PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (**http://bmjopen.bmj.com/site/about/resources/checklist.pdf**) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Early detection of Covid-19 in China and the United States: Summary of the implementation of a digital decision-support and disease surveillance tool
AUTHORS	Hswen, Yulin; Brownstein, John; Xu, Xiang; Yom-Tov, Elad

VERSION 1 – REVIEW

REVIEWER	Amaryllis Mavragani University of Stirling, UK
REVIEW RETURNED	15-Aug-2020

GENERAL COMMENTS	In the strengths and limitations bullet points (under the abstract), the authors have not included any limitations. I believe that Thermia should be mentioned in the title of the manuscript, which should also be modified to something to provide a concise description of the nature of this work, also including the term “digital disease surveillance”. In addition, this is an infodemiology study, and the authors should provide a concise description of the topic in their introduction, along with similar previous approaches of such tools/platforms in digital disease surveillance. Thermia is not a widely known tool in this line of research. Thus, the authors should provide a more elaborate description of the platform in their methods section, as well as provide detailed explanation of how data are retrieved, adjusted etc. This way, even the non-expert reader will be able to follow the flow of the manuscript and understand the aims and contribution of this work. Cite Thermia in the reference list rather than inserting the url in the main body of the manuscript. Include the descriptive statistics in a Table (now included as Table 1 in the Supp.) rather than in the text. Why did the authors not explore COVID-19 predictability using the ratio of deaths to cases, which is, from a statistical perspective, a more standard approach? In their limitations section, the authors state that they cannot verify that Thermia users correspond to actual COVID-19 cases (which is the general case with online tools). However, this is not the only limitation of this work. The authors should also address the issue of Thermia popularity amongst Internet users, the relatively low

	observation count (as is the case with all COVID-19 studies at this point), etc. Finally, the authors should make ac clear statement of their contribution to the advancement of the field, what the novelty of their approach is, and also discuss their results along with the results of similar approaches in COVID-19 predictability.

REVIEWER	Smadar Shilo Weizmann Institute of Science, Rambam Health Care Campus
REVIEW RETURNED	24-Aug-2020

GENERAL COMMENTS	Thank you for the opportunity to review the manuscript Early detection of Covid-19 in China and the United States using a digital health decision making tool. The manuscript presents a modification in an online health decision-making tool "Thermia" that could potentially supplement Covid-19 surveillance. The fact that Thermia was deployed to identify infections and advise the public about their potential risks of having Covid-19 already In late January 2020, before the World Health Organization declared Covid-19 as a Public Health Emergency is impressive. However, I have several comments related to the study and to the overall conclusions: Major comments: In this study, Thermia sessions served as a proxy for queries for symptoms related to Covid-19 , since those who quarry were presented with a web campaign and users who clicked on the ad were referred to the Thermia website. The authors themselves hypothesized that the reason for the correlation observed is most likely patients seeking information about their Covid-19 symptoms on the Internet before they were tested for Covid-19. It is therefore not clear to me whether using this proxy to identify infections is better than simply analyzing symptoms queried themselves? Does the advertising campaign coupled with a digital health decision-making provides any added value here or we can simply analyze related symptoms internet search? Does using Thermia specifically had any added value? One of the main strength of the study is the Cross-country analysis. However I have several comments on that: (1) in both countries the number of the mean daily sessions is very low compared to the mean number of patients (2) the maximal correlation observed in the USA was relatively low (r = 0.41) and (3) Thermia sessions to Covid-19 case reports were 3-days in China and 19-days in the United States. This large difference emphasis the difficulty of generalized the usage of this tool for its suggested purpose. As persons who queried for symptoms related to Covid-19 were presented with a web campaign and COVID-19 related symptoms also appear in Influenza infection. How do the authors explain the negative cross-correlation between influenza A and B and Thermia sessions? The suthors cita a study showing it was previously validated for early detection of influenza in China, what modifications were done to make it more specific to COVID-19 detection? I find it most likely to be a result of different trajectories of the prevalence of these two diseases in this time

	period (an increase in the prevalence of COVID-19 and a decrease in the prevalence of Influenza) and therefore the question of whether or not this tool will be valid if both diseases (as well as other infectious diseases that causes similar symptoms) will be similarly or more prevalent at the same time period Minor comments: Abstract Line 24- I think the authors should specify more on “Thermia” in the first time this system is mentioned, as the readers are not necessarily familiar with this system Strengths and limitations of this study Only strengths are mentioned, please also refer to the limitations of the study Background Page 4, line 42 – Ref 4 is regarding digital health application for influenza surveillance in China. However, there are already many examples of these application on the current pandemic, for example as reviewed here Ting, D. S. W., Carin, L., Dzau, V., & Wong, T. Y. (2020). Digital technology and COVID-19. Nature medicine, 26(4), 459-461.. Methods Page 5: Line 14 - Thermia is a web-based decision-making tool that provides clinical advice on how to treat febrile illness. Was it used here also for non-febrile individuals who had other symptoms? Line 38 - “Persons who queried for symptoms related to Covid-19 were presented with a web campaign” – Was this process updated throughout the pandemic? For example, the importance of the symptom Anosmia was discovered during the pandemic, and it was not considered as a COVID-19 symptom early on. Was the web campaign presented to people with Anosmia? Line 52- What was considered to be a sufficient amount of web traffic to the platform to preform your analysis? Were these countries the only ones who achieved this “amount”? What about other English speaking countries such as England and Canada? Page 6, line 6: Is it possible that the same person will have several Themia sessions? If so, can you analyze the data when each individual is only considered once? Page 7, line 40: The period in which data were analyzed is limited to April 2020. Since then, a lot of new data has probably accumulated (more Thermia sessions and more COVID-19 cases). Does the result of the study still hold in light of these new data?

	Results Page 8, line 23 – Its is worthy to note that in the USA the mean session number is X1.5 higher and the SD is X3.3 higher. Do you think it had an effect on the different results obtained in both countries? Page 9, line 4 - It seems that in the USA the maximal correlation between Thermia sessions and COVID-19 cases was not strong, with r= 0.41. line 12 – See my comment above Discussion Page 10, Line 28 – I think that the comparison to number of test in south Korea is not relevant here as the differences between the time lag in china and the USA is discussed. Can you cite the number of daily tests in China compared to the USA? Line 49 – Worth mentioning that the first confirmed case of 2019-nCoV infection in the United States, reported on January 20, 2020 (Holshue, Michelle L., et al. "First case of 2019 novel coronavirus in the United States." New England Journal of Medicine (2020)) Page 11, line 15 – Please also refer in the limitation sections to the biases in using your proposed method (for example, as it relays on internet search it can only capture individuals who have access to computers and requires a certain educational level) Line 47 – I would rephrase since from my understanding, the predictive signal was not due to the usage of this web platform rather it was due to symptoms related search on the web. I don't think that it was proven here that the platform itself played any role or improved the prediction ability. Figure 1B – It seems that in one specific state (California?) the amount of sessions was exceptionally high. How do you explain it? This can further be related to the study limitation and the bias in the data. Figure 2- While the trends and general shape of Thermia sessions and COVID-19 patients looks similar in China, they do not look similar at all in the USA. Would be good to continue this timeline with the new data and see if the relatively weak correlation found still holds in future data. Supplemental data In the result section, it states that in China, the mean Thermia daily sessions was 17.7 (SD=7.0) and the mean Covid-19 daily cases was 9346.0 (SD=12881.1). In the United States, the mean Thermia daily sessions was 27.1 (SD=23.4) and the mean Covid-19 daily cases was 839.6 (SD=2885.2).

	In this table, the result shown are different, which is correct? In both countries, the minimal number of sessions is very low- 4 sessions a day…

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

In the strengths and limitations bullet points (under the abstract), the authors have not included any limitations.
RESPONSE: We added the following limitations:

- "Data could not be linked to actual cases so as to preserve privacy. Therefore, only correlational evidence can be provided to support findings."
- "Our data comprised of relatively few user interactions. This is partly due to awareness on the side of people and partly to the budget allocated to creating such awareness through ads. Finally, although the internet is increasingly ubiquitous in use, there are still biases in our data due to differences in access due to social status and income."

I believe that Thermia should be mentioned in the title of the manuscript, which should also be modified to something to provide a concise description of the nature of this work, also including the term "digital disease surveillance".
RESPONSE: Following this comment we modified the title to "Early detection of Covid-19 in China and the United States: Summary of the implementation of a digital decision-support and disease surveillance tool"

In addition, this is an infodemiology study, and the authors should provide a concise description of the topic in their introduction, along with similar previous approaches of such tools/platforms in digital disease surveillance.
REPSONSE: We added the following overview of past and current infodemiology systems, stating: "Digital surveillance tools have been used in the past, beginning with tools for surveillance of influenza5 and later expanded to other conditions.[6,7] However, experience has shown that such tools need careful tuning to successfully track cases of illness.[8] Notwithstanding these limitations, tools based on search queries[9], advertising[10] and other digital signals[11] have been proposed. Currently, these tools are used for tacking Covid-19[12], for example in England[13,14] using search engine data."

Thermia is not a widely known tool in this line of research. Thus, the authors should provide a more elaborate description of the platform in their methods section, as well as provide detailed explanation of how data are retrieved, adjusted etc. This way, even the non-expert reader will be able to follow the flow of the manuscript and understand the aims and contribution of this work.
RESPONSE: We added more information on Thermia to the Methods. We also provide a screenshot of the platform. The relevant text now states that: "Thermia[4] is a digital decision-support tool, that was developed by researchers at Harvard Medical School, to provide clinical advice on how to treat febrile illness based on an evidence-based algorithm. User are referred to the Thermia platform based on their symptoms queries related to febrile illness on a web search engine. At Thermia, users are directed through a series of questions about their temperature, symptoms and biometric characteristics (see Figure 1) and are given recommendations on how to further proceed with medical care based on their provided answer. Thermia has also been used for digital surveillance, and was previously validated for early detection of influenza in China.[4]"

Cite Thermia in the reference list rather than inserting the url in the main body of the manuscript.
RESPONSE: Thank you. We have updated this so that Thermia is in the reference list.

Include the descriptive statistics in a Table (now included as Table 1 in the Supp.) rather than in the text.
RESPONSE: We have moved the Supplemental Table to be included as Table 1 in the manuscript.

Why did the authors not explore COVID-19 predictability using the ratio of deaths to cases, which is, from a statistical perspective, a more standard approach?
RESPONSE: We agree with the reviewer that this may be a more robust approach. However, as information on deaths at the beginning of the epidemic are noisy due to the dearth of tests, we think that such a comparison will be noisier than comparison to case numbers.

In their limitations section, the authors state that they cannot verify that Thermia users correspond to actual COVID-19 cases (which is the general case with online tools). However, this is not the only limitation of this work. The authors should also address the issue of Thermia popularity amongst Internet users, the relatively low observation count (as is the case with all COVID-19 studies at this point), etc.
RESPONSE: Thank you for this comment. We added the following limitations after the Abstract:

- "Data could not be linked to actual cases so as to preserve privacy. Therefore, only correlational evidence can be provided to support findings."
- "Our data comprised of relatively few user interactions. This is partly due to awareness on the side of people and partly to the budget allocated to creating such awareness through ads."

Furthermore, we added the following sentence to the Discussion: "Another limitation of our study is the relatively few user interactions we were able to obtain. This is partly due to awareness on the side of people and partly to the budget allocated to creating such awareness through ads."

Finally, the authors should make ac clear statement of their contribution to the advancement of the field, what the novelty of their approach is, and also discuss their results along with the results of similar approaches in COVID-19 predictability.
RESPONSE: We added the following paragraph to the Conclusion section of the manuscript: "Here we demonstrated the ability to rapidly respond to a novel disease outbreak by quickly creating a system which provided people with decision support and, through the data collected by it, provide surveillance information that could be used by health authorities. The use of both Thermia and advertising allowed us to go gain several advantages, including directly approaching people who may yet be invisible to the health system, obtain information beyond searches themselves (because of the use of Thermia), and create awareness through a targeted advertising campaign."

Reviewer 2:

Major comments:

1. In this study, Thermia sessions served as a proxy for queries for symptoms related to Covid-19 , since those who quarry were presented with a web campaign and users who clicked on the ad were referred to the Thermia website. The authors themselves hypothesized that the reason for the correlation observed is most likely patients seeking information about their Covid-19 symptoms on the Internet before they were tested for Covid-19. It is therefore not clear to me whether using this proxy to identify infections is better than simply analyzing symptoms queried themselves? Does the advertising campaign coupled with a digital health decision making provides any added value here or we can simply analyze related symptoms internet search? Does using Thermia specifically had any added value?

RESPONSE: Thank you for this excellent question. The problem we faced at the start of the pandemic was that, even though limited understanding of the symptoms associated with Covid-19

existed, there was insufficient data to train a query-based model. Hence, we had to resort to the combination of ads and Thermia, which together provided a much richer understanding of people's experience (including travel habits). We added an explanation of this to the beginning of the Discussion, stating that:

"Most epidemiological monitoring tools, especially those dependent on online interactions (e.g., search based), rely on a combination of factors for their success. The first of these is that there is a significant lag between the appearance of symptoms and the first time that people visit the medical symptoms or that only a small part of the infected population visits the medical system. When symptoms are severe enough to warrant an urgent visit to a hospital and when most infected people visit a medical provider, data from the medical system will be superior to that of search-based infodemiological systems. When no symptoms exist, people will not query about them, making the monitoring tools ineffective. Second, good ground-truth data is needed to calibrate these systems. In the case of flu, for example, researchers often use past seasons to tune the models.

Covid-19 has the first set of attributes (e.g., lag between symptoms and visit to the medical system[13]), but, especially at the beginning of the epidemic, while there was an understanding of the symptoms, which allowed us to modify Thermia to provide information to people on their condition, there was insufficient ground truth to tune a symptom-search model. Stated differently, multiple questions were needed to ascertain the severity of symptoms, and these were obtained through the use of Thermia. Thus, the combination of ads and questionnaire allowed us to go beyond simple searches. Later in the pandemic, as more ground-truth data became available, tuning was made possible, as shown in Lampos et al..[14]"

2. One of the main strength of the study is the Cross-country analysis. However I have several comments on that: (1) in both countries the number of the mean daily sessions is very low compared to the mean number of patients (2) the maximal correlation observed in the USA was relatively low (r = 0.41) and (3) Thermia sessions to Covid-19 case reports were 3-days in China and 19-days in the United States. This large difference emphasis the difficulty of generalized the usage of this tool for its suggested purpose.

RESPONSE: We agree with the reviewer that this a slightly lower correlation than traditionally seen in previous studies related to using online digital tools for digital detection. However, these past correlations have ranged between (r = 0.42) and (r = 0.88) for surveillance systems related to influenza using Google Flu Trends. The fact that we detect a significant correlation of (r = 0.41) in the U.S. with Thermia, a survillance tool with Covid-19, a novel disease, is meaningful. Additionally, this lower correlation and the greater lag time in the United States compared to China may be a result of greater testing in China at the beginning of the pandemic. We have included this in the limitation section of the manuscript.

3. As persons who queried for symptoms related to Covid-19 were presented with a web campaign and COVID-19 related symptoms also appear in Influenza infection. How do the authors explain the negative cross-correlation between influenza A and B and Thermia sessions? The suthors cita a study showing it was previously validated for early detection of influenza in China, what modifications were done to make it more specific to COVID-19 detection? I find it most likely to be a result of different trajectories of the prevalence of these two diseases in this time period (an increase in the prevalence of COVID-19 and a decrease in the prevalence of Influenza) and therefore the question of whether or not this tool will be valid if both diseases (as well as other infectious diseases that causes similar symptoms) will be similarly or more prevalent at the same time period

RESPONSE: As we explain in the Results, the negative cross-correlation between influenza A and B and Thermia sessions was seen because Thermia sessions were potentially detecting COVID-19 cases and not influenza. Weekly influenza data for the 2019-2020 season is provided from the Centers for Disease Control and Prevention (CDC) and showed that influenza A began to decline mid-February while influenza B declined in early January and both had close to zero cases mid-March. Conversely, Thermia sessions begin to rise in mid-Feb and peak mid-March. These results highlight that Thermia was likely capturing cases of Covid-19 and not influenza A or B since there

were no cases of influenza shortly after mid-March. It is unclear if Thermia would be valid to differentiate Covid-19 to influenza if the prevalence of these disease were parallel over the same time period. Although our validation is based on the different trajectories of the prevalence of influenza and Covid-19, the advertising campaign for Thermia was specifically for symptoms and travel related to Covid-19. Therefore, there is less of a likelihood of user visiting Thermia who were cases of influenza and our web campaign advertisement for Thermia may have supported detection of cases of Covid-19 from persons who had recently traveled outside the United States. Future studies should test Thermia's ability to differentiate influenza from Coivid-19 during the seasonal influenza season.

Minor comments:

Abstract
Line 24- I think the authors should specify more on "Thermia" in the first time this system is mentioned, as the readers are not necessarily familiar with this system
RESPONSE: Thank you. We have added a more detailed description Thermia the first time it is mentioned stating that: "Thermia[4] is a digital decision-support tool, that was developed by researchers at Harvard Medical School, to provide clinical advice on how to treat febrile illness based on an evidence-based algorithm. User are referred to the Thermia platform based on their symptoms queries related to febrile illness on a web search engine. At Thermia, users are directed through a series of questions about their temperature, symptoms and biometric characteristics (see Figure 1) and are given recommendations on how to further proceed with medical care based on their provided answer. Thermia has also been used for digital surveillance, and was previously validated for early detection of influenza in China.[4]"

Strengths and limitations of this study
Only strengths are mentioned, please also refer to the limitations of the study
RESPONSE: We added the following limitations:

- "Data could not be linked to actual cases so as to preserve privacy. Therefore, only correlational evidence can be provided to support findings."
- "Our data comprised of relatively few user interactions. This is partly due to awareness on the side of people and partly to the budget allocated to creating such awareness through ads."

Background
Page 4, line 42 – Ref 4 is regarding digital health application for influenza surveillance in China. However, there are already many examples of these application on the current pandemic, for example as reviewed here Ting, D. S. W., Carin, L., Dzau, V., & Wong, T. Y. (2020). Digital technology and COVID-19. Nature medicine, 26(4), 459-461..
RESPONSE: We added the following overview of past and current infodemiology systems, stating: "Digital surveillance tools have been used in the past, beginning with tools for surveillance of influenza5 and later expanded to other conditions.[6,7] However, experience has shown that such tools need careful tuning to successfully track cases of illness.[8] Notwithstanding these limitations, tools based on search queries[9], advertising[10] and other digital signals[11] have been proposed. Currently, these tools are used for tracking Covid-19[12], for example in England[13,14] using search engine data."

Methods
Page 5:
Line 14 - Thermia is a web-based decision-making tool that provides clinical advice on how to treat febrile illness. Was it used here also for non-febrile individuals who had other symptoms?

RESPONSE: Although Thermia was a web-based decision tool to support febrile illness, it was adapted to include symptoms and human mobility behaviors related to Covid-19 and provided recommendations for treatment of Covid-19.

Line 38 - “Persons who queried for symptoms related to Covid-19 were presented with a web campaign” – Was this process updated throughout the pandemic? For example, the importance of the symptom Anosmia was discovered during the pandemic, and it was not considered as a COVID-19 symptom early on. Was the web campaign presented to people with Anosmia?
RESPONSE: Anosmia was described as a symptom of Covid-19 only later (beginning of April). Hence, they were not added to the system. As described in the Methods, the only change we made to the ads was to refer to general travel outside the US, rather than only travel to China, when CDC made the relevant change.

Line 52- What was considered to be a sufficient amount of web traffic to the platform to preform your analysis? Were these countries the only ones who achieved this “amount”? What about other English speaking countries such as England and Canada?
RESPONSE: We do not think there is a clear-cut threshold above which we could conduct our analysis. However, we would expect to have daily traffic from each country on, at least, most days. That said, we selected the US because our ads were targeted to that country. China was added because it had more sessions to Thermia (without ads) than the US. Although there was traffic from Canada and the United Kingdom, it was relatively sparse and we did not have sessions on enough days to perform the cross-correlation analysis.
To reflect this, we added the following text to the manuscript: “The volume of daily web traffic to Thermia from the US and China enabled us to conduct cross-correlation analysis of traffic volume with case counts. We restricted this study to investigating visits to the Thermia website in China and the United States because the campaign was placed in both English and Mandarin and the ads were shown in the US and related to travel from China and outside the US. We also observed the largest data volume in sessions from those two countries. Therefore, although web traffic was seen from Canada and the United Kingdom, the amount of daily web traffic was not sufficient to conduct our temporal analyses.”

Page 6, line 6: Is it possible that the same person will have several Themia sessions? If so, can you analyze the data when each individual is only considered once?
RESPONSE: It is possible that the same user will have several Thermia sessions. We are unable to isolate unique users because of limitations imposed to preserve privacy (e.g., data is provided in an aggregated format. However, since Thermia is decision-support tool that provided recommendations for Covid-19, it is unlike a user would revisit unless it was to retrieve information about another user whom they were using the Thermia to get information for. We have added this as a limitation in the limitation section of the manuscript.

Page 7, line 40: The period in which data were analyzed is limited to April 2020. Since then, a lot of new data has probably accumulated (more Thermia sessions and more COVID-19 cases). Does the result of the study still hold in light of these new data?
RESPONSE: We stopped the advertising campaign and data collection on Thermia at the end of April, as more and more decision support and epidemiological tools became available. Therefore, we cannot comment on how these data would appear today.

Results
Page 8, line 23 – Its is worthy to note that in the USA the mean session number is X1.5 higher and the SD is X3.3 higher. Do you think it had an effect on the different results obtained in both countries?
RESPONSE: The fact that there were more sessions in the US is not surprising, since in this country people could reach Thermia through either organic search or ads, whereas in China they found the

site only through organic search. We are unaware of a way in which we could quantify if this variance could affect our results. If the reviewer can suggest such a quantifier, we would be happy to apply it.

Page 9, line 4 - It seems that in the USA the maximal correlation between Thermia sessions and COVID-19 cases was not strong, with r= 0.41.
line 12 – See my comment above
RESPONSE: We agree with the reviewer that the cross-correlation of r = 0.41 in the U.S. is slightly lower than traditionally research that evaluates the use of online digital tools for earlier detection. However, these correlations between Google Flu Trends and influenza case counts have ranged between (r = 0.42) and (r = 0.88) (cite) Therefore, the fact that we detect a significant correlation of (r = 0.41) in the U.S. with Thermia, a specific surveillance tool adapted for Covid-19, a novel disease, is meaningful. This lower correlation and the greater lead time in the United States of 19 days is seen compared to a 3-day lead China this may be a result of greater testing capacity in China at the beginning of the pandemic. (cite)0.41

Discussion
Page 10, Line 28 – I think that the comparison to number of test in south Korea is not relevant here as the differences between the time lag in china and the USA is discussed. Can you cite the number of daily tests in China compared to the USA?
RESPONSE: As of June 29, 2020, China has carried out one test for every 15 people, compared to one in 11 in the U.S. These number were more divergent in the early phase of the pandemic because of the limited tested that was provided by the U.S. . We added this to the paper, stating: “As of June 29, 2020, China has carried out one test for every 15 people, compared to one in 11 in the U.S..[25] These number were more divergent in the early phase of the pandemic because of the limited tested that was provided by the U.S.[26]”

Line 49 – Worth mentioning that the first confirmed case of 2019-nCoV infection in the United States, reported on January 20, 2020 (Holshue, Michelle L., et al. "First case of 2019 novel coronavirus in the United States." New England Journal of Medicine (2020))
RESPONSE: We thank the reviewer for this citation and have added it in the manuscript.

Page 11, line 15 – Please also refer in the limitation sections to the biases in using your proposed method (for example, as it relays on internet search it can only capture individuals who have access to computers and requires a certain educational level)
RESPONSE: We added this to our limitations, stating that: “Although the internet is increasingly ubiquitous in use, there are still biases in our data because of differences in access due to social status and income disparities.”

Line 47 – I would rephrase since from my understanding, the predictive signal was not due to the usage of this web platform rather it was due to symptoms related search on the web. I don’t think that it was proven here that the platform itself played any role or improved the prediction ability.
RESPONSE: The reviewer is correct to the extent that health-seeking behavior in the form of queries to online search engines often precede provider visits (cite). Thus, search queries related to Covid-19 symptoms on the web may have also played a role in generating predictive signal by visiting Thermia. However, as we now know, symptom searches are not sufficient for predicting the level of illness. The additional validation and enrichment of symptoms provided by Thermia (and by the response to the ads) provides a better predictive signal. Moreover, as we now discuss in the text, symptom searches require ground-truth data in order to tune a predictive model. At the beginning of the pandemic such data was not available. Hence, the use of Thermia could assist in this respect.

Figure 1B – It seems that in one specific state (California?) the amount of sessions was exceptionally high. How do you explain it? This can further be related to the study limitation and the bias in the data.

RESPONSE: The occurrence of cryptic cases of Covid-19 from updated evidence show the virus was circulating in early February. This data indicates a sustained community transmission had started before the detection of the first U.S. cases. Our results, showing a greater number of sessions in California may have been Covid-19 that were not originally detected, for example, due to (Very) limited texting.
We added to the paper: “Furthermore, the occurrence of cryptic cases of Covid-19 from updated evidence show the virus was circulating in the U.S. in early February.[27] These data indicate a sustained community transmission had started before the detection of the first U.S. cases. Our results, showing a greater amount of sessions in California, may be due to Covid-19 cases that were not originally detected. The findings here suggest that the 19-day lead time of Thermia session in the United States is a signal for early Covid-19 cases that were not captured by traditional public health monitoring.”

Figure 2- While the trends and general shape of Thermia sessions and COVID-19 patients looks similar in China, they do not look similar at all in the USA. Would be good to continue this timeline with the new data and see if the relatively weak correlation found still holds in future data.
RESPONSE: Since the advertising campaign and data collection on Thermia no longer continued after the end of April, we are unable to comment on the trajectory of the timeline of Thermia after that date.

Supplemental data
In the result section, it states that in China, the mean Thermia daily sessions was 17.7 (SD=7.0) and the mean Covid-19 daily cases was 9346.0 (SD=12881.1). In the United States, the mean Thermia daily sessions was 27.1 (SD=23.4) and the mean Covid-19 daily cases was 839.6 (SD=2885.2).
In this table, the result shown are different, which is correct? In both countries, the minimal number of sessions is very low- 4 sessions a day…
RESPONSE: We very much appreciate the review for pointing out this descripancy. We have corrected the text to be in align with the Table 1.

VERSION 2 – REVIEW

REVIEWER	Amaryllis Mavragani University of Stirling, UK
REVIEW RETURNED	20-Sep-2020

GENERAL COMMENTS	The authors have adequately addressed all concerns raised during the first review round. In order for all limitations to have been discussed, please include your response to the following comment in the discussion section: Why did the authors not explore COVID-19 predictability using the ratio of deaths to cases, which is, from a statistical perspective, a more standard approach? RESPONSE: We agree with the reviewer that this may be a more robust approach. However, as information on deaths at the beginning of the epidemic are noisy due to the dearth of tests, we think that such a comparison will be noisier than comparison to case numbers.

REVIEWER	Smadar Shilo Weizmann Institute of Science, Israel Rambam health care campus, Israel
REVIEW RETURNED	17-Oct-2020

GENERAL COMMENTS	Thank you for the opportunity to review the revised manuscript "Early detection of Covid-19 in China and the United States: Summary of the implementation of a digital health decision-making tool" The manuscript presents a modification of an online health decision-making tool "Thermia" that could potentially supplement Covid-19 surveillance. Overall, I still think that one of the major limitations of the work is the fact that all analyses are based on very few user interactions that were collected (a minimum of 4 a day and low average of interaction in both USA and China). I think this should be better emphasized in the manuscript. Also, the very different lag time between countries, which may change in time as user interactions also change with time, will challenge the ability of health authorities to make practical use of this system. Abstract I would revise the result of "A negative cross-correlation between influenza A and B and Thermia sessions was obtained, implying that Thermia sessions were probably not related to cases of influenza" to better reflect that it is probably due to the decreasing prevalence of influenza at this time of year and possibly not related to the specificity of the system for COVID-19 identification at all (as implied from the text now). Methods In your response to one comment you stated that "although Thermia was a web-based decision tool to support febrile illness, it was adapted to include symptoms and human mobility behaviors related to Covid-19" Other than the travel history, what additional changes were made in the app to distinguish the symptoms of COVID-19 from Flu? Results "In China, the average number of daily Thermia sessions was 27.1 (SD=23.4) and the average Covid-19 number of daily cases was 839.6 (SD=2885.2). In the United States, the average number of Thermia sessions was 17.7 (SD=7.0) and the average number of Covid-19 cases was 9346.0 (SD=12881.1)" – please add a min and max for these results Discussion – "the first time that people visit the medical symptoms" – this sentence is not clear; do you mean medical system? "especially at the beginning of the epidemic, while there was an understanding of the symptoms, which allowed us to modify Thermia to provide information to people on their condition, there was insufficient ground truth to tune a symptom-search model" – In my opinion, in the beginning of the epidemic there was not enough understanding of the symptoms and the sentence contradicts itself. For example, loss of taste and smell which were later shown to be with the highest OR for COVID-19 infection were still not identified then. multiple questions were needed to ascertain the severity of symptoms – which questions were they? In Fig 1 it's hard to see the questions included, were there any on disease severity? (please improve the Fig visibility) "The advertising campaign for Thermia was specifically for symptoms and travel related to Covid-19" – This is not accurate as all the symptoms in the app can also be present in other infectious diseases and they are not specific for COVID-19 at all.

VERSION 2 – AUTHOR RESPONSE

Reviewer(s)' Comments to Author:

Reviewer: 1

Reviewer Name: Amaryllis Mavragani
Institution and Country: University of Stirling, UK
Competing interests: None declared

Comments to the Author
The authors have adequately addressed all concerns raised during the first review round. In order for all limitations to have been discussed, please include your response to the following comment in the discussion section:

We thank you very much for this acknowledgment. We have included this following comment in this discussion section.

Why did the authors not explore COVID-19 predictability using the ratio of deaths to cases, which is, from a statistical perspective, a more standard approach?
RESPONSE: We agree with the reviewer that this may be a more robust approach. However, as information on deaths at the beginning of the epidemic are noisy due to the dearth of tests, we think that such a comparison will be noisier than comparison to case numbers.

Thank you again for your time and efforts. We have made this addition to the discussion section.

Reviewer: 2

Reviewer Name: Smadar Shilo
Institution and Country:
Weizmann Institute of Science, Israel
Rambam health care campus, Israel
Competing interests: None declared

Comments to the Author
Thank you for the opportunity to review the revised manuscript “Early detection of Covid-19 in China and the United States: Summary of the implementation of a digital health decision-making tool”
The manuscript presents a modification of an online health decision-making tool “Thermia” that could potentially supplement Covid-19 surveillance.
Overall, I still think that one of the major limitations of the work is the fact that all analyses are based on very few user interactions that were collected (a minimum of 4 a day and low average of interaction in both USA and China). I think this should be better emphasized in the manuscript. Also, the very different lag time between countries, which may change in time as user interactions also change with time, will challenge the ability of health authorities to make practical use of this system.

We thank the reviewer for her comment. We understand that one of the major limitations is that there are few user interactions with a minimum of 4 a day and low average of interaction in the USA and China. We also acknowledge that the different lag times between countries could change in time as user interactions also change with time. This will be a challenge for health authorities to use. We have integrated this into the discussion and limitation section of the manuscript, stating that: "However, the different lag time between countries could also change in time as user interactions with Thermia develop. This presents a challenge for the ability of health authorities to make practical use of the data from this system. Another limitation of our study is the relatively few user interactions we were able to obtain. This is partly due to awareness on the side of people and partly to the budget allocated to creating such awareness through ads. Thus, analyses are based on very few user interactions that were collected with a minimum of four per day and low average of interaction in both USA and China."

Abstract
I would revise the result of "A negative cross-correlation between influenza A and B and Thermia sessions was obtained, implying that Thermia sessions were probably not related to cases of influenza" to better reflect that it is probably due to the decreasing prevalence of influenza at this time of year and possibly not related to the specificity of the system for COVID-19 identification at all (as implied from the text now).

We have revised this section based on the reviewers' comments, stating that: "We found negative cross-correlation between the number of Thermia sessions and rates of influenza A and B, possibly due to the decreasing prevalence of influenza and the lack of specificity of the system for identification of Covid-19."

Methods
In your response to one comment you stated that "although Thermia was a web-based decision tool to support febrile illness, it was adapted to include symptoms and human mobility behaviors related to Covid-19"
Other than the travel history, what additional changes were made in the app to distinguish the symptoms of COVID-19 from Flu?

The reviewer is correct to note that the main adaptation was to include symptoms of human mobility behaviors related to Covid-19 which at the time was the major identifying characteristics of Covid-19. The additional modifications were made to the recommendations for treatment of Covid-19 based on the CDC guidelines which at the time was to stay at home and notify your provider of your symptoms and travel. We feel this is an important point that the reviewer makes and have added this to the Methods section to clarify. Thus, we added the following: "At the beginning of the Coivd-19 pandemic differences between the flu and Covid-19 were not fully understood and travel history was prominent in helping identifying possible Covid-19 infection as defined by the CDC. As well, questions about severity of symptoms were also not included because at the time of the study the CDC had only provided a list of potential symptoms of Covid-19 and did not expand on the severity of the symptoms and their differential diagnosis of the disease. Thus, the modification of Thermia from influenza to Covid-19 comprised of adding questions on travel history and a change to the response to users, which was based on CDC guidelines."

Results
“In China, the average number of daily Thermia sessions was 27.1 (SD=23.4) and the average Covid-19 number of daily cases was 839.6 (SD=2885.2). In the United States, the average number of Thermia sessions was 17.7 (SD=7.0) and the average number of Covid-19 cases was 9346.0 (SD=12881.1)” – please add a min and max for these results

Thank you for the suggestion. We have added the min and max for these results.

In China, the average number of daily Thermia sessions was 27.1 (SD=23.4, Min=4, Max=123) and the average Covid-19 number of daily cases was 839.6 (SD=2885.2, Min=3, Max=15136). In the United States, the average number of Thermia sessions was 17.7 (SD=7.0, Min=4, Max=42) and the average number of Covid-19 cases was 9346.0 (SD=12881.1, Min=0, Max=34126).

Discussion –
“the first time that people visit the medical symptoms” – this sentence is not clear; do you mean medical system?

Indeed, we meant the medical system. Thank you for picking this up. We have changed it to “medical system”.

“especially at the beginning of the epidemic, while there was an understanding of the symptoms, which allowed us to modify Thermia to provide information to people on their condition, there was insufficient ground truth to tune a symptom-search model” – In my opinion, in the beginning of the epidemic there was not enough understanding of the symptoms and the sentence contradicts itself. For example, loss of taste and smell which were later shown to be with the highest OR for COVID-19 infection were still not identified then.

We adjusted this sentence to be clearer, stating that: “especially at the beginning of the epidemic, there was a limited understanding of the symptoms and although we modified Thermia to provide information to people on their condition, there was insufficient ground truth to tune a symptom-search model. For instance, most recent evidence has shown that symptoms such as loss of taste and smell are highly indicative of a Covid-19 infection.”

multiple questions were needed to ascertain the severity of symptoms – which questions were they? In Fig 1 it’s hard to see the questions included, were there any on disease severity? (please improve the Fig visibility)

We improved the visibility of Figure 1 so as to make it clearer. There were no questions on disease severity as in the beginning of the pandemic, as at that stage the CDC only provided guidelines on the types of symptoms of Covid-19 and not the degree of symptom severity. We have added this to the Methods section of the study (see also above).

“The advertising campaign for Thermia was specifically for symptoms and travel related to Covid-19” – This is not accurate as all the symptoms in the app can also be present in other infectious diseases and they are not specific for COVID-19 at all.

We have corrected this sentence to be more accurate.

“The advertising campaign for Thermia was adapted to include symptoms of Covid-19 and travel related to the primary areas of emergence of Covid-19”

VERSION 3 – REVIEW

REVIEWER	Smadar Shilo Rambam healthcare campus Weizmann institute of science
REVIEW RETURNED	17-Nov-2020

GENERAL COMMENTS	The authors have adequately addressed all concerns raised during the previous review rounds