Pandemic Metric with Confidence (PMC) Model to Predict Trustworthy Probability of Utilized COVID-19 Pandemic Trajectory across the Global

Zhengkang Zuo[1, *], and Hongying Zhao[1, *]

[1] School of Earth and Space Science, Peking University, Beijing, 100871, China.

Corresponding authors: Zhengkang Zuo (1801110646@pku.edu.cn); Hongying Zhao (Zhaohy@pku.edu.cn)

Key Points:

- PMC model is proposed in the hypothesis of Bernoulli Distribution of pandemic trajectory to dig the real trajectory data.
- The average confidence of COVID-19 pandemic trajectory across the global is not in excess of 12.1%.
- Trade-off strategy in PMC model supports 61% countries to predict realer trajectory data with confidence beyond 50%.

Abstract

Lots of works aim to reveal the driving factors of COVID-19 pandemic trajectory yet ignore the confidence of utilized trajectory data, making consequent results suspicious. Hereby, we proposed a pandemic metric with confidence (PMC) model in the hypothesis of Bernoulli Distribution of nine trajectories reported from 113 countries. Results exhibit the average confidence of trajectories across the global not in excess of 12.1% with the error threshold configuration of 1E-5. In contrast, the 95% high confidence setting also failed to predict the trajectory containing the acceptable error not beyond 1E-3. Thus, a proposed trade-off strategy between two contradictory expections (>50% confidence, <1E-3 error) supports 61% of investigated countries to predict the varying trajectory with confidence beyond 50%. Moreover, PMC model recommend the remanent 39% countries to extend the proportion of populaces in COVID-19 detecting-pool to a suggested-value (>1% of populations), ensuing the average confidence up to 70%.

Plain Language Summary

There are two main obstacles to reveal the real climatic role in the COVID-19 pandemic, that is to separate coupling effects of climatic and non-climatic factors on its trajectory, and to raise the confidence of utilized trajectory. Our previous work published on Earths' Future overcame the first obstacle, and this work focuses on the second obstacle with two steps. First, a pandemic metric with confidence (PMC) model is proposed in the hypothesis of Bernoulli Distribution of nine trajectories reported from 113 countries, containing two tunable variables (confidence and error) as well as two observed variables (trajectory data and test strategy). One capacity of PMC model is to predict either the confidence or according error of existing trajectory data. The other capacity is either to dig the real trajectory data with the certain confidence and error threshold, or to adjust present test strategy for more reliable trajectory data. Second, a trade-off strategy is proposed given the coupling relationship of four PMC variables. Results reveal that error threshold within scope of 1E-5 to 1E-3, confidence in excess of 50%, and the number of populaces participating COVID-19 detecting pool beyond 1E+8 together yield the optimal prediction of reliable trajectory data.

1 Introduction

The coronavirus disease 2019 (COVID-19) pandemic is a rapidly evolving global emergency that continues to strain healthcare systems (Gallo Marin et al., 2021). To curb pandemic, lots of works were carried out to uncover the potential driving factors to its trajectory. NATURE and SCIENCE published articles affirming the positive effect of non-natural factors on mitigating the pandemic (Lai, 2020; Tian and Liu, et al., 2020), but still the possibility cannot be ruled out that the up-to-date varying trajectory is partially attribute to other unknown climatic factors. Hereby, to promote the profounder cognition on COVID-19, various natural driving factors such as temperature, humidity, solar radiation, aerosol, wind speed and vegetation were used to discover associations between the pandemic trajectory and those factor. However, different studies showed diverged and even contradictory results. Previous studies support temperature a key driving factor (Wang and Jiang et al., 2020; Wang and Tang et al., 2020) whereas the other studies (Yao et al., 2020; Ma et al., 2020) do not support that. Besides, such diverged conclusions also occur in humidity driving factor. References (Ahmadi et al., 2020; Wang and Tang et al., 2020) support high humidity reduces the transmission of COVID-19. However, reference (Luo et al., 2020) concludes that the role of absolute humidity in transmission of

COVID-19 has not been established. In addition, COVID-19 may transmit through aerosol (Liu et al., 2020; Wang and Du, 2020), whereas there are also important reasons to suspect it plays a role in the high transmissibility of virus (Sima, 2020). Further, a study shows that an outbreak at low wind speed is remarkable (Islam et al., 2020), but this result is nullifled by another study (Oliveiros et al., 2020). Recently, EARTHS FUTURE published an article summarizing 46 contradictory conclusions (Zuo et al., 2021).

There are two main snags to disclose the climatic role in the urgent human public health events. On the one hand, the spreading mechanism of virus is very complex, coupling numerous factors. However, most of aforesaid public researches only consider single climatic factors and ignore their coupling relationship. Besides, the utilized COVID-19 trajectory is the interaction outcome of both natural factors and human interventions. Thus, the direct correlation analysis between climatic factors and such COVID-19 trajectory yet cannot reveal the real climatic influence on the pandemic, because the human interventions are the dominant factors and the accessorial climatic influence will certainly be inhibited (ignored) in such correlation analysis. Owing to this, the influence separation of these two types of driving factors is quite important for the better understanding and authenticity of the climatic influence on the pandemic. On the other hand, we cannot deny that the uncertainty existing in the reported data concerning pandemic trajectory makes the capacity of published researchs questionable for delivering the accurate results. For example, references (Cordes and Castro, 2020; Morrison et al., 2020) summarized two classes of pandemic trajectories concerning the transmission of COVID-19 that are quite sensitive to distinct COVID-19 test strategy adopted by the global (Onder et al., 2020). There are two main test technologies such as reverse transcription-polymerase chain reaction (RT-PCR) (Freeman et al., 1999) and nucleic-acid test (NAT) (Wu et al., 2020). However, due to high false-positive rate of NAT (Zhuang et al, 2020) and high false-negative rate of RT-PCR (Li and Yao et al., 2020), traces of virus detected by NAT and RT-PCR are not properly correlated with the real trajectory (Xiao et al., 2020). The similar uncertainties are also observed in pandemic trajectory concerning the other metrics (Section 2.2). In a short, the first snag was solved by our previous work (Zuo et al., 2021), and the main contribution of this paper is to overcome another snag through the pandemic metrics with confidence (PMC) model.

2 Materials and Methods

2.1 Data Collection

This work first aims to investigate the effect of region-specific test strategy on the reported pandemic data across the global. Hereby, we selected 113 countries as a research region (e.g., US, India, China, etc. in which their testing rate data are public avaliable), and the corresponding data are avaliable from the world info meter of COVID-19 (https://www.worldometers.info/coronavirus/). Besides, the time series of pandemic raw data reported from January 21 and August 31, 2020 are respectively collected from the COVID-19 data repository (https://github.com/CSSEGISandData/COVID-19), i.e., total confirmed cases, total deaths, total recoveries and total confirmed cases with critical symptoms. Finally, all data for this study are available in the Zenodo repository (PMC-2021. (2021, May 28). PKU-2021/PMC-Model: PMC-Model (Version v1.0.0). Zenodo. http://doi.org/10.5281/zenodo.4831821).

2.2 Pandemic Metrics

Derivative pandemic metrics are widely used in many statistical researchs concerning COVID-19, i.e., the proportion of the positive tests (Cordes and Castro, 2020) and positivity rate (Morrison et al., 2020) to uncover the infection-specific characteristic of the COVID-19; case fatality rate (Yang et al., 2020), mortality rate (Baud et al., 2020) and closed case fatality rate (Pueyo et al., 2020) for death-specific characteristic; discharge rate (Tian and Hu et al., 2020), recovery rate (Li and Huang et al., 2020) and survival case discharge rate (Khafaie and Rahim, 2020; Bhatraju et al., 2020) for recovery-specific characteristic; ICU case rate (Bhatraju et al., 2020; Felice et al., 2020) and clinical deterioration rate (Felice et al., 2020) for worsening-specific characteristic, respectively. The details about those metrics calculation are listed in Table S1 and the geometric probability representation of the Bernouli Distribution in investigated pandemic metrics is drawn on Figure S1.

2.3 Pandemic Metrics with the Confidence (PMC) Model

To guarantee a fine match between the reported pandemic metrics and the actural pandemic trajectory, and to provide a probability that the reported metrics could be trusted, a PMC model is developed to predict the reported metric with certain confidence (Figure 1), and the area of blue in Figure 1-a represents the according confidence with the allowable maximum of error (Figure 1-b). Consequently, the PMC model could be expressed with the function of $\bar{x}$, n, and ε in Equation 1, where $1-\alpha$ is the final output of confidence, $\bar{x}$ is the mean of all sample values for each metric, n is the sample size, and ε is the allowable maximum of error. The detailed formula derivation can be found in the file of Supporting Information (Equation S1~S6).

$$\begin{cases} 1-\alpha = \frac{z_{\alpha/2}+0.2021}{1.972} \\ z^2_{\alpha/2} = \frac{\sqrt{b^2-4ac}-b}{2a} \\ a = 1-\varepsilon^2 \\ b = 4n\bar{x}(1-\bar{x}) - 2n\varepsilon^2 \\ c = -\varepsilon^2 n^2 \end{cases} \quad (1)$$

For instance, we set the allowable maximum of error as ε=1E-5 (see the detailed threshold selection in Section 3.1), and the reported value of the first metric (Table S1) in USA (reported date is 8-31, 2020) is about 7.57% (i.e., $\bar{x}$=7.57%) with the total number of 81,102,397 tests. Therefore, this reported data with the error that satisfies the configured threshold could be trusted with the confidence of only 18.88% (more results can be found in Section 3.3). Owing to this, it is significant to add the **confidence** and **errors** into the existing statistical conclusions and further relevant sciences which were (or would) drawn on those reported raw metrics.

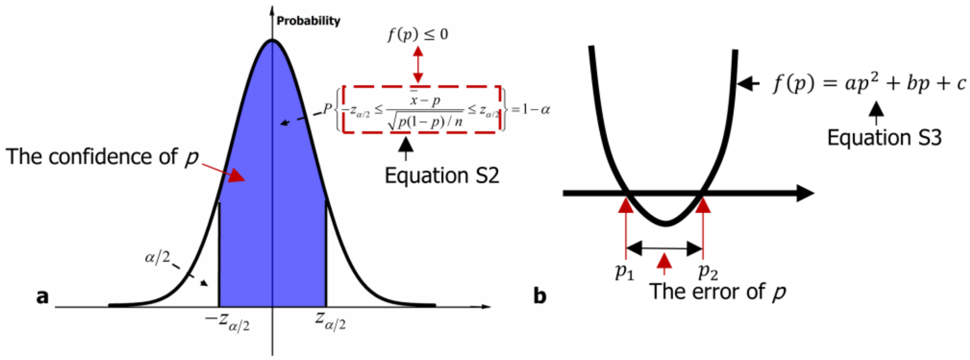


Figure 1. The data confidence and the allowable maximum error of the expected value of the investigated pandemic metrics. (**a**) shows the probability density curve of z statistical quantity where the area of blue represents the data confidence (the second equation in Equation S3); Besides, the black curve and two stationary points in (**b**) correspond to the first equation in Equation S3 and two boundaries of the confidence interval, respectively. The distance between two boundaries describes the allowable maximum of error.

3 Results

3.1 Relationship between Confidence and Error Threshold Configure of the Reported Pandemic Metrics

Figure 2 shows a response of the confidence for the reported pandemic metrics to the allowable maximum error configure. Except for metrics in Figure 2-a, -g, -h, the confidence of the other metrics with the 1E-05 of error threshold (red line) exhibits the notable differentiation across the global countries. Meanwhile, the confidence of the worldwide reported data exhibited the significant decrease when the error threshold to be more minor, especially in the decline from 1E-5 (red line) to 1E-06 (blue line). Whereafter, the confidence tends to be robust (insensitive) whatever the decline of the error threshold (1E-7 with green, 1E-8 with black, and 1E-9 with yellow line). Accordingly, the error threshold of setting in excess of 1E-5, it is difficult to distinguish the country-wise confidence for the reported data. Therefore, the error threshold of 1E-05 is preferred in this work.

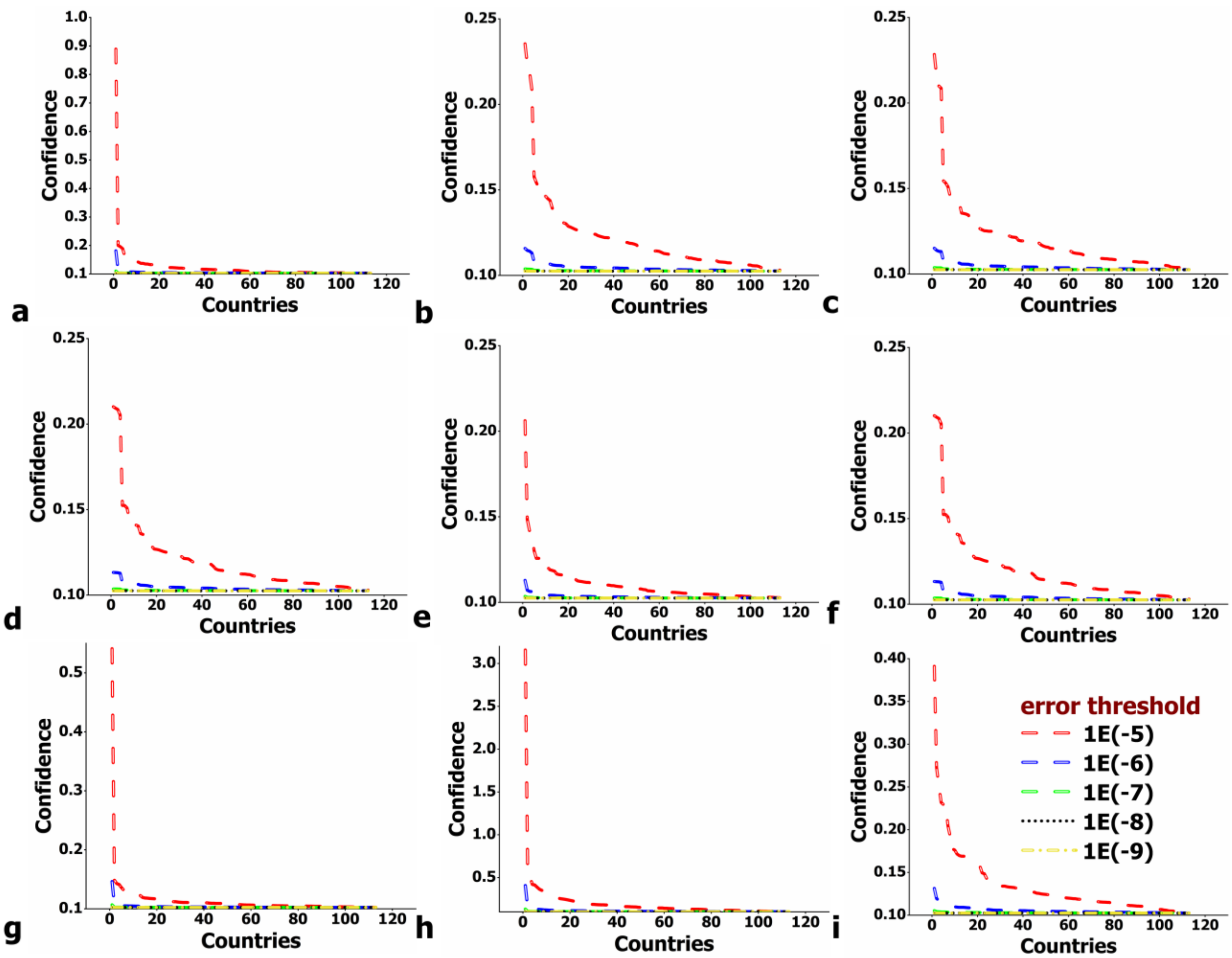


Figure 2. The response of the confidence for the reported pandemic metrics to the allowable maximum error configure (i.e., error threshold), such as 1E-5, 1E-6, 1E-7, 1E-8 and 1E-9 which individually corresponds to the red, blue, green, black and yellow dotted line in each subfigure (Figure 2-i shares the legend with the other subfigures). **a**) represents the aforementioned response for the proportion of the positive tests, **b**) for the case fatality rate, **c**) for the mortality rate, **d**) for the closed case fatality rate, **e**) for the discharge rate, **f**) for the recovery rate, **g**) for the survival case discharge rate, **h**) for the ICU case rate and **i**) for the clinical deterioration rate, respectively. Onward, in each subfigure, the horizontal axes represents the number of 113 investigated countries and the vertical axes represents the confidence of each reported pandemic metric (calculated based on the method in Section 2.3).

3.2 Relationship between Confidence and Country-wise Test Strategy

Apart from the error threshold, the varying strategies to detect COVID-19 cases also effect the region-specific data confidence (Figure 3). It was observed that the confidence of all pandemic metrics raises as the COVID-19 detection pool widens. Meanwhile, circles on box charts distributed more dispersed along the x-axis direction of Figure 3, which reveals the consequently enlarged differentiation of data credibility among diverged countries with the widen COVID-19 detection pool. Ignoring the outliers beyond the red line in Figure 3-h and -i, we can observe the larger confidence and more notable region-wise differentiation contributed by the detecting strategy of 1E+8.

However, Figure S3 shows that no one country carryed out the detecting strategy of 1E+8 until the day of August-31, 2020. Accordingly, tuning up the detecting strategy to 1E+8 supports the prediction of more credicable pandemic trajectory.

Prior to the credicable prediction, Figure S4 also demonstrates the region-wise sensitivity of data confidence to the distinct COVID-19 detecting strategy. Specifically, the data confidence—concerning the pandemic metric -1, -8 and -9—reported from China exhibits the most sensitivity to the detecting strategy, respectively. Subsequently, regarding the pandemic metric from -3 to -7, the data confidence reported from Qatar indicates the most sensitivity. Moreover, Guatemala reveals the most sensitivity for the pandemic metric -10. Also, it could be observed the more detailed sequence of aforementioned sensitivity in Figure S4.

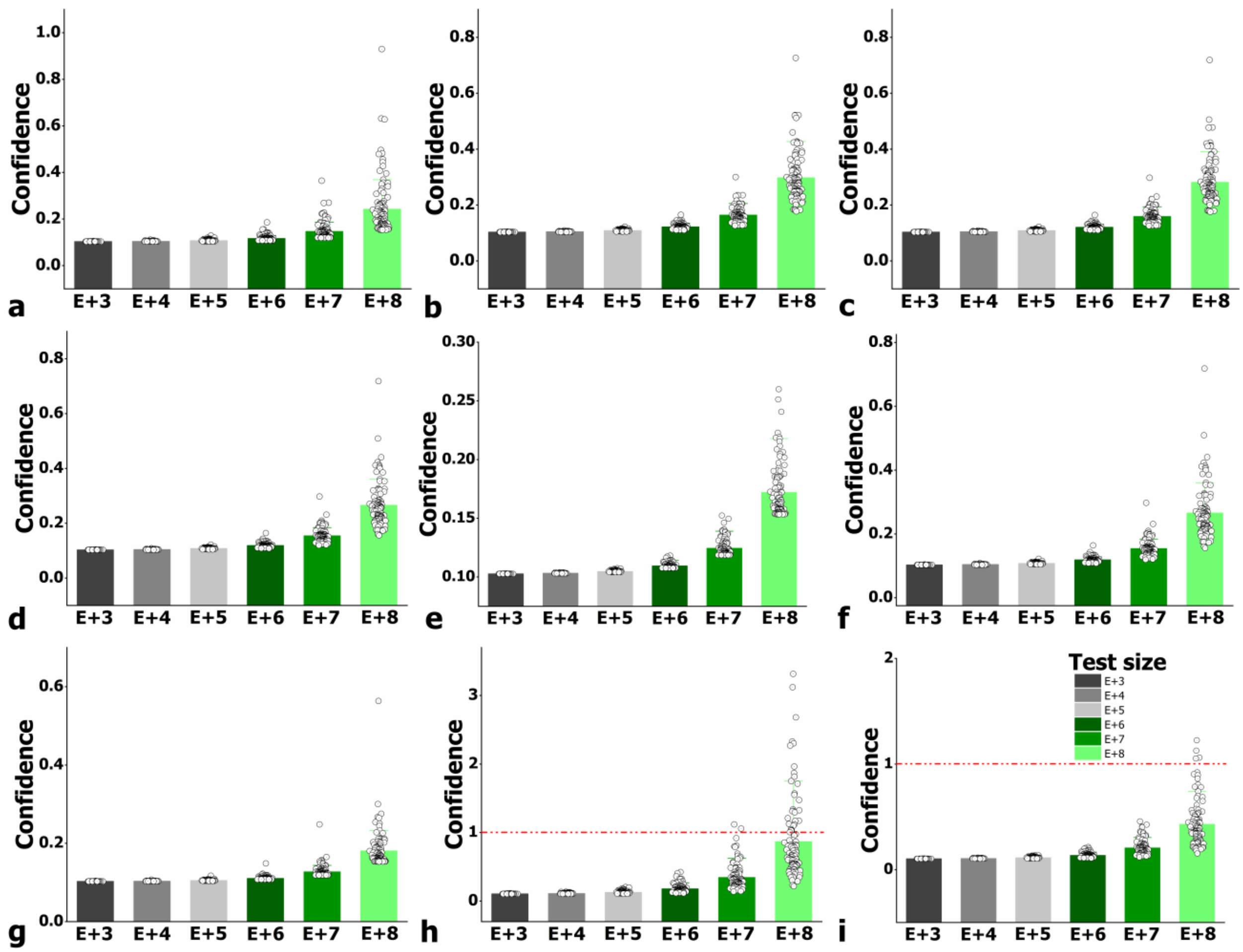


Figure 3. The response of the confidence for the reported pandemic metrics to the total test size, such as 1E+3, 1E+4, 1E+5, 1E+6, 1E+7 and 1E+8 which individually corresponds to the dark-black, moderate-black, grayish-black, dark-green, moderate-green and grayish-green boxes in each subfigure (Figure 3-i shares the legend with the other subfigures). Additionally, the vertical locations of the circles on each box chart exhibites the data confidence at each country (n=113). Specifically, **a**) represents the aforementioned response for the proportion of the positive tests, **b**) for the case fatality rate, **c**) for the mortality rate, **d**) for the closed case fatality rate, **e**) for the discharge rate, **f**) for the recovery rate, **g**) for the survival case discharge rate, **h**) for the ICU case rate and **i**) for the clinical deterioration rate, respectively.

3.3 Global Pandemic Metric with Data Confidence

Section 2.2 combined with Section 2.3 (set error threshold as 1E-5) supported the estimation of the data confidence for each reported pandemic metric across the global. Figure 4-a shows that almost all countries (China-except) exhibited the data confidence of 20% below for the first metric. Specifically, China exhibited the highest confidence of 89% for this metric, followed by Australia (20.0%), Russia (19.6%), USA (18.9%), Germany (16.1%), India (15.9%), and Jordan (15.0%), etc. whereas this metric value reported from the Turks and Caicos exhibited the lowest confidence (10.3%). Apart from that, a quite few countries (only about 4%) exhibited the data confidence in excess of 20% for the other three metrics in Figure 4-b, -c and -d, respectively. For instance, the confidence sequence of the metric in Figure 4-b is USA (23.5%), India (22.4%), Russia (21.9%) and China (20.9%) in turns. Analogously, Figure 4-c shows the sequence of USA (22.8%), Russia (21.6%), India (21.0%), and China (20.9%) in turns. However, Figure 4-d exhibited the sequence of India (21.0%), Russia (20.9%), China (20.8%), and USA (20.5%) in turns. As a whole, these three metrics exhibited the similar yet low confidence (24% below), and the reported metric values in Figure 4-d is much higher than the other two (Figure 4-b and -c).

Subsequently in Figure 4-e, -f and -g, three metrics (along with the estimated data confidence) in terms of the pandemic recovery are here. Likewise, metrics in Figure 4-e and -f exhibited the low confidence (21% below), whereas the maximum of the confidence for the metric in Figure 4-g was in excess of 54%, which could be, to some extent a preferable metric in the estimation of the global recovery state from the COVID-19 pandemic. Specifically, China exhibited the largest confidence of 54.08%, followed by USA (14.86%), Russia (14.31%), India (14.14%), Germany (13.61%), and Canada (13.16%), etc. It is worthy to mention that China (the red patch in Figure 4-h) exhibited the outlier of the confidence in excess of 1 with the error threshold of 1E-5. In this regard, combined with Figure 2-h, the allowable maximum of error for the metric in Figure 4-h should be replaced with 1E-6 for the reasonable confidence output. In the last metric (Figure 4-i), USA (green patch) exhibited the highest confidence of 39.11%, followed by China (yellow patch, 27.27%), India (orange patch, 25.23%), Russia (gray-green patch, 23.18%), Italy (green patch, 22.93%), and France (green patch, 22.90%), etc.

Despite that a few countries exhibited a relatively high confidence for the certain reported metrics, its average value is not in excess of 12.1%, i.e., metric in Figure 4-a (mean=12.36%), -b (mean=12.13%), -c (mean=12.02%), -d (mean=11.91%), -e (mean=11.04%), -f (mean=11.91%), -g (mean=11.40%), and -i (mean=13.39%), respectively. Hereby, to estimate and predict the pandemic trajectory in terms of those metrics with high confidence is quite necessary and significant.

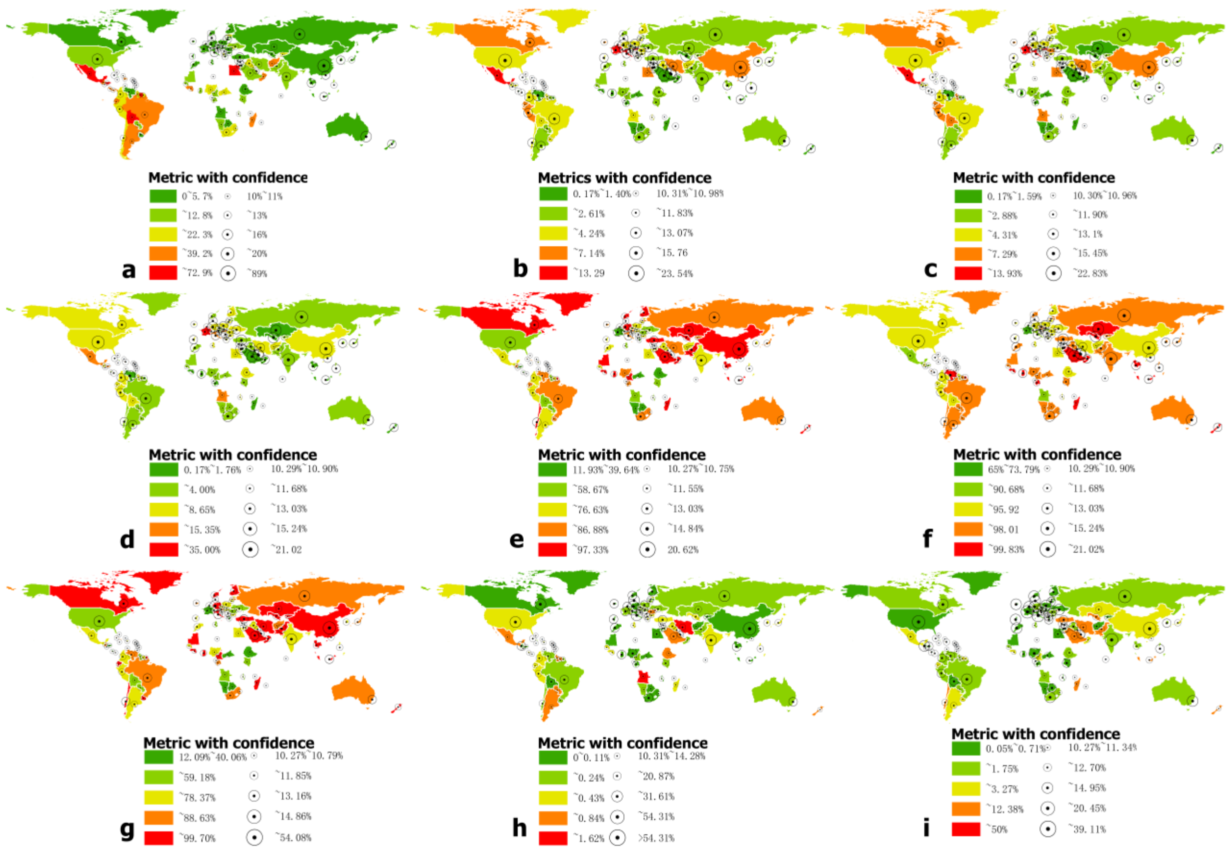


Figure 4. The nine of reported pandemic metrics along with the corresponding data confidence in our research region (i.e., 113 countries in map), where the graduated colors at each subfigure represent the value of certain pandemic metric reported from a specific country, and the circle size on each colored-patch represents the corresponding confidence. Onward, the detailed value could be linked to each legend, in which the first colum and the second column exhibited the value range of the certain pandemic metric and the data confidence, respectively. Specifically, **a**) represents the proportion of the positive tests with the data confidence, **b**) for the case fatality rate, **c**) for the mortality rate, **d**) for the closed case fatality rate, **e**) for the discharge rate, **f**) for the recovery rate, **g**) for the survival case discharge rate, **h**) for the ICU case rate and **i**) for the clinical deterioration rate, respectively. Note, the color sequence (top to down) in legend are describe as green, gray-green, yellow, orange, and red, respectively in the main text.

4 Discussion

4.1 Impossible to Predict Pandemic Trajectory with the High Confidence and the Minor Error Simultaneously

Based on the PMC model in Section 2.3, given that the allowable maximum of error is 1E-5, the confidence (credibility) of nine investigated pandemic metrics, which were reported from 113 countries respectively, was predicted in Section 3.3. The result exhibited the average confidence not in excess of 12.1% for all the COVID-19 pandemic metric (Figure 4). Owing to this, the setting of the allowable maximum error of 1E-5 (minor error) failed to predict the pandemic trajectory with the high credibility. Meanwhile, the setting of the confidence of 95% (high credibility) also failed to predict the trajectory with the minor error (Figure 5). For instance, the trajectory simulated with the tenth metric in Antigua and Barbuda was predicted with the the highest error of 76.34% across the global (Figure 5-i), followed by the eighth (11.18%, Figure 5-g) and sixth metric (11%, Figure 5-e) in Saint Martin, then the fifth and seventh metric (10.88%) in Martinique (Figure 5-d and -f), the third (6.62%, Figure 5-c) and fourth metric (6.55%, Figure 5-b) in Antigua and Barbuda, the ninth metric in Saint Martin (5.29%, Figure 4-h), and the first metric in Turks and Caicos (2.78%, Figure 5-a). Moreover, China exhibited the minimal error of 1.08E-5 and 0.06% in Figure 5-a and –g, while India presented 0.03% in Figure 5-d and –f, and USA displayed 0.02%, 0.03%, 0.07% and 7.1E-5 in Figure 5-b, -c, -e, and –h, respectively.

What will happen if the confidence of 95% (high credibility) and the allowable maximum error of 1E-5 (minor error) are together input into the PMC model to predict the pandemic trajectory simulated with different metrics? The result exhibited the predicted values in excess of 99% for all metrics across the global (Figure S5), which is quite an impossible outcome in the pandemic trajectory. Meanwhile, the predicted sample size for each pandemic trajectory simulated by aforementioned metrics is faraway the real in the condition of 95% confidence and the allowable maximum error of 1E-5 (Figure S6). For example, the pandemic trajectory (simulated by the first metric) with 95% of confidence and 1E-5 of error existed only in the hypothesis that the average population of COVID-19 detection in excess of 9.49E+9 (Figure S6-a), which exceed the total population on the earth. Likewise, neigher of the existing pandemic trajectories simulated by metrics in Figure S6-b to -i satisfies such conditions. Owing to this, the expection of the high confidence and the low error in the pandemic trajectory predicted through the PMC model are contradictory, and we should make a trade off between them.

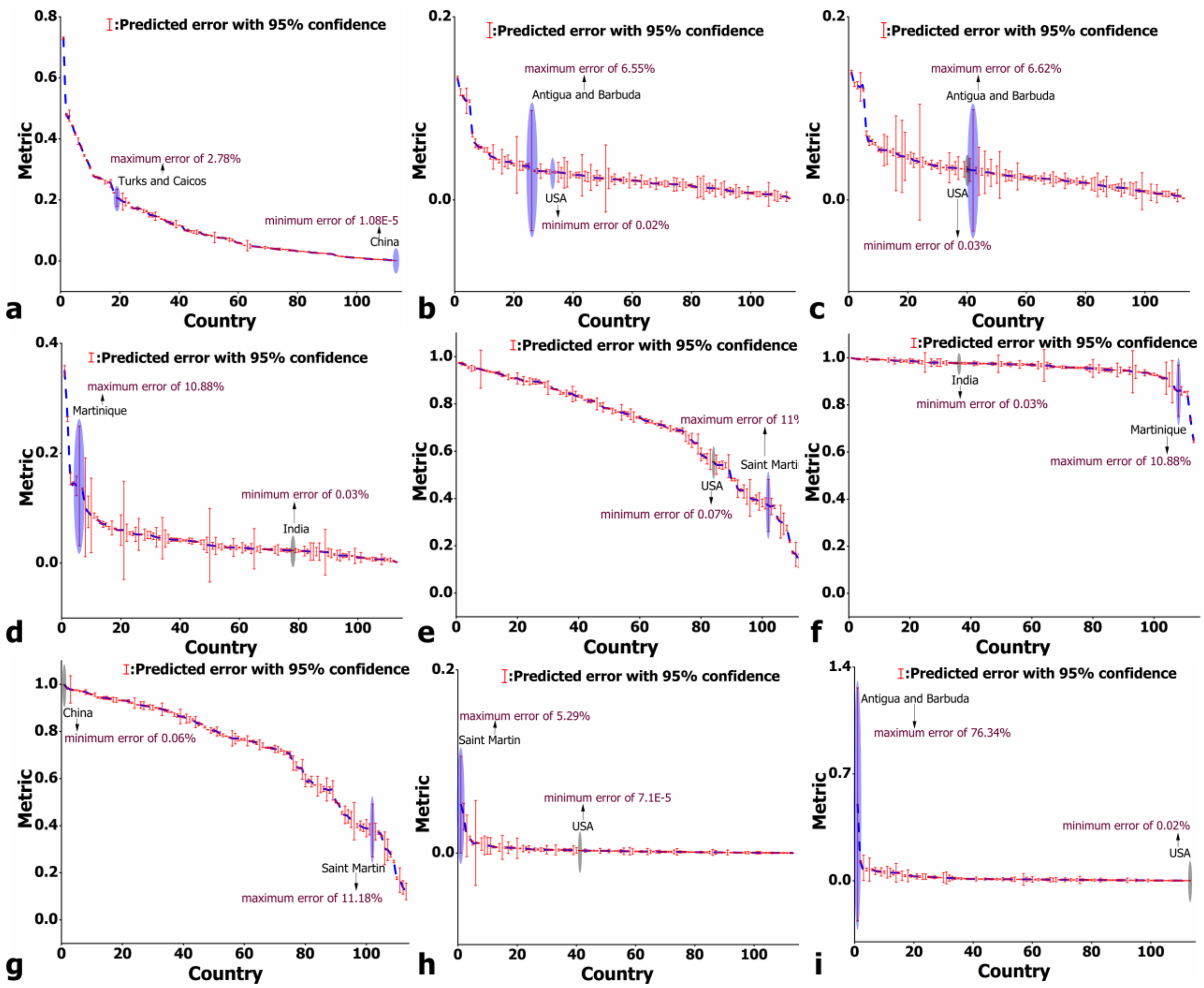


Figure 5. The error with the confidence of 95% which is predicted for each pandemic metric through the PMC model. Onward, the blue dotted line and the red error bar represent the country-specific reported value of each metric and the corresponding predicted errors, respectively. In each subfigure, the maximum and the minimum of error are marked with the ellipse, respectively.

4.2 Trade off between the Confidence and the Allowable Maximum of Error

Tracking back Section 3.1, despite the capacity of the setting of allowable maximum error of 1E-5 to distinguish the confidence of the simulated pandemic trajectory across the global (Figure 2), the predicted confidence not in excess of 12.1% is impossible to support any futher conclusions drawn on those trajectories. Owing to this, a trade off between those two contradictory expections is significant. Subsequently, the acceptable confidence (>50%) with the maximum error not in excess of 1E-3 was predicted in Figure 6 based on the trade-off strategy. We observe the maximum error scope of 1E-5 to 1E-3 supports the prediction of acceptable confidence in most of countries. Setting the maximum error of 1E-5, reported data (0.09%) only from China with respect to the first pandemic metric exhibits the acceptable confidence (e.g., 88.90%). However, with the boost of maximum error to 9E-5 , eight countries support the prediction of acceptable confidence, such as Australia, Russia, USA, Germany, India, Jordan, Denmark and New

Zealand. Besides, the maximum error setting of 1E-4 to 1E-3 with the interval of 1E-4 suppots the prediction of acceptable confidence of 7, 17, 26, 34, 27, 23, 20, 18, 13 and 13 countries, respectively.

On the contrary, due to the current low proportion of country-specific populaces comprised in the COVID-19 detection pool, the remaining 44 countries fail to support the trade-off strategy between the confidence beyond 50% and the maximum error within 1E-3. However, the upsurge of the percentage of populaces as the member of the COVID-19 detection pool could make those countries support the trade-off strategy. Specifically, fixing the maximum error of 1E-3, the raise of detecting-rate to 1% aids Guatemala and Oman give prediction of the confidence in excess of 50%, that is 85.14% and 68.08%, respectively. Onward, the promotion of detecting-rate to 2% also helps the other five countries to forecast the confidence beyond 50%, such as Senegal (60.74%), Cameroon (65.57%), Afghanistan (56.60%), Malawi (57.80%) and Madagascar (53.44%). Subsequently, there are more countries (the number of 17, 7 and 7 countries) with the predicted confidence upstairing over to 50% since the upgrade of detecting-rate to 10%, 20% and 100%, respectively.

Nevertheless, eight regions still cannot predict the confidence in excess of 50% even with all populaces as the member of the COVID-19 detection pool, such as French Guiana, Aruba, Mayotte, Antigua and Barbuda, Suriname, Saint Martin, Sint Maarten and Turks and Caicos. To promote the trade-off strategy work, the maximum error is tuned up to 2E-3 in five regions and to 3E-3 in three regions. Consequently, Suriname and French Guiana separately exhibit 88.08% and 83.78% of data confidence along with the maximum error of 2E-3, followed by Antigua and Barbuda (83.02%), Mayotte (71.22%) and Aruba (66.87%). Also, given the maximum error of 3E-3, Saint Martin, Sint Maarten and Turks and Caicos show 95.19%, 66.33% and 50.07% of data confidence, respectively. Similarily, the trade-off strategy could also be applied in other pandemic metrics shown in Table S1.

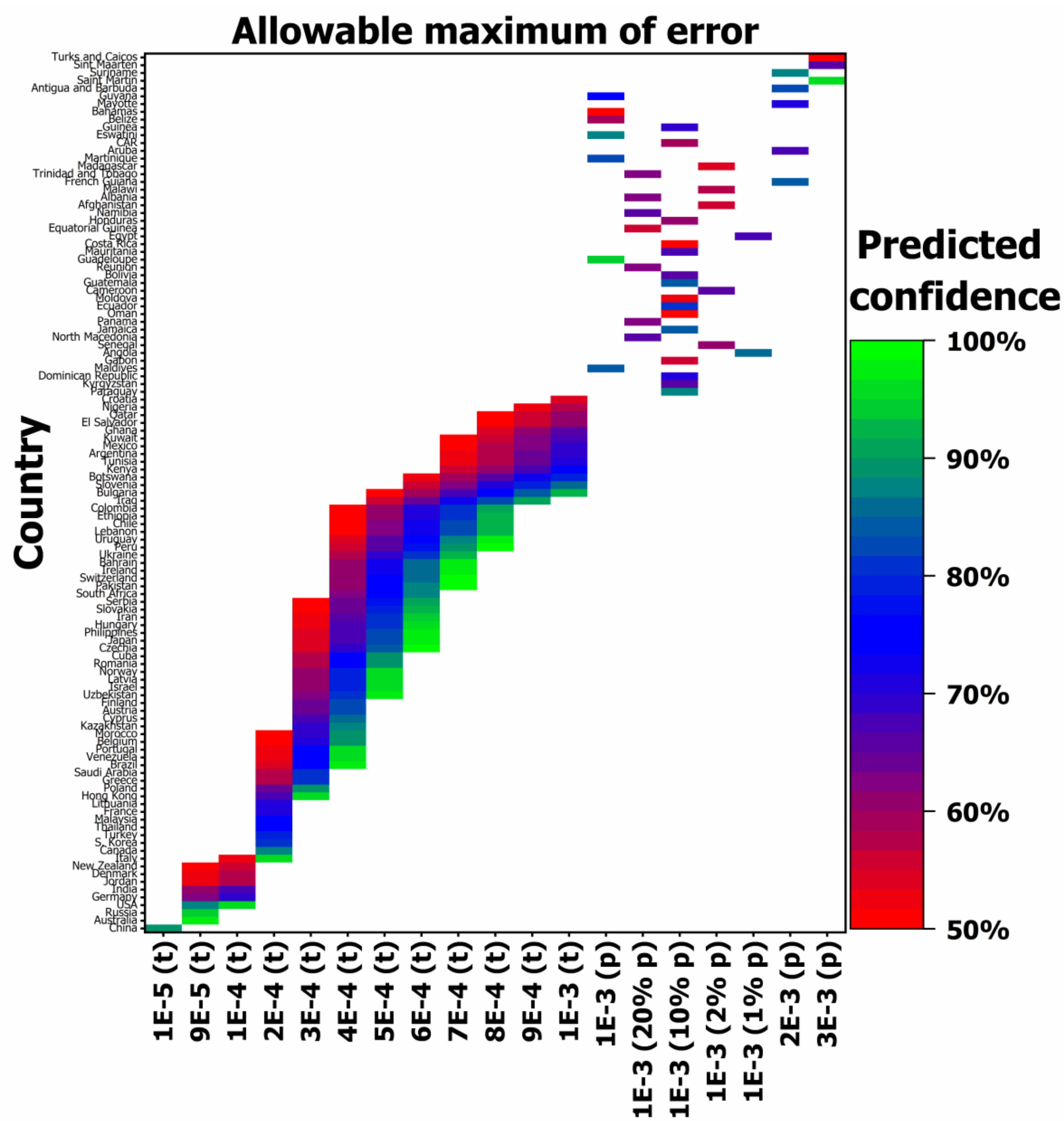


Figure 6. The trade off between the confidence and the allowable maximum error for the first pandemic metric. The different color represents the predicted confidence (i.e., exceeds 50%) that is acceptable in the trade off strategy. Onward, values on the horizontal axis describes the tunable adaptive setting of the allowable maximum of error. Moreover, the term #% *p* and *t* in the bracket represent the denominator of the first pandemic metric (Table S1), that is the number of certain percentage of populace, and the number of populace who participated the COVID-19 detection, respectively. Onward, the vertical axis exhibits the name of 113 investigated countries / regions.

5 Conclusions

Many works aim to reveal the driving factors of COVID-19 pandemic trajectory yet ignore the data confidence of utilized trajectory. In this work, we proposed the PMC model in the hypothesis of Bernoulli Distribution of nine utilized trajectories announced from 113 countries. The **first merit** of PMC model is to predict the region-specific confidence of utilized nine trajectories and the predicted average confidence across the global is not in excess of 12.1% with the error threshold configuration of 1E-5, cutting the reliability of relevant researchs. **Another merit** of PMC model is to dig the real trajectory data with the certain confidence and error threshold, but the purpose of the high confidence (>95%) and the minor error (<1E-5) threshold is impossible to achieve simultaneously. Therefore, a proposed trade-off strategy between those two contradictory expections (>50% confidence; <1E-3 error) supports 61% of investigated countries to predict the varying trajectory with confidence in excess of 50%. The **third merit** of PMC model is to recommend the remanent 39% countries to extend the proportion of populaces in COVID-19 detecting-pool to a suggested-value (at least 1% of populations), ensuing the average confidence up to 70%. Finally, PMC model in combination with trade-off strategy jointly reveal that error threshold within scope of 1E-5 to 1E-3, confidence in excess of 50%, and the number of populaces participating COVID-19 detecting pool beyond 1E+8 together produce the optimal prediction of the reliable trajectory data. It is significant to add the **confidence** and **error** concepts into the existing statistical conclusions and future relevant studies.

Nevertheless, the study also has a limitation that the hypothesis of Bernoulli Distribution of pandemic trajectory in PMC model brings some uncertainty to the predicted outcomes of the confidence and according trajectory. Apart from extending PMC model to global-city level in COVID-19 pandemic trajectory, a more difficult task that awaits our community is using reliable trajectory data predicted by PMC model to reveal the real climatic role in this urgent human public health events.

Acknowledgments

We thank the funding from the National Key R & D Program of China (2017YFB0503003).

Data Availability Statement

All original and intermediate data for this work are available in the Zenodo repository (PMC-2021. (2021, May 28). PKU-2021/PMC-Model: PMC-Model (Version v1.0.0). Zenodo. http://doi.org/10.5281/zenodo.4831821).

Conflict of Interest

The authors declare no conflicts of interest. Moreover, all authors emphasize that Taiwan province is a part of the People's Republic of China, and the reason why Taiwan province did not be drawn in the map of China (Figure 4) is due to the difficulty of Taiwan related data collection.

References

Ahmadi, M., et al. (2020), Investigation of effective climatology parameters on COVID-19 outbreak in Iran. *Science of The Total Environment*, 138705. https://doi.org/10.1016/j.scitotenv.2020.138705.

Baud, D., Qi, X., Nielsen-Saines, K., Musso, D., Pomar, L., et al. (2020). Real estimates of mortality following COVID-19 infection. *The Lancet infectious diseases*, 7(20): 773. https://doi.org/10.1016/S1473-3099(20)30195-X.

Bhatraju, P.K., Ghassemieh, B.J., Nichols, M., Kim, R., Jerome, K. R., Nalla, A. K., et al. (2020). Covid-19 in critically ill patients in the Seattle region—case series. *New England Journal of Medicine*, 382(21), 2012-2022. https://doi.org.10.1056/NEJMoa2004500.

Cordes, J., Castro, M.C. (2020). Spatial analysis of COVID-19 clusters and contextual factors in New York City. *Spatial and Spatio-temporal Epidemiology*, 34, 100355. https://doi.org/10.1016/j.sste.2020.100355.

Freeman, W. M., Walker, S. J., & Vrana, K. E. (1999). Quantitative RT-PCR: pitfalls and potential. *Biotechniques*, 26(1), 112-125. https://doi.org/10.2144/99261rv01.

Felice, C., Nardin, C., Di Tanna, G.L., Grossi, U., Bernardi, E., Scaldaferri, L., et al. (2020). Use of RAAS inhibitors and risk of clinical deterioration in COVID-19: results from an Italian cohort of 133 hypertensives. *American Journal of Hypertension*. https://doi.org/10.1093/ajh/hpaa096.

Gallo Marin, B., Aghagoli, G., Lavine, K., Yang, L., Siff, E. J., Chiang, S. S., ... & Michelow, I. C. (2021). Predictors of COVID-19 severity: A literature review. *Reviews in medical virology*, 31(1), 1-10. https://doi.org/10.1002/rmv.2146.

Khafaie, M.A., Rahim, F. (2020). Cross-country comparison of case fatality rates of COVID-19/SARS-COV-2. *Osong Public Health and Research Perspectives*, 11(2), 74. https://doi.org/10.24171/j.phrp.2020.11.2.03.

Lai, S., Ruktanonchai, N.W., Zhou, L., Prosper, O., Luo, W., Floyd, J.R., et al. (2020). Effect of non-pharmaceutical interventions to contain COVID-19 in China. *Nature*. http://nrs.harvard.edu/urn-3:HUL.InstRepos:42661263.

Luo, W., Majumder, M., Liu, D., Poirier, C., Mandl, K., Lipsitch, M., & Santillana, M. (2020). The role of absolute humidity on transmission rates of the COVID-19 outbreak. *medRxiv*. https://doi.org/10.1101/2020.02.12.20022467.

Liu, Y., Ning, Z., Chen, Y. et al. (2020), Aerodynamic analysis of SARS-CoV-2 in two Wuhan hospitals. *Nature*. https://doi.org/10.1038/s41586-020-2271-3.

Li, Y., Yao, L., Li, J., et al. (2020). Stability issues of RT-PCR testing of SARS-CoV-2 for hospitalized patients clinically diagnosed with COVID-19. *J Med Virol*. https://doi.org/10.1002/jmv.25786.

Li, L.Q., Huang, T., Wang, Y.Q., Wang, Z.P., Liang, Y., Huang, T.B., et al. (2020). COVID-19 patients' clinical characteristics, discharge rate, and fatality rate of meta-analysis. *Journal of medical virology*, 92(6), 577-583. https://doi.org/10.1002/jmv.25757.

Ma, Y., Zhao, Y., Liu, J., He, X., Wang, B., Fu, S., ... & Luo, B. (2020). Effects of temperature variation and humidity on the death of COVID-19 in Wuhan, China. *Science of the total environment*, 724, 138226. https://doi.org/10.1016/j.scitotenv.2020.138226.

Morrison, C.N., Rundle, A.G., Branas, C.C., Chihuri, S., Mehranbod, C., Li, G. (2020). The Unknown Denominator Problem in Population Studies of Disease Frequency. *Spatial and Spatio-temporal Epidemiology*, 100361. https://doi.org/10.1016/j.sste.2020.100361.

Onder, G., Rezza, G., Brusaferro, S. (2020). Case-Fatality Rate and Characteristics of Patients Dying in Relation to COVID-19 in Italy. *JAMA*. 323(18):1775–1776. https://doi.org/10.1001/jama.2020.4683.

Pueyo, T. Coronavirus: Why you must act now. Politicians, community leaders and business leaders: what should you do and when? 2020 Mar 10. https://medium.com/@tomaspueyo/coronavirus-act-today-or-people-will-die-f4d3d9cd99ca.

Sima A., et al. (2020). The coronavirus pandemic and aerosols: Does COVID-19 transmit via expiratory particles? *Aerosol Science and Technology*, 54:6, 635-638. https://doi.org/10.1080/02786826.2020.1749229.

Tian, H., Liu, Y., Li, Y., Wu, C. H., Chen, B., Kraemer, M.U., et al. (2020). An investigation of transmission control measures during the first 50 days of the COVID-19 epidemic in China. *Science*, 368(6491), 638-642. http://doi.org/10.1126/science.abb6105.

Tian, S., Hu, N., Lou, J., Chen, K., Kang, X., Xiang, Z., et al. (2020). Characteristics of COVID-19 infection in Beijing. *Journal of Infection*, 80.4, 401-406. https://doi.org/10.1016/j.jinf.2020.02.018.

Wang, M., Jiang, A., Gong, L., Luo, L., Guo, W., Li, C., ... & Chen, Y. (2020). Temperature significant change COVID-19 Transmission in 429 cities. *MedRxiv*. https://doi.org/10.1101/2020.02.22.20025791.

Wang, J., Tang, K., Feng, K., & Lv, W. (2020). High temperature and high humidity reduce the transmission of COVID-19. Available at SSRN 3551767. https://doi.org/10.2139/ssrn.3551767.

Wang, J., Du, G. Q. (2020), COVID-19 may transmit through aerosol. *Irish Journal of Medical Science*, 1-2. https://doi.org/10.1007/s11845-020-02218-2.

Wu, J., Liu, J., Li, S., Peng, Z., Xiao, Z., Wang, X., Luo, J. (2020). Detection and analysis of nucleic acid in various biological samples of COVID-19 patients. *Travel medicine and infectious disease*, 101673. https://doi.org/10.1016/j.tmaid.2020.101673.

Xiao, A.T., Tong, Y.X., Zhang, S. (2020). False-negative of RT-PCR and prolonged nucleic acid conversion in COVID-19: rather than recurrence. *Journal of medical virology*. https://doi.org/10.1002/jmv.25855.

You, Y., & Pan, S. (2020). Urban Vegetation Slows Down the Spread of Coronavirus Disease (COVID-19) in the United States. *Geophysical Research Letters*, 47(18), e2020GL089286. https://doi.org/10.1029/2020GL089286.

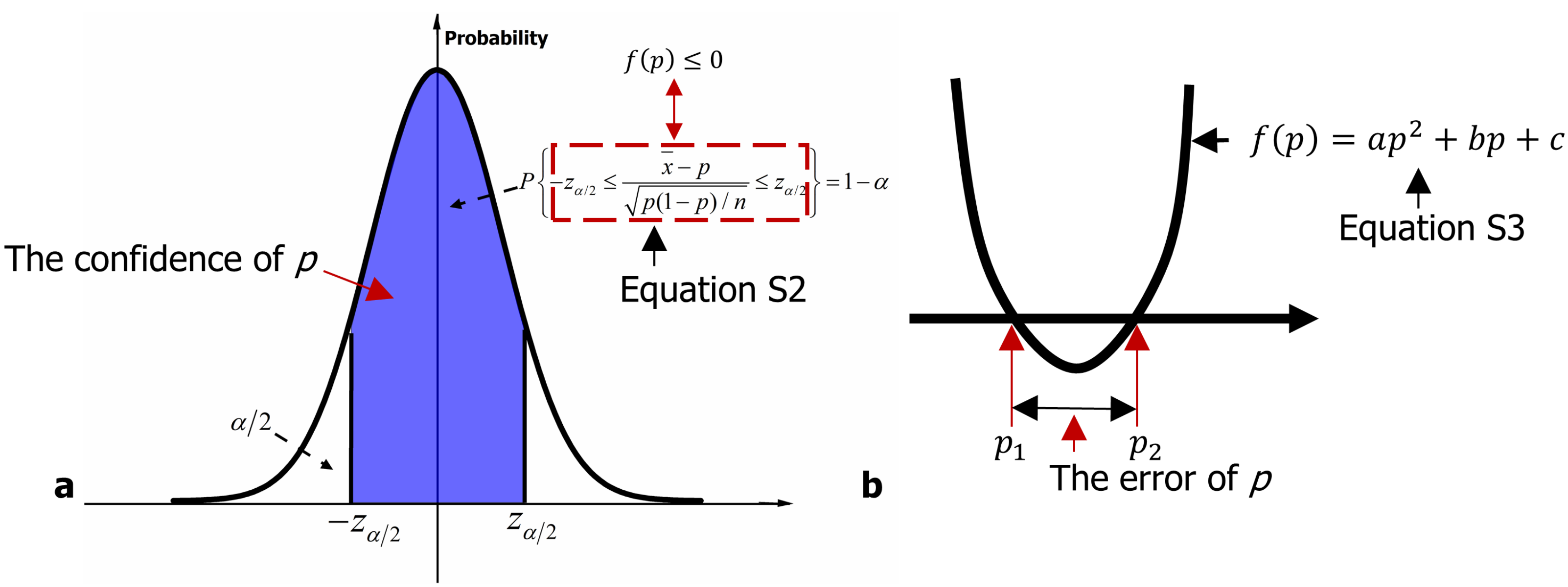
Yao, Y., et al. (2020), No Association of COVID-19 transmission with temperature or UV radiation in Chinese cities. *European Respiratory Journal*. https://doi.org/10.1183/13993003.00517-2020.

Yang, S., Cao, P., Du, P., Wu, Z., Zhuang, Z., Yang, L., ... & Li, W. (2020). Early estimation of the case fatality rate of COVID-19 in mainland China: a data-driven analysis. *Annals of translational medicine*, 8(4). https://doi.org/10.21037/atm.2020.02.66.

Zuo, Z., Ullah, S., Yan, L., Sun, Y., Peng, F., Jiang, K., & Zhao, H. (2021). Trajectory Simulation and Prediction of COVID-19 via Compound Natural Factor (CNF) Model in EDBF Algorithm. *Earth's Future*, 9(4), e2020EF001936. https://doi.org/10.1029/2020EF001936.

Zhuang, G. H., et al. (2020). WITHDRAWN: Potential false-positive rate among the'asymptomatic infected individuals' in close contacts of COVID-19 patients. *Zhonghua liu xing bing xue za zhi*, 41.4, 485-488. https://doi.org/10.3760/cma.j.cn112338-20200221-00144.

Figure 1.



Probability
$f(p) \le 0$
$P\left\{-z_{\alpha/2} \le \frac{\bar{x}-p}{\sqrt{p(1-p)/n}} \le z_{\alpha/2}\right\} = 1-\alpha$
Equation S2
The confidence of p
$\alpha/2$
a
$-z_{\alpha/2}$
$z_{\alpha/2}$
b
$f(p) = ap^2 + bp + c$
Equation S3
p_1
p_2
The error of p

Figure 2.

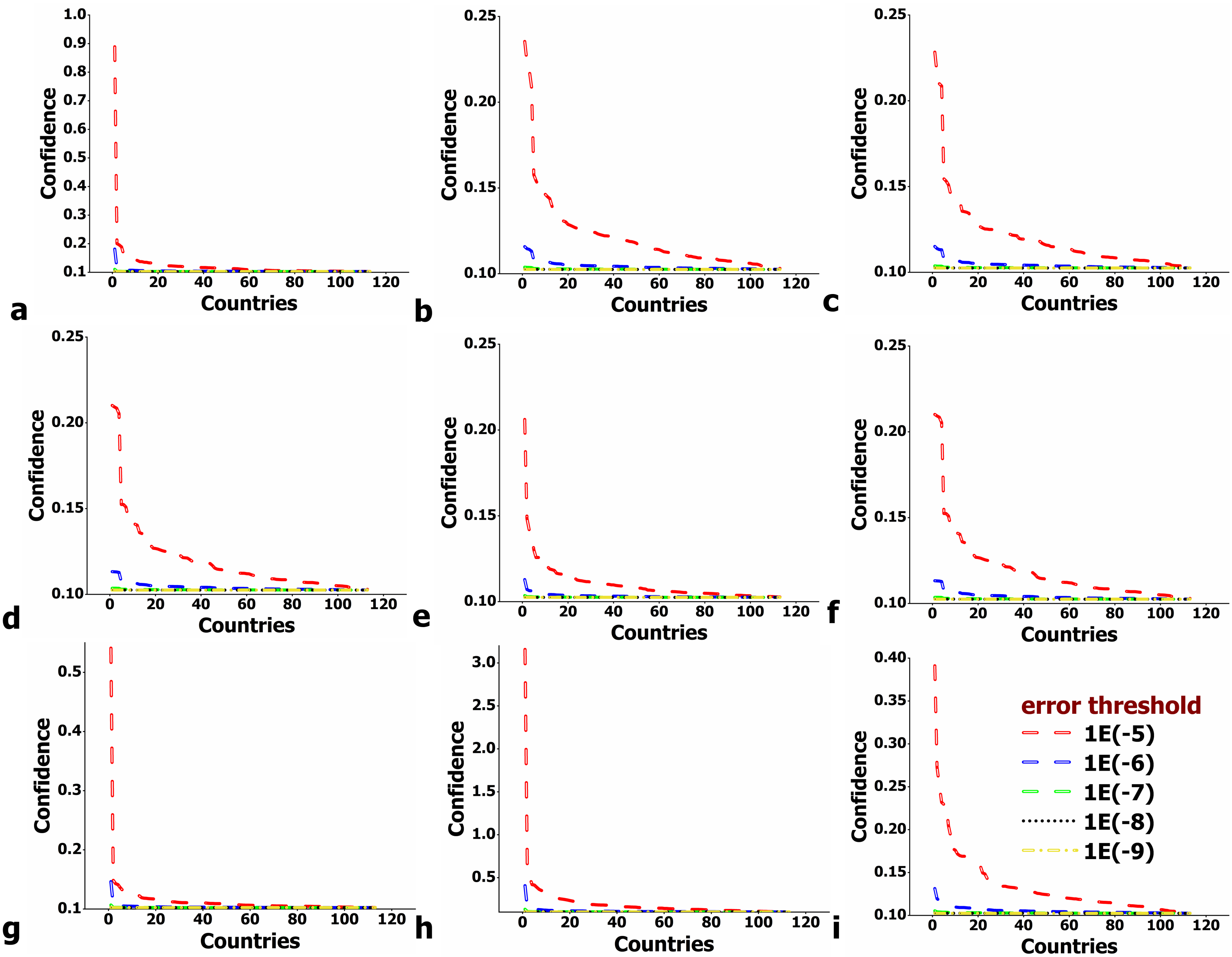

Confidence
Countries
error threshold
1E(-5)
1E(-6)
1E(-7)
1E(-8)
1E(-9)
a
b
c
d
e
f
g
h
i

Figure 3.

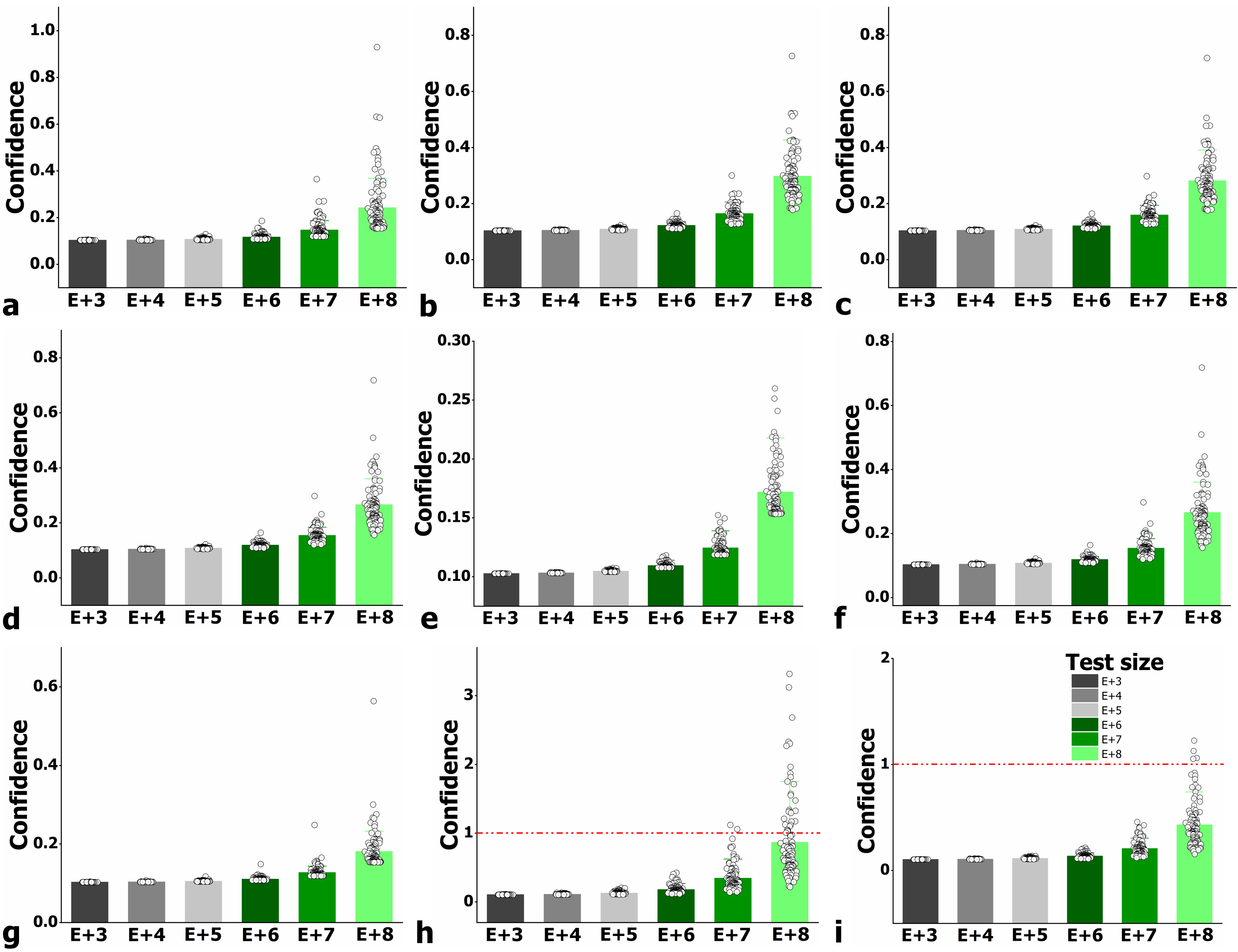
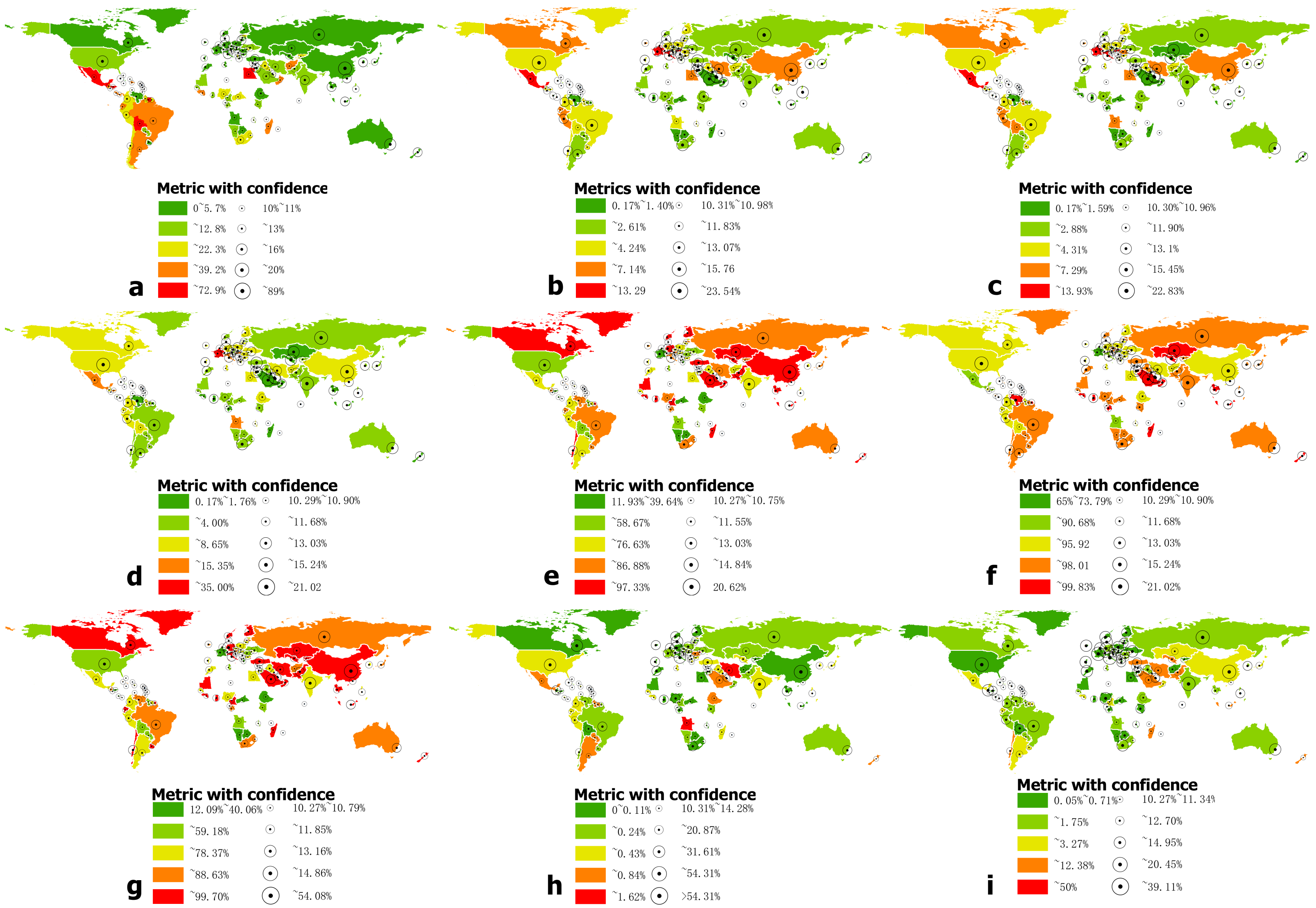
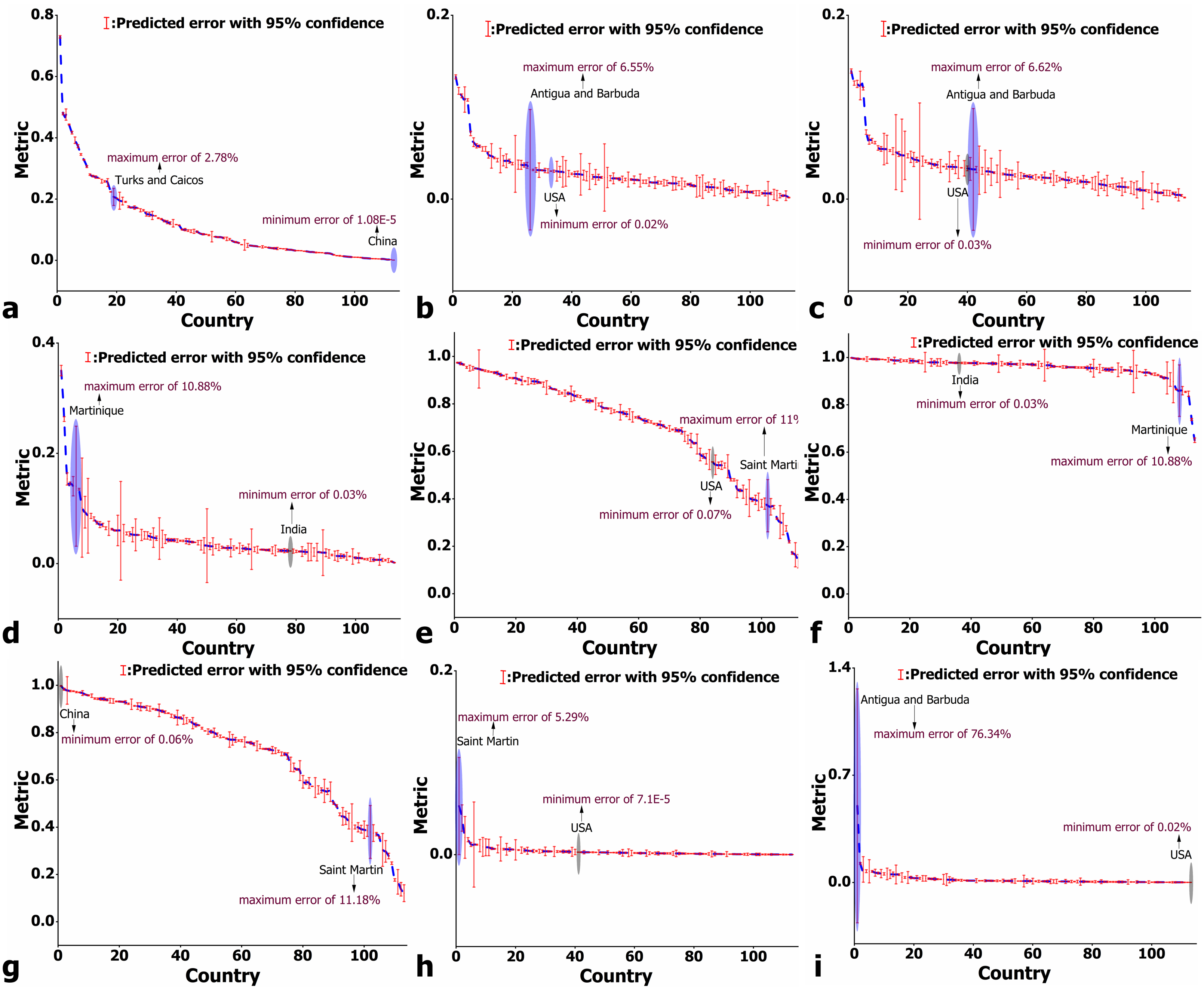

Confidence
E+3
E+4
E+5
E+6
E+7
E+8
a
b
c
d
e
f
g
h
i
Test size

Figure 4.



Metric with confidence
0~5. 7%
10%~11%
~12. 8%
~13%
~22. 3%
~16%
~39. 2%
~20%
~72. 9%
~89%
a
Metrics with confidence
0. 17%~1. 40%
10. 31%~10. 98%
~2. 61%
~11. 83%
~4. 24%
~13. 07%
~7. 14%
~15. 76
~13. 29
~23. 54%
b
Metric with confidence
0. 17%~1. 59%
10. 30%~10. 96%
~2. 88%
~11. 90%
~4. 31%
~13. 1%
~7. 29%
~15. 45%
~13. 93%
~22. 83%
c
Metric with confidence
0. 17%~1. 76%
10. 29%~10. 90%
~4. 00%
~11. 68%
~8. 65%
~13. 03%
~15. 35%
~15. 24%
~35. 00%
~21. 02
d
Metric with confidence
11. 93%~39. 64%
10. 27%~10. 75%
~58. 67%
~11. 55%
~76. 63%
~13. 03%
~86. 88%
~14. 84%
~97. 33%
20. 62%
e
Metric with confidence
65%~73. 79%
10. 29%~10. 90%
~90. 68%
~11. 68%
~95. 92
~13. 03%
~98. 01
~15. 24%
~99. 83%
~21. 02%
f
Metric with confidence
12. 09%~40. 06%
10. 27%~10. 79%
~59. 18%
~11. 85%
~78. 37%
~13. 16%
~88. 63%
~14. 86%
~99. 70%
~54. 08%
g
Metric with confidence
0~0. 11%
10. 31%~14. 28%
~0. 24%
~20. 87%
~0. 43%
~31. 61%
~0. 84%
~54. 31%
~1. 62%
>54. 31%
h
Metric with confidence
0. 05%~0. 71%
10. 27%~11. 34%
~1. 75%
~12. 70%
~3. 27%
~14. 95%
~12. 38%
~20. 45%
~50%
~39. 11%
i

Figure 5.



a
:Predicted error with 95% confidence
maximum error of 2.78%
Turks and Caicos
minimum error of 1.08E-5
China
Metric
Country
b
:Predicted error with 95% confidence
maximum error of 6.55%
Antigua and Barbuda
USA
minimum error of 0.02%
Metric
Country
c
:Predicted error with 95% confidence
maximum error of 6.62%
Antigua and Barbuda
USA
minimum error of 0.03%
Metric
Country
d
:Predicted error with 95% confidence
maximum error of 10.88%
Martinique
minimum error of 0.03%
India
Metric
Country
e
:Predicted error with 95% confidence
maximum error of 11%
Saint Marti
USA
minimum error of 0.07%
Metric
Country
f
:Predicted error with 95% confidence
India
minimum error of 0.03%
Martinique
maximum error of 10.88%
Metric
Country
g
:Predicted error with 95% confidence
China
minimum error of 0.06%
Saint Martin
maximum error of 11.18%
Metric
Country
h
:Predicted error with 95% confidence
maximum error of 5.29%
Saint Martin
minimum error of 7.1E-5
USA
Metric
Country
i
:Predicted error with 95% confidence
Antigua and Barbuda
maximum error of 76.34%
minimum error of 0.02%
USA
Metric
Country

Figure 6.

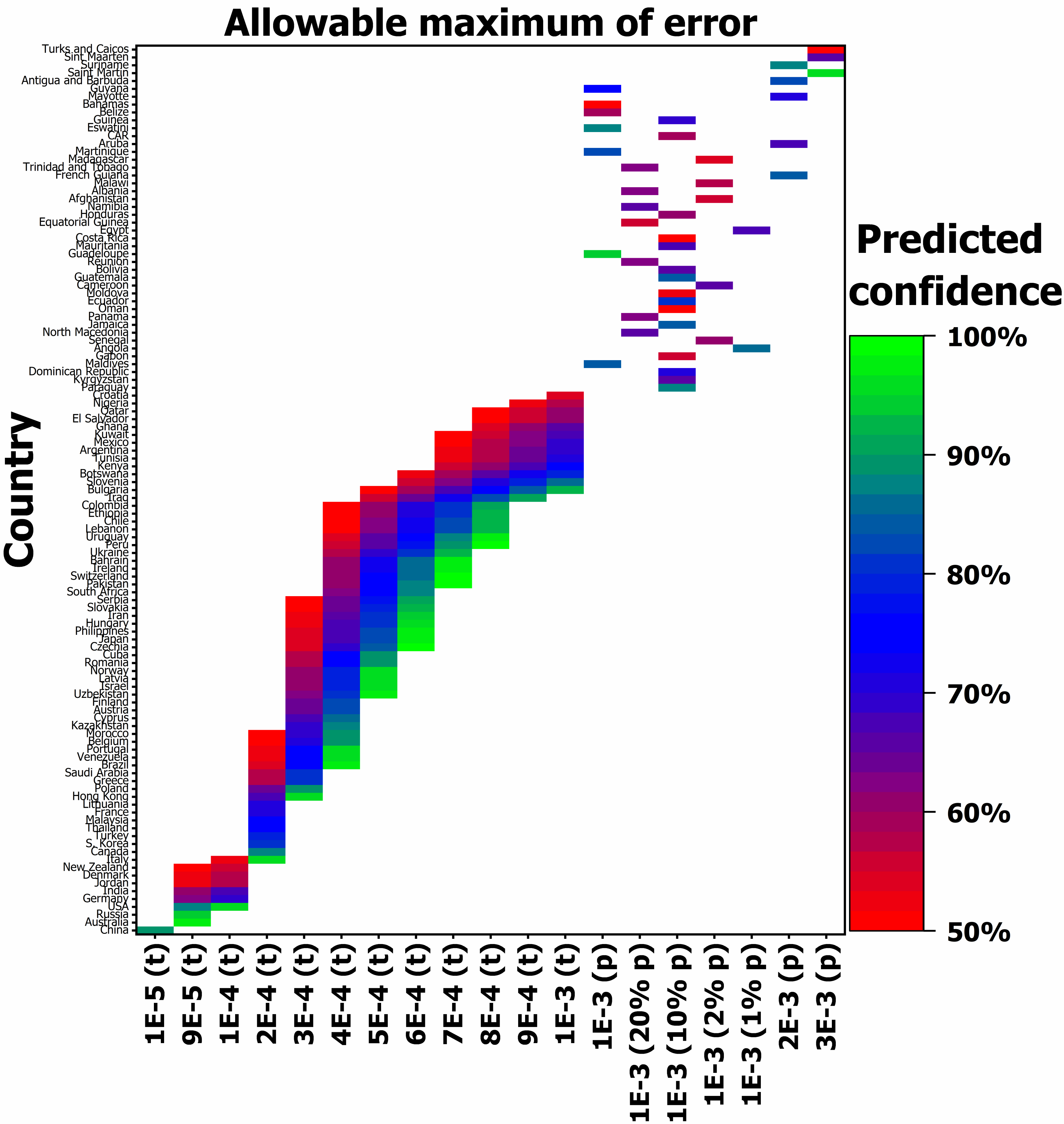

Allowable maximum of error
Country
Predicted confidence
100%
90%
80%
70%
60%
50%
1E-5 (t)
9E-5 (t)
1E-4 (t)
2E-4 (t)
3E-4 (t)
4E-4 (t)
5E-4 (t)
6E-4 (t)
7E-4 (t)
8E-4 (t)
9E-4 (t)
1E-3 (t)
1E-3 (p)
1E-3 (20% p)
1E-3 (10% p)
1E-3 (2% p)
1E-3 (1% p)
2E-3 (p)
3E-3 (p)
Turks and Caicos
Sint Maarten
Suriname
Saint Martin
Antigua and Barbuda
Guyana
Mayotte
Bahamas
Belize
Guinea
Eswatini
CAR
Aruba
Martinique
Madagascar
Trinidad and Tobago
French Guiana
Malawi
Albania
Afghanistan
Namibia
Honduras
Equatorial Guinea
Egypt
Costa Rica
Mauritania
Guadeloupe
Reunion
Bolivia
Guatemala
Cameroon
Moldova
Ecuador
Oman
Panama
Jamaica
North Macedonia
Senegal
Angola
Gabon
Maldives
Dominican Republic
Kyrgyzstan
Paraguay
Croatia
Nigeria
Qatar
El Salvador
Ghana
Kuwait
Mexico
Argentina
Tunisia
Kenya
Botswana
Slovenia
Bulgaria
Iraq
Colombia
Ethiopia
Chile
Lebanon
Uruguay
Peru
Ukraine
Bahrain
Ireland
Switzerland
Pakistan
South Africa
Serbia
Slovakia
Iran
Hungary
Philippines
Japan
Czechia
Cuba
Romania
Norway
Latvia
Israel
Uzbekistan
Finland
Austria
Cyprus
Kazakhstan
Morocco
Belgium
Portugal
Venezuela
Brazil
Saudi Arabia
Greece
Poland
Hong Kong
Lithuania
France
Malaysia
Thailand
Turkey
S. Korea
Canada
Italy
New Zealand
Denmark
Jordan
India
Germany
USA
Russia
Australia
China

Figure S1.

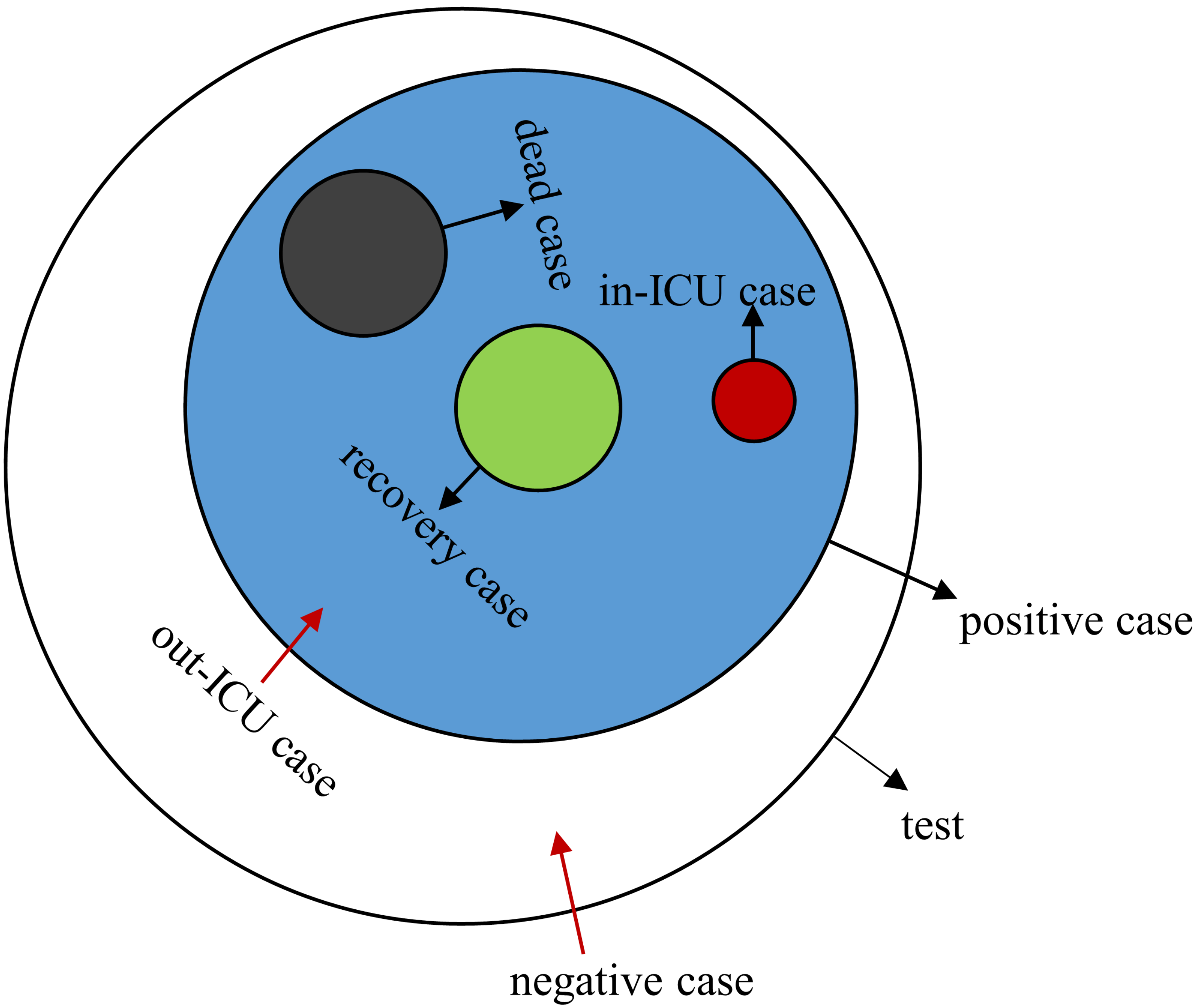
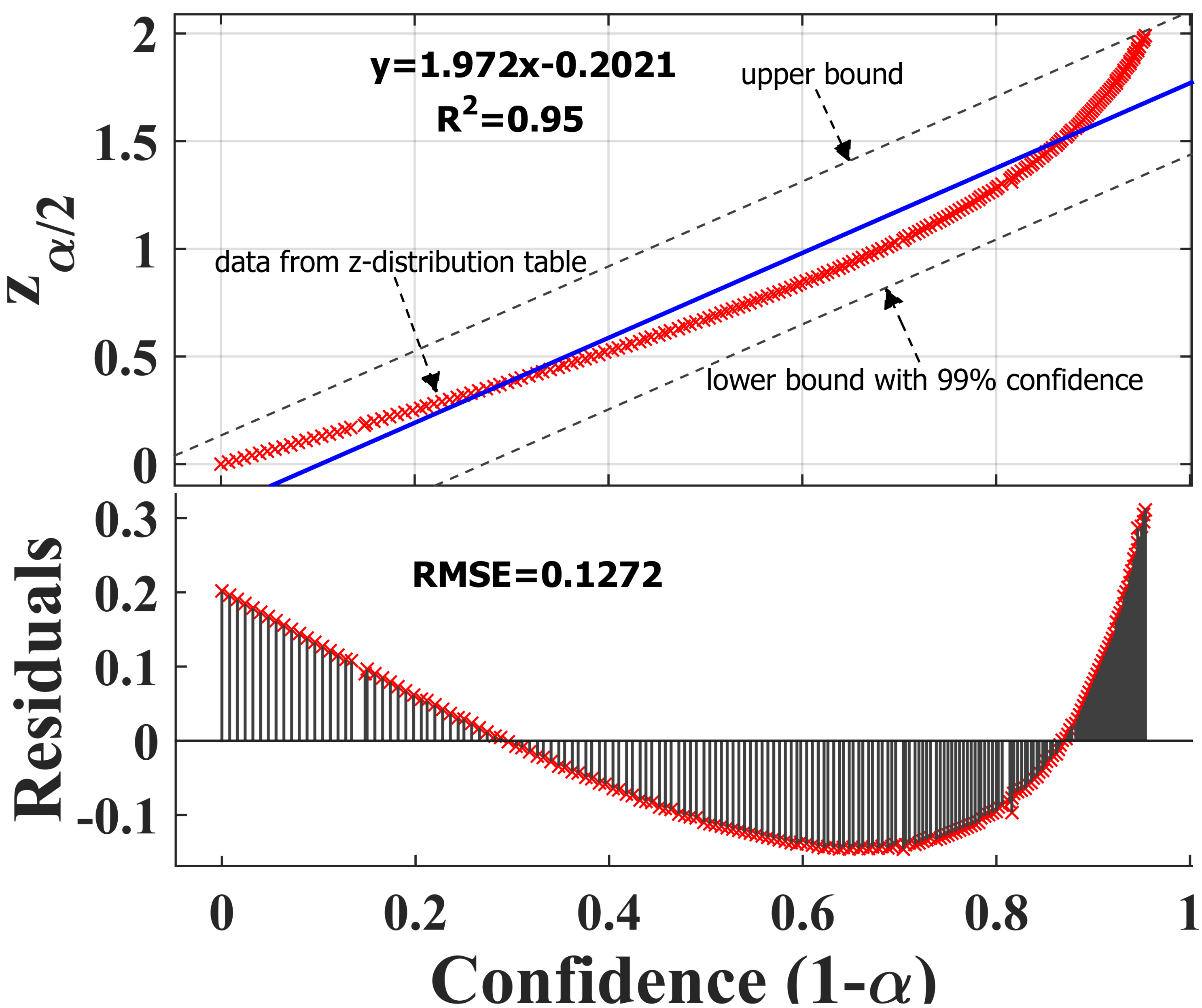

dead case
in-ICU case
recovery case
positive case
out-ICU case
test
negative case

Figure S2.



y=1.972x-0.2021
R2=0.95
upper bound
data from z-distribution table
lower bound with 99% confidence
RMSE=0.1272
Residuals
Confidence (1-α)

Figure S3.

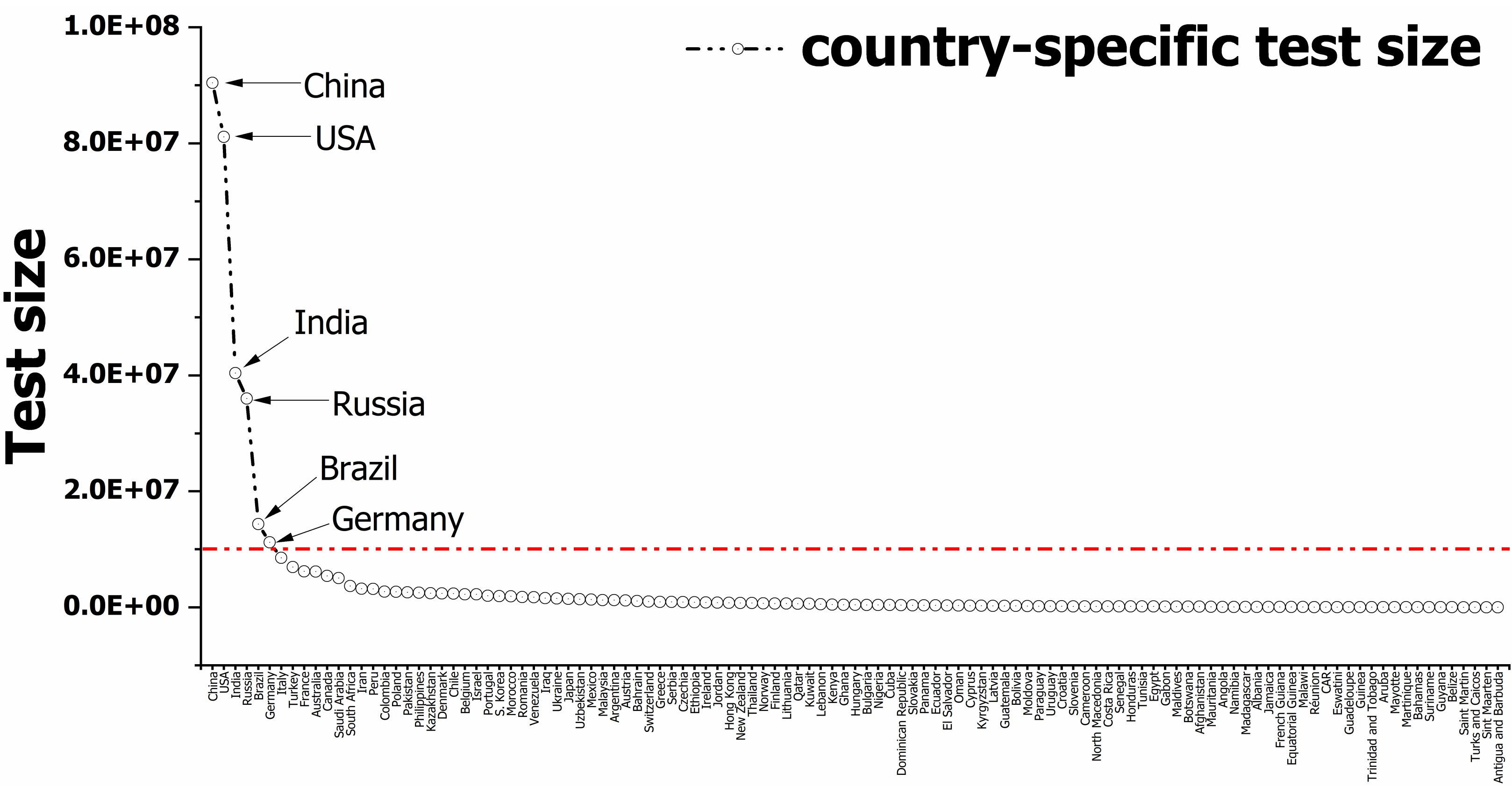

country-specific test size
Test size
1.0E+08
8.0E+07
6.0E+07
4.0E+07
2.0E+07
0.0E+00
China
USA
India
Russia
Brazil
Germany
China
USA
India
Russia
Brazil
Germany
Italy
Turkey
France
Australia
Canada
Saudi Arabia
South Africa
Iran
Peru
Colombia
Poland
Pakistan
Philippines
Kazakhstan
Denmark
Chile
Belgium
Israel
Portugal
S. Korea
Morocco
Romania
Venezuela
Iraq
Ukraine
Japan
Uzbekistan
Mexico
Malaysia
Argentina
Austria
Bahrain
Switzerland
Greece
Serbia
Czechia
Ethiopia
Ireland
Jordan
Hong Kong
New Zealand
Thailand
Norway
Finland
Lithuania
Qatar
Kuwait
Lebanon
Kenya
Ghana
Hungary
Bulgaria
Nigeria
Cuba
Dominican Republic
Slovakia
Panama
Ecuador
El Salvador
Oman
Cyprus
Kyrgyzstan
Latvia
Guatemala
Bolivia
Moldova
Paraguay
Uruguay
Croatia
Slovenia
Cameroon
North Macedonia
Costa Rica
Senegal
Honduras
Tunisia
Egypt
Gabon
Maldives
Botswana
Afghanistan
Mauritania
Angola
Namibia
Madagascar
Albania
Jamaica
French Guiana
Equatorial Guinea
Malawi
Réunion
CAR
Eswatini
Guadeloupe
Guinea
Trinidad and Tobago
Aruba
Mayotte
Martinique
Bahamas
Suriname
Guyana
Belize
Saint Martin
Turks and Caicos
Sint Maarten
Antigua and Barbuda

Figure S4.

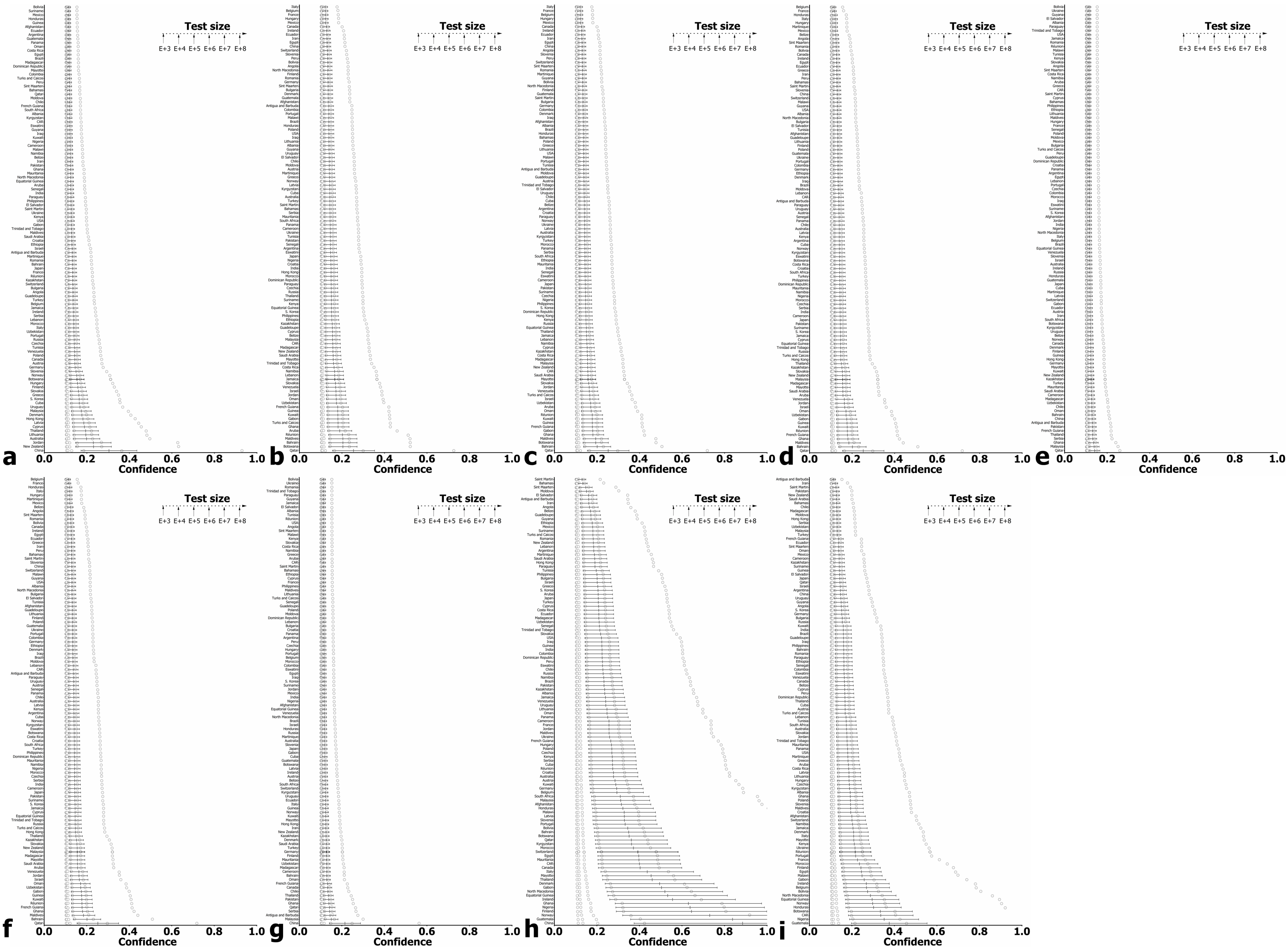

a
b
c
d
e
f
g
h
i
Test size
E+3 E+4 E+5 E+6 E+7 E+8
0.0 0.2 0.4 0.6 0.8 1.0
Confidence

Figure S5.

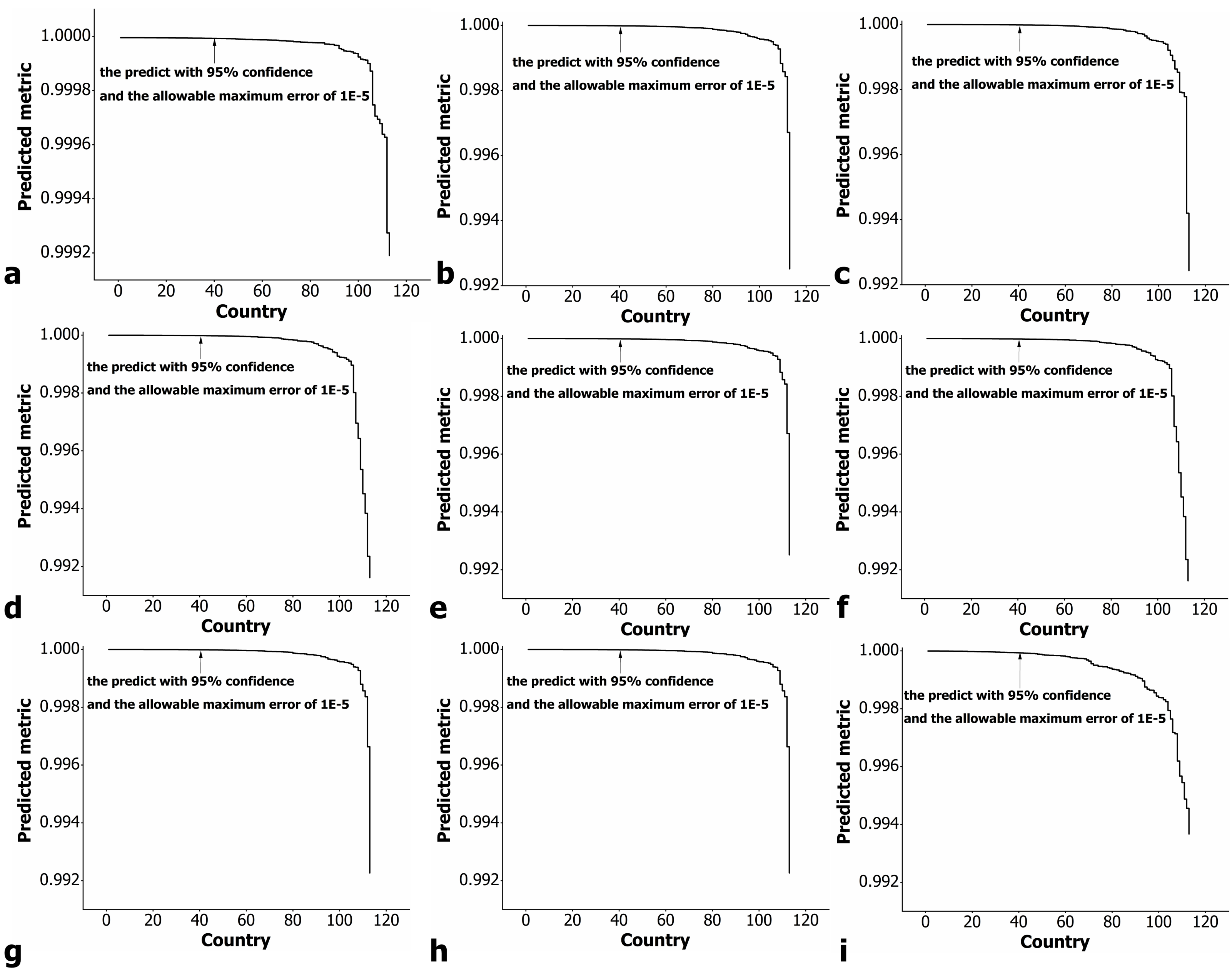
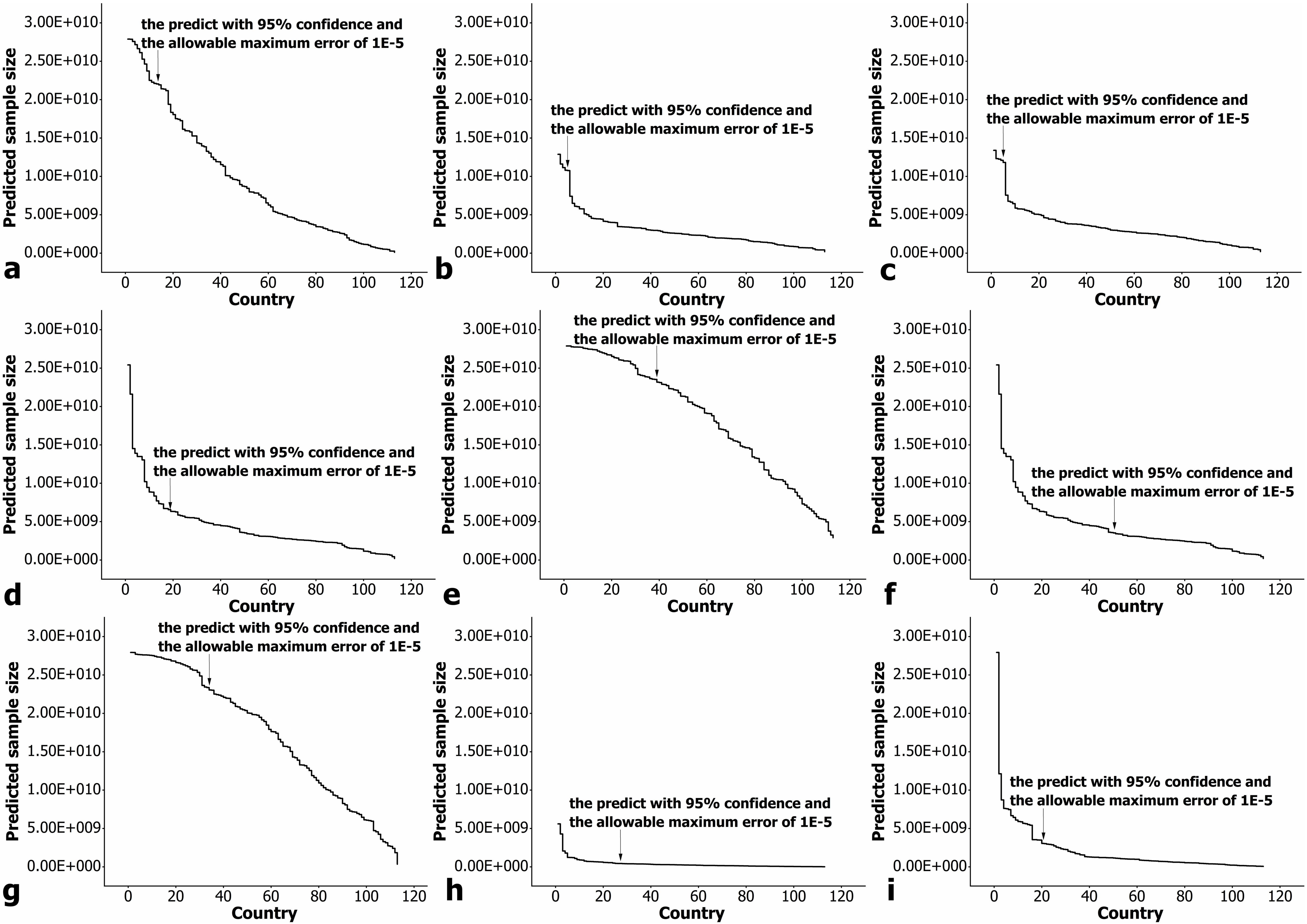

a
Predicted metric
1.0000
0.9998
0.9996
0.9994
0.9992
Country
0
20
40
60
80
100
120
the predict with 95% confidence
and the allowable maximum error of 1E-5
b
Predicted metric
1.000
0.998
0.996
0.994
0.992
Country
the predict with 95% confidence
and the allowable maximum error of 1E-5
c
Predicted metric
Country
the predict with 95% confidence
and the allowable maximum error of 1E-5
d
Predicted metric
Country
the predict with 95% confidence
and the allowable maximum error of 1E-5
e
Predicted metric
Country
the predict with 95% confidence
and the allowable maximum error of 1E-5
f
Predicted metric
Country
the predict with 95% confidence
and the allowable maximum error of 1E-5
g
Predicted metric
Country
the predict with 95% confidence
and the allowable maximum error of 1E-5
h
Predicted metric
Country
the predict with 95% confidence
and the allowable maximum error of 1E-5
i
Predicted metric
Country
the predict with 95% confidence
and the allowable maximum error of 1E-5

Figure S6.


a
b
c
d
e
f
g
h
i
Predicted sample size
Country
3.00E+010
2.50E+010
2.00E+010
1.50E+010
1.00E+010
5.00E+009
0.00E+000
0
20
40
60
80
100
120
the predict with 95% confidence and
the allowable maximum error of 1E-5