Introduction

The Cornell University paper, "An Observational Study of the Effect of Nike Vaporfly Shoes on Marathon Performance," by Joseph Guinness et al. examines the Vaporfly performance enhancement on a group of elite and sub-elite marathon runners (308 men and 270 women) over the period 2015 to 2019.[1] The Cornell study applied mixed effects linear regression and found that, when controlling for runner and marathon course random effects (in levels only), the improvement in marathon times from wearing Vaporflys ranges between 2.1 and 4.1 minutes for men (3.12 minutes on average) and between 1.2 and 4.0 minutes for women (2.59 minutes on average). My study extends the Cornell model by incorporating interaction terms for individual runners and for marathon courses, and it evaluates the uncertainty in the Vaporfly effect under a Bayesian framework. The objective of this work is to determine the magnitude of the improvement under this extended approach and to evaluate whether the improvement varies across (1) gender, (2) runner and/or (3) course.

Methods

The extended methodology is applied to the same data as used in the Cornell study. The data consists of 1618 marathon performances ($n = 840$ for men, $n = 778$ for women) where the shoe type was noted.[2] Similar to the Cornell paper, each runner is assigned an index between 1 and R ($R = 308$ men, $R = 270$ women), and each marathon course is assigned an index between 1 and M ($M = 21$ for men, $M = 22$ for women). I also followed the same notation as used in the Cornell paper:

Y_i = marathon time in minutes for performance i

$$X_i = \begin{cases} 1 & \text{if Vaporfly shoes are worn in performance } i \\ 0 & \text{if Vaporfly shoes are not worn in performance } i \end{cases}$$

$j(i)$ = runner who completed performance i

$k(i)$ = marathon course associated with performance i

The Bayesian models evaluated in my study are presented in Table 1. As in the Cornell paper, the slope β and intercept α represent *fixed effects*, meaning they represent random draws from the population and do not vary by runner or marathon course. *Random effects*, on the other hand, are meant to capture features that could potentially vary by "group" in which the data was collected. The variable A_j represents the random effect for runner $j = 1, \cdots, R$, while the variable C_k represents the random effect for marathon course $k = 1, \cdots, M$. These variables impact the model's intercept directly; hence, they are random effects *in levels*. The new slope parameters β_j and β_k represent random effects associated with the Vaporfly variable X_i. β_j represents the runner-Vaporfly interaction, while β_k represents the course-Vaporfly interaction. Part of the objective of my study is to determine if these slope parameters are significant. This can be determined by looking at their posterior distributions to see if zero is contained in the 95% credible set. Whether or not the Vaporfly effect varies by gender can be determined by comparing the predicted posterior distributions (PPDs) between the final model run on men's data and the same model run on women's data.

[1] See https://www.researchers.one/article/2020-02-14.

[2] The raw data set consists of 1690 performance; however, 72 performances (40 men and 32 women) were excluded as the type of shoe worn was not identified.

Computation

Once the data was prepared and loaded in R, the models were run using JAGS with a burn-in (B) and total iterations (T) shown in Table 1. Next, I checked model convergence by looking at the Gelman-Rubin DIC mean divergence statistic (< 1.10) and effective sample size (>1000-2000).[3] In the case of convergence problems, I looked at trace plots and either reran the algorithm or increased the number of iterations.[4] This was a necessary approach to evaluate convergence with so many variables. For instance, while model 1 (the most complex among the models tested) has seven main parameters and 11 hyperparameters, it has 673 parameters in total when taking indexing into consideration (e.g., $A_1, \ldots, A_R$).

The MCMC simulation chains from JAGS were then collected and combined for each parameter to analyze their posterior distributions and for use in generating the PPDs. Beside looking at the posterior distribution for the β_j and β_k parameters to evaluate whether the Vaporfly effect varies across runner and/or course, I evaluated the PPD of $\Delta\widehat{Y}_i = \left(\hat{\beta} + \beta_j + \beta_k\right)\Delta X$ to (1) determine the magnitude of improvement for men and women separately and (2) to assess whether the improvement varies between the genders.[5] The PPD was generated by simulating 50,000 random draws for $\beta_j \sim Normal(0,1/\hat{\tau}_A)$, $\beta_k \sim Normal(0,1/\hat{\tau}_C)$, and $\varepsilon_i \sim N(0,1/\hat{\tau}_\varepsilon)$; and then drawing $\Delta\widehat{Y}_i \sim Normal\left(\left[\hat{\beta} + \beta_j + \beta_k\right]\Delta X, \varepsilon_i^2\right)$. The "hat" variables were estimated from the means of their MCMC posterior distributions.

Model Comparisons

All models were compared using their DIC mean divergence (see the last two columns in Table 1): the lower the number, the better. See the descriptions for the different models attempted. I started with the most complex model and then simplified it systematically. Model 4 is the final model selected; it produces a low DIC and has good convergence. There is also no strong need for informative (or semi-informative) priors on β as in model 3, for example.

Table 1. Models

No.	Model	Description	DIC (multi. psrt)[6]	
			Men	Women
1	$Y_i = \alpha + \left(\beta + \beta_{j(i)} + \beta_{k(i)}\right)X_i + A_{j(i)} + C_{k(i)} + \varepsilon_i$ <u>Fixed effects:</u> $\alpha \sim N(\mu_\alpha, 1/\tau_\alpha)$, $\beta \sim N(\mu_\beta, 1/\tau_\beta)$ $\varepsilon_i \sim N(0,1/\tau_\varepsilon)$, $\mu_\alpha, \mu_\beta \sim N(0,100)$ $\tau_\alpha, \tau_\beta, \tau_\varepsilon \sim Gamma(0.1, 0.1)$ <u>Random effects:</u> $A_j \sim N(\mu_A, 1/\tau_A)$ for $j = 1, \cdots, R$ $C_k \sim N(\mu_C, 1/\tau_C)$ for $k = 1, \cdots, M$ $\beta_j \sim N(\mu_a, 1/\tau_a)$ for $j = 1, \cdots, R$ $\beta_k \sim N(\mu_c, 1/\tau_c)$ for $k = 1, \cdots, M$ $\mu_A, \mu_C, \mu_a, \mu_c \sim N(0,100)$ $\tau_A, \tau_C, \tau_a, \tau_c \sim Gamma(0.1, 0.1)$	Main parameters ($\alpha, \beta, \beta_j, \beta_k, A_j$, and C_k) have *informative* priors (i.e., stochastic means and variances based on the pooled data) $B = 100{,}000$ $T = 750{,}000$	4862 (1.03)	4826 (1.09)

[3] Deviance Information Criterion (DIC).
[4] At some point, this required increasing the memory allocated to R by using the `memory.limit()` command.
[5] Note that X_i was scaled to apply Bayesian analysis; hence, ΔX represents the difference in scaled values for Vaporfly vs. no Vaporfly.
[6] Multivariate Potential Scale Reduction Factor (PSRT).

No.	Model	Description	DIC (multi. psrt)[6]	
			Men	Women
2	$Y_i = \alpha + (\beta + \beta_{j(i)} + \beta_{k(i)})X_i + A_{j(i)} + C_{k(i)} + \varepsilon_i$ Fixed effects: $\alpha \sim N(0,1/\tau_\alpha)$, $\beta \sim N(\mu_\beta, 1/\tau_\beta)$ $\varepsilon_i \sim N(0,1/\tau_\varepsilon)$, $\mu_\beta \sim N(0,100)$ $\tau_\alpha, \tau_\beta, \tau_\varepsilon \sim Gamma(0.1, 0.1)$ Random effects: $A_j \sim N(0,1/\tau_A)$ for $j = 1, \cdots, R$ $C_k \sim N(0,1/\tau_C)$ for $k = 1, \cdots, M$ $\beta_j \sim N(\mu_a, 1/\tau_a)$ for $j = 1, \cdots, R$ $\beta_k \sim N(\mu_c, 1/\tau_c)$ for $k = 1, \cdots, M$ $\mu_a, \mu_c \sim N(0,100)$ $\tau_A, \tau_C, \tau_a, \tau_c \sim Gamma(0.1, 0.1)$	Slopes ($\beta, \beta_j,$ and β_k) have *informative* priors (i.e., stochastic means and variances based on pooled data); levels (α, A_j, and C_k) have *semi-informative* priors (i.e., zero means and random variance based on the pooled data) $B = 100{,}000$ $T = 750{,}000$	4862 (1.05)	4827 (1.06)
3	$Y_i = \alpha + (\beta + \beta_{j(i)} + \beta_{k(i)})X_i + A_{j(i)} + C_{k(i)} + \varepsilon_i$ Fixed effects: $\alpha \sim N(0,1/\tau_\alpha)$, $\beta \sim N(0, 1/\tau_\beta)$ $\varepsilon_i \sim N(0,1/\tau_\varepsilon)$, $\mu_\beta \sim N(0,100)$ $\tau_\alpha, \tau_\beta, \tau_\varepsilon \sim Gamma(0.1, 0.1)$ Random effects: $A_j \sim N(0,1/\tau_A)$ for $j = 1, \cdots, R$ $C_k \sim N(0,1/\tau_C)$ for $k = 1, \cdots, M$ $\beta_j \sim N(0, 1/\tau_a)$ for $j = 1, \cdots, R$ $\beta_k \sim N(0, 1/\tau_c)$ for $k = 1, \cdots, M$ $\tau_A, \tau_C, \tau_a, \tau_c \sim Gamma(0.1, 0.1)$	Main parameters have *semi-informative* priors (i.e., zero means and stochastic variances based on the pooled data) $B = 100{,}000$ $T = 500{,}000$	4861 (1.00)	4827 (1.00)
4	$Y_i = \alpha + (\beta + \beta_{j(i)} + \beta_{k(i)})X_i + A_{j(i)} + C_{k(i)} + \varepsilon_i$ Fixed effects: $\alpha, \beta \sim N(0,1000)$ $\varepsilon_i \sim N(0,1/\tau_\varepsilon)$, $\tau_\varepsilon \sim Gamma(0.1, 0.1)$ Random effects: $A_j \sim N(0,1/\tau_A)$ for $j = 1, \cdots, R$ $C_k \sim N(0,1/\tau_C)$ for $k = 1, \cdots, M$ $\beta_j \sim N(0, 1/\tau_a)$ for $j = 1, \cdots, R$ $\beta_k \sim N(0, 1/\tau_c)$ for $k = 1, \cdots, M$ $\tau_A, \tau_C, \tau_a, \tau_c \sim Gamma(0.1, 0.1)$	Parameters α and β have *uninformative* priors; parameters A_j, C_k, β_j, and β_k have *semi-informative* priors (i.e., zero means and stochastic variances based on the pooled data) $B = 100{,}000$ $T = 400{,}000$	4861 (1.00)	4827 (1.00)
5	$Y_i = \alpha + (\beta + \beta_{j(i)} + \beta_{k(i)})X_i + A_{j(i)} + C_{k(i)} + \varepsilon_i$ Fixed effects: $\alpha, \beta \sim Normal(0,1000)$ $\varepsilon_i \sim N(0,1/\tau_\varepsilon)$, $\tau_\varepsilon \sim Gamma(0.1, 0.1)$ Random effects: $A_j \sim Normal(0,1000)$ for $j = 1, \cdots, R$ $C_k \sim Normal(0,1000)$ for $k = 1, \cdots, M$ $\beta_j \sim Normal(0,1000)$ for $j = 1, \cdots, R$ $\beta_k \sim Normal(0,1000)$ for $k = 1, \cdots, M$	All parameters have *uninformative* priors $B = 100{,}000$ $T = 400{,}000$	4885 (1.03)	4810 (1.02)

Results

The posterior distributions of the model's main parameters are presented in Table 2. This indicates that the fixed effects parameters α and β are both significant, whereas the remaining variables are not significant. In particular, the runner-Vaporfly interaction parameter β_j and the course-interaction parameter β_k are not significant since their 95% credible intervals contain zero and the parameters' mean values are extremely small. This means that the improvement in marathon performance does not vary by runner or course.

Table 2. Posterior Distribution Statistics

Parameter	Men			Women		
	mean	2.5%	97.5%	mean	2.5%	97.5%
α	139.1342	137.9511	140.4180	159.4883	157.6964	161.3121
β	-0.8258	-1.3635	-0.3126	-0.5356	-1.149	0.0699
β_j	0.0000	-1.3220	1.3279	-0.0001	-1.3142	1.3173
β_k	0.0000	-1.0516	1.0579	0.0006	-1.1511	1.1446
A_j	0.0076	-7.7082	8.2679	0.0230	-12.3811	11.6756
C_k	0.0297	-3.3032	5.0520	0.0762	-5.0133	7.1875

$B = 50{,}000;\ T = 250{,}000$

In addition, Figure 1 and Table 2 indicates that there is no significant difference in the level of improvement between men and women due to wearing Vaporfly shoes. Their 95% credible intervals have a lot of overlap. Based on this analysis, I can conclude that the Vaporfly effect is a fixed effect, and not a random effect that varies by gender, runner or course. Pretty much any long-distance runner should expect to see a performance improvement from wearing these shoes!

Figure 1. ΔY Posterior Probability Distributions (Men vs. Women)

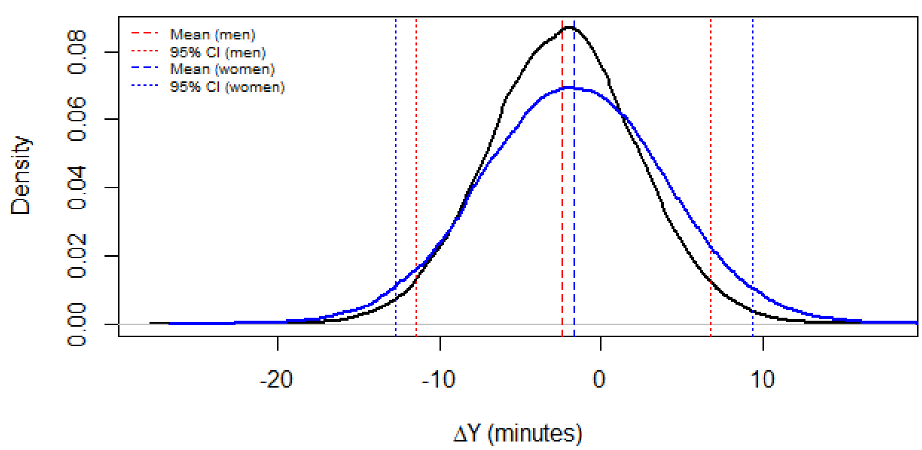

Table 3. ΔY PPD Statistics

	mean	2.5%	97.5%
Men	-2.3625	-11.4228	6.7727
Women	-1.7142	-12.932	9.3943