Supplementary information

Content table

Supplementary Text 1: Daily survey in The Tick App (2019). We used data collected from this survey from April 1st 2019 to April 27th 2020. The 2020 version was deployed on April 28th 2020

<u>Block: Outdoor activities and preventative behaviors</u>

1- Did you do any outdoor activities today?
  - No, I was indoors all day [SKIP to Q6]
  - Yes
2- Which of these outdoor activities did you do ______ [fill in date/ yesterday or today, depending on timing of survey taking] (select all that apply)
  - Garden (vegetable/flower)
  - Mow the lawn
  - Picnic/grill/eat outdoors on the lawn
  - Picnic/grill/eat outdoors in woodlands or park
  - Hike/walk/run on nature trails
  - Visit the beach (river, lake or ocean)
  - Camp
  - Bike/ATV/motorcycle on nature trails
  - Other outdoor activity
3- [if other in Q2] What other outdoor activities did you do?
4- Today, did you use any of these measures to reduce exposure to ticks or tick bites? (select all that apply)
  - Applied tick repellent (ex. DEET, picaridin)
  - Adjusted clothing (ex. light-colored, long-sleeved, tucking pants in socks, boots)
  - Shower or bathe to remove ticks
  - Wore insecticide-treated clothing (ex. permethrin treated pants)
  - Checked myself for ticks
  - Other
  - None
5- [if other in Q4] What other measure to reduce exposure to ticks did you use?

<u>Block: Ticks on self</u>

6- How many ticks did you find on yourself today?
  - 0 [SKIP to Q15]
  - 1
  - 2
  - 3 or more
7- Was the tick attached to your skin?
  - No
  - Yes
  - I don't know
8- When do you think you picked up the tick?
  - Today
  - Yesterday
  - Earlier than yesterday
9- [if Earlier than yesterday in Q8]: What date do you think the tick was picked up? (month / day / year)
10- What do you think you were doing when you picked up the tick?
  - I don't know
  - Mowing the lawn

- Gardening / weeding
- Trimming trees / removing brush / raking leaves
- Playing in the yard / park
- Grilling / eating / sitting outdoors
- Working (my job as listed in my profile)
- Hiking / walking / running on unpaved paths
- Camping
- Fishing
- Other activity

11- [if other in Q10] What do you think you were doing when you picked up the tick?

12- Where do you think you picked up the tick?

- In my yard
- On my neighbor's property
- In a recreational area (ex. forest, park)
- Other
- I don't know

13- [if Other in Q12] Where do you think the tick was picked up?

- Name of park/forest /area: ______________________________
- Zipcode: ____
- City/Township: ______________________________________
- State:____

14- Next we have some questions regarding what your tick looks like. Pick the image that looks most like your tick.

14a- Was your tick big or small? [pick between three images: one of a big, medium and small tick]

14b-Was you tick flat or round (filled with blood)? [pick between an image of a flat and an engorged tick]

14c- What kind of colors does your tick have? [Pick between 4 images with text]

14e- Please upload a picture of your tick.

- Add picture
- I don't have the tick anymore

<u>Block: Ticks on others</u>

15- Did another person in your household find tick(s) on themselves?

- Not applicable
- No
- Yes. Please complete a tick report by clicking Report-a-Tick in the homescreen.

16- How many ticks did you find on your pet(s)?

- Not applicable
- 0
- 1
- 2-5
- More than 5

Supplementary Text 2: Daily survey in The Tick App (2020). We used data collected from this survey from April 28th, 2020 to July 31st 2020. We only included modified and new questions compared to the 2019 daily survey version. We indicate whether the question or section has been *[MODIFIED]*, is *[NEW]* or the *[SAME AS 2019]*. The block for Outdoor activities was split into 3 questions to facilitate user's data entry and posterior analysis.

<u>Block: Outdoor activities and preventative behaviors</u>

1- Did you do any outdoor activities today?
    - No, I was indoors all day [SKIP to Q12]
    - Yes
2- Today, which of these outdoor activities did you do in your yard (lawn and garden) or someone else's yard? (select all that apply) [*MODIFIED*]
    - I did not do yard activities
    - Mowed the lawn
    - Gardened / weeded
    - Removed brush / trimmed trees / raked leaves
    - Grilled / ate / sat in the yard
    - Played in the yard
    - Other outdoor activity in the yard
3- [if Other in Q2] What other activities did you do in your yard?
4- Today, which of these outdoor activities did you do in natural areas and parks? (select all that apply) [*MODIFIED*]
    - I did not do outdoor activities in natural areas and parks
    - My work (as listed in my profile)
    - Walked / walked the dog
    - Hiked / ran
    - Played with dogs / kids, did sports
    - Rode a bike / ATV / motorcycle
    - Grilled / ate / sat
    - Camped
    - Other outdoor activity
5- [if Other in Q4] What other activities did you do in natural areas and parks?
6- Today, did you do any other outdoor activities? [*MODIFIED*]
    - None
    - Walked on sidewalks, pavement
    - Spent time on courts / concrete, for example played basketball, washed the car
    - Other
7- [if Other in Q6] What other outdoor activities did you do?
8- Today, did you use any of these measures to reduce exposure to ticks or tick bites? (select all that apply)
    - Applied tick repellent (for example DEET, picaridin)
    - Adjusted clothing (for example long sleeves, tucking pants in socks)
    - Showered or bathed to remove ticks
    - Wore permethrin-treated clothing
    - Checked myself or had someone check me for ticks
    - Avoided areas with ticks
    - Other measure
    - None

9- [If Other in Q8] What other measure to reduce exposure to ticks did you use?

10- Today, has the need for social distancing or other COVID-19 measures impacted your outdoor activity? [*NEW*]

- No impact
- I spent more time in my yard
- I spent more time in public parks and natural areas
- I shortened time spent in public parks and natural areas
- I avoided going to public natural areas and parks
- I chose to visit a less crowded park / natural area
- I spent time outside of maintained paths, areas I would normally avoid
- I increased the frequency of outdoor activities
- Other

11- [if Other in Q10] Please specify, how the need for social distancing impacted your outdoor activity?

12- [if No in Q4] What was the most important reason you decided not to visit parks or natural areas today? [*NEW*]

- No time
- No access / parks closed
- The weather
- I was sick
- I was anxious / emotionally overwhelmed
- Avoid ticks
- Avoid mosquitoes
- Avoid people, because of COVID-19
- Other

13- [if Other in Q12] Please specify, what the most important reason was you decided not to visit parks or natural areas?

Block: Ticks on self [*SAME AS 2019*]

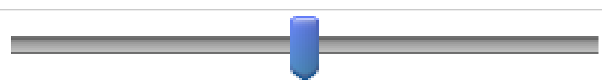
Block: Ticks on others [*SAME AS 2019*]

Block: emotional wellbeing [*NEW SECTION*]

1- Given the current situation with COVID-19, we would like to know how it affects your wellbeing. We greatly value your participation. Would you be willing to complete two additional questions?

- No [end survey]
- Yes

2- How did you feel today?

	0	1	2	3	4	5	6	7	8	9	10
0 = bad (anxious, stressed), 10 = awesome (happy, calm) ()											

3- How much do the impacts of COVID-19 influence your response?

- A lot
- Somewhat
- Neutral
- A little
- Not at all

For information on managing mental health, click here. [insert hyperlink: https://www.cdc.gov/coronavirus/2019-ncov/daily-life-coping/managing-stress-anxiety.html]

Supplementary Methods

Pre-processing independent variables.

We constructed two types of models:

a- Generalized Linear Models (GLMs) to analyze temporal changes in activity patterns as daily proportions of the activities reported by users during the spring and summer. We used these models to compare patterns in 2020 versus 2019, and to analyze the effect of COVID-19 cases and mitigation efforts during 2020.

b- Generalized Linear Mixed Models (GLMMs) to analyze individual self-reported responses to the impact of COVID-19 on activity patterns, emotional well-being, and tick encounters.

The variables included in the models analyzing activity patterns as daily proportions of activities reported by users (a), were summarized as a daily variable for active users on a given day (i.e., only users that submitted a daily survey and provided information on their activities and tick encounters that day). All proportions were estimated if the number of active users were 5 or more.

- ***Mean daily temperature.*** This variable was estimated as the median of the mean daily temperature (°C) experienced by active users on a given day. The mean daily temperature (°C) was obtained from PRISM climate dataset[1] at a county level and assigned to active users according to their place of residence.
- ***Proportion of users on a rainy day.*** This variable was estimated as the proportion of active users experiencing a rainy day. Daily precipitation values (mm) were obtained from PRISM climate dataset[1] at a county level and assigned to active users according to their county of residence. Using precipitation values we created a binary variable defining a rainy day if precipitation was > 2.54 mm. This threshold was decided based on the amount of rain that could have an impact on outdoor activities. Although precipitation > 1 mm would be considered a rainy day[2], this amount would slightly wet the surface[3]. By comparison, precipitation of at least 2.54 mm corresponds to light rain for 30–45 min, moderate rain for 10 min, or heavy rain for 5 min[3]. This variable doesn't take into account geographic variability within the county and during the day.

  Proportion of users in rural, small or medium metro, or large metro counties. These three variables were estimated for each day considering only active users and their county of residence. Rural-urban classification of counties were obtained from the National Center of Health Statistics[4]. Because they are mutually exclusive variables, we estimated the Spearman correlation matrix to decide which combination to include in the model to avoid multicollinearity issues. According to the results, the proportion of users living in small or medium metro county was highly correlated to the percent of users living in rural (Spearman ρ= -0.84, *P*-value <.001) or large metro counties (Spearman ρ= -0.54, *P*-value <.001), thus it was not included as a variable. The proportion of users living in rural vs. large metro counties showed low correlation coefficients (Spearman ρ= -0.004, *P*-value = 0.96). For those models considering individual responses (b), urbanization level was considered as a categorical variable (urbanicity, 3 levels: living in a rural county, living in a small/medium metropolitan county, living in a large county) for each user.

- ***Proportion of females.*** This variable was estimated based on the gender (Female or Male) reported by active users at a given day. A third option (i.e., "other/prefer not to say") was excluded given its very low proportion. Given that the variables proportion of females and proportion of males were mutually exclusive, we kept the former in our models. For those models considering individual responses (b), we considered gender as a categorical variable (3 levels) for each user.

- ***Median age.*** This variable was estimated as the median of the reported age by active users on a given day.
- ***Statewide stay-at-home orders.*** This is a binary variable that determines the "before" and "after" periods determined by stay-at-home orders issued by the different States in the Northeast and Midwest in response to COVID-19 in March and April[5,6]. Given that states have different dates for which the orders were issued (from March 21st to April 2nd), we classified all surveys completed between February 1st and March 31st as "before" and those completed between April 1st and July 31st as "after".
- ***Shelter-in-Place Index.*** This variable was estimated at a county level and represents the change in the proportion of people staying at home compared to the baseline value (between February 6th and February 12th, 2020). For all counties in the Northeast and Midwest, we estimated the Shelter-in-Place Index for each day during the study period, using social distancing metrics provided by SafeGraph COVID-19 Data Consortium[7], which are derived from a panel of GPS pings from anonymous mobile devices. The "stay at home" behavior was determined by defining the "home" as the most common nighttime location in recent months identified to a precision of about 100 square meters from a given anonymous device[7]. The data was aggregated and provided by SafeGraph at a census block level. We used this data to derive the Shelter-in-Place Index by first calculating the proportion of the population staying at home as the number of devices at "home" divided by the total number of devices at a county level, and then subtracting the baseline value (0 means no difference from the baseline, while positive values means an increase of people staying at home and negative values means increased mobility)[8]. The variable included in the model was the daily median value of Shelter-in-Place Index from counties where active users lived. For those models considering individual responses (b), the Shelter-in-Place Index for a user on a given day, corresponds to the value estimated for their county of residence. The variable was rescaled before being included in the model to represent an 10% increase instead of a unit of increase.
- ***Number of COVID-19 cases per 100k county and nationwide.*** The number of COVID-19 per day were estimated as the median number of cases nationwide (per 100k) and the number of cases per county (per 100k) estimated as the median number of cases where the active users were located. The data was obtained from the CDC National Environmental Public Health Tracking Network, as the rolling average over 7 days[6]. For those models considering individual responses (b), the number of COVID-19 cases for a user on a given day, corresponds to the value estimated for their county of residence.

Model assessment.

For all models, we assessed multicollinearity using the Variance Inflation Factor (VIF) and corroborated that all VIF<4[9]. For the GLMs assessing daily variation in the proportion of outdoor activities, we assessed temporal autocorrelation of residuals using the autocorrelation function (ACF) in R. We conducted graphical analysis of residuals for all models to evaluate misspecifications. GLMs were evaluated using standard residual plots, while GLMM residual plots were created using a simulation-based approach, similar to the Bayesian p-value or the parametric bootstrap, which transforms the residuals to a standardized scale and can be interpreted as standard residual plots[10], implemented in the DHARMa package in R. Classification performance was analyzed for models with binary outcomes using the ROC curve and the Area Under the Curve (AUC). To assess model sensitivity and specificity, we employed an optimal threshold value that minimized the sum of error frequencies[11]. This value was obtained by finding the maximum sum of sensitivity and specificity for all threshold values t (sens(t)+spec(t)) using the pROC R-package[12].

To compare nested models with different specifications for the same dataset (e.g., evaluating interactions, including quadratic terms, etc.), we used the Akaike Information Criterion (AIC) and the Log-likelihood ratio test using the R package car[13]. If comparing two models, AIC would be lower for the model with the highest likelihood and models within 2 delta AIC are considered equivalent[14]. We also estimated the McFadden's pseudo R^2 for GLMs in R as 1- logLik(fitted.model) / logLik(null.model), which indicates the level of improvement from the null to fitted model. This was used as an additional metric to compare models estimated for the same dataset and outcome, and evaluate model fit[15]. If comparing two models, McFadden's pseudo R^2 would be higher for the model with the highest likelihood. McFadden's pseudo R^2 values between 0.2 and 0.4 are considered to be indicative of good model fits (comparable to R^2 values 0.7–0.9 in linear models)[16].

References

1. Oregon State University. PRISM Climate Group. Accessed December 23, 2020. https://prism.oregonstate.edu/

2. USGS. US Rain/Dry Days Readme | Early Warning and Environmental Monitoring Program. Accessed December 28, 2020. https://earlywarning.usgs.gov/usraindry/rdreadme.php

3. Weather Insurance Agency. Rain Guidelines. Accessed December 28, 2020. https://weatherins.com/rain-guidelines/

4. National Center for Health Statistics, CDC. Urban Rural Classification Scheme for Counties. Published December 2, 2019. Accessed December 28, 2020. https://www.cdc.gov/nchs/data_access/urban_rural.htm

5. Littler Mendelson P.C. Stay on Top of "Stay At Home" – A List of Statewide Orders. Published March 24, 2020. Accessed December 28, 2020. https://www.littler.com/publication-press/publication/stay-top-stay-home-list-statewide

6. Center for Disease Control. National Environmental Public Health Tracking Network Query Tool. Accessed December 28, 2020. https://ephtracking.cdc.gov/DataExplorer/

7. SafeGraph. Shelter in Place Index: The Impact of Coronavirus on Human Movement. SafeGraph. Accessed December 28, 2020. https://safegraph.com/data-examples/covid19-shelter-in-place/

8. SafeGraph. Data Analysis Methodology for SafeGraph's Stay-At-Home Index. Accessed December 28, 2020. https://docs.google.com/document/d/1k_9LGQn95P5gHsSeuBdzgtEWGGCmzXdcOkcphWi0Cas/edit

9. Zuur AF, Ieno EN, Elphick CS. A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution*. 2010;1(1):3-14. doi:https://doi.org/10.1111/j.2041-210X.2009.00001.x

10. DHARMa: residual diagnostics for hierarchical (multi-level/mixed) regression models. Accessed December 23, 2020. https://cran.r-project.org/web/packages/DHARMa/vignettes/DHARMa.html

11. Schisterman EF, Perkins NJ, Liu A, Bondell H. Optimal Cut-point and Its Corresponding Youden Index to Discriminate Individuals Using Pooled Blood Samples. *Epidemiology*. 2005;16(1):73-81. doi:10.1097/01.ede.0000147512.81966.ba

12. Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*. 2011;12(1):77. doi:10.1186/1471-2105-12-77

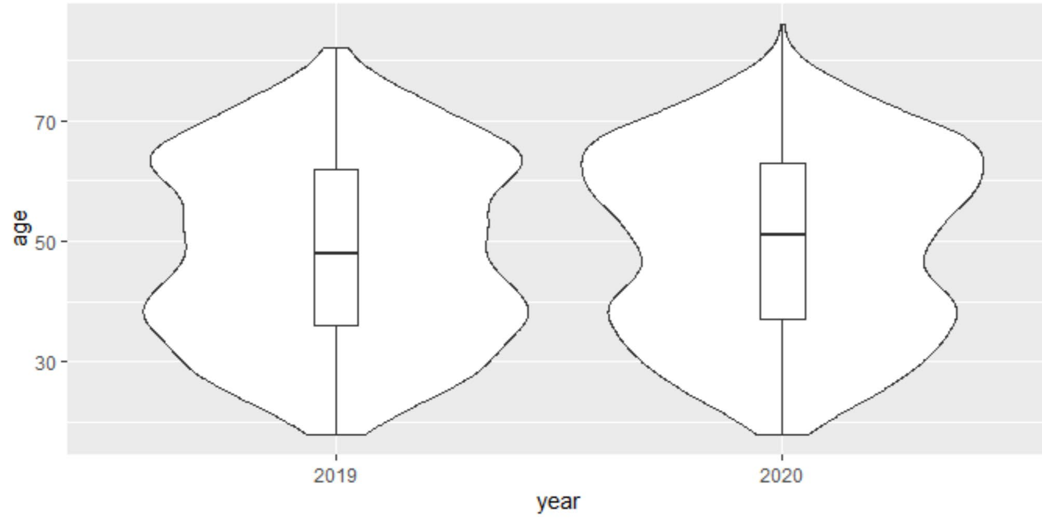
13. Fox J, Weisberg S. *An R Companion to Applied Regression*. Third. Sage; 2019. https://socialsciences.mcmaster.ca/jfox/Books/Companion/

14. Burnham KP, Anderson DR. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach.* Springer-Verlag; 2002.

15. Freese J, Long S. *Regression Models for Categorical Dependent Variables Using Stata.* College Station: Stata Press; 2006.

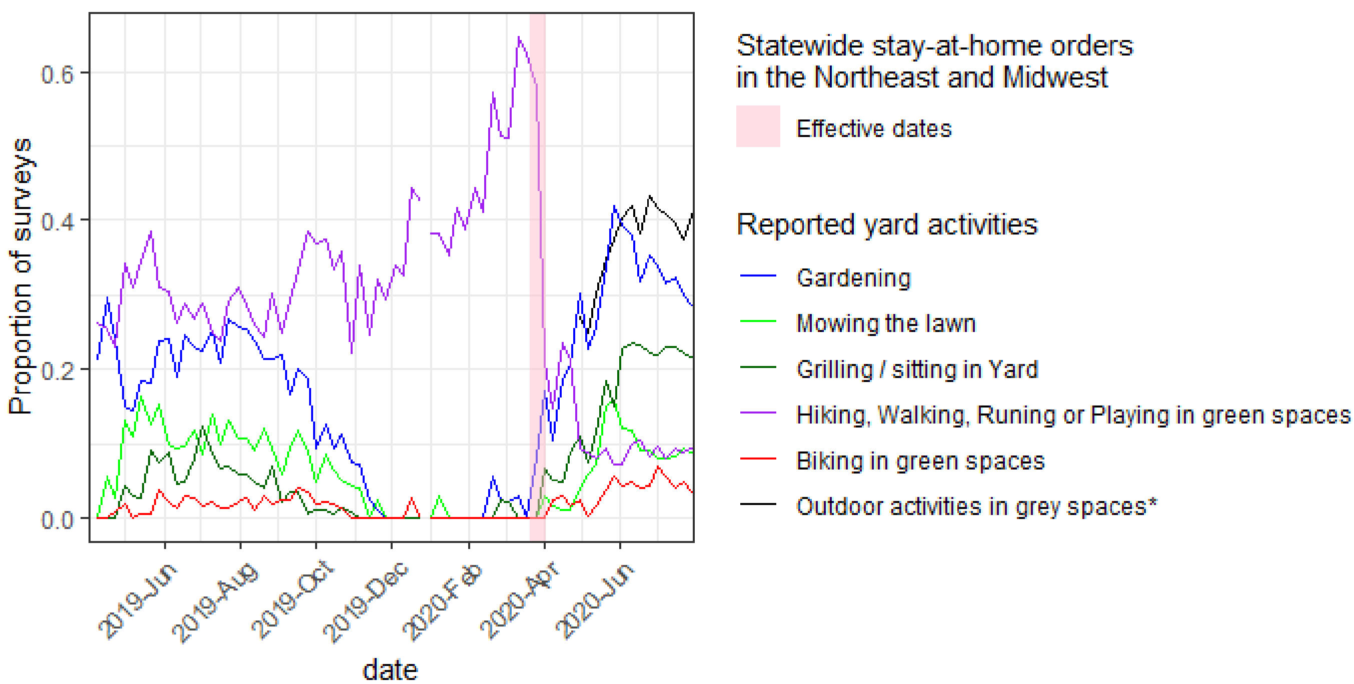
16. Louviere J, Hensher D, Swait J, Adamowicz W. Stated Choice Methods: Analysis and Applications. In: *Stated Choice Methods: Analysis and Applications.* Cambridge University Press; 2000:55. 10.1017/CBO9780511753831

Supplementary Figure 1: Age distribution of users in 2019 and 2020.



Supplementary Figure 2: Unadjusted proportion of surveys reporting outdoor activities per week, by type.

The effective dates for statewide stay-at-home orders are indicated for the weeks in light red. *Outdoor activities in grey spaces refer to any recreational activities not done in parks or green spaces (for example, jogging on sidewalks).



Supplementary Table 1: Outdoor Activity Model (2020 vs. 2019)

Generalized linear model for the proportion of The Tick App users reporting any type of outdoor activity during April 1–July 31st, 2020 versus the same period in 2019. The model examined whether the daily proportion of users reporting outdoor activities varied between years, accounting for monthly variation, temperature and precipitation, urbanicity, gender, and age of reporting users. Models were run for all users and for returning users (those who completed surveys for both years). We used a binomial regression with a logit link function, with the effect of each independent variable expressed as odds ratios. Reference categories for Year and Month were 2019 and April, respectively.

Users	Variables	Odds Ratio	95% CI (Lower, Upper)		P-value
All users	Intercept	4.29	1.59	11.68	<.001*
	Year				
	2019	1			
	2020	2.65	2.33	3.00	<.001*
	Month				
	April	1			
	May	0.97	0.77	1.22	.78
	June	0.77	0.60	1.00	.05
	July	0.70	0.53	0.93	.01*
	Mean daily temperature	1.09	1.04	1.14	<.001*
	Mean daily temperature^2 (quad. term)	9.98E-01	9.97E-01	9.99E-01	.02*
	Proportion of users on a rainy day	0.41	0.34	0.50	<.001*
	Proportion of users in rural counties	0.63	0.25	1.56	.32
	Proportion of users in large metro counties	0.31	0.11	0.9	.03*
	Proportion of females	1.05	0.52	2.11	.89
	Median age	0.99	0.97	1	.05*
Returning users	Intercept	4.12	1.07	16.53	.04*
	Year				
	2019	1			
	2020	1.96	1.49	2.58	<.001*
	Month				
	April	1			
	May	1.38	0.95	1.98	.09
	June	0.84	0.55	1.28	.42
	July	0.65	0.40	1.07	.09
	Mean daily temperature	1.04	1.01	1.06	<.001*
	Proportion of users on a rainy day	0.53	0.38	0.73	<.001*
	Proportion of users in rural counties	2.11	0.59	7.57	.25
	Proportion of users in large metro counties	0.50	0.13	2.00	.33
	Proportion of females	0.45	0.18	1.15	.10
	Median age	0.99	0.97	1.02	.56

**P-value* < .05

For all users, adding a quadratic term for temperature improved the fit of the model compared to the model without ($AIC_{QuadraticModel}$=1292, $AIC_{ReducedModel}$=1296; log-likelihood ratio test (d.f.=1), *P* = 0.02), and the former showed a significant reduction in deviance compared to the null model (log-likelihood ratio test (d.f.=10), *P* <.001; AIC_{null}=1969). The McFadden Pseudo-R^2 for the model with the quadratic term was 0.35, while the reduced model had a McFadden Pseudo-R^2 of 0.36.

For returning users, adding a quadratic term for temperature did not improve the fit of the model ($AIC_{QuadraticModel}$=860, $AIC_{ReducedModel}$=859; log-likelihood ratio test (d.f.=1), *P* = 0.64), and both models had a McFadden Pseudo-R^2= 0.18. Thus, the reduced model is presented and showed a significant reduction in deviance compared to the null model (log-likelihood ratio test (d.f.=10), *P* <.001; AIC_{null}=1018).

Upon evaluating the interaction between year and month, the interaction term was not significant for either model (all users and returning users).

Supplementary Table 2: Peridomestic Activity Model (2020 vs. 2019)

Generalized linear model for the proportion of The Tick App users reporting peridomestic activities during April 1–July 31st, 2020 versus the same period in 2019. The model examined whether the daily proportion of users reporting peridomestic outdoor activities varied between years, accounting for monthly variation, temperature and precipitation, urbanicity, gender, and age of reporting users. Models were run for all users and for returning users (those who completed surveys for both years). We used a binomial regression with a logit link function, with the effect of each independent variable expressed as odds ratios. Reference categories for Year and Month were 2019 and April, respectively.

Users	Variables	Odds Ratio	95% CI (Lower, Upper)		P-value
All users	**Intercept**	0.43	0.20	0.92	.03*
	Year				
	2019	1			
	2020	2.26	2.05	2.51	<.001*
	Month				
	April	1			
	May	1.07	0.89	1.28	.47
	June	0.92	0.76	1.12	.42
	July	0.85	0.68	1.07	.16
	Mean daily temperature	1.03	1.02	1.04	<.001*
	Proportion of users on a rainy day	0.55	0.47	0.64	<.001*
	Proportion of users in rural counties	1.00	0.50	2.04	.99
	Proportion of users in large metro counties	0.23	0.10	0.52	<.001*
	Proportion of females	2.44	1.41	4.24	<.001*
	Median age	1.00	0.99	1.01	.62
Returning users	Intercept	0.64	0.20	2.03	.45
	Year				
	2019	1			
	2020	2.39	1.93	2.97	<.001*
	Month				
	April	1			
	May	1.18	0.89	1.58	.25
	June	0.84	0.59	1.19	.32
	July	0.67	0.45	1.01	.05*
	Mean daily temperature	1.04	1.02	1.06	<.001*
	Proportion of users on a rainy day	0.56	0.43	0.75	<.001*
	Proportion of users in rural counties	5.28	1.91	14.67	<.001*
	Proportion of users in large metro counties	1.26	0.42	3.78	.68
	Proportion of females	0.78	0.37	1.65	.51
	Median age	0.98	0.97	1.00	.10

*P-value < 0.05

For all users, adding a quadratic term for temperature did not improve the fit of the model compared to the model without ($AIC_{QuadraticModel}$=1382, $AIC_{ReducedModel}$=1383; log-likelihood ratio test (d.f.=1), P = 0.11), and both models had a McFadden Pseudo-R^2= 0.37. Thus, we present the reduced model. This model showed a significant reduction in deviance compared to the null model (log-likelihood ratio test (d.f.=10), P <.001; AIC_{null}=2149).

For returning users, adding a quadratic term for temperature did not improve the fit of the model ($AIC_{QuadraticModel}$=1013, $AIC_{ReducedModel}$=1012, log-likelihood ratio test (d.f.=1), P = 0.64), and both models had a McFadden Pseudo-R^2= 0.24. Thus, we present the reduced model. This model showed a significant reduction in deviance compared to the null model (log-likelihood ratio test (d.f.=10), P <.001; AIC_{null}=1300).

Upon evaluating the interaction between year and month for all users, the model with the interaction term had a significant reduction in deviance compared to the model without ($AIC_{Interaction}$=1377, $AIC_{ReducedModel}$=1393; log-likelihood ratio test (d.f.=3), P = 0.01; McFadden Pseudo-$R^2_{interaction}$ = 0.37).

However, for returning users, no differences were observed ($AIC_{Interaction}$=1012, $AIC_{ReducedModel}$=1012.2; log-likelihood ratio test (d.f.=3), *P* = 11; McFadden Pseudo-$R^2_{Interaction}$ = 0.24). Results of the interactions are shown on eTable 3.

Supplementary Table 3: Interactive Effects of Month and Year on Peridomestic Activities (2020 vs. 2019)

Comparative results of the interactive effects of month and year on peridomestic activities for all users. April 2019 is the reference category for the interaction.

Year / **Month**	2019 OR (95% CI)	2020 OR (95% CI)
April	1	**1.82 (1.20, 2.79)***
May	0.77 (0.55, 1.07)	**1.26 (1.01, 1.55)***
June	0.76 (0.53, 1.08)	0.97 (0.77, 1.22)
July	0.74 (0.51, 1.08)	0.87 (0.67, 1.14)

*P-value < 0.05

Supplementary Table 4: Recreational Activity Model (2020 vs. 2019)

Generalized linear model (GLM) for the proportion of The Tick App users reporting recreational activities during April 1 – July 31st, 2020 versus the same period in 2019. The model examined whether the daily proportion of users reporting (non-peridomestic) outdoor recreational activities varied between years, accounting for monthly variation, temperature and precipitation, urbanicity, gender, and age of reporting users. Models were run for all users and for returning users (those who completed surveys both in 2019 and 2020). We used a binomial regression with a logit link function, with the effect of each independent variable expressed as odds ratios. Reference categories for Year and Month were 2019 and April, respectively.

Users	Variables	Odds Ratio	95% CI (Lower, Upper)		P-value
All users	Intercept	1.79	0.78	4.11	.17
	Year				
	2019	1			
	2020	0.26	0.23	0.30	<.001*
	Month				
	April	1			
	May	0.78	0.63	0.96	.02*
	June	0.58	0.47	0.73	<.001*
	July	0.64	0.49	0.83	<.001*
	Mean daily temperature	1.00	0.99	1.02	.47
	Proportion of users on a rainy day	0.70	0.58	0.58	<.001*
	Proportion of users in rural counties	3.30	1.45	7.51	<.001*
	Proportion of users in large metro counties	1.86	0.71	4.86	.21
	Proportion of females	0.97	0.50	1.86	.92
	Median age	0.98	0.97	0.99	<.001*
Returning users	Intercept	0.60	0.17	2.11	.43
	Year				
	2019	1			
	2020	0.15	0.11	0.19	<.001*
	Month				
	April	1			
	May	0.83	0.58	1.19	.30
	June	0.67	0.44	1.02	.06
	July	0.52	0.32	0.85	.01*
	Median Daily Temperature	1.01	0.99	1.03	.39
	Daily Proportion of Users on a Rainy Day	0.84	0.61	1.14	.26
	Proportion of users in rural counties	1.31	0.39	4.40	.67
	Proportion of users in large metro counties	1.57	0.42	5.88	.51
	Proportion of females	0.50	0.20	1.20	.12
	Median Age	1.01	0.99	1.03	.26

*P-value < 0.05

For all users, adding a quadratic term for temperature did not improve the fit of the model compared to the model without ($AIC_{QuadraticModel}$=1237, $AIC_{ReducedModel}$=1236; log-likelihood ratio test (d.f.=1), P = 0.32), and both models had a McFadden Pseudo-R^2= 0.51. Thus, the reduced model is presented, and showed a significant reduction in deviance compared to the null model (AIC_{null}=2469.7; log-likelihood ratio test (d.f.=10), P <.001).

For the return users, adding a quadratic term for temperature did not improve the fit of the model either ($AIC_{QuadraticModel}$=835, $AIC_{ReducedModel}$=834, ; log-likelihood ratio test (d.f.=1), P = 0.30), and both models had a McFadden Pseudo-R^2= 0.43. Thus, the reduced model is presented and showed a significant reduction in deviance compared to the null model (AIC_{null}=1433; log-likelihood ratio test (d.f.=10), P <.001).

Upon evaluating the interaction between year and month for all users, the model with the interaction term had a significant reduction in deviance compared to the reduced model without the interaction ($AIC_{Interaction}$=1214, $AIC_{ReducedModel}$=1236, log-likelihood ratio test (d.f.=3), P < 0.01; McFadden Pseudo-

$R^2_{Interaction}$ = 0.52). The same was observed when considering the returning users ($AIC_{Interaction}$=804, $AIC_{ReducedModel}$=834; log-likelihood ratio test (d.f.=3), *P* < .001; McFadden Pseudo-$R^2_{Interaction}$ = 0.45). Results of the interactions are shown on eTable 5.

Supplementary Table 5: Interactive Effects of Month and Year on Recreational Activity (2020 vs. 2019)

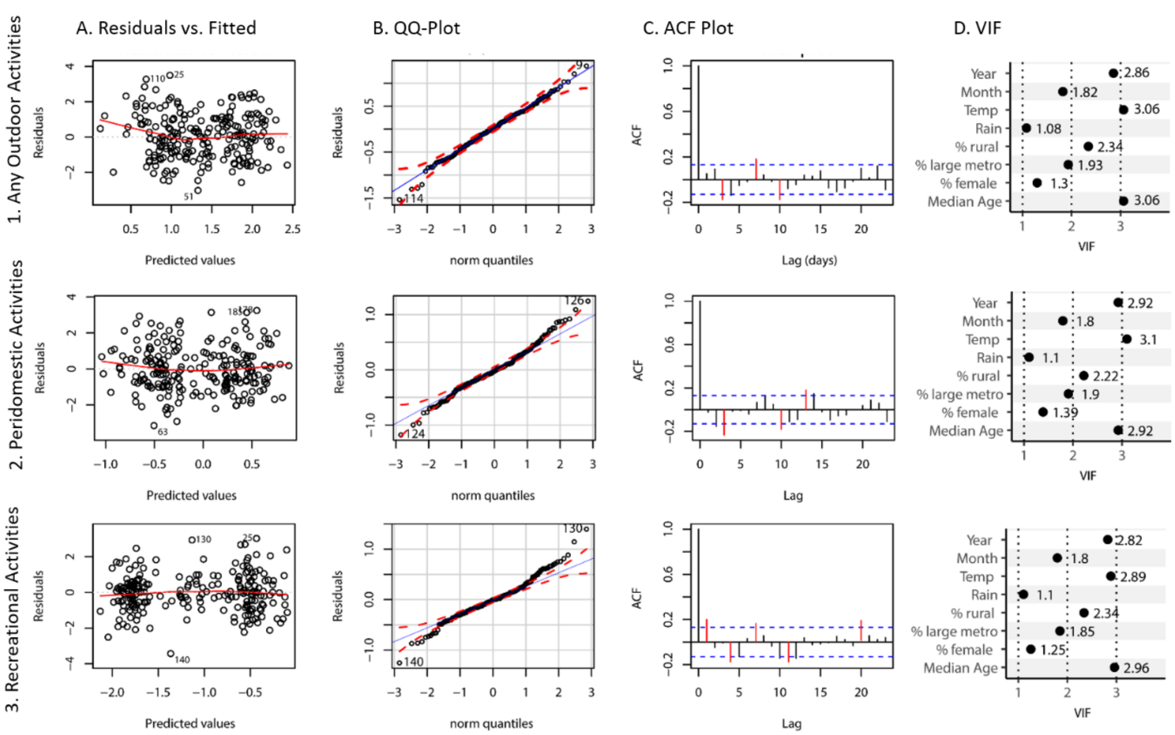
Comparative results of the interactive effects of month and year on recreational activity for all users and returning users. April 2019 is the reference category for the interaction.

	All users		Returning users	
Year / Month	2019 OR (95% CI)	2020 OR (95% CI)	2019 OR (95% CI)	2020 OR (95% CI)
April	1	0.77 (0.48, 1.23)	1	1.40 (0.62, 3.27)
May	**1.54 (1.08, 2.20)***	**0.53 (0.40, 0.68)***	**3.93 (1.99, 8.13)***	**0.45 (0.29, 0.70)***
June	1.20 (0.83, 1.75)	**0.39 (0.30, 0.51)***	**3.16 (1.59, 6.56)***	**0.33 (0.19, 0.54)***
July	1.18 (0.79, 1.77)	**0.43 (0.31, 0.59)***	**2.58 (1.22, 5.69)***	**0.23 (0.13, 0.41)***

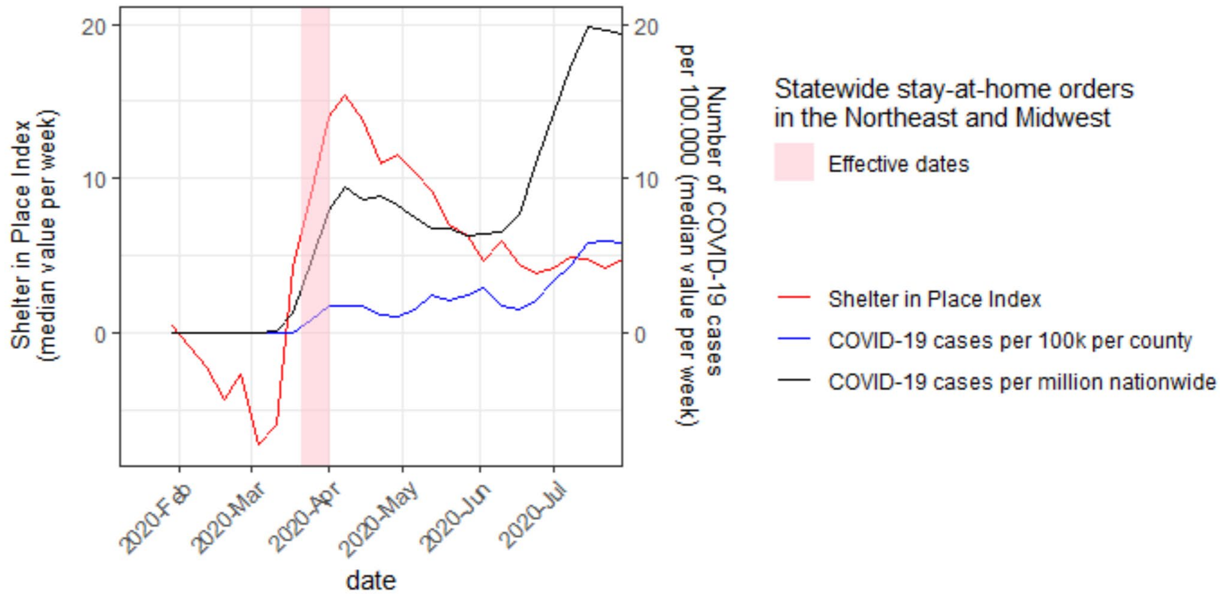
*P-value < 0.05

Supplementary Figure 3: GLMs Comparing Activity Patterns in 2020 vs. 2019 – Diagnostic Plots

The Residuals vs. Fitted (Panel A) and QQ-Plots (Panel B) did not detect any significant issues with the distribution of residuals, although small deviations from normality are observed for the recreational activity model (Plots A3, B3). We assessed temporal autocorrelation of residuals using the autocorrelation function (ACF) (Panel C). Although some small temporal autocorrelation was observed at 3-4 days, 7 days and 10-11 days (shown by red line in ACF plot) (Plots C1–C3), no major patterns were observed. VIF values were <4 for all models, indicating no multicollinearity issues (Panel D).



Supplementary Figure 4: Median Shelter-in-Place Index and COVID-19 cases per week in Northeast and Midwest Counties. (Left y-axis): For a given week, we show the median value of the Shelter-in-Place Index. (Right y-axis): For a given week, we show the number of COVID-19 cases per 100,000 using the median value per week. County-level numbers are shown in blue while nationwide numbers are shown in black. Weeks shaded in light red indicate the effective dates for stay-at-home orders in Northeast and Midwest counties.



Supplementary Table 6: Outdoor Activity Model (2020)

Generalized linear model (GLM) for the proportion of The Tick App users reporting outdoor activity during February 1st – July 31st, 2020. The model examined whether daily outdoor activity was higher among The Tick App users before versus after stay-at-home orders were put into place (see Figure 2), accounting for temperature and precipitation, urbanicity, gender, and age of reporting users per day. The Month variable was dropped because of multicollinearity with climate variables (VIF>4). We used a binomial regression with a logit link function, with the effect of each independent variable expressed as odds ratios. The reference category for Statewide stay-at-home orders was “Before” stay-at-home order implementation (from February 1st – March 31st, 2020).

Variables	**Odds Ratio**	**95% CI (Lower, Upper)**		**P-value**
Intercept	23.90	3.39	171.14	<.001*
Statewide stay-at-home orders				
Before	1			
After	2.01	1.21	3.29	.01*
Mean daily temperature	1.02	1.00	1.04	.07
Proportion of users on a rainy day	0.30	0.22	0.40	<.001*
Proportion of users in rural counties	0.70	0.21	2.33	.42
Proportion of users in large metro counties	0.51	0.11	2.24	.18
Proportion of females	1.25	0.45	3.51	.54
Median age	0.97	0.95	1.00	.02*
Shelter in Place Index	0.69	0.55	0.88	.01*
Number of COVID-19 cases per 100,000 nationwide	0.99	0.98	1.00	.06

*P-value < 0.05

The model including the number of COVID-19 per 100k per county (in which active users were located) and the model including the number of cases per 100k nationwide had the same model fit ($AIC_{PerCountyModel}$ = 709, $AIC_{NationwideModel}$ = 709; McFadden Pseudo-R^2= 0.15). We therefore present the nationwide model.

For the nationwide model, adding a quadratic term for the temperature did not improved the fit of the model compared to the model without ($AIC_{QuadraticModel}$= 709, $AIC_{ReducedModel}$ = 709; log-likelihood ratio test (d.f.=1); P = 0.20) and both models had a McFadden Pseudo-R^2= 0.15. Thus, we present the reduced model. This model showed a significant reduction in deviance compared to the null model (log-likelihood ratio test (d.f.=9), P <.001; AIC_{null}=816).

Supplementary Table 7: Peridomestic Activity Model (2020)

Generalized linear model (GLM) for the proportion of The Tick App users reporting peridomestic activities during February 1st – July 31st, 2020. The model examined whether daily outdoor activity was higher among The Tick App users before versus after stay-at-home orders were put into place (Main Text, Figure 2), accounting for temperature and precipitation, urbanicity, gender, and age of reporting users per day. The Month variable was dropped because of multicollinearity with climate variables (VIF>4). We used a binomial regression with a logit link function, with the effect of each independent variable expressed as odds ratios. The reference category for Statewide stay-at-home orders was "Before" stay-at-home order implementation (from February 1st – March 31st, 2020).

Variables	**Odds Ratio**	**95% CI (Lower, Upper)**		**P-value**
Intercept	0.51	0.11	2.28	.38
Statewide stay-at-home orders				
Before	1			
After	11.25	6.73	19.55	<.001*
Mean daily temperature	1.01	0.99	1.02	0.20
Proportion of users on a rainy day	0.57	0.45	0.71	<.001*
Proportion of users in rural counties	0.32	0.13	0.82	.02*
Proportion of users in large metro counties	0.25	0.08	0.74	.01*
Proportion of females	2.27	1.04	4.92	.04*
Median age	0.99	0.97	1.00	.16
Shelter in Place Index	0.68	0.56	0.82	<.001*
Number of COVID-19 cases per 100k per county	0.96	0.93	0.99	.03*

*P-value < 0.05

The model including the number of COVID-19 per 100k per county (in which active users were located) and the model including the number of cases per 100k nationwide had the same model fit (for both models AIC = 803, McFadden Pseudo-R^2= 0.25).

For the county model, adding a quadratic term for the temperature did not improve the fit of the model compared to the model without ($AIC_{QuadraticModel}$= 803, $AIC_{ReducedcModel}$= 803; log-likelihood ratio test (d.f.=1), P = 0.12; McFadden Pseudo-R^2= 0.26). Thus, we present the reduced model. This model showed a significant reduction in deviance compared to the null model (log-likelihood ratio test (d.f.=9), P <.001; AIC_{null}=1054).

Supplementary Table 8: Recreational Activity Model (2020)

Generalized linear model (GLM) for the proportion of Tick App users reporting outdoor activities during February 1st – June 31st, 2020. The model examined whether daily outdoor activity was higher among The Tick App users before versus after stay-at-home orders were put into place (Main Text, Figure 2), accounting for temperature and precipitation, urbanicity, gender, and age of reporting users per day. The Month variable was dropped because of multicollinearity with climatic variables (VIF>4). We used a binomial regression with a logit link function, with the effect of each independent variable expressed as odds ratios. The reference category for Statewide stay-at-home orders was "Before" stay-at-home order implementation (from February 1st – March 31st, 2020).

Variables	**Odds Ratio**	**95% CI (Lower, Upper)**		**P-value**
Intercept	5.25	0.81	34.18	.08
Statewide stay-at-home orders				
Before	1			
After	0.19	0.12	0.30	<.001*
Mean daily temperature	0.94	0.90	0.98	.01*
Mean daily temperature^2 (quad. term)	1.002	1.0001	1.003	.04*
Proportion of users on a rainy day	0.73	0.54	1.00	.05~
Proportion of users in rural counties	2.17	0.70	6.79	.18
Proportion of users in large metro counties	0.44	0.11	1.76	.24
Proportion of females	0.79	0.29	2.16	.65
Median age	0.98	0.95	1.00	.07
Shelter in Place Index	1.08	0.86	1.35	.50
Number of COVID-19 cases per 100k nationwide	1.00	0.99	1.03	.22

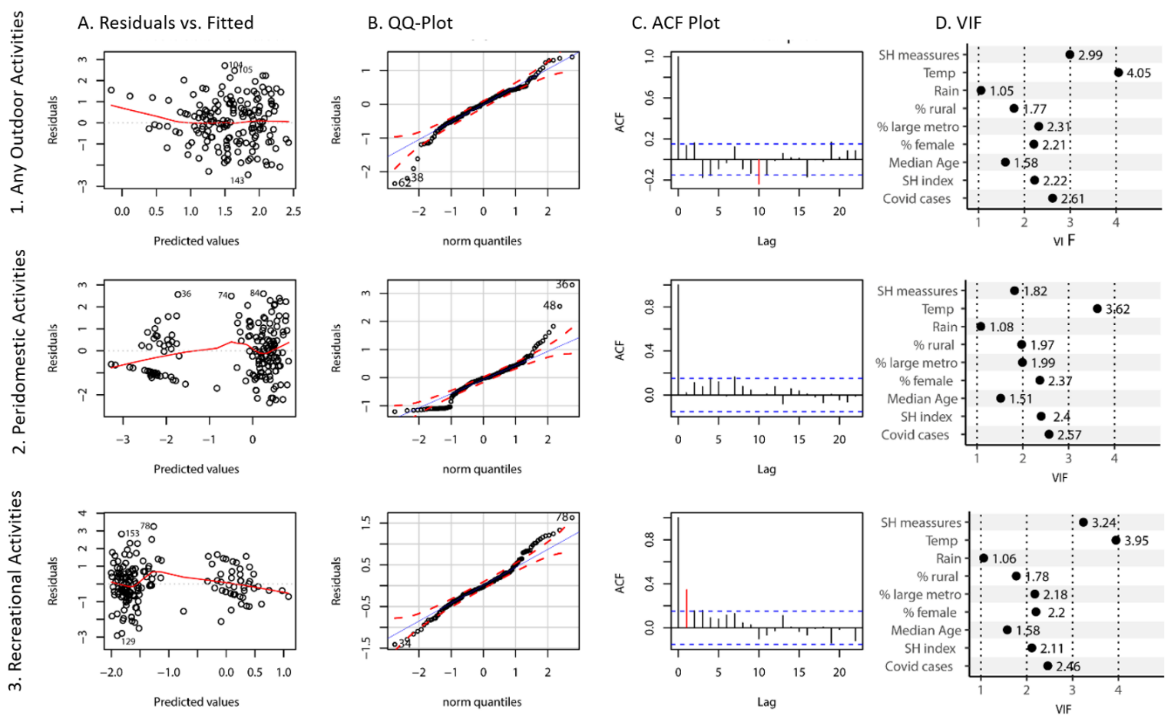
*P-value < 0.05 ~P-value = 0.05

The model including the number of COVID-19 per 100k nationwide had a better fit than the model including the number of cases per 100k per county (in which active users were located) ($AIC_{NationwideModel}$ = 734, $AIC_{PerCountyModel}$ = 736), and both models had a McFadden Pseudo-R^2= 0.20. Thus, we present the nationwide model. Nonetheless, both models had qualitatively similar results.

For the nationwide model, adding a quadratic term for temperature improve the fit of the model compared to the model without ($AIC_{QuadraticModel}$=731; log-likelihood ratio test (d.f.=1), *P* = 0.04). The McFadden Pseudo-R^2 for the model with the quadratic term was 0.21. Thus, we present the model with the quadratic term. This showed a significant reduction in deviance compared to the null model (log-likelihood ratio test (d.f.=10), *P* <.001; AIC_{null}=898).

Supplementary Figure 5: GLMs Evaluating COVID-19 Impact on Activity Patterns (2020) — Diagnostic Plots.

The Residuals vs. Fitted and QQ-Plots detected small deviations from normality and outliers probably introduced by small sample sizes prior to March 2020, but no major issues were observed (Panels A and B). We assessed temporal autocorrelation of residuals using the autocorrelation function (ACF) (Panel C). Only slight temporal autocorrelation was observed at 10 days for any type of outdoor activities (Plot C1) and 2 days for recreational activities (shown by red line in ACF plot) (Plot C3). VIF values were <4 for all models, indicating no multicollinearity issues (Panel D).



Supplementary Table 9: Self-reported Impacts of COVID-19 on Outdoor and Recreational Activities

Using daily survey data, we show the number and percent of surveys that reported the impact of COVID-19 mitigation measures on outdoor activities. We also show the number and percent of surveys that report reasons for not doing outdoor or recreational activities in a park or natural area.

Today, has the need for social distancing or other COVID-19 measures impacted your outdoor activity?	**Number of surveys (%)**
No impact	5512 (70.0%)
Spent more time in yard	1202 (15.3%)
Increased the frequency of outdoor activities	553 (7.0%)
Chose to visit a less crowded park or natural area	553 (7.0%)
Spent more time in public parks and natural areas	387 (4.9%)
Avoided going to public parks and natural areas	330 (4.2%)
Shortened the time in public parks and natural areas	155 (2.0%)
Spent time outside of maintained paths, areas I would normally avoid	102 (1.3%)
Other	109 (1.4%)
What was the most important reason you decided not to visit parks or natural areas today?	
No time / other activity	2650 (65.2%)
Weather	640 (15.8%)
Avoid crowds	231 (5.7%)
No particular reason or desire to go to the park	137 (3.4%)
I chose to stay in my yard / do outdoor activities elsewhere / rural property	121 (3.0%)
I was sick / injured / tired or was caregiver	106 (2.6%)
To avoid ticks, mosquitoes or other insects	52 (1.3%)
No access / parks closed	37 (0.9%)
I was anxious / emotionally overwhelmed	32 (0.8%)
Other	56 (1.4%)

Supplementary Table 10: COVID-19 Impact on Recreational Activities Model (2020)

Generalized linear mixed model (GLMM) for the association between self-reported COVID-19 impact (yes/no) reported by The Tick App users in the daily surveys between April 28th – July 31st, 2020. The model examined whether any impact on recreational activities, regardless of the type of impact (positive or negative), was associated to the Shelter-in-Place Index as a proxy for governmental interventions, the number of COVID-19 cases in the county of residence and if there were monthly variations, accounting for urbanicity, gender and age. We used a binomial regression with a logit link function and the user identity as a random variable. The effect of each independent variable is expressed as odds ratios. We rescaled the age variable so the unit change is 5 years.

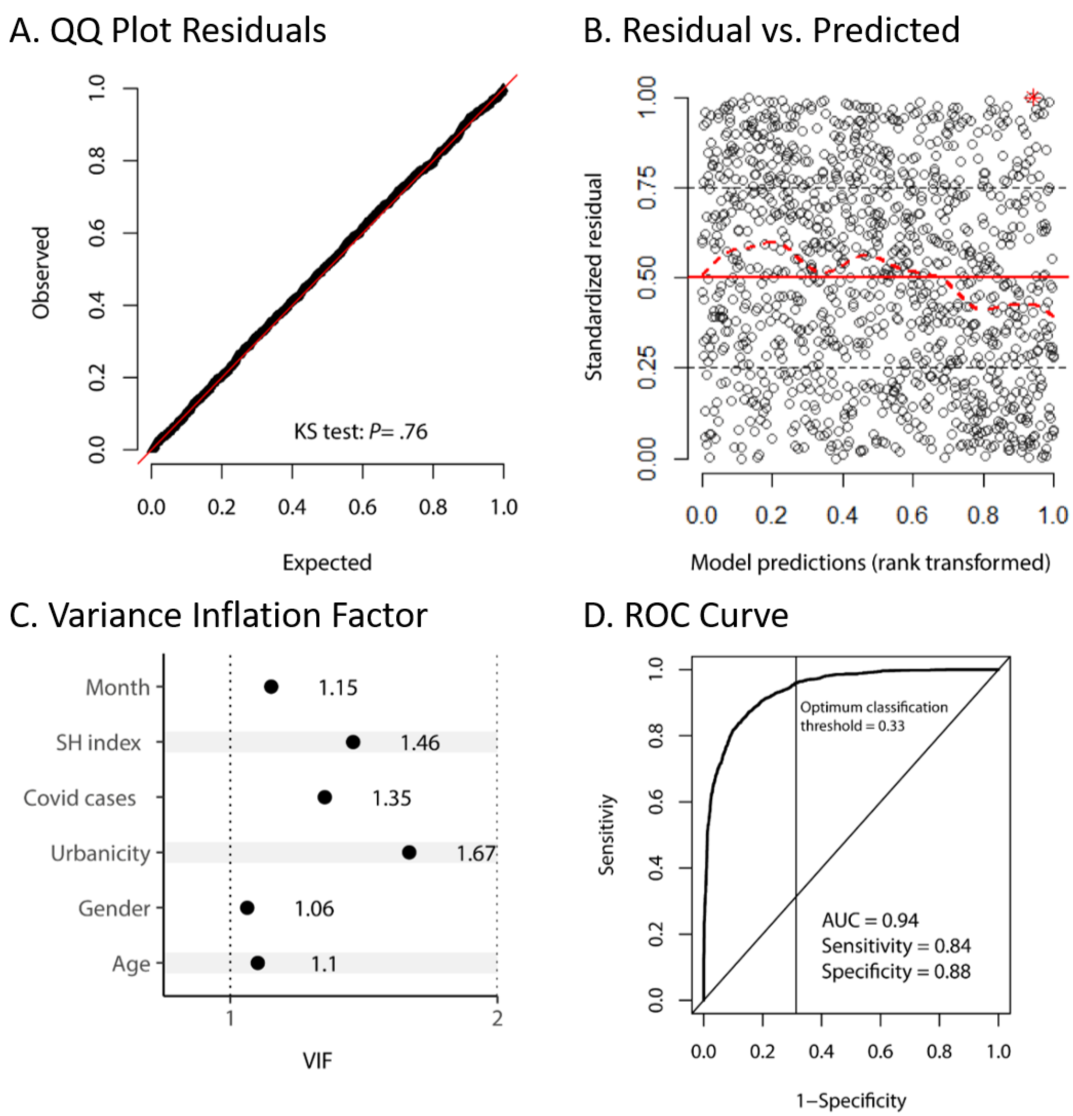
Variables	**Odds Ratio**	**95% CI (Lower, Upper)**		**P-value**
Intercept	0.34	0.15	0.78	.01*
Month				
May	1			
June	0.88	0.66	1.17	.38
July	0.85	0.53	1.38	.52
Shelter in Place Index	1.37	1.09	1.73	.01*
Number of COVID-19 cases per 100k nationwide	0.99	0.97	1.01	.31
Urbanicity				
Lives in a rural county	1			
Lives in a small or medium metro area	1.76	1.12	2.78	.01*
Lives in a metro area	1.69	1.03	2.76	.04*
Gender				
Female	1			
Male	0.31	0.21	0.46	<.001*
Other/prefer not to say	2.32	0.32	16.60	.40
Age	1.04	0.97	1.10	.27

*P-value < 0.05

The GLMM including user ID as a random variable and including the number of COVID-19 nationwide had a similar fit compared to the GLMM model including the number of cases per county (in which active users where located) (for both models, AIC = 6159). The former was also significantly different from the GLM (log-likelihood ratio test(1), $P < .001$; $AIC_{GLM} = 8826$) and significantly different from the null model (log-likelihood ratio test(9), $P < .001$; $AIC_{null} = 6213$).

Supplementary Figure 6: COVID-19 Impact on Recreational Activities Model (2020) — Diagnostic Plots.

The QQ Plot Residuals did not detect significant issues with residuals. The Kolmogorov-Smirnov (KS) test for goodness of fit for testing normality had a value of *P*=.76 (A). The Residual vs. Predicted plot also did not detect significant issues with residuals (B). Variance Inflation Factor (VIF) values were <4 for all models, indicating no multicollinearity issues (C) and classification performance was good with an AUC = 0.94 (D).



Supplementary Table 11: Tick Encounter Model (2020 vs. 2019)

Generalized linear mixed model (GLMM) for the proportion of The Tick App users with a self-reported tick encounter during April 1st – July 31st, 2020 versus the same period in 2019. The model examined whether individual-level tick encounters differed between years or in association with outdoor activities that day, accounting for month, urbanicity, gender, and age of reporting users. We did not consider dates prior to April 1st because of the lower likelihood of encountering a tick due to tick phenology. We used a binomial regression with a logit link function and user ID as a random variable. The effect of each independent variable was expressed as odds ratios. Reference categories for Year, Month, Outdoor Activities, Urbanicity, and Gender were 2019, April, No outdoor activity that day, Lives in a rural county, and Female, respectively.

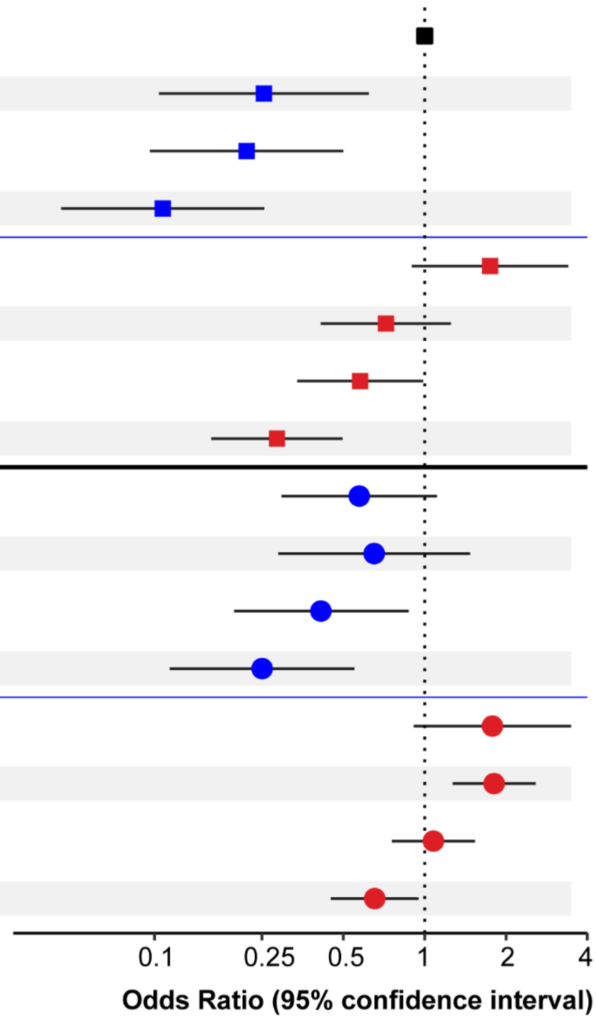
Variables	Odds Ratio	95% CI (Lower, Upper)		P-Value
Intercept	0.05	0.03	0.10	<.001*
Year				
2019	1			
2020	1.36	1.12	1.64	<.001*
Month				
April	1			
May	1.36	1.01	1.83	.04*
June	0.90	0.67	1.21	.49
July	0.50	0.37	0.68	<.001*
Did any outdoor activities that day?				
No	1			
Yes	4.35	3.47	5.44	<.001*
Urbanicity				
Lives in a rural county	1			
Lives in a small or medium metro area	0.52	0.39	0.68	<.001*
Lives in a metro area	0.56	0.41	0.77	<.001*
Gender				
Female	1			
Male	1.23	0.96	1.58	.09
Other/prefer not to say	1.14	0.36	3.65	.83
Age	0.99	0.98	1.00	.06

The GLMM including user ID as a random variable was significantly different from the GLM (AIC_{GLMM}=10046, AIC_{GLM}=11390; log-likelihood ratio test (1), *P* < .001) and significantly different than the null model (AIC_{null} = 10412; log-likelihood ratio test(10), *P* <.001).

The model with the interactions between year and month, and month and doing outdoor activities, showed a significant reduction in the model deviance compared to the reduced model ($AIC_{Interactions}$ = 10040; log-likelihood ratio test(6), *P* = .004). Odds ratios for the pairwise interactions are shown on eFigure 10.

Supplementary Figure 7: Tick Encounter Model (2020 vs. 2019) — Forest plot for the two-way interactions between Year – Month and Outdoor Activity (Yes/No) – Month. The reference categories are April 2019 and reporting "No" outdoor activity that day (OR=1).

Interaction	OR (95% CI)
2019 − No − April	1
2019 − No − May	0.25 (0.10 − 0.62)
2019 − No − June	0.22 (0.10 − 0.50)
2019 − No − July	0.11 (0.05 − 0.25)
2019 − Yes − April	1.75 (0.90 − 3.41)
2019 − Yes − May	0.72 (0.41 − 1.25)
2019 − Yes − June	0.58 (0.34 − 0.99)
2019 − Yes − July	0.28 (0.16 − 0.50)
2020 − No − April	0.57 (0.29 − 1.11)
2020 − No − May	0.65 (0.29 − 1.47)
2020 − No − June	0.41 (0.20 − 0.87)
2020 − No − July	0.25 (0.11 − 0.55)
2020 − Yes − April	1.78 (0.91 − 3.49)
2020 − Yes − May	1.81 (1.27 − 2.57)
2020 − Yes − June	1.08 (0.76 − 1.54)
2020 − Yes − July	0.65 (0.45 − 0.95)



Supplementary Figure 8: Tick Encounter Model (2020 vs. 2019) —Diagnostic Plots

The QQ plot did not demonstrate significant issues with residuals. The Kolmogorov-Smirnov (KS) test for goodness of fit for testing normality had a value of *P*=.20 (A). The Residual vs. Predicted plot did not show significant issues with residuals (B). Variance inflation factor (VIF) values were <4 for all models, indicating no multicollinearity issues (C) and classification performance was good with an AUC = 0.91 (D).