions, there was no significant difference in MV between the 1b growth rate and the $0.5b$ growth rate. Once again, there was significant intra- and inter subject variance. A paired t-test and one-way, repeated measures ANOVA corroborated the results of the linear mixed effects model.

An LMEM model was also implemented to discern group level effects in terms of age and sex. Participants were split into 4 age groups ranging from Group 1: 18-30 yrs, Group 2: 31-42 yrs, Group 3: 43-60 yrs, and Group 4: 61-80 yrs. Each group did not have the same number of participants; there were five subjects for group 1, four subjects for group 2, two subjects for group 3, and three subjects for group 4. Age

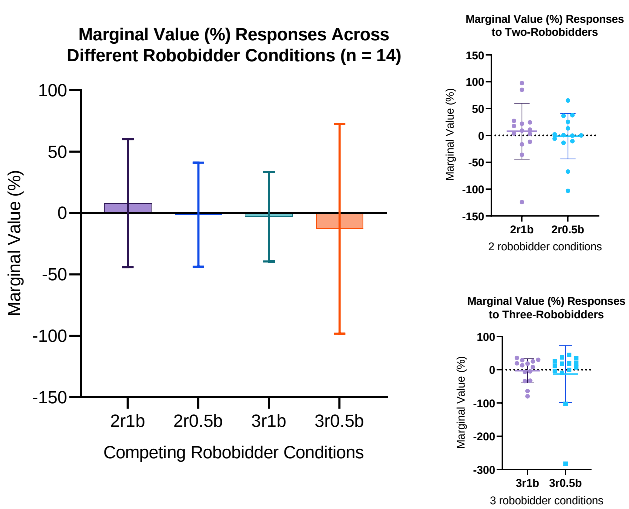

Fig. 7. Results of the sensitivity analysis are compiled for each robobidder model condition across all 14 subjects. It is evident through the overlapping standard deviation bars that there is no significant difference in MV across all the robobidder models with a significance threshold of $p > 0.05$.

was denoted as a fixed variable. When comparing all other age groups to group 4, none had significantly different MV responses to group 4. Group 2 differed the most in terms of MV from group 4, with a difference of -36.545% and a p-value of 0.080776, where $p > 0.05$ denotes non-significance.

When comparing MV across sex, there was no significant difference in MV between male and female exoskeleton users.

R-squared values were computed to determine the correspondence between the linear form of the data and linear model fits, acting as an estimator for the strength between the raw data and the regression model. Usually, a value close to '1' indicates that significant variability in the data is explained by the model. R^2 values as low as 0.5 were deemed acceptable to account for the unpredictability of human behavior.

R^2 values of the exponential least square fits through the raw bidding data ranged from 0.10 to 1, indicating that the bidding behavior of certain subjects was not aptly modelled by the assumed exponential fatigue response. Specifically, subject 8 was an outlier that had particularly low fidelity fits as evidenced by R^2 values as low as 0.18. Subject 2 also had poor to modest model fitting with R^2 values ranging from 0.41 to 0.88.

For price-to-walk curves from repeater subjects, R^2 values for the exoskeleton-powered-on condition ranged from 0.74 to 0.97, while the R^2 values for the walking-without-exoskeleton condition ranged from 0.72 to 0.98. These values indicate high fidelity trends and reinforced that subjects fatigued according to the exponential model repeatably. Out of all the subjects, subject 3 consistently had the lowest fidelity fatigue model fits compared to the other subjects with R^2 values of 0.72 to 0.79, and the standard deviation was also smaller than the average of the deviations experienced by other subjects. In contrast, subject 4 had a standard deviation far larger than the average of the deviations experienced by the other subjects, while model fidelity was quite high reflected in the high R^2 values of 0.94 to 0.96. Both subjects were not considered outliers as they followed instructions of bidding honestly and walked for at least 10 minutes, which served as the only criteria for retaining data.

TABLE III
REPEATER MV DATA FOR 3R1B ROBOBIDDER CONDITION

Subject	MV(%)	Std(%)	WNE R^2	EPO R^2
1	-4.78 28.38	23.45	0.93 0.96	0.97
3	35.59 34.11	1.05	0.79 0.72	0.88
4	13.63 -50.06	45.03	0.94 0.96	0.96
5	19.40 42.24	16.15	0.92 0.94	0.89
6	-32.73 -12.79	14.10	0.96 0.88	0.74
11	18.38 45.35	19.07	0.98 0.98	0.88
14	-6.61 -30.25	16.72	0.98 0.97	0.89

Participant fatigue rates for the $1b$ and $0.5b$ WNE and EPO conditions were subtracted from the pre-set robobidder fatigue rates to determine whether participant fatigue corresponded with the robobidder's artificial fatigue. The difference between participant fatigue rate and robobidder fatigue rate across all conditions was not significantly different from 0 as determined by a two-tailed paired t-test ($p > 0.05$), indicating that participants tracked and mimicked the robobidder rate of fatigue. The average difference in participant and robo-bidder fatigue rate for the WNE condition appeared to be greater than than experienced during the EPO condition across all $1b$ cases. However, a two-tailed paired t-test ($p > 0.05$) concluded that the difference in fatigue rates for the WNE case compared to the EPO case was not significantly different. This may indicate that participants tracked with robo-bidder fatigue regardless of using the exoskeleton or not. Participant initial bidding value (k) was also compared to that of the robobidder's pre-set initial bidding value for all 4 conditions and for both WNE and EPO, although this value was kept as a constant throughout all cases. The difference between participant k and robobidder k was found to be not significantly different from 0 across all conditions according to another two-tailed paired t-test ($p > 0.05$). This indicates that participants adapted their initial bidding value to match that of the robo-bidder's initial bidding value. Histograms of the difference between nominal robobidder parameters and participant parameters can be visualized in **Figure 8**.

Additional mixed models were implemented to determine whether win rates differed across robobidder parameter combinations. Win rate was designated as a response variable while number of robobidders and robobidder growth rate was designed as fixed effects with potential interaction effects. Subjects were accounted for as random effects. Within both WNE and EPO conditions, the win rate did not significantly

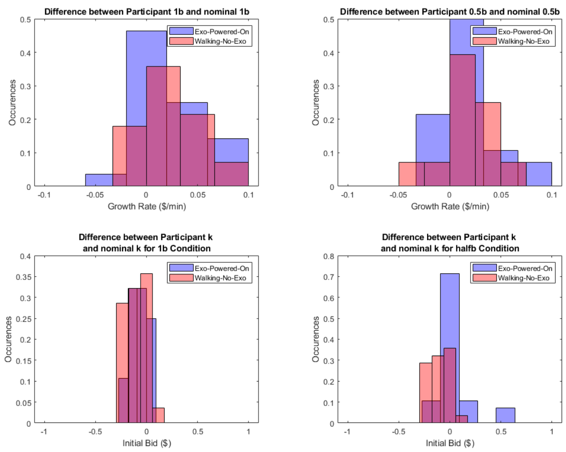

Fig. 8. A histogram depicting the distribution of differences between computed participant parameters k and b versus pre-set robobidder parameters for both WNE and EPO conditions.

differ between robobidder parameter combinations. Further, the LMEM model determined that win rates between the WNE and EPO conditions across all robobidder parameter combinations, did not significantly differ between each other. This may indicate that despite added exoskeleton assistance participants did not deem its assistance valuable enough to walk more often with it, as encoded in the win rate.

IV. Discussion

A sensitivity analysis was conducted to evaluate the effects of varying robobidder models on perceived exoskeleton value while engaged in the Vickrey Auction paradigm. Robobidder models were varied in terms of growth rate as well as the number, which were anticipated to impact the convergence behavior and thus user-perceived value of the exoskeleton. Four robobidder model combinations were implemented - 2r0.5b, 2r1b, 3r0.5b and 3r1b - to compete against 14 participants during uphill walking with the exoskeleton powered-on and walking without the exoskeleton. Fourteen subjects were recruited following a power analysis and were asked to participate in the Seller's Vickrey auction paradigm. Information about the robobidders was withheld from subjects to prevent attempts to learn the behavior of the robobidders, and they were encouraged to bid according to their internal value system i.e. as honestly as possible. Marginal value, or the dollar value added through exoskeleton use, was then computed for each subject and for each robobidder condition to observe if there were differences in MV within a certain detectable threshold when competing against a certain robobidder model.

A detectable threshold was established based off of repeater data and variance taken across seven randomly selected subjects. These subjects completed an additional session during which they competed against the 3r1b robobidder model and their marginal value was computed. Both within subject and across-subject standard deviation of the repeatability data was quite high, suggesting that individual MV's are highly variable. This variability may be due to the differences in perceived exoskeleton value as well as differences in how the exoskeleton assistance was leveraged. This variability is also greater than that observed across participants by Medrano et al. A distinction between the previous work and this study lies in the fact that this study collected repeatability data for the 3r1b model as compared to the original repeatability data that was gathered when using the 2r1b model. Further, Medrano and team conducted up to a maximum of three repeatability trials for a smaller pool of 4 subjects, while this study deemed 2 trials to be sufficient for a larger pool of 7 subjects. Another protocol-based distinction that could account for this greater degree of MV variability is use of handrails in this study during uphill walking, which could enable certain participants to leverage this additional assistance to unknown degrees, injecting some unpredictability into the repeater data.

Linear mixed effects models and t-tests conducted on the gathered data indicate that robobidder numbers and growth rate don't appear to be changing marginal value of exoskeleton assistance above a detectable threshold of $\pm$ 84.17%. However, it was noted that out of the 4 combinations of growth rates and number of robobidders - 2r0.5b, 2r1b, 3r0.5b and 3r1b - only the 2r1b case experienced a positive averaged MV, indicating that on average participants competing with the 2r1b condition perceived the exoskeleton assistance to be valuable. It is possible that participants may have perceived the 2r1b case to be financially rewarding compared to the other conditions, potentially leading to artificial suppression of price-to-walk responses in order to win and earn more money. Explicitly, the 2r0.5b and 3r0.5b conditions both had lower growth rates meaning that the robobidders would not fatigue as fast and would not only be harder to under-bid if participants were looking to win, but they would also be suppressing the winning payout, or second lowest bid, that the participants could earn. Although the 3r1b condition had the same growth rate as the 2r1b condition, the time it took to randomly perturb and force all of the robobidder models to traverse the fatigue curve was longer (it took 3 consecutive losses of the Vickrey Auction instead of 2 if a participant was looking to increase the incentive that would be most in-line with their task valuation). This greater interval to increase the incentive to walk uphill may have caused the exoskeleton to be viewed as less valuable as it's use was not improving chances of earning a financial payout. Thus, when participants competed against the 2r1b model, the exoskeleton may have been viewed as a means to an end since there was a greater monetary incentive on the line, and it's assistance was deemed more valuable to achieve that goal, resulting in a pronounced positive MV valuation.

While it was expected that 3r1b condition would be the most analogous to the 2r1b condition due to the similarity in growth rates and would have the second largest MV, the second largest MV was actually from competing against the 2r0.5b robobidder models. This may indicate that the number of robobidders may have a greater impact on MV than the growth rate in perceiving valuable exoskeleton assistance, as the time it takes to forcibly increase the robobidders price-to-walk as they traverse the model is reduced in the 2r0.5b condition.

Despite the different average MV for each of the conditions, given the power of the study and sample size, these differences are not statistically significant and still preserve MV as being independent from robobidder model effects. This result has implications for perception-dependent metrics and indicates that MV has the potential to be an objective measure to evaluate the efficacy of wearable systems. Participant bidding independence from robobidder behavior enables the standardization of models across various labs to allow for cross comparisons of exoskeleton design and controls choices.

Additional linear mixed effects models revealed that additional confounds to perceived exoskeleton value, such as age and sex, also did not impact MV valuations. It was hypothesized that the various age groups would have significantly different valuations of the exoskeleton due to differing degrees of dependencies on the device for lower limb support, as well as varying monetary value that would likely mature with age and experience. While the independence of MV to these confounds is promising, future studies may investigate whether socioeconomic factors as well as baseline fitness impact MV. Future studies may also look into how MV correlates with established quantitative metrics to evaluate exoskeleton efficacy, as well as qualitative measures of device perception like the Likert scale.

A limitation of this study was the use of treadmill handrails during the uphill walking task. While subjects were made aware of and encouraged to utilize the handrails, their use was not enforced. External support provided by the handrails could potentially impact the marginal value collected from participants, but it's effect was deemed negligible as participants still experienced fatigue reflected by their increasing price-to-walk. The bidding trends conformed to the exponential fatigue model as shown by low SER values indicative of good model fits to the collected data, with the exception of one participant - subject 8. The use of the handle bars, while modifying the original protocol presented by Leo et al, did not alter the impact of the experiment significantly, and thus participant sensitivity to robobidder impact was still verifiable.

V. Conclusion

In this study, we validated a novel metric and associated methodology to quantify the economic value provided by a bilateral ankle exoskeleton during a fatiguing uphill walking task. Specifically, we conducted a sensitivity analysis to determine whether the economic value metric that participants reported was sensitive to the automated robobidders involved in the methodology. Our results indicate that participant bidding behavior and perceived exoskeleton value are unaffected by the various competing robobidder models and are also not impacted by confounding variables such as age and sex which may impact perceived economic value. These results serve to validate the reliability of our proposed metric - marginal value - and highlights the fact that such nuanced, perception-dependent metrics can be used as an objective measure to evaluate the efficacy of wearable systems. The lack of dependency of economic value on the robobidder models allows for these automated bidders to be standardized so that research labs can uniformly compare exoskeleton performance. Our results also show that participants are highly variable in terms of receiving valuable exoskeleton assistance, emphasizing the need for further studies investigating why some participants are able to derive greater value through exoskeleton use. Ultimately, the validated economic value metric can be used to develop more valuable exoskeletons by informing control system and hardware design according to a standard that surpasses the perceptual threshold of exoskeleton assistance.

REFERENCES

[1] T. G. . R. E. Medrano, R.L., "Can humans perceive the metabolic benefit provided by augmentative exoskeletons?" *Journal of NeuroEngineering and Rehabilitation*, vol. 19, 2022.

[2] J. L. Pons, "Rehabilitation exoskeletal robotics," *IEEE Engineering in Medicine and Biology Magazine*, vol. 29, no. 3, pp. 57–63, 2010.

[3] G. Chen, C. K. Chan, Z. Guo, and H. Yu, "A review of lower extremity assistive robotic exoskeletons in rehabilitation therapy," *Critical Reviewstrade; in Biomedical Engineering*, vol. 41, no. 4-5, pp. 343–363, 2013.

[4] B. O. K. I. e. a. Sawicki, G.S., "The exoskeleton expansion: improving walking and running economy." *Journal of NeuroEngineering Rehabilitation*, 2020.

[5] S. M. Siedl and M. Mara, "Exoskeleton acceptance and its relationship to self-efficacy enhancement, perceived usefulness, and physical relief: A field study among logistics workers," *Wearable Technologies*, vol. 2, p. e10, 2021.

[6] S. Upasani, R. Franco, K. Niewolny, and D. Srinivasan, "The potential for exoskeletons to improve health and safety in agriculture—perspectives from service providers," *IISE Transactions on Occupational Ergonomics and Human Factors*, vol. 7, no. 3-4, pp. 222–229, 2019. [Online]. Available: https://doi.org/10.1080/24725838.2019.1575930

[7] S. Kim, A. Moore, D. Srinivasan, A. Akanmu, A. Barr, C. Harris-Adamson, D. M. Rempel, and M. A. Nussbaum, "Potential of exoskeleton technologies to enhance safety, health, and performance in construction: Industry perspectives and future research directions," *IISE Transactions on Occupational Ergonomics and Human Factors*, vol. 7, no. 3-4, pp. 185–191, 2019. [Online]. Available: https://doi.org/10.1080/24725838.2018.1561557

[8] D. Hill, C. S. Holloway, D. Z. Morgado Ramirez, P. Smitham, and Y. Pappas, "What are user perspectives of exoskeleton technology? a literature review," *International Journal of Technology Assessment in Health Care*, vol. 33, no. 2, p. 160–167, 2017.

[9] P. C. B. J. M. B. M. J. Wolff, J., "A survey of stakeholder perspectives on exoskeleton technology," *Journal of NeuroEngineering Rehabilitation*, 2014.

[10] D. F. N. Gordon, G. Henderson, and S. Vijayakumar, "Effectively quantifying the performance of lower-limb exoskeletons over a range of walking conditions," *Frontiers in Robotics and AI*, vol. 5, 2018. [Online]. Available: https://www.frontiersin.org/articles/10.3389/frobt.2018.00061

[11] G. M. P. A. Pesenti M, Antonietti A, "Towards a functional performance validation standard for industrial low-back exoskeletons: State of the art review." *Sensors (Basel)*, 2021.

[12] H. A. W. J. Zheng L, Lowe B, "Evaluation and test methods of industrial exoskeletons in vitro, in vivo, and in silico: A critical review." *Critical Reviews in Biomedical Engineering*, 2021.

[13] D. Lee, B. J. McLain, I. Kang, and A. J. Young, "Biomechanical comparison of assistance strategies using a bilateral robotic knee exoskeleton," *IEEE Transactions on Biomedical Engineering*, vol. 68, no. 9, pp. 2870–2879, 2021.

[14] P. W. Franks, G. M. Bryan, R. M. Martin, R. Reyes, A. C. Lakmazaheri, and S. H. Collins, "Comparing optimized exoskeleton assistance of the hip, knee, and ankle in single and multi-joint configurations," *Wearable Technologies*, vol. 2, p. e16, 2021.

[15] H. A. Richardson S, "Endurance fitness and blood lactate concentration during stepping exercise in untrained subjects." *Br J Sports Med.*, 1989, pMID: 2620238.

[16] A.-H. H. El-Sobkey SB, "Heart rate and perceptual responses to graded leg and arm ergometry in healthy college-aged saudis: Effects of gender and exercise mode." *J Nov Physiother Phys Rehabil*, 2014.

[17] P. E. S. P. W. A. Seip RL, Snead D, "Perceptual responses and blood lactate concentration: effect of training state." *Med Sci Sports Exerc.*, 1991, pMID:1997816.

[18] S. Marcora, "Perception of effort during exercise is independent of afferent feedback from skeletal muscles, heart, and lungs," *Journal of Applied Physiology*, vol. 106, no. 6, pp. 2060–2062, 2009, pMID: 18483166. [Online]. Available: https://doi.org/10.1152/japplphysiol.90378.2008

[19] S. H.-W. Chuah, P. A. Rauschnabel, N. Krey, B. Nguyen, T. Ramayah, and S. Lade, "Wearable technologies: The role of usefulness and visibility in smartwatch adoption," *Computers in Human Behavior*, vol. 65, pp. 276–284, 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0747563216305374

[20] N. Thatte, H. Duan, and H. Geyer, "A method for online optimization of lower limb assistive devices with high dimensional parameter spaces," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE Press, 2018, p. 1–6. [Online]. Available: https://doi.org/10.1109/ICRA.2018.8460953

[21] M. Tucker, E. Novoseller, C. Kann, Y. Sui, Y. Yue, J. W. Burdick, and A. D. Ames, "Preference-based learning for exoskeleton gait optimization," in *2020 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2020, pp. 2351–2357.

[22] M. Tucker, M. Cheng, E. Novoseller, R. Cheng, Y. Yue, J. W. Burdick, and A. D. Ames, "Human preference-based learning for high-dimensional optimization of exoskeleton walking gaits," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 3423–3430.

[23] K. A. Ingraham, C. D. Remy, and E. J. Rouse, "The role of user preference in the customized control of robotic exoskeletons," *Science Robotics*, vol. 7, no. 64, p. eabj3487, 2022. [Online]. Available: https://www.science.org/doi/abs/10.1126/scirobotics.abj3487

[24] J. L. Lusk, T. Feldkamp, and T. C. Schroeder, "Experimental auction procedure: impact on valuation of quality differentiated goods," *American Journal of Agricultural Economics*, vol. 86, no. 2, pp. 389–405, 2004.

[25] T. Sandholm, "Issues in computational vickrey auctions," *International Journal of Electronic Commerce*, vol. 4, pp. 107 – 129, 2000.

[26] C. M. Koo, "Homegrown values in repeated auctions: the vickrey auction versus the random n-th price auction," 1995.