

On the motivational power of pride: a field study with participants in a road-running event – Pre-Analysis Plan

Jessica Milligan and Severine Toussaert*

April 2019

1 Introduction

Pride is a powerful motivational force, which often propels individuals to take on personal challenges and push beyond their limits. One context in which the demand for personal challenges manifests itself very clearly is participation in road-running events such as marathons, which have become increasingly popular over the recent years. We conduct a two-part online study with participants in a ten-mile race in Washington DC, which typically attracts 15,000 – 20,000 (mostly amateur) runners each year. We survey volunteer participants before and after the event in order to study their motivation to perform at their best during the race as well as the pride they derive from their achievement.

Of particular interest to us is the motivation that emerges from perceived threats to one’s (self or social) image triggered by others’ assessments. Consider Ben who was told by his dad that he should forget about applying to Ivy League schools and concentrate instead on schools “at his level”. Or think about Carol whose colleagues think she does not have what it takes to be promoted. It is not difficult to think of examples where the desire to prove others wrong pushes people to perform beyond others’ (and possibly, their own) expectations. Examples in a sports context abound.¹ We investigate the power of this motivational force in the context of the ten-mile race.

In the pre-race survey, 431 participants were asked to tell us about their time goal for the race and were challenged/encouraged to achieve it. Participants saw one of 6 different versions of the pre-race survey, which differed in whether we (i) expressed doubts about their ability to achieve their time goal; (ii) rewarded them for achieving their time goal (with either a \$10 or a \$25 gift card). The post-race survey collected data about participants’ satisfaction regarding their race performance and their valuation of several aspects of the event such as the race medal. Both surveys have been completed by study participants at the time we are submitting this Pre-Analysis Plan. However, we have not received yet the performance data as well as other key elements of the dataset, which we detail below.

The Pre-Analysis Plan is organized as follows. In Section 2, we present an overview of our study, detail and justify some of our design choices, and describe the various sources of data that are (or will be) available to us. In Section 3, we present the various families of outcomes we will study and define all outcomes within each family. In Section 4, we detail our empirical strategy and formulate some hypotheses. Information about the recruitment material, power calculations, and a dictionary for the explanatory variables can be found in the appendix.

*We acknowledge financial support from the ESRC (Research Training Support Grant) as well as St John’s College, University of Oxford. Contact: jessica.milligan@economics.ox.ac.uk and severine.toussaert@economics.ox.ac.uk.

¹See for instance, “Why He Kayaked Across the Atlantic at 70 (for the Third Time),” The New York Times, March 22, 2018.

2 Experimental Design

2.1 Experiment Overview

2.1.1 Running event and population

We study the motivation and race performance of amateur runners who entered a large road-running event in the US. The event is the Credit Union Cherry Blossom Ten Mile Run (CUCB), a race which has been taking place in Washington DC every year since 1973.² This year, the race took place on Sunday April 7, 2019. CUCB is a very popular event and the field of entrants is very large for this type of distance. It is usually oversubscribed and race entries are allocated through a lottery mechanism. In 2018, there were 21,855 race applicants for the ten-mile distance, 17,697 who received a place to run, and 16,959 completed the race on the day.

There are more women than men in the field of participants. In 2018, 59% of participants in the ten-mile distance were women. This is consistent with historical figures: 2017 had 60% and 2016 had 62% female participants. The median age is around 35 (about three years below the US median age) and virtually all entrants are American. Finally, around 50% of participants each year are first-time finishers. Descriptive statistics of the population of applications, successful lottery winners, and race finishers has not yet been made available to us from the race organizers for the 2019 event. Below, we report summary statistics for the subsample of study participants.

For this study, we use a combination of several data sources described in more detail in the next sections. First, we use race registration and finisher’s performance data available for all ten-mile runners in the 2019 race (approximately 17,000). Second, we collect survey data before and after the race from a subsample of over 400 runners who volunteered to take part in our study. The recruitment procedures are detailed in the next section.

2.1.2 Recruitment procedures and sample size

We recruited 431 runners to take part in a pre- and post- race survey of less than 10 minutes each. The study was advertised as a “study about participation of runners in races” conducted by social science researchers at the University of Oxford in partnership with the race organizers.³ As an incentive, we advertised that participants would have an opportunity to win prizes via a lottery but that (i) they would only be eligible to win once they had completed both surveys, and that (ii) they would have only 7 days to complete the post-race survey upon receiving it via email. The prizes were as follows:

4 free, guaranteed entry to next year’s race (donated by the CUCB race organizers)

\$10, \$25 and \$150 e-gift card vouchers to buy Under Armour gear and apparel⁴

Our sample of 431 runners represents 2.4% of the entire population of race finishers. In an appendix to the paper, we will document potential differences between our study sample and the rest of the population using the registration data available to us for all participants (see Section 2.3 for more information about the available data). The choice of sample size was partly dictated by budget constraints. Besides the lottery prizes, participants in two of our treatments could receive a voucher payment for meeting their goal time for the race (further details in the next section). Our limited budget together with power considerations, the anticipation of at least a 10% attrition rate between the pre- and post-race surveys, and the expectation that not all runners would meet their target race time resulted in our decision to recruit as close to 430 runners into the study as we could.⁵

The choice of sample size — resulting in the comparison between groups of approximately 100 to 110 survey participants in the control and each treatment respectively — dictates a minimum effect size that we will be

²Runners can also participate in a shorter distance (a 5k Walk/Run). For the purpose of our study, we exclusively analyze data related to the ten-mile race participants.

³The wording refers to the email communication sent out to all runners by the race organizers before the event; see below for more details. The various communications and reminders can be found in Appendix A.

⁴We selected Under Armour as they were the official clothing brand sponsor of the CUCB in 2019.

⁵In our recruitment messages, we announced that participation would be capped at 400. We estimated at least a 10% attrition rate between pre- and post-race survey, and hence we over-recruited for the pre-race survey.

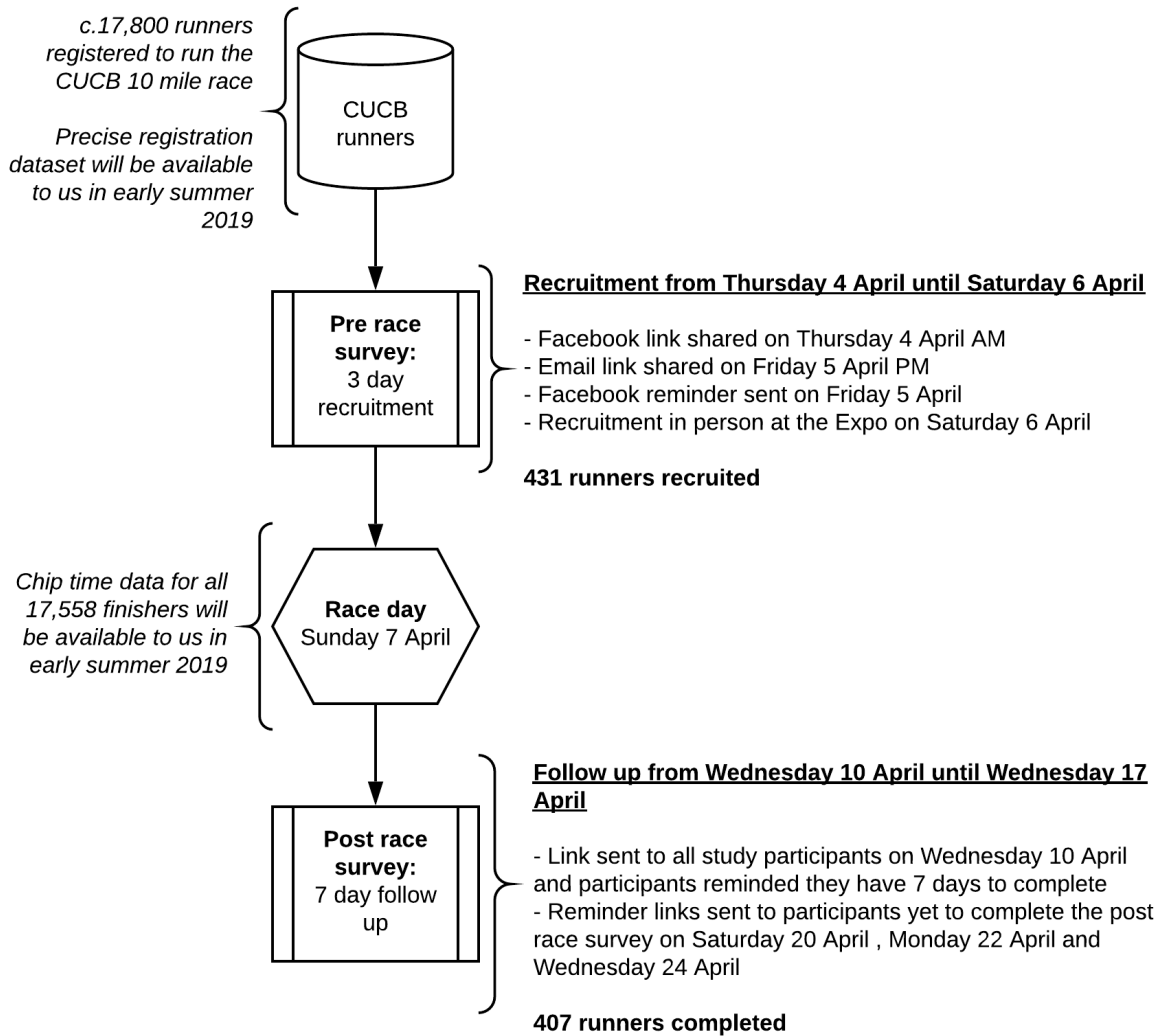


Figure 1: Data collection timeline

able to detect. Our main outcome variable is the % of runners who meet their goal time. Based on prior evidence, we estimate that 25% of runners in the control group will do so.⁶ With $\alpha = 0.05$, a control mean of 25%, 100 runners per condition, and 80% power, we will be able to detect a 19 percentage point difference (or larger) between treatment and control. To be able to detect smaller effect sizes such as a 10 percentage point difference, one would need over 300 runners per treatment group (thus, 1,200 overall). Power calculations can be found in Appendix B.

Participants were invited to sign up and complete the pre-race survey from Thursday 4 April 2019 (3 days before the race). Before starting, participants were taken to a consent form which explained the structure of the study, the incentives, the type of data collected and protection of this data. Those who completed the pre-race survey were sent a follow up post-race survey on Wednesday 10 April 2019 (3 days after the race took place), and were given 7 days to complete it. We recruited participants through 3 channels:

⁶Markle et al. (2018) measure the success rate of runners across 15 US marathons in meeting their goal time, finding that an average of 25.2% of runners are successful. Krawczyk and Wilamowski (2017) present evidence from 1,000,000 marathon runners suggesting that less than a third of runners meet their goal time.

1. A Facebook ad on the official CUCB race event page (16,096 individuals follow this page) in the morning of Thursday 4 April 2019, with a follow up reminder in the morning on Friday 5 April 2019.
2. An ad in the Pre-Race E-News from the race director which was sent to all participants in the evening on Friday 5 April 2019.
3. In-person recruitment at a stand at the Race Expo on Saturday 6 April 2019, from 9am until 4pm.

In total, 75% of participants were recruited via email or Facebook, with the remainder recruited at the Expo.⁷ To participate in the study, runners had to be at least 18 years of age and officially registered in the 2019 CUCB ten-mile race. Although we estimated based on historical data that the CUCB race would have about 60% female participants, we made an effort to recruit a 50:50 split between female and male survey respondents as we anticipate that gender will be an important moderator. Female runners were much faster to complete the first survey than men: at the end of the first recruitment day, we already had 214 female runners but fewer than 50 male participants who had completed the pre-race survey. Therefore, the survey was closed to female runners from Friday 5 April and recruitment at the Expo on Saturday 6 April was exclusively targeted at male runners. We had successfully recruited 217 men by the end of the Expo. Table 1 provides descriptive statistics about the survey respondents (pre-race survey responses). Just under half of the participants were running the CUCB race for the first time (close to the 2018 figure).

Table 1: Survey participants

Age Quartiles	Q1	Q2	Q3	Q4	1 st CUCB Race
Male	60 (28.6%)	45 (20.7%)	41 (18.9%)	71 (32.7%)	86 (39.6%)
Female	54 (25.2%)	62 (29.0%)	63 (29.4%)	35 (16.4%)	111 (51.9%)
Total	114 (26.5%)	107 (24.8%)	104 (24.1%)	106 (24.6%)	197 (45.7%)

Notes: Q1: 22–31; Q2: 31–48; Q3: 38–48; Q4: 48–77. These statistics are based on responses to the pre-race survey (N = 431).

2.2 Structure of the study

2.2.1 Pre-race survey

After consenting to the study, participants were asked basic socio-demographic questions (gender, age, education, income) as well as more specific questions about their prior participation in the ten-mile race (number of completed CUCB races, most recent time, personal best). The next set of questions focused on the 2019 race, asking about level of preparation for the race and time goal. More precisely, participants were asked what time (hh:mm) they were aiming for on race day, how much it would mean for them to achieve their goal, how *committed* they were to their goal, and how *con dent* they were that they would meet their goal.

Participants were then randomized into one of 4 different treatments, which differed in whether (i) we expressed doubts about participants’ ability to meet their time goal; (ii) we offered them monetary incentives for meeting their goal:

Control (C): *Good luck with the race! Don't forget to pace yourself. Let's see how things go on Sunday!*

Money (M): *Good luck with the race! Don't forget to pace yourself. If you meet your goal, we will send you a \$X Under Armour gift card by email once you completed Survey 2!*

Pride (P): *Good luck with the race! Don't forget to pace yourself. Runners are often too optimistic about the goals they set.*

We have to say that we would be surprised if you met the goal you chose. Although we're certain you can make it in your dreams. ☺

Let's see how things go on Sunday!

⁷We anticipate there may be demographic differences between those recruited via the different channels.

Pride and Money (PM): *Good luck with the race! Don't forget to pace yourself. Runners are often too optimistic about the goals they set.*

We have to say that we would be surprised if you met the goal you chose. Although we're certain you can make it in your dreams. ☺

In fact, we are willing to bet money on your race performance: if we are wrong and you meet your goal, we will send you a \$X Under Armour gift card by email once you completed Survey 2!

Let's see how things go on Sunday!

In the money treatments (M and PM), we cross randomized the level of incentives to be either $X = \$10$ or $X = \$25$. This additional variation will be used to estimate the responsiveness of race performance to incentives and compute a monetary equivalent to the pride effect (using the M treatment). With the PM treatment, we will test whether the pride manipulation has an effect even when the message is cheap talk or if putting money on the line is necessary for motivation to increase and affect performance. The level of incentives was selected based on budget and sample size considerations, with an effort to maximize the difference between the two amounts. The number of observations by treatment cell is presented in Table 2.

In the pride treatments (P and PM), our choice of words attempted to balance several considerations. First, although our message was deliberately provocative, we tried to minimize the negative wording and instead adopt a teasing tone. We acknowledge that not all participants might have perceived the message as merely teasing and we will carefully document any feedback we received in the post-race survey (see Section 3.6). Second, we wanted our evaluation to be perceived as a noisy signal of future race performance i.e., as somewhat informative (otherwise our message might have been too easily discarded), but not so informative as to make participants systematically revise their expectations downwards. We therefore chose to remain ambiguous about the basis of our assessment, which was not person-specific, but instead pertained to the population of runners as a whole.

The last part of the survey aimed to test whether runners in the pride treatments would form more pessimistic beliefs about their finish time than runners in the control and money treatments.⁸ To do so, we elicited respondents' probabilistic beliefs about their finish time on race day, using their time goal as the reference point. More precisely, we asked respondents to estimate the chances that they would be N minutes faster/slower than their time goal of hh:mm, with time brackets $\sqrt{\text{Less than one minute, 1:00-4:59 minutes, 5:00-9:59 minutes, At least 10:00 minutes}}$. Participants could enter any number between 0 and 100 for each time bracket with the constraint that their answers sum to 100.

Table 2: Sample size

Treatment	Incentive	Males	Females	Total	(incl. post-race)
Control (C)	(\$0)	54	54	108	(102)
Money (M)	\$10	26	28	54	(53)
	\$25	26	29	55	(51)
Pride (P)	(\$0)	58	50	108	(102)
Pride & Money (PM)	\$10	26	27	53	(49)
	\$25	27	26	53	(50)
Total		217	214	431	(407)

Notes: Numbers correspond to responses to the pre-race survey. The total number of participants who also completed the post-race survey in each treatment is presented in brackets in the last column.

⁸Right before the belief elicitation question, we asked two questions about time tracking i.e., whether the participant planned to keep track of his/her time during the race using a watch/phone and whether s/he usually does so. Those questions were mainly used as buffer questions to increase distance between the treatment manipulation and the belief question in an attempt to minimize demand effects.

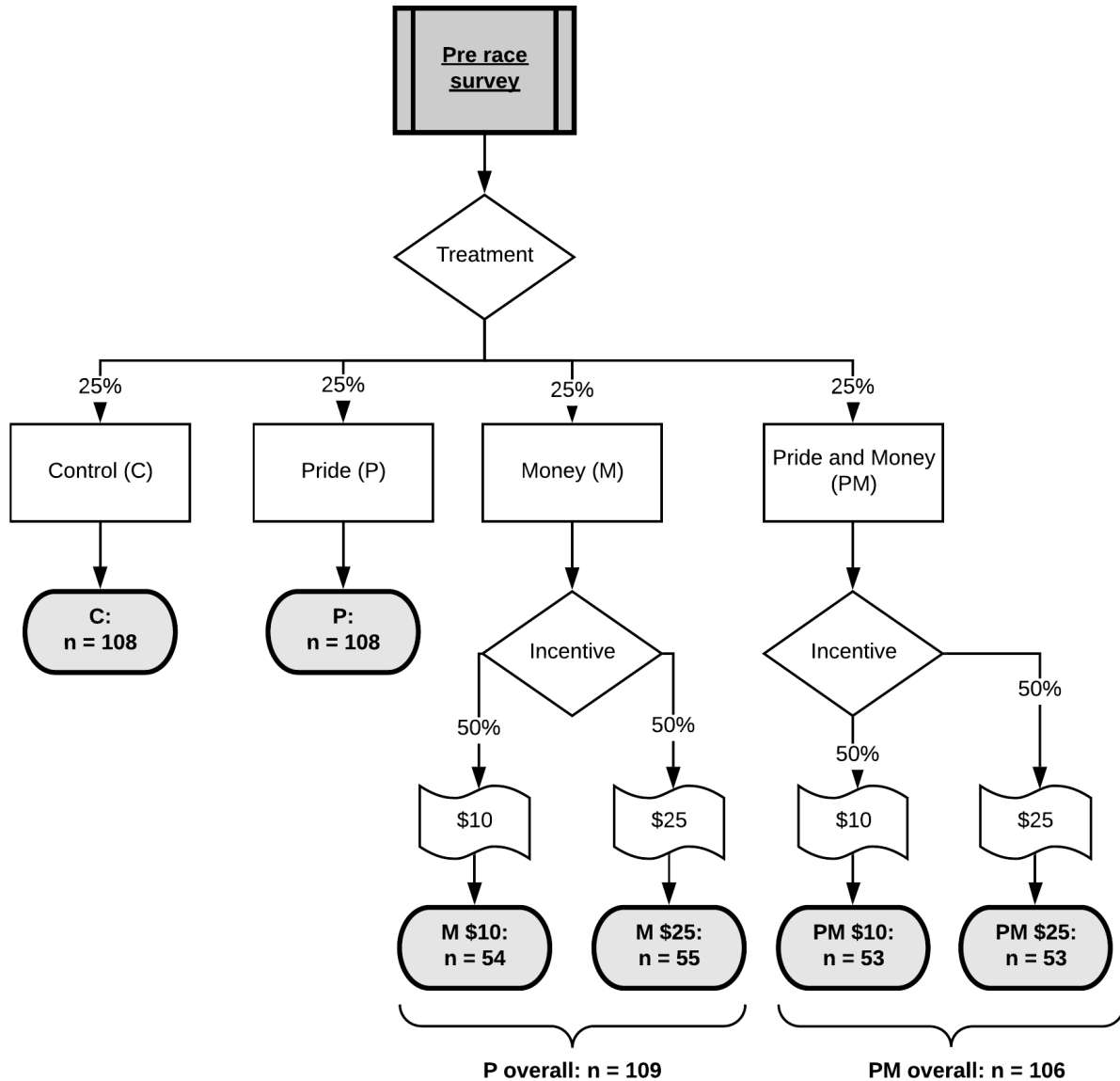


Figure 2: Pre-race Survey

2.2.2 Post-race survey

Study participants were sent the post-race survey via email 3 days after the race. The first block of questions tested respondents' self-reported knowledge/recall of (i) their time goal, (ii) their performance, and (iii) the treatment they were assigned. First, they were asked to enter the time goal they set in the pre-race survey as well as their actual finish time to the nearest minute (hh:mm). These questions allow us to evaluate respondents' accuracy relative to their actual goal and finish time. To help us evaluate the quality of their responses, we also recorded response times in both questions. Second, we asked participants whether they recalled "any information or comment they received about their time goal" from the pre-race survey in order to ascertain how well they remembered the treatment message, if at all.

The second block of questions was exclusively for participants who were in the money treatments (M or PM). If participants met their goal, we offered them the opportunity to make a counter signal: they could turn down the clothing voucher and instead choose that we make a donation to charity on their behalf.⁹ Offering this option allows us to investigate whether successful participants in the PM condition may choose to turn down the incentives as a way to deny the effect of our teasing comment. Note that by making the donation on their behalf (and not in their own name), we eliminate the possibility of a social image benefit vis-à-vis the charity.

The third block was common to all pre-race treatments. First, we included direct questions about participants' feeling of pride and satisfaction regarding their race performance. To obtain a revealed preference measure of pride about race performance, we asked about (i) intentions to buy the race medal (if not purchased at registration) and (ii) decisions to purchase the race photographs. To understand participants' attachment to the medal, we asked about the number of medals they already owned and tracked their clicks on links to websites that offer ways of displaying vs. disposing of race medals (medal hangers vs. recycling / donation options).¹⁰ Finally, we administered the Tracy and Robins (2007) psychological pride scale, which captures two facets of pride: *authentic* pride (positive or likeable pride feelings and emotions) and *hubristic* pride (negative or dislikable pride feelings and emotions). Our goal is to test for differences across treatments in the expression of pride as captured by these various measures.

The final block offered participants the chance to sign up to a commitment contract using a website called stickK.com (<https://www.stickk.com/>). Full details are in the instructions document. We included a treatment: for 50% of participants, we suggested that they “wouldn't bother” to make the commitment, whilst for the other 50% we did not express any view on whether they would take up the opportunity or not. The aim of this final block is to study whether the various treatment manipulations had a differential impact on participants' commitment to running and willingness to take on new challenges.

2.3 Other data sources

We will exploit 3 other sources of data, which we expect to receive from the race director in early summer.

2.3.1 Registration data

Runners register in early December preceding the race and are notified in mid December whether they have been chosen in the lottery to receive a place. At the time of registration, runners complete a detailed form, which contains the following types of data:

Race type (called **race name** in the database) refers to the various events that runners can sign up for: 5K run/walk, 10 mile race (as an individual or as a team)

Socio-demographics: gender, age, location (city, state/province, country, postal code). We will use gender and age as a check when merging the different sources of data, since both variables are also available in the pre-race survey; in case of inconsistency, we will use the registration data. The geographical information will be used to construct a variable for whether the registered participant is a local runner (based in DC, VA or MD).

Racing experience: whether the participant ran CUCB previously, (if yes) last year ran, best 5K, 10K and half marathon times in any race, most recent 10 mile time. This information is used to compute a predicted time (called **ten mile equivalent** in the database), which is available for nearly all runners and is used to assign a starting pen for each runner (bib number and bib color).

Additional purchases: a simple race entry costs \$45 and includes a basic finisher t-shirt. At the time of registration, runners can pay to purchase additional items. Of particular interest to us, runners can choose to buy a standard finisher medal (\$20), an enhanced finisher medal which has the participant's finish time engraved on the back (\$25), or a t-shirt upgrade (\$20). Various additional add-ons, such as training packages, “good luck” boxes, or a donation for carbon offset are also proposed each year; we will not make use of this data due to the low demand for these add-ons in previous years.

⁹If participants opted to donate to charity, they were able to specify the charity of their choice by directly entering the name of the charity or by making a selection from the following list: Children's Miracle Network Hospitals (the official charity supported by CUCB 2019), World Wildlife Fund, Disabled American Veterans Charitable Service Trust or American Youth Soccer Organisation.

¹⁰We randomized the order of the list items in order to control for order effects.

Table 3: Components of the dataset

Registration data	<ul style="list-style-type: none"> - predicted time/corral - medal purchase: regular, enhanced, no medal - purchase of T-shirt
Pre-race survey	<ul style="list-style-type: none"> - socio-demographics - race experience and elicitation of time goal - treatment manipulation: control (C), money (M), pride (P), pride & money (PM) - elicitation of belief about race performance
Performance data	<ul style="list-style-type: none"> - official race time - overall rank - split times
Post-race survey	<ul style="list-style-type: none"> - recall questions about goal, race time, and treatment - measures of pride/satisfaction with performance - (M/PM only) choice between gift card and donation to charity - decision to buy medal and race pictures - measures of interest in medals - interest in and take up of race commitment contract

Notes: Additional post-race purchase data (medals and race pictures) may also become available.

2.3.2 Race performance

The official race time of participants is measured through a chip located at the back of their bib number, which records their individual start time and finish time, as well as the time they cross the 5K (3.11 mi), 10K (6.21 mi), and 15K (9.32 mi) marks. We will refer to this official record as the “chip time” (also called “net time”).¹¹ Race results are publicly available on the website (<https://results.sporhive.com/events/6513547676544403712>). Additional information such as overall position in the race, within gender and age group, and a computed Age Grade score is also available (more details in Section 3.1). A screenshot of the information displayed is presented in Appendix A. Results posted on the website will remain unofficial until April 30th. The official race results will be sent to us by the race director in early summer.

Additional data on race performance can be found using a database called RunPix, which communicates results in a more interactive way (<http://www.runpix.co/arace11/57/chy18/rp.php>). For instance, participants can see how many runners they passed during the race and how many runners passed them (overall, before the 6 mile mark, and after the 6 mile mark). We will make use of this data to compute measures of pace relative to other runners.

2.3.3 Post-race purchases

Participants were able to purchase medals when they registered to run the race in December, and then those who received places through the lottery were offered the chance to decide to purchase medals up to February 28th (this was communicated via the January and the February race newsletters). After this cut-off, participants were informed they could only purchase a medal after the race if there were remaining stocks. We have requested data on post-race purchases (medal and race photographs), but we have not received confirmation yet from the race director that this data will be available to us. The race photographs can be ordered through the official race photographers Marathon Photos from midday the day after the race.

The various components of our dataset are summarized in Table 3.

¹¹For large races, the “gun time” (also called “clock time”) is a much less accurate measure since runners cross the start line at different times.

3 Outcome variables

3.1 Race performance

Primary outcomes Our main outcome measures focus on race performance in relation to participants' goals. In the presentation of our results, goal times and actual finish times will be reported in minutes. We will use the official chip time as our measure of finish time (see Section 2.3.2 for more information). As a benchmark, the median finish time in 2018 was 98.3 minutes (91.1 for men and 102.4 for women), very close to the mean of 99.3 minutes (92.7 for men and 103.7 for women). We will present histograms of the distribution of finish times for the 2019 race (comparing our study sample to the entire population of finishers) and report the quartiles of the distribution broken down by treatment. For runners in our study sample, the distribution of actual finish times will be contrasted with the distribution of goal times. We will study the effect of our treatments on the following set of outcome variables:

goal attained is an indicator for whether the participant met the time goal s/he provided in the pre-race survey. Participants were asked to select a time goal hh:mm (without specifying the seconds). On the other hand, the official race time has a precision to the second. We will therefore consider that a participant met his time goal if his chip time was less than a minute away from the goal (for instance, running 1:39:45 if his goal was 1:39). Runners in the incentive conditions (M and PM) will receive the gift card if they meet this rule. For the purpose of our analysis, we will check the robustness of our findings to a more stringent definition of meeting one's goal, which requires to be no slower than the goal time up to the second (so, running 1:39:00 or under if the goal was 1:39).

deviation from goal is defined as the difference (chip time - goal time) measured in minutes. If runners tend to be overconfident as suggested in Markle et al. (2018) and Krawczyk and Wilamowski (2017), we expect this difference to be positive on average, with most participants being slower than their goal (at least in the control group).

We will also consider a relative measure of deviation from the goal time, *% deviation from goal* = (chip time - goal time)/goal time, as one might expect faster runners to be closer to their goal in absolute value than slower runners.

Secondary outcomes For our secondary outcomes, we consider measures of absolute and relative performance, as well as measures of pacing:

age grade and *chip time* will be used as measures of absolute performance. In descriptive statistics, we will use *age grade*, which produces a score for race times expressed as a percentage of the world-best time for the distance for a given age and gender.¹² In regressions, we will use chip time and control for age, gender, and other observables described in Section 4.2.

overall ranking (based on chip time) will be used as a measure of relative performance.

slowed down is an indicator for whether the participant was slower between kilometers 10 and 15 than s/he was between kilometers 0 and 5 as measured using the chip time information.

pace difference is computed as the difference in minutes between the time taken to complete the distance between kilometers 10 and 15 and the time taken between kilometers 0 and 5, again as measured using the chip time information.

relative pace before mile 6 is defined as the ratio $\frac{n_{i,-i}}{n_{-i,i}}$ where $n_{i,-i}$ (respectively, $n_{-i,i}$) is the number of runners who were passed by (respectively, who passed) participant i over the first 6 miles of the race.

relative pace after mile 6 is defined as the ratio $\frac{N_{i,-i}}{N_{-i,i}}$ where $N_{i,-i}$ (respectively, $N_{-i,i}$) is the number of runners who were passed by (respectively, who passed) participant i after the first 6 miles of the race.

Δ *relative pace* is defined as the relative ratio (*relative pace after mile 6* / *relative pace before mile 6*).

¹²<https://www.runnersworld.com/advanced/a20801263/age-grade-calculator/>

3.2 Expected performance

After the treatment manipulation in the pre-race survey, we asked participants to form beliefs about their chances of reaching their time goal t plus or minus N minutes, for different values of N (see Section 2.2.1). More precisely, participants formed beliefs over 7 possible finish time intervals i.e., $I := [t^-; t - 10]; (t - 10; t - 5]; (t - 5; t - 1]; (t - 1; t + 1); [t + 1; t + 5); [t + 5; t + 10); [t + 10; t^+]$ where t^+ and t^- are some upper and lower bounds.¹³ We use participants' answers to this question as well their actual finish time in order to construct measures of expected performance, belief uncertainty, and belief accuracy, which we will compare across treatments. To introduce those measures, let p_k denote the subjective probability of ending with a finish time t located in interval I_k , that is, $p_k := P(T = t \text{ for some } t \in I_k)$ where T is the random variable of finish times. In addition, let t_k be the median finish time among those who finished in interval I_k ; we use this median value to obtain proxies for a participant's expectation and variance in beliefs.¹⁴ We consider the following measures:

expected finish time is computed as $E(T) = \sum_{k=1}^7 p_k t_k$

belief accuracy : j actual finish time - expected finish time j

belief entropy : is the entropy of a participant's belief over potential finish times, which is computed as

$$H(T) = - \sum_{k=1}^7 p_k \log(p_k)$$

belief spread measures the variance in a participant's belief over potential finish times and is computed as

$$V(T) = \sum_{k=1}^7 p_k (t_k - E(T))^2$$

We will favor the entropy as our measure of uncertainty for it does not require making arbitrary assumptions about the expected finish time. In our main analysis, we will present figures for each treatment that contrast the average belief \bar{p}_k for each interval I_k across all participants with the actual proportion of those participants who finished in interval I_k .

3.3 Behavioral measures of pride

To study variation across treatments in participants' feelings of pride from completing the race, we consider behavioral measures of pride based on post-race purchases of memorabilia (medal, race pictures). We also use donation decisions in the money treatments to further study whether participants feel differently about their achievement in the PM and M treatments:

purchased medal is an indicator for whether the participant ordered the medal *after* the race among those who did not order the medal when they registered. In 2018, about 32% (under 5%) of finishers had chosen to buy the standard (enhanced) medal at the time of registration. Based on those figures, we expect that the number of medal orders after the race will be very low, particularly for the enhanced medal. We will therefore pool purchases of the two types of medals into one category.

intends to buy medal is an indicator for whether the participant expressed an intention to buy the medal after the race conditional on not having purchased one at the time of registration (collected in the post-race survey). As for *purchased medal*, we will not make a distinction between the two types of medals (regular vs. enhanced). We will use this variable as our primary outcome measure if we do not obtain data on actual purchases.

¹³The first (last) interval only specified "being at least 10:00 minutes faster (slower)" than the time goal.

¹⁴All times will be expressed in minutes and rounded to the nearest 1 decimal place. We use the median finish time in each interval rather than the mean or the midpoint of the interval in order to limit the sensitivity of our calculations to outliers; furthermore, taking the midpoint of the first and last intervals would require us to make arbitrary assumptions on the bounds t^- and t^+ . We will test the robustness of our findings to alternative procedures such as fitting a parametric distribution on the histograms of predicted times in order to compute the expectation and variance of beliefs.

bought race picture is an indicator for whether the participant reported in the post-race survey having bought/planning to buy at least one of the pictures taken by professional photographers during the race. If we obtain the purchase data from the race director, we will use actual purchases rather than self-reports in order to construct this variable.

donated to charity is an indicator for whether a participant who met his/her goal in treatments M or PM chose to make a donation to charity instead of receiving the voucher payment.

3.4 Self-reported measures of pride

Our proposed behavioral measures of pride are indirect and likely capture other dimensions besides pride. We therefore also rely on participants' personal assessment of their race performance and self-reported feelings of pride in the post-race survey. Our first two variables in the list below are specific to the race context. To go beyond domain-specific measures of pride/satisfaction, we also administered the Tracy and Robins psychological pride scale. For 14 different feelings and emotions, participants were asked to rate on a scale from 1 to 5 the extent to which they felt this way (1- not at all; 5- extremely). The first 7 expressions were positive (accomplished, like I am achieving, confident, fulfilled, productive, like I have self-worth, successful), while the remaining 7 terms were negative (arrogant, conceited, egotistical, pompous, smug, snobbish, stuck-up). We aggregate these ratings to construct two scores capturing the positive and negative facets of pride. Mean ratings for each emotion broken down by treatment will be presented in an appendix to the paper.

pride in the race is a respondent's response from 1 to 5 (1- Nothing to be proud of; 2- Somewhat proud; 3- Moderately proud; 4- Very proud; 5- Extremely proud) to the question asked in the post-race survey "How proud do you feel of your participation in the 2019 Cherry Blossom 10 mile run?". In an appendix to the paper, we will present the histogram of responses broken down by treatment, separating those who met their goal from those who did not.

performance rating is a respondent's rating of his/her performance. Participants were asked to rate their performance on a letter scale with 13 options (F, D-, D, D+, C-, C, C+, B-, B, B+, A-, A, A+). We will convert the rating into a score from 1 (for F) to 13 (for A+).

positive pride score is a respondent's mean rating on the 7 positive feelings and emotions in the Tracy and Robins psychological scale.

negative pride score is a respondent's mean rating on the 7 negative feelings and emotions in the Tracy and Robins psychological scale.

3.5 Recall of performance

At the beginning of the post-race survey, we collected data on participants' recall/knowledge of their pre-race goal time (hh:mm) as well as their actual finish time (hh:mm). Participants could either select a time hh:mm from a dropdown menu or the option "I do not know/I cannot remember". For both questions, we also recorded response times. Our objective is to understand whether runners selectively recall their goal time and/or finish time depending on (i) the treatment they were assigned and (ii) their actual performance relative to the goal set in the pre-race survey (see hypotheses below).

We will consider the following variables as our primary outcomes:

recall bias for goal is measured as the difference in minutes between the goal recalled in the post-race survey and the goal entered in the pre-race survey.

recall bias for finish time is measured as the difference in minutes between the reported finish time and the actual finish time. The difference will be rounded down to 0 if the difference is less than a minute.

In addition, we will consider the following secondary outcome measures:

forgot goal is an indicator for whether the participant reported not knowing/not recalling his or her goal.

forgot_nish_time is an indicator for whether the participant reported not knowing/not recalling his or her finish time.

correctly recalled goal is an indicator for whether the participant entered the same goal time as in the pre-race survey.

correctly recalled_nish_time is an indicator for whether the finish time reported by a respondent was within 59 seconds of his or her official finish time. We allow for a buffer of ± 59 seconds since the official finish time as recorded by the chip has a precision up to the second.

recalled success is an indicator for whether the participant remembered having met his or her goal. Since participants were asked to specify their goal and finish time up to the minute, we define meeting one's goal as finishing with a time no higher than (but possibly equal to) the goal time. We will contrast this measure with actual success.

3.6 Recall of treatment

In the post-race survey, we asked all respondents if they remembered having seen any information or comment regarding their time goal in the first survey and, if so, what the information was. Respondents were encouraged to enter as much information as possible. At the end of the post-race survey, we also asked if they had any comments for us about the study. We will make use of participants' comments in both questions to construct two variables:

remembered treatment is an indicator for whether the participant accurately remembered his or her treatment by providing details about the content of the message (P, PM, M) or reporting having no recall of any comment (control).

reaction to treatment is a categorical variable capturing a participant's emotional reaction to the treatment. The variable takes a value of -1 if the participant expressed a negative emotion (e.g., upset, irritated, angry, etc...) or provided a negative evaluation of the treatment (e.g., unethical, mean, etc...), a value of 1 if s/he provided a positive or neutral opinion (e.g., challenging, amusing, etc...) and 0 if no emotion/opinion was expressed.

We will document the coding scheme developed to build those variables once we have analyzed the distribution of answers.

3.7 Measures of interest in commitment

At the end of the post-race survey, we challenged participants to create a commitment contract for their next race on stickK.com and to nominate us as a referee.¹⁵ The idea was to test whether runners would choose to make a commitment to complete a new race or achieve a new time goal for an upcoming race depending on the treatment they were assigned and their race performance.¹⁶ Creating a commitment contract was entirely voluntary and preliminary numbers suggest that very few participants took up the offer; we will therefore not analyze this outcome. To measure expression of interest, we tracked whether participants clicked on a link "it's time to prove it", which pointed to the website (indicator *clicked on stickK.com link*). We will report the percentage of clicks by treatment in an appendix to the paper.

4 Empirical Analysis

4.1 Empirical strategy

For each outcome variable Y_i presented in Section 3, we estimate the effect of the treatment manipulations (P, M and PM) relative to the control group (C). Let $treat_P_i$ (respectively, $treat_M_i$ and $treat_PM_i$) be an indicator

¹⁵The website allows participants to make personal commitments to a goal of their choice, put money on the line in case of failure, and nominate somebody to verify their success.

¹⁶We presented one of two versions to participants, which only slightly differed in terms of wording. See instructions for more details.

variable taking value 1 if participant i was assigned to treatment P (respectively, M and PM). Let Z_i^{25} be an indicator variable taking value 1 if the incentive level (voucher payment) given to participant i in treatments M and PM was \$25 (instead of \$10). For each outcome Y_i , we consider the following baseline regressions

$$Y_i = \beta_0 + \beta_1 \text{treat_P}_i + \beta_2 \text{treat_M}_i + \beta_3 \text{treat_PM}_i + \mathbf{X}_i^0 + \beta_s + \epsilon_i \quad (1)$$

where \mathbf{X}_i is a vector of controls (possibly outcome-specific) and β_s is a vector of dummies for the mode of recruitment (Email, Facebook, Expo). We detail our choice of controls for each outcome in Section 4.2.

To study the responsiveness to incentives for meeting one’s goal, let pride_i be an indicator variable taking value 1 if participant i was in one of the pride treatments (P or PM) and let Z_i be a continuous variable for the level of incentives offered (\$10 or \$25 in M and PM, \$0 in C and P). We consider the following specifications, which assume a linear response to incentives:

$$Y_i = \beta_0 + \beta_1 \text{pride}_i + \beta_2 Z_i + \beta_3 \text{pride}_i \cdot Z_i + \mathbf{X}_i^0 + \beta_s + \epsilon_i \quad (2)$$

While these specifications allow for a straightforward monetization of the pride effect, the responsiveness to incentives might be non linear. We will therefore also report the results from regressions of the form:

$$Y_i = \beta_0 + \beta_1 \text{treat_P}_i + \beta_2 \text{treat_M}_i + \beta_3 \text{treat_PM}_i + \beta_4 \text{treat_M}_i \cdot Z_i^{25} + \beta_5 \text{treat_PM}_i \cdot Z_i^{25} + \mathbf{X}_i^0 + \beta_s + \epsilon_i$$

We anticipate a possible lack of power for the estimation of the interaction terms (coefficients β_4 and β_5) so we will make the linear specification (2) our primary specification.

Since we will run regressions on a large number of outcomes, we will adjust our p-values for multiple hypothesis testing as deemed appropriate within classes of primary outcomes.

4.2 Control variables

For each type of outcome, we will run regressions with a minimum set of controls and with a richer set of covariates described below. A description of the various control variables is presented in Appendix C.

Race performance outcomes For the outcomes pertaining to race performance, all regressions will control for the goal chosen by the participant in the pre-race survey, *time goal*, as well as (possibly) the time predicted by an algorithm based on a participant’s gender, age and past performance (available in the registration data and called *predicted time*).¹⁷ In addition, regressions will include socio-demographic variables and race-specific variables:

Race-specific variables: whether the participant is a first timer, training level for the race, importance of/commitment to the time goal, confidence in success.¹⁸

Socio-demographics: controls for gender, age, state of residence, education and income.

We will present regressions with and without the socio-demographic controls in order to focus on the race-specific factors first.

Expected performance We will use the same controls as for actual race performance.

Behavioral and self-reported measures of pride For all outcomes that use answers to the post-race survey, we will include the indicator variable *S2 response date*, which is equal to 1 if the participant responded to the survey on the first day it was sent (nearly 75% of respondents). All regressions will include this control.

In addition, for our behavioral and self-reported measures of pride, we will include measures of race performance (both absolute and relative to the goal and other runners) as well as socio-demographic controls:

¹⁷We will use *predicted time* only if available for all our participants.

¹⁸Two variables, *goal importance* and *goal commitment*, capture a participant’s attachment to his or her time goal. If the two variables are highly correlated, we will take the mean score as our measure of goal attachment.

Race performance: whether the respondent reached his goal, chip time/age grade, whether s/he slowed down in the second half of the race.¹⁹

Socio-demographics: controls for gender, age, state of residence, education and income.

Interest in medals (for medal purchase only): number of medals owned, interest in medal hangers, interest in donating medals.

Recall of performance Controls will include *S2 response date*, *goal attained*, response time when recalling the goal/finish time, socio-demographics.

Recall of treatment Controls will include *S2 response date*, *goal attained*, socio-demographics.

4.3 Heterogeneity

We will estimate additional specifications to test for heterogeneous treatment effects. For all our outcome variables, we will present separate estimates for 3 moderators: gender (*male* vs. *female*), past experience with the race (*rst-timer* vs. *not rst-timer*), and attachment to the event as measured through pre-race purchase of the race medal (*medal pre-ordered* vs. *medal not pre-ordered*). All 3 moderators split the sample into approximately equal groups.²⁰ In addition, we will test whether the strength of the response depends on a participant's attachment to the goal (*goal importance* and *goal commitment*), as well his confidence that the goal will be achieved (*goal confidence*). For all post-race outcomes, we will present separate results for those who met their goal and those who did not, although we do not know at this time how balanced the two groups will be.

4.4 Attrition

For the analysis of race outcomes detailed in Section 3.1, we anticipate virtually no attrition, with a participation in the race very close to 100% among our study participants.²¹ If we have participants for whom the performance data is not available (for instance, because they did not show up or because they forgot to wear their bib number), we will either drop them from the estimation sample or assume that they would have missed their goal had their race time been recorded. For the outcomes constructed from responses to the post-race survey, the attrition rate is very low at 6.6% (24/431) in the full sample and is nearly identical across all treatments (see Table 2). As a robustness check, we will compute Lee (2009) bounds on the treatment effects to account for attrition.

4.5 Hypotheses

To conclude this PAP, we would like to formulate a few hypotheses:

1. We anticipate that participants' performance will be higher in the incentive treatments (M and PM) relative to the control group (C). We do not have a strong prior regarding the mean effect in the pure pride treatment (P). We do expect a positive interaction effect of pride and money.
2. We expect a higher variance in the distribution of performance in the pride treatments (P and PM) due to heterogeneity in the emotional reaction to the message.
3. We are unsure about the effect of the pride message on participants' beliefs about their performance relative to the goal. We expect that some participants will update downwards, thus anticipating lower performance than the control; however, we also expect that some participants will not update at all (thus

¹⁹In place of an indicator for whether the goal was reached, we will use the continuous variable *% deviation from goal* if it has higher explanatory power. Similarly, we will consider the continuous measure Δ *relative pace* as an alternative to the indicator *slowed down*. If the various measures of performance are very highly correlated, we will focus on performance relative to the goal.

²⁰There were 49.7% of females and 45.0% of first-timers (using the pre-race survey). Based on responses to the post-race survey, there were also 50.0% of participants who pre-ordered the medal.

²¹This is based on the following two observations. First, using the registration data available to us for years 2016, 2017 and 2018, we estimate the likelihood of not showing up on race day conditional on having picked up the race number at the Expo to be around 6.5% (averaged over all 3 years). Second, we expect that registered participants who signed up for our study are particularly committed to completing the race.

ignoring the message) or possibly update upwards if they anticipate a boosting effect of the message. We are also unsure about the effect of incentives on beliefs. On the one hand, participants might infer from incentives being offered that meeting one's goal is hard, which could have a demotivating effect. On the other hand, they might internalize the effect of incentives on their own motivation, leading them to update upwards.

4. Ex post, we expect that participants who failed to meet their goal will try to self-enhance by recalling a harder goal and/or a better finish time, or simply by reporting that they do not remember. We suspect that selective recall by those who failed will be stronger in the pride and incentive treatments relative to the control (we do not have an hypothesis about the relative strength of incentives vs. pride or the interaction between the two). In contrast, we expect that successful participants will remember their goal time and finish time with great accuracy, particularly so in the pride and incentive treatments (relative to the control).
5. We conjecture that there might be two types of individuals who feel differently about acknowledging their own pride:

Individuals who feel very good about telling others that “they did it and they were wrong”. We expect those to be more likely to remember the treatment manipulation and express their thoughts on it, more likely to take the gift card (“I earned it”), and more likely to acknowledge their own pride (including possibly its dark side).

Individuals who do not feel proud of their own pride and need counter-signals. We expect that those will be less likely to remember the treatment manipulation and express their thoughts on it, more likely to donate the money they earned, and less likely to acknowledge their pride (especially when it has a negative connotation).

To test our last point, we will study potential correlations across our various outcome measures.

References

- Krawczyk M. and Wilamowski W. (2017), “Are we all overconfident in the Long Run? Evidence from One Million Marathon Participants,” *Journal of Behavioral Decision Making*, 30, 719—730.
- Lee D.S. (2009), “Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects,” *The Review of Economic Studies* 76 (3), 1071—1102.
- Markle A., Wu G., White R. and Sackett A. (2018), “Goals as reference points in marathon running: A novel test of reference dependence,” *Journal of Risk and Uncertainty*, 56, 19—50.
- Tracy J. and Robins R. (2007), “The Psychological Structure of Pride: A Tale of Two Facets,” *Journal of Personality and Social Psychology*, 92.3, 506—525.

Appendix A: Detail of communications

Table 4: Timeline of communications sent to survey participants

Pre{race survey	
Facebook invite	Thurs, April 4th
Facebook update (no more women)	Thurs, April 4th
Email invite	Fri, April 5th
Facebook reminder	Fri, April 5th
Post{race survey	
Email invite	Wed, April 10th
Email reminder	Sat, April 13th
Email reminder	Mon, April 15th
Email reminder	Wed, April 16th
Prizes and Thank you	
Email notification	Sun, April 28th

Notes: In addition, oral communications with race participants occurred during the in-person recruitment at the CUCB Race Expo on Saturday 6 April.

Figure 3: Pre—race survey: Email Advert

Help Support Social Science Research (limited to 400 respondents)!

The race is partnering with a group of social science researchers at the University of Oxford to conduct a study about participation of runners in races. Runners in the 10 mile can help this effort by completing a pre- and a post-race survey (no 5K participants, sorry!). Survey respondents will be entered into a drawing for a chance to win great prizes including: \$150, \$25 and \$10 vouchers from Under Armour, and four free, guaranteed entries into next year's Credit Union Cherry Blossom race! You will only be eligible to win the prizes if you complete both the pre- and the post-race survey. You'll have up to 7 days to complete Survey 2 once we send you the link. We expect each survey to take no more than 10 minutes. If you are interested in participating, [CLICK HERE](#). If you have questions or concerns, please email cucbresearch@gmail.com.

Notes: Message included as part of a general email sent by the race director to all registered participants before the race.

Figure 4: First Facebook Advert



Figure 5: Facebook Reminder

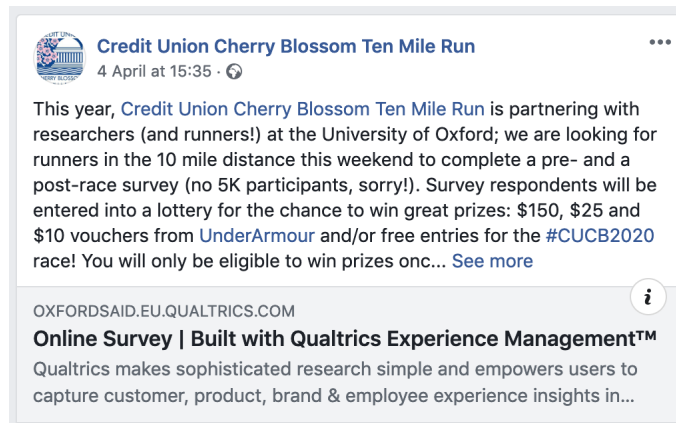


Figure 6: Facebook update (no more women)



Figure 7: Post—race survey: Email link

Dear Cherry Blossom runner,

We hope you had a fantastic race day on Sunday! Thank you for participating in the research study. It was great to meet so many of you at the Expo on Saturday.

Please follow this link to Survey 2: [Take the Survey](#)

Please complete Survey 2 within 7 days of receipt of this email. Note that Survey 2 is *even shorter* than Survey 2 :-)

It is really important for the validity of the research that everyone completes Survey 2.

And remember, you will not be eligible to win any prizes until you have submitted your responses to Survey 2. A reminder of the fantastic prizes up for winning...

- Free, guaranteed entry to next year's race
- \$10, \$25 and \$150 vouchers for Under Armour

We greatly appreciate your time, and look forward to sharing the results with you and all road runners!

Any queries or concerns, please do not reply to this email but direct them to cucbresearch@gmail.com.

Best wishes,

Cherry Blossom 10 mile & University of Oxford Research Team

[Click here to unsubscribe](#)

Figure 8: Post—race survey: Email reminder 1

Dear all,

Just a reminder to complete Survey 2: [Take the Survey](#)

We really appreciate your contribution to the study.

And, don't forget that you won't be eligible to win any prizes until you've completed Survey 2. The link will be active until Wednesday morning.

Thank you very much again for your participation,

Cherry Blossom Research Study

[Click here to unsubscribe](#)

Figure 9: Post—race survey: Email reminder 2

Dear Cherry Blossom Runner,

Thank you very much for agreeing to be in our study!

This is a reminder that Survey 2 will remain open only for another **48 hours**.

To preserve the credibility of the study, we would really appreciate your contribution. Survey 2 should take even less time than Survey 1 and you will have the opportunity to give us feedback on the study.

Complete it here: [Take the Survey](#).

All participants who completed both surveys will be entered in a lottery for the chance to win a free, guaranteed entry for next year's race or Under Armour vouchers.

Thank you again for your participation in this research!

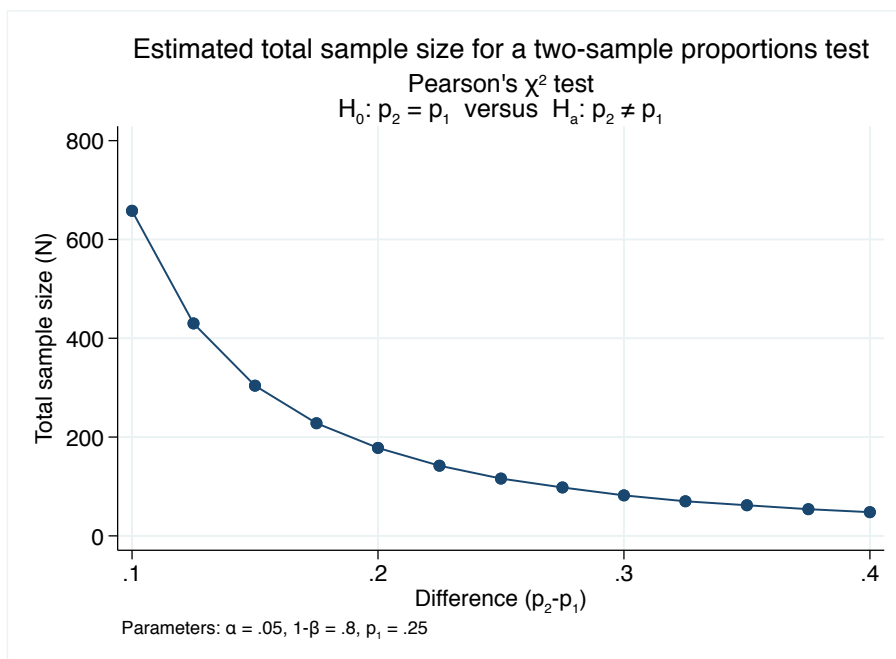
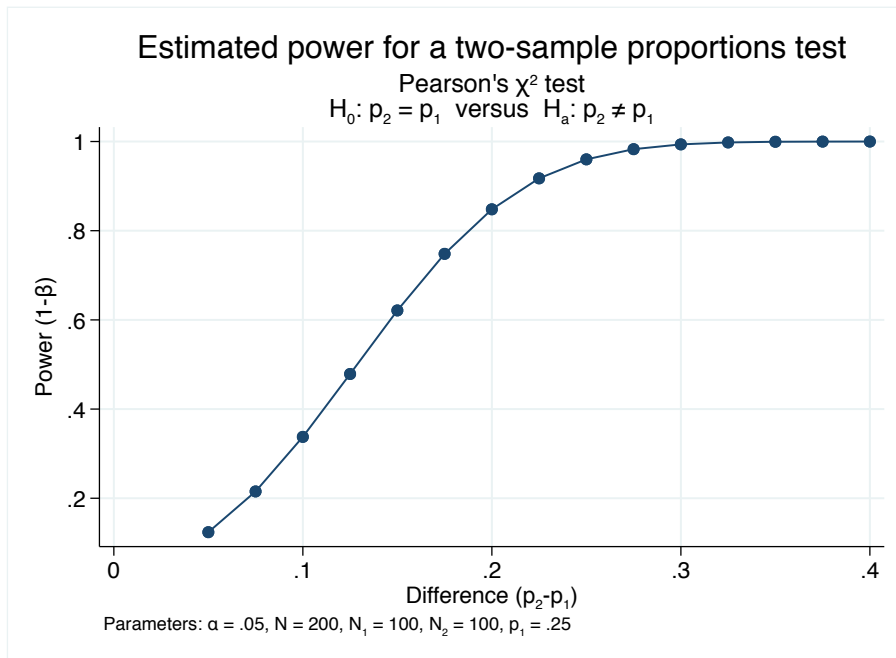
With best wishes,

Cherry Blossom Research Study

[Click here to unsubscribe](#)

Appendix B: Power calculations

Below we present power calculations for differences in proportions between the control and any of treatment groups (focusing on our main outcome, *goal attained*, which is binary). The first figure shows power for different treatment differences, fixing the total sample size at 200 (100 in each group) and setting α at 5%. The control group mean is set at 25% as we estimated that only 1 in 4 runners would meet their time goal. To obtain 80% power, the treatment difference between two groups must be at least 19%. The second figure shows how the minimum detectable effect size varies with the size of the sample, holding power at 80% and α at 5%. To detect a treatment difference of 10 percentage points, one would need at least 300 participants in each group.



Appendix C: Explanatory variables

Note: The data source is indicated in brackets after the definition of each variable where R: registration data, S1: Survey 1 (pre-race), S2: Survey 2 (post-race).

Socio{demographic characteristics	
<i>female</i>	Takes value 1 if the participant is a female (R)
<i>age</i>	Years of age on race day (R)
<i>local</i>	Takes value 1 if participant lives in DC, VA or MD
<i>education</i>	Number of years of education post high school (S1); constructed from question: “What is the highest level of school you have completed or the highest degree you have received?” 1: Less than high school degree 2: High school graduate 3: Some college but no degree 4: Associate degree in college (2-year) 5: Bachelor’s degree in college (4-year) 6: Master’s degree 7: Doctoral degree 8: Professional degree (JD, MD) Conversion: 1,2: 0 years, 3: 1 year, 4: 2 years, 5: 4 years, 6: 6 years, 7: 9 years, 8: 7 years
<i>income</i>	Set of 6 indicator variables for household income (S1); constructed from question: “How much total combined money did all members of your household earn last year? Please indicate the answer that includes your entire household income in 2018 before taxes:” 1: Less than \$20,000 2: \$20,000 to \$29,999 3: \$30,000 to \$39,999 4: \$40,000 to \$49,999 5: \$50,000 to \$59,999 6: \$60,000 to \$69,999 7: \$70,000 to \$79,999 8: \$80,000 to \$89,999 9: \$90,000 to \$99,999 10: \$100,000 to \$149,999 11: \$150,000 to \$199,999 12: \$200,000 or more 13: Prefer not to answer To create more equal bins, we will regroup income categories 1-6 into a category “Less than \$70,000” and categories 7-9 into “\$70,000 to \$99,000”.
Race-speci c variables	
<i>rst-timer</i>	Takes value 1 if the participant took part in the CUCB race for the first time in 2019 (R)
<i>training level</i>	Score from 1 to 7 assessing personal training level (S1); answer to question: “How well have you trained for this year’s race?” (1= did not train at all, 7 = trained extremely well)
<i>time goal</i>	Time goal for the race in minutes (S1); based on answer to question: “What time are you aiming for on Sunday?” (hh:mm)
<i>goal importance</i>	Score from 1 to 7 measuring attachment to goal (S1); answer to question: “How much would it mean to you to achieve this goal?” (1= Nothing at all, 7 = A great deal)
<i>goal commitment</i>	Score from 1 to 7 measuring commitment to goal (S1); answer to question: “How committed are you to this goal?” (1= Not committed at all, 7 = extremely committed)
<i>goal confidence</i>	Score from 1 to 7 measuring confidence in reaching the goal (S1); answer to question: “How confident are you that you will meet your goal?” (1= not confident at all, 7 = certain)
<i>predicted time</i>	Finish time in minutes as predicted by an algorithm based on a participant’s gender, age, and past performance in one or more race distances (R)

(continued)

Interest in medals	
<i>pre-ordered medal</i>	Equal to 1 if the respondent pre-ordered the race medal (whether standard or enhanced) at the time of registration (R)
<i>number of medals</i>	Set of indicator variables for the number of medals owned from previous races (S2) with options “none”, “between 1 and 5”, “between 5 and 10”, “between 10 and 20” and “more than 20”
<i>interest in medal display</i>	Equal to 1 if the respondent clicked on one or both of the medal display idea links (S2)
<i>interest in medal donation</i>	Equal to 1 if the respondent clicked on the medal donation and/or recycling idea links (S2)

Figure 10: Results database

Jemal Yimer
Bib #1, 22 years , ETH , Male

[Photo](#) [Certificate](#) [Share](#)

CHIP TIME	NAME	SPLITS MI.	POSITION	PACE	TIME
00:45:36 Gun time 00:45:36 Finished 10.0 Miles Speed 13.11 miles/h Pace 04:35 min/miles Age Grade 0.9649 Country ETH	5K	3.11 mi	2	04:44 min/mi	00:14:42
	10K	6.21 mi	2	04:35 min/mi	00:28:27
	15K	9.32 mi	3	04:34 min/mi	00:42:36
	Finish	9.96 mi		04:35 min/mi	00:45:36

Overall position **1** Out of 17437
Men **1** Out of 7146
M 20-24 **1** Out of 360

[Do you have a question about this result?](#)