medRxiv preprint doi: https://doi.org/10.1101/2022.06.06.22275944; this version posted March 28, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Novel Findings in Travel-related Thrombosis and COVID-19 Related Cardiovascular Data: Data Mining Method for Previously Published Data

Author list

Keiichiro Kimoto, M.Sc.[1, 2 *], Munekazu Yamakuchi, M.D., Ph.D.[1], Kazunori Takenouchi, M.D., Ph.D. [1], Teruto Hashiguchi, M.D., Ph.D.[1]

Author affiliations

[1] Department of Laboratory and Vascular Medicine, Kagoshima University Graduate School of Medical and Dental Sciences, 8-35-1 Sakuragaoka, Kagoshima 890-8520, Japan

[2] Present Affiliation: External Advisor for Data Strategy Research Institute, Yokohama, Japan

Corresponding author information

*****Corresponding author: Keiichiro Kimoto

E-mail: k1974142@kadai.jp

Department of Laboratory and Vascular Medicine, Kagoshima University, Graduate School of Medical and Dental Sciences, 8-35-1 Sakuragaoka, Kagoshima 890-8544, Japan.

Phone: /Fax: +81-99-275-5437/+81-99-275-2629

Author contributions

Keiichiro Kimoto takes responsibility for this research, making study concepts, data analysis, interpretation of analysis results, and drafting the manuscript. Dr. Yamakuchi contributed to interpreting analysis results, manuscript drafting, supervision, and administrative role. Dr. Takenouchi contributed to the supervision. Dr. Hashiguchi contributed to the study concept, interpretation of analysis results, manuscript drafting, supervision, and administrative role.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2022.06.06.22275944; this version posted March 28, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Competing interest statement

Keiichiro Kimoto has been in charge of external advisor for the Data Strategy Research Institute but received no financial support for this study. Except for this, the authors have no conflicts of interest or financial disclosures that should be disclosed.

Classification: Research article, Medical Sciences

Keywords

Traveler's thrombosis, cardiovascular biomarker, COVID-19, Oral Contraceptives (OC), and Hormone Replacement Therapy (HRT)

This file includes Abstract, Significant statement, Main Text, Figures 1 to 7, and Table 1. Also, this article has supporting information that contains supporting figures, tables, method details, and supporting discussions.

medRxiv preprint doi: https://doi.org/10.1101/2022.06.06.22275944; this version posted March 28, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Abstract

Image analysis software is used in the review process to detect research misconduct (*1*), but not for positive motives (e.g., for missing findings). Also, scientists analyze "raw images" (e.g., MRI), but there are no methods for published "charts." So, we conceived an analysis method. We applied our idea to two cardiovascular data related to thrombosis (traveler's or travel-related thrombosis and COVID-19) because there are still controversial discussions that might be due to some overlooked information. We found overlooked an approximately 28 days cycle of thrombosis onset over several weeks after travel in a figure (Cannegieter et al., *PLoS Med.* 2006, in the top 25% of all research scored by Altmetric) (*2*). Also, we found an eighteen-day cycle of thrombosis onset in another chart (Kelman et al., *BMJ.* 2003, in the top 5% of research) (*3*). In COVID-19, we found overlooked subgroup patterns in a scatterplot of Troponin T and NT-proBNP (Guo et al. *JAMA Cardiol.* 2020, top 5% research) (*4*), which has already appeared in the cardiomyopathy study in 2016 (Budnik et al., *Int J Cardiol.* 2016) (*5*). The above cyclic patterns suggested that travel-related oral contraceptive initiation (e.g., honeymoon and OC initiation) or hormone replacement therapy caused thrombosis. Also, the in-flight environment can likely not explain the cyclic patterns. In conclusion, women's life-stage-related events, which are also related to travel, may be the underlying cause. In COVID-19, biomarker data should require a re-analysis. Our findings in the well-known figures suggest that our tips can be an essential review technique.

Significance statement

Raw images (raw data), such as MRI, are often analyzed, but the chart is not. We developed a method for review charts and found about 28 days of thrombosis onset rhythm after travel in a published figure, which suggests travel-related oral contraceptive (OC) initiation exists (e.g., honeymoon and starting birth control OC). The OC has already been discussed as a risk (aggravating) factor in the context that the thrombosis is caused by increased blood viscosity due to dehydration in a dry cabin or blood stasis due to restricted movement in a narrow seat (e.g., economy class). However, our result may require reversing the root cause because the rhythm suggests that thrombus (blood clot) is not developed in the cabin but is related to drug concentration.

medRxiv preprint doi: https://doi.org/10.1101/2022.06.06.22275944; this version posted March 28, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Main text

Introduction

Recently, some image-checking AI detects research misconduct but not for helpful, latent findings (1). So, we first made questioned whether something new exploratory data analysis (data mining) method assisted us in finding important knowledge from "previously published" figures **(Fig. S1, a)**.

Besides, scientists analyze "raw images" (e.g., MRI, CT, or X-ray), but there are no methods for "charts" previously published. Also, software improvements have allowed us to make a beautiful chart, but an understanding chart is required (*6*). So, we took the idea of "reading a chart carefully" one step further, and we made a secondary question that something new data mining method assisted us in finding necessary knowledge from "previously published" "chart" **(Fig. S1, b)**.

Meanwhile, in the 20th century, statistics based on probability theory have applied to biology and medicine (e.g., biostatistics and p-value), and in the 21st-century differential equation often used in clinical research (e.g., mathematical biology and application of Susceptible-Infected-Recovered (SIR) model for COVID-19 pandemic to estimate the number of an infected person) (*7*).

We turned our attention from statistics and differential equation to geometry regarding its importance in science history. For example, Isaac Newton elicited the theory of gravitation by analyzing an elliptical orbit with the geometrical technique after German astronomer Johannes Kepler found that the planet's orbit was elliptical in the 17th century. So, we finally decided to develop a data mining method that assisted us in finding essential knowledge from the "previously published" "chart" by "geometrical approach" **(Fig. S1, c)**.

Considering the above three concepts, we were interested in a research field, traveler's thrombosis, because many controversies (*2, 8-24*) in this field seemed to be associated with some overlooked information **(Supporting discussion 1: history of traveler's thrombosis)**. In the literature review, we focused on a figure that showed a result of meta-regression analysis **(Fig. S1, b)**(Chandra et al., *Ann Intern Med.* 2009;151(3):180-90., Figure 3) (*16*). In this figure, we saw a geometrical shape (layered hyperbolic shapes) that seemed to be formed by overlapping two subgroups **(Fig. 1, Fig. 3, a)**.

Finally, we conceived a methodology, which was a kind of pattern recognition, searching layered hyperbolic patterns in a figure showing a result of meta-regression analysis to detect latent sub-groups. Our basic idea is that the confidence interval length for the odds ratio is inversely proportional to the "potential patient population size" **(see Fig. S2, d-f)**.

medRxiv preprint doi: https://doi.org/10.1101/2022.06.06.22275944; this version posted March 28, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

To be precise, the ability of patient recruitment, in other words, the "sample size in each study," affects the confidence interval length. However, we speculated that the effect of the differences in sample size could not be a practical problem for two reasons. The first reason is that stratified analysis, often performed in clinical studies, can be a force in reducing the differences. The second reason is that two types of boundary conditions in the hyperbolic shape fitting, which are the "S-curve" and the "error bar length," reduce the degree of freedom in searching and increase the possibility of finding the desired shape (details in **Materials and methods**).

We were aware that only published data would not be able to prove that meaningful subgroups formed overlapping hyperbolic patterns. Also, some researchers might argue that the pattern alone is not enough basis for a discussion to be significantly considered. However, we thought we could show the discussion was meaningful by showing that the curves derived from theoretical considerations can be applied to the hyperbolic patterns.

To validate our ideas, we worked on consecutive exploratory data analyses **(the overall study framework is shown in Fig. 2).**

Results

Numerical Experiment

Usually, a sigmoid function is used for curve fitting to a dose-response relationship and outline of the cumulative distribution. Also, the curve expressing the end of a confidence interval is a sum of the S-shaped and U-shaped curves **(see Fig. 1)**.

To decide on a formula to express a U-shaped curve, we performed a numerical experiment. The result indicated that a parabola could approximate a U-shaped curve **(Fig. 1, c & d).** We constructed three mathematical formulas for fitting hyperbolic patterns to data **(see Fig. 1 and Materials and methods).**

Analysis 1A (dataset: Chandra et al.) (*16*)

Before redrawing the figures of the meta-regression analysis, we reviewed the cited four studies (Martinelli et al., 2003; Parkin et al., 2006; Cannegieter et al., 2006; Kuipers et al., 2007) (*2, 8, 11, 25*). The main objective of Chandra et al.'s research (*16*) was estimating a risk, but the main of our analysis is detecting the subgroups. So, we are considering the possibility of some problems with inclusion in our study, even if there were no problems in Chandra et al.'s research.

medRxiv preprint doi: https://doi.org/10.1101/2022.06.06.22275944; this version posted March 28, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

In this assessment, two study data (Martinelli et al., 2003; Cannegieter et al., 2006) (*2*, *8*) were judged inappropriate and excluded from our analysis. Cannegieter et al. and Martinelli et al. used a unique research design, which uses the patient's partner as a control (Martinelli et al.: friend and partner). In the introduction, we described that although the properties of each study (the ability of patient recruitment, sample size) disrupt the effect of the size of the potential patient population, the stratified analysis could reduce the disruption. However, the impact of adding a unique study design was unclear.

Additionally, the odds ratio reported by Parkin et al. (*25*) was suspected to be affected by the misplacement of the cells in the cross table for calculating the odds ratio. Still, qualitatively, it could be used **(see Fig. 2, Table S1, Table S2, and Supporting information: Method details)**.

In the residual dataset (Parkin et al., 2006 & Kuipers et al., 2007)(*11*, *25*), layered hyperbolic patterns appeared, and applied regression analysis with the above three formulas. The inflection points of the S-curves (center of distributions) were 7.1 hours and 11.8 hours, respectively **(Fig. 3, b).**

Additionally, it seemed that S-curve could be fitted to the group of points reported by Cannegieter et al. and Martinelli et al., respectively. Also, those of inflection points would position at positions of less than 10 hours and more than 10 hours. This result seems consistent with the two hyperbolic shapes fitting **(Fig. 3, b, orange, and green dots)**.

Notably, we observed an overlooked cyclic pattern of thrombosis onset in the figure reported by Cannegieter et al. (2) during the above data review process. The cycle period was 27.4 days **(Fig. 4, a-e)**. The wave of about 28 days consisted of a gradually decaying wave and a monotonically decreasing function **(Fig. 4, e)**. The wave disappeared after about three months (90 days). The monotonically decreasing part estimated that it continued for over 90 days after the travel **(Fig. 4, f)**. Cannegieter et al. (2) did not mention this wave pattern.

Analysis 1B (dataset: Philbrick et al.) (*13*)

Considering that only a few pieces of data were available for the first analysis, we performed a similar analysis using another dataset gathered by Philbrick et al. for a systematic review. As a preparation, we conducted a data review to confirm eligibility for our regression analysis **(see Fig. 2)**. The dataset did not contain risk ratio data. In this case, hyperbolic shapes could not be fitted (for the details, see **Supporting information: Method details**). So, only S-curves were applied. This analysis found two S-curves in the stratified datasets by pulmonary embolism (PE) and deep vein thrombosis (DVT). The inflection point of PE was about 9.2 hours, and the point of DVT was about 12.1 hours, respectively **(Fig. S3, b)**, which were similar to the first analysis.

medRxiv preprint doi: https://doi.org/10.1101/2022.06.06.22275944; this version posted March 28, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Additionally, we found that unrecognized periodic pattern in the data reported by Clérel and Caillard (*26*) during the data review, which was suspected to be the effect of circadian rhythm **(Fig. S4, c).**

We also found a cyclic pattern after travel in the data reported by Kelman et al. (*3*). However, the period of the wave was 17.9 days, and the wave was less clear than the wave observed in the data reported by Cannegieter et al. In the waveform, there was a raw-risk period just before 90 days. Also, in this data, there was an uptrend **(Fig. 5)**.

Furthermore, we found a correlation between the stepped stages of increasing PE patients and the issuance of the guidelines (*27-30*) for preventive hormone replacement therapy (HRT) to a menopause woman in a figure reported by Clérel and Caillard (*26*) **(Fig. S5)**. Especially the thrombosis onset in Paris airports has increased since the late 1990s **(Fig. S5)**.

Analysis 1B-related additional analysis

Trend analysis and history of sex ratio change

We performed the trend analysis of the literature to assess the generalization on the increasing number of patients at Paris airports reported by Clérel & Caillard (*26*) **(Fig. S5)**. We searched the literature in PubMed (keywords: “Economy Class Syndrome,” “Traveler’s Thrombosis,” and “Travel-related thrombosis”). The number of references increases rapidly after the 2000s **(Fig. S6)**.

The difference between the above result of trend analysis and the result in the data reported by Clérel and Caillard (*26*) can be explained by the difference between publication day and onset day. So, there is no practical difference between the two results.

The literature search found that two articles (Symington & Stack, 1977, Cruickshank, Gorlin, & Jennett, 1988) (*31, 32*) contained each patient’s details. We already obtained each patient’s data from the other two articles in analyses 1A and 1B (Clérel & Caillard, 1999, Parkin et al., 2006) (*25, 26*). So, we confirm changing the sex ratio using those data to assess the effect of HRT. We found that the rate of women in the PE data gathered from the above four studies (*25, 26, 31, 32*) gradually increased but was most pronounced during the late 1990s **(Table 1).**

Literature review targeting early literature and analysis focusing on men

We also conducted a literature review targeting early literature because we strongly questioned the “economy class syndrome” concept from our findings in Analysis 1A and 1B. We found that the word appeared

medRxiv preprint doi: https://doi.org/10.1101/2022.06.06.22275944; this version posted March 28, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

in Symington & Stack, 1977 (*31*). However, judging from the citation history checked on the Europe PMC website (https://europepmc.org/), the article by Cruickshank, Gorlin, & Jennett, 1988 (*32*) might influence the word as evangelists. The paper was a "personal paper," a type of article category of the Lancet, and the authors prepared case reports of their own experiences.

Dr. Cruickshank was middle-aged (*32*). Usually, the mid-life crisis is concerned around this range of age. Dr. Jennett was a dean of medicine at Glasgow (*33*). Generally, heavy psychological stress based on heavy social responsibility is concerned. So, we hypothesized that men's life event-related psychological stress could be one of the risk factors **(see Fig. S7, b).**

To assess our hypothesis, we merged data from the above four studies (*25*, *26*, *31*, *32*) and made a bar chart for 1-year. The onset of PE was concentrated in certain age stages, such as around 30, the 40s, late 50s, and early 70s **(Fig. S7, e)**.

We performed the same analysis in the women's data except for data reported by Clérel & Caillard (*26*). Although the peak range was 65 to 70 in the data of Clérel & Caillard (*26*) **(Fig. S4, d),** PE was concentrated in the 40s to 50s (menopausal stage) in this analysis **(Fig. S8)**.

Additionally, we found an inverted-U-shaped (mountainous) relationship in the scatter plot: age on the horizontal axis and air travel time on the vertical axis **(Fig. S10, d)**. However, it was possible to fit a straight line as an up-right linear relationship or a down-right linear relationship**(Fig. S9, a, and Fig. S10, a, c)**.

These analyses were also performed on the female data reported by Clérel & Caillard (*26*), and we found a U-shaped relationship **(Fig. S9, c)**.

Analysis 2 (dataset: COVID-19) (*34*)

We decided to apply our ideas to the current complex COVID-19 pandemic problem to find some overlooked things because there was some controversy (*35-41*) about COVID-19-related thrombosis. This situation was similar to that of the traveler's thrombosis. **(Supporting discussion 2: research situations of COVID-19-related thrombosis)**, and we started this analysis with the inspiration from the famous mathematician Polya who stated that solving a similar problem helped us solve a more complex problem (*42*).

As a note, we described that stratified is necessary for the introduction section, but clinical studies of COVID-19 did not perform it. However, Stratification seemed unnecessary because the size of the studies was not huge and roughly the same due to no mega trial during the pandemic.

medRxiv preprint doi: https://doi.org/10.1101/2022.06.06.22275944; this version posted March 28, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

One of us (KK) searched for the layered hyperbolic pattern using the search engine Google and extracted a figure reported by Matsushita et al. (*34*) **(Fig. 6, a-c)**. After evaluating eligibility **(see Fig. 2)**, patterns that allowed to be fitted hyperbolic shapes appeared **(Fig. 6, d-f, & g-i)**.

Considering the cause of these patterns, we found the matching between grouping the data by hyperbolic patterns and grouping by data cut-off days **(Fig. S11)**. We calculated weighted averages of age by grouping. In the earlier days group, non-severe was 46.5, and severe was 57.7, which were middle age (age 45-64). In the late date group, non-severe was 40.9, and severe was 68.6, which were mature age (age 25-44) and elderly (age > 65), respectively.

Surprisingly, we found unrecognized patterns during the data review process in a figure reported by Guo et al. (*4*), which showed a relationship between N-terminal pro-brain natriuretic peptide (NT-proBNP) and Troponin T (TnT). There were three clusters of subgroup patterns, and each cluster could be fitted parabola **(Fig. 7 a, b, Fig. S12, a, b & Fig. S13, a, d, e-h)**.

The third subgroup pattern was similar to ST-elevating myocardial infarction in the figure report by Budnik et al. (*5*). That study was a comparative study between Takotsubo cardiomyopathy (also known as stress cardiomyopathy or broken heart syndrome) and ST-elevating myocardial infarction. This tilted parabola appeared in another study on COVID-19 patients **(Fig. S12, c)** (*43*), and the pattern had already appeared in other studies (*44, 45*) before the pandemic **(Fig. S13, b & c)**. However, patterns in COVID-19 patients were not recognized by each author.

Additionally, In the case of dropping points to the horizontal axis (troponin axis), the histogram on the troponin axis results in bimodality. The above results matched the patterns reported by other studies **(Fig. S12, d & e)** (*43, 46, 47*).

Furthermore, by mounting the data of high-sensitivity C-Reactive protein (hsCRP) onto the third subgroup in the data reported by Guo et al. (*4*) and constructing a parabolic cylinder, we found a pattern that might further be subdivided into two subgroups on the side surface of this parabolic cylinder **(Fig. 7, c, d, & Fig. S14)**.

Discussion

Hyperbolic shape fitting in traveler's thrombosis

In traveler's thrombosis, we obtained consistent results from two analyses (Analysis 1A: 7.1 hours and 11.8 hours, Analysis 1B: 9.2 hours and 12.1 hours) **(Fig. 3, b, Fig. S3, b)**. The discrepancy between the value of

medRxiv preprint doi: https://doi.org/10.1101/2022.06.06.22275944; this version posted March 28, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

7.1 hours and 9.2 hours seems to be explained by the miscalculation found in the data review process **(see Table S1, Table S2, and Supporting information: Method details)**. Our results implied two S-curves. Also, we hypothesized two high-risk periods and two types of high-risk groups, which may have to be considered with the "factor V Leiden paradox" (*48*) **(Supporting discussion 3: two high-risk periods and two types of high-risk groups?)**.

Cyclic patterns newly founded in this research and effect of OC – life event of young women-

Another noteworthy point is the cyclic patterns in the data reported by Cannegieter et al. (*2*) **(Fig. 4).** In Cannegieter et al. (*2*), the cyclic pattern disappeared around three months (90 days), consistent with the high-risk period in oral contraceptive (OC) use was the first three months (*49*). Also, there is multi-phase OC in the category of OC. So, our findings imply that the 28 days cycle OC is attributable to the cyclic pattern.

Additionally, the law-risk period just before 90 days in the data reported by Kelman et al. (*3*) **(Fig. 5)**. This law-risk period could be explained that extended-use type OC, which has a planned drug withdrawal, such as 84 active days and seven placebo days (*50, 51*).

The uptrend in the data by Kelman et al.(*3*) might be due to depot agent type OC administrated in 90 days cycle, which injects drugs periodically after decreasing the drugs in the body.

Based on the above results, types of OC use (multi-phase type, extended-use type, and depot agent type) might be triggered by travel.

Importantly, thromboembolism onset is caused by passing two stages. The first stage is developing a "thrombus (blood clot)." The second stage is "embolization." Suppose the in-flight environment caused thrombus formation, and the onset of thrombosis (thrombus release and embolization) occurs after travel. In that case, some mechanism must promote the release of thrombus periodically after travel. However, such a mechanism is not known. Also, the cyclic pattern found in this study suggested that OC and HRT are involved in the cyclic pattern. Therefore, it is more plausible that thrombus formation occurred related to changes in drug concentrations rather than related to the in-flight environment.

The data reported by Cannegieter et al. contained many car trips (air travel, train, bus, and car trip). These cars might be honeymoon cars.

medRxiv preprint doi: https://doi.org/10.1101/2022.06.06.22275944; this version posted March 28, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Notable descriptions in the previous work – valley of the wave -

Kelman et al. described that the risk increased within 2 to 4 weeks of flight (hazard period) but the risk was low after the period because travelers were healthier than the average of a group (healthy traveler effect) (*3*). Also, Rothman mentioned the "persistence of travel-related risk" and "spline regression" (*10*). They perhaps intuitively realized the importance of the pattern at the position around 3-4 weeks, which we have identified as the valley of the wave. If Cannegieter et al. used a more appropriate visualization technique instead of a 3-dimensional bar graph **(Fig. 4, a)**, someone might point out the above cyclic patterns at a faster stage.

Difference of the cyclic patterns and effect of HRT history – life event of matured women-

Interestingly, there was a difference between the period of cyclic patterns (Cannegieter et al.: about 28 days, Kelman et al.: about 18 days). Also, the wave in the data reported by Cannegieter et al. was clear, but the wave in the data reported by Kelman et al. was not clear despite the exposures in Cannegieter et al. being more complex (air travel, train, bus, and car trip) (*2*) than Kelman et al. (only air travel) (*3*).

Regarding the halfway cycle (about 18 days) in Kelman et al. (*3*), it is notable that sequential type drug for HRT is used as daily estrogen dosing and 10–14 days progestogen (*52*). Also, in 1999, Clérel and Caillard mentioned the possibility of both the effects of OC and HRT (*26*).

It is also notable that thrombosis onset in the data reported by Cannegieter et al. was "March 1999-March 2000" (only most late 1990s to early 2000) (*2*). Thrombosis onset in the data reported by Kelman et al. was "1981-1999" (including the whole of the 1990s) (*3*).

We found a slowdown of the increase of thrombosis onset at Paris airports in 1998, the year of HERS result was issued **(Fig. S5)**. Also, our trend analysis showed that the number of literature on thrombosis decreased since around 2000 **(Fig. S6)**. Moreover, our merged data from four studies (*25, 26, 31, 32*) on PE from the 1970s to the 2000s showed that the rate of women was most pronounced during the late 1990s **(Table 1)**.

Therefore, the difference in waves can be explained as follows: the halfway cycle (17.9 days) in Kelman et al. (*3*) was a mixture of the 28 days OC cycle and a type of HRT cycle. However, only OC users might remain in the data reported by Cannegieter et al. (*2*) after the issue of the HERS study in 1998 (*29*).

Although the existence of the women patients not "menopausal" stage women (40s to 50s) but among around 70-75 years of age seems to disagree with HRT use **(Fig. S4, d, & Fig. S8, d)**, the subjects of the WHI study were "postmenopausal" women (age 50-79) for prophylactic use for cardiovascular problems.

medRxiv preprint doi: https://doi.org/10.1101/2022.06.06.22275944; this version posted March 28, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Men's life stage and travel related phycological stress – life event in the men's life stage -

Analysis 1B-related additional analysis suggested that the development of travel-related thrombosis occurs at specific ages **(Fig. S7)**. On the other hand, Kelman et al. reported the incidence of VTE per 100,000 passengers (*3*). In this data (*3*), the incidence of VTE sharply increased after age 40, and it was a temporary decrease in the early 60s. Although it should be noted that these results are not limited to men, they are consistent with our results.

We also found that the relationship between male age and travel duration seems to be inverted-U-shaped (mountainous) **(Fig. S9 & Fig. S10)**. Regarding this result, men's life event-related psychological stress may be one of the risk factors.

Moreover, every four studies (*25, 26, 31, 32*) showed inconsistent results (correlated or inversely correlated). This situation seems to be the same as the parable of "blind men and an elephant." Considering the correlation between age and the risk of thrombosis, the above phenomena may cause controversial discussions **(Supporting discussion 1: history of traveler's thrombosis)**.

Not travel-related but life event-related thrombosis? - how to improve QOL in women and men -

Taking OC or HRT can be regarded as a women's life events-related phenomenon. In the case of men, periods of life events that increase psychological stress or flight time linked to psychological stress may be associated with the development of thrombosis. From a comprehensive view, "traveler's thrombosis" or "travel-related thrombosis" may require to be regarded as "life-event-related thrombosis."

For women, even if a substantial risk factor is not the in-flight environment, it is not neglectable. So, our thought that starting OC or HRT was triggered by travel suggests that usage based on risk diversification is important to avoid overlapping risks of drug and in-flight environment. Also, well-planned may allow more safe use of OC and HRT. Additionally, for men, perhaps we should focus more on "stress reduction."

Application of our method for COVID-19 cardiovascular data

In appreciation of our ideas for COVID-19, the dataset grouping by Matsushita et al.(*34*) matched the grouping by data cut-off dates **(Fig. S11)**. Age structure diverged from middle age to mature and elderly. Middle-aged people (age 45-64) might enter the workforce for manual labor treating fish containers, mature

medRxiv preprint doi: https://doi.org/10.1101/2022.06.06.22275944; this version posted March 28, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

people (age 25-44) might want to select IT jobs, and elderly people (age > 65) might stay in the house. So, our results were consistent with the hypothesis that COVID-19 spread from the seafood market.

In our data review process **(Supporting discussion 4: linear regression analysis)**, our analysis of the data reported by Guo et al. (*4*) showed three clusters of subgroups and formed the tilted parabola. Also, the patterns matched the other studies **(Fig. 7 a, b, Fig. S12 & Fig. S13)** (*5, 43-47*). Considering the study by Budnik et al. (*5*), those subgroups may have different biological mechanisms. Also, we found patterns on the side surface of this parabolic cylinder **(Fig. 7, c, d, & Fig. S14)**.

On the other hand, a paradoxical result on troponin was obtained just before the pandemic (*53*). Meisel et al. mentioned that the CRP to troponin ratio (CRP/troponin) could serve to differentiate between myopericarditis and acute myocardial ischemia (AMI), although not the study on COVID-19 patients. In COVID-19, Caro-Codón et al. reported interesting behavior of CRP (*46*). A meta-analysis by Lagunas-Rangel reported that the lymphocyte□to□C□reactive protein ratio (LCR) level, which was not a simple measurement value but a ratio, might be related to an inflammatory process (*54*).

The above studies might imply complex data structures in 2-dimensional and 3-dimensional scatter plots consisting of biomarker values, and our findings may serve the cardiac biomarkers' research field.

Supplementary, our result suggested that the treatment of implicit function, which is unfamiliar in the medical and biological fields, is required. In this research, we used our method to fit the curves. However, a Fortran program created by Dr. Timmes (*55*), a researcher in astronomy, may also be helpful for future research.

Proposed data mining method based on geometry

From the consistent results of Analysis 1A and 1B, and also the explainable result of analysis 2, it is conceivable that the discussion based on the concept of layered hyperbolic shapes was not nonsense. Surprisingly, we discovered many overlooked patterns in evaluating the process of our ideas unexpectedly. In this process, the geometrical viewpoint allowed us to catch the patterns. Therefore, approaching medical data from geometry and exploring data focusing on layered hyperbolic patterns can be a practical option for researchers. We imagine that the concept of "medical research test" or "inspection of medical data analysis" can be developed analogically as an extension of the medical inspection, including laboratory tests for each patient. Like medical inspection using equipment, detecting overlayed hyperbolic shape patterns may be a tool or marker of inadequate data analysis or some overlooked information that requires re-analysis **(Fig. S1, e)**.

medRxiv preprint doi: https://doi.org/10.1101/2022.06.06.22275944; this version posted March 28, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Materials and methods

Deriving the exact formula directly was difficult. So, firstly we focused only on the U-shaped curve. The length of the confidence interval for the odds ratio is inverse proportional-like to the size of the potential patient population, which onset thrombosis by exposure to risk factors **(see Fig. 1)**. Although, there are multiple possibilities (e.g., 1/N, $1/N^2$, or 1/sqrt(N)), regarding the form of equations (1) and (2) below, we intuitively predicted that it would be approximately equal to the inverse of the root (equation (3)).

$$Log_e(95\%CI) = Log_e\,(OR) \pm 1.96\sqrt{\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d}} \tag{1}$$

$$\sqrt{\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d}} = \frac{\sqrt{bcd+acd+abd+abc}}{\sqrt{abcd}} = \frac{complex\ combination\ of\ numbers}{\sqrt{abcd}} \tag{2}$$

$$Log_e(95\%CI) = Log_e\,(OR) \pm 1.96\sqrt{\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d}} \cong Log_e(OR) \pm 1.96 \times \frac{1}{\sqrt{N}} \tag{3}$$

To be precise, the "sample size in each study," which depends on the ability to recruit patients, also affects the confidence interval length. It is similar to Hooke's law on spring (*56*), which states that the pull-back force of spring is proportional to the length of the pull. In this case, the power also depends on the spring type (e.g., thick spring or thin spring). Thus, the formula is as follows, with spring constant (k) depending on the kind of spring (equation (4)). Similarly, equation (3) express equation (5) precisely.

$$Force = k \times Distance \tag{4}$$

$$Log_e(95\%CI) = Log_e\,(OR) \pm 1.96\sqrt{\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d}} \cong Log_e(OR) \pm 1.96 \times \frac{k}{\sqrt{N}} \tag{5}$$

However, we speculated that the effect of the sample size and the difference between values of the k in each study could not be practical problems for two reasons, described in the introduction section.

Considering the numerical experiment, we decided on the equations for the S-shaped curve, the upper end of the confidence limit, and the lower end of the confidence limit (equation (1), (2), and (3)). Note that in the

medRxiv preprint doi: https://doi.org/10.1101/2022.06.06.22275944; this version posted March 28, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

below equations, the coefficient L is generally set as 1. So, we used this equation under the condition of L=1 unless there is some reason.

$$y = \left\{\frac{K_1}{1+e^{L(x-M)}} + K_2\right\} \tag{6}$$

$$y = \left\{\frac{K_1}{1+e^{L(x-M)}} + K_2\right\} + (ax^2 + bx + c) \tag{7}$$

$$y = \left\{\frac{K_1}{1+e^{L(x-M)}} + K_2\right\} - (ax^2 + bx + c) \tag{8}$$

The whole process, including, data review, regression analysis, and tools for the analysis, is described in the supplementary information (**Supporting information: Method details**).

medRxiv preprint doi: https://doi.org/10.1101/2022.06.06.22275944; this version posted March 28, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Acknowledgments and funding sources

One of the authors, Keiichiro Kimoto, appreciates the kindful encouragement of Hideo Yoshioka, MEcon., who is in charge of the Data Strategy Research Institute representative.

Statement of our intention for data analysis results on previous published articles.

In addition to general acknowledgments, to clarify our stance, we mention this statement of intention for results. In this article, we pointed out many overlooked information and errors. However, we have no intention to attack previous works because our analysis results owing to their original works, including original research articles, reported valuable data, and writings of meta-analysis synthesized valuable datasets. Since just a scientist had better reconfirm the previous studies with no preconceptions and being grateful to those studies' researchers, we carefully reviewed the data reported by previous studies. We highly respect previous works intending to solve medical issues.

Data availability statement

We analyzed clinical data published by other studies (third parties). Used data is identified by indicated information of citation (reference numbers and list of references). The corresponding author responds to inquiries in the case of measured values from published figures requested by reviewers or readers.

Code availability statement

Correspondence author (KK) can respond to inquiries for the corresponding author's email address on offering the Python source code and spreadsheet software files for statistical analysis. Also, we have already uploaded our Python programs for the following GitHub repository to keep traceability perfectly.

https://github.com/Keiichiro-KIMOTO-Kagoshima-University/Novel-Findings-in-Travel-related-Thrombosis-and-COVID-19-Related-Cardiovascular-Data

medRxiv preprint doi: https://doi.org/10.1101/2022.06.06.22275944; this version posted March 28, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

References

1. R. Van Noorden, Publishers launch joint effort to tackle altered images in research papers. *Nature*, (2020).
2. S. C. Cannegieter, C. J. Doggen, H. C. van Houwelingen, F. R. Rosendaal, Travel-related venous thrombosis: results from a large population-based case control study (MEGA study). *PLoS Med* **3**, e307 (2006).
3. C. W. Kelman *et al.*, Deep vein thrombosis and air travel: record linkage study. *BMJ* **327**, 1072 (2003).
4. T. Guo *et al.*, Cardiovascular Implications of Fatal Outcomes of Patients With Coronavirus Disease 2019 (COVID-19). *JAMA Cardiol* **5**, 811-818 (2020).
5. M. Budnik *et al.*, Simple markers can distinguish Takotsubo cardiomyopathy from ST segment elevation myocardial infarction. *Int J Cardiol* **219**, 417-420 (2016).
6. A. Cairo. (W. W. Norton & Company, New York, 2019).
7. I. Cooper, A. Mondal, C. G. Antonopoulos, A SIR model assumption for the spread of COVID-19 in different communities. *Chaos Solitons Fractals* **139**, 110057 (2020).
8. I. Martinelli *et al.*, Risk of venous thromboembolism after air travel: interaction with thrombophilia and oral contraceptives. *Arch Intern Med* **163**, 2771-2774 (2003).
9. Y. Adi, S. Bayliss, A. Rouse, R. S. Taylor, The association between air travel and deep vein thrombosis: systematic review & meta-analysis. *BMC Cardiovasc Disord* **4**, 7 (2004).
10. K. J. Rothman, Thrombosis after travel. *PLoS Med* **3**, e300 (2006).
11. S. Kuipers *et al.*, The absolute risk of venous thrombosis after air travel: a cohort study of 8,755 employees of international organisations. *PLoS Med* **4**, e290 (2007).
12. S. Kuipers *et al.*, Travel and venous thrombosis: a systematic review. *J Intern Med* **262**, 615-634 (2007).
13. J. T. Philbrick, R. Shumate, M. S. Siadaty, D. M. Becker, Air travel and venous thromboembolism: a systematic review. *J Gen Intern Med* **22**, 107-114 (2007).
14. F. A. Spencer, Review: Flight times >8 hours and presence of risk factors for VTE increases travel related VTE. *Evid Based Med* **12**, 105 (2007).

medRxiv preprint doi: https://doi.org/10.1101/2022.06.06.22275944; this version posted March 28, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

15. A. J. Trujillo-Santos, A. Jiménez-Puente, E. Perea-Milla, Association between long travel and venous thromboembolic disease: a systematic review and meta-analysis of case-control studies. *Ann Hematol* **87**, 79-86 (2008).

16. D. Chandra, E. Parisini, D. Mozaffarian, Meta-analysis: travel and risk for venous thromboembolism. *Ann Intern Med* **151**, 180-190 (2009).

17. J. M. Gore, in *NEJM Jounal Watch*. (2009).

18. J. P. Vandenbroucke, S. C. Cannegieter, F. R. Rosendaal, Travel and venous thrombosis: an exercise in thinking about bias. *Ann Intern Med* **151**, 212-213 (2009).

19. American College of Chest Physicians. (2012).

20. M. Crowther, A. C. o. C. Physicians, Ed. (2012).

21. S. R. Kahn *et al.*, Prevention of VTE in nonsurgical patients: Antithrombotic Therapy and Prevention of Thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. *Chest* **141**, e195S-e226S (2012).

22. H. J. Schünemann *et al.*, American Society of Hematology 2018 guidelines for management of venous thromboembolism: prophylaxis for hospitalized and nonhospitalized medical patients. *Blood Adv* **2**, 3198-3225 (2018).

23. S. L. Clark, S. Onida, A. Davies, Long-haul travel and venous thrombosis: What is the evidence? *Phlebology* **33**, 295-297 (2018).

24. N. L. Reyes, M. G. Beckman, K. Abe, in *CDC Yellow Book 2020 Health Information for International Travel*, Centers for Disease Control and Prevention (CDC), G. W. Brunette, J. B. Nemhauser, Eds. (Oxford University Press, New York, United States of America, 2019), chap. 8.

25. L. Parkin, M. L. Bell, G. P. Herbison, C. Paul, D. C. Skegg, Air travel and fatal pulmonary embolism. *Thromb Haemost* **95**, 807-814 (2006).

26. M. Clérel, G. Caillard, [Thromboembolic syndrome from prolonged sitting and flights of long duration: experience of the Emergency Medical Service of the Paris Airports]. *Bull Acad Natl Med* **183**, 985-997; discussion 997-1001 (1999).

27. Guidelines for counseling postmenopausal women about preventive hormone therapy. American College of Physicians. *Ann Intern Med* **117**, 1038-1041 (1992).

28. S. C. Smith *et al.*, AHA consensus panel statement. Preventing heart attack and death in patients with coronary disease. The Secondary Prevention Panel. *J Am Coll Cardiol* **26**, 292-294 (1995).

medRxiv preprint doi: https://doi.org/10.1101/2022.06.06.22275944; this version posted March 28, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

29. S. Hulley *et al.*, Randomized trial of estrogen plus progestin for secondary prevention of coronary heart disease in postmenopausal women. Heart and Estrogen/progestin Replacement Study (HERS) Research Group. *JAMA* **280**, 605-613 (1998).

30. J. E. Rossouw *et al.*, Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results From the Women's Health Initiative randomized controlled trial. *JAMA* **288**, 321-333 (2002).

31. I. S. Symington, B. H. Stack, Pulmonary thromboembolism after travel. *Br J Dis Chest* **71**, 138-140 (1977).

32. J. M. Cruickshank, R. Gorlin, B. Jennett, Air travel and thrombotic episodes: the economy class syndrome. *Lancet* **2**, 497-498 (1988).

33. Professor Bryan Jennett. *British Journal of Neurosurgery* **22**, 305-306 (2008).

34. K. Matsushita *et al.*, The Relationship of COVID-19 Severity with Cardiovascular Disease and Its Traditional Risk Factors: A Systematic Review and Meta-Analysis. *Glob Heart* **15**, 64 (2020).

35. S. Ahmed, O. Zimba, A. Y. Gasparyan, Thrombosis in Coronavirus disease 2019 (COVID-19) through the prism of Virchow's triad. *Clin Rheumatol* **39**, 2529-2543 (2020).

36. A. Kushner, W. P. West, L. S. Pillarisetty. (2021).

37. C. D. Barrett, H. B. Moore, M. B. Yaffe, E. E. Moore, ISTH interim guidance on recognition and management of coagulopathy in COVID-19: A comment. *J Thromb Haemost* **18**, 2060-2063 (2020).

38. J. Thachil *et al.*, ISTH interim guidance on recognition and management of coagulopathy in COVID-19. *J Thromb Haemost* **18**, 1023-1026 (2020).

39. E. Callaway, H. Ledford, S. Mallapaty, Six months of coronavirus: the mysteries scientists are still racing to solve. *Nature* **583**, 178-179 (2020).

40. S. Tal, G. Spectre, R. Kornowski, L. Perl, Venous Thromboembolism Complicated with COVID-19: What Do We Know So Far? *Acta Haematol* **143**, 417-424 (2020).

41. M. Marietta, V. Coluccio, M. Luppi, COVID-19, coagulopathy and venous thromboembolism: more questions than answers. *Intern Emerg Med* **15**, 1375-1387 (2020).

42. G. Polya, *How to Solve It: A New Aspect of Mathematical Method (Princeton Science Library).* (Princeton University Press, Princeton, New Jersey, 1957).

43. Y. Wang *et al.*, Cardiac Injury and Clinical Course of Patients With Coronavirus Disease 2019. *Front Cardiovasc Med* **7**, 147 (2020).

medRxiv preprint doi: https://doi.org/10.1101/2022.06.06.22275944; this version posted March 28, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

44. S. Sugawa, I. Masuda, K. Kato, M. Yoshimura, Increased Levels of Cardiac Troponin I in Subjects with Extremely Low B-type Natriuretic Peptide Levels. *Sci Rep* **8**, 5120 (2018).

45. S. Satyan, R. P. Light, R. Agarwal, Relationships of N-terminal pro-B-natriuretic peptide and cardiac troponin T to left ventricular mass and function and mortality in asymptomatic hemodialysis patients. *Am J Kidney Dis* **50**, 1009-1019 (2007).

46. J. Caro-Codón *et al.*, Characterization of NT-proBNP in a large cohort of COVID-19 patients. *Eur J Heart Fail* **23**, 456-464 (2021).

47. O. M. Demir *et al.*, Impact and Determinants of High-Sensitivity Cardiac Troponin-T Concentration in Patients With COVID-19 Admitted to Critical Care. *Am J Cardiol* **147**, 129-136 (2021).

48. K. van Langevelde, L. E. Flinterman, A. van Hylckama Vlieg, F. R. Rosendaal, S. C. Cannegieter, Broadening the factor V Leiden paradox: pulmonary embolism and deep-vein thrombosis as 2 sides of the spectrum. *Blood* **120**, 933-946 (2012).

49. A. van Hylckama Vlieg, F. M. Helmerhorst, J. P. Vandenbroucke, C. J. Doggen, F. R. Rosendaal, The venous thrombotic risk of oral contraceptives, effects of oestrogen dose and progestogen type: results of the MEGA case-control study. *BMJ* **339**, b2921 (2009).

50. I. Wiegratz *et al.*, Effect of extended-cycle regimen with an oral contraceptive containing 30 mcg ethinylestradiol and 2 mg dienogest on bleeding patterns, safety, acceptance and contraceptive efficacy. *Contraception* **84**, 133-143 (2011).

51. R. S. Legro, "A Review of Extended-Cycle Oral Contraception," *CLINICAL / CONTRACEPTIVE TECHNOLOGY UPDATE* (Morrisville, North Carolina, USA, 2003).

52. I. W. G. o. t. E. o. C. R. t. Humans, Combined estrogen-progestogen contraceptives and combined estrogen-progestogen menopausal therapy. *IARC Monogr Eval Carcinog Risks Hum* **91**, 1-528 (2007).

53. E. Pérez-Rodríguez *et al.*, Incidence of air travel-related pulmonary embolism at the Madrid-Barajas airport. *Arch Intern Med* **163**, 2766-2770 (2003).

54. F. A. Lagunas-Rangel, Neutrophil-to-lymphocyte ratio and lymphocyte-to-C-reactive protein ratio in patients with severe coronavirus disease 2019 (COVID-19): A meta-analysis. *J Med Virol* **92**, 1733-1734 (2020).

55. F. X. Timmes. (2021).

56. W. contributors.

medRxiv preprint doi: https://doi.org/10.1101/2022.06.06.22275944; this version posted March 28, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Figures and Tables

Fig. 1

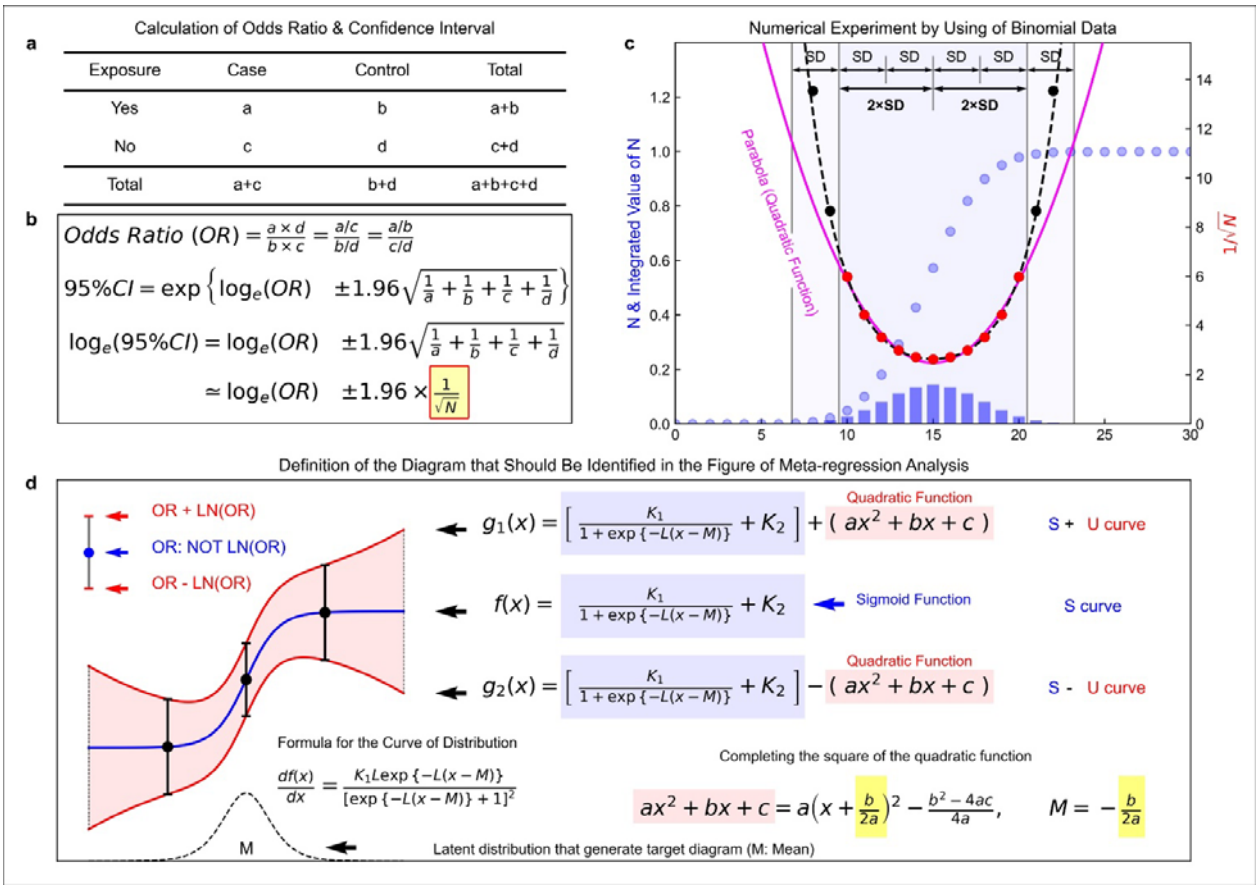

Fig. 1. A hyperbolic shape formed by confidence limits.

a, A two-way cross-tabulation (contingency table) for calculating an odds ratio (OR). **b**, Calculation of OR and its 95% confidence interval. **c**, Histogram of the data following a binomial distribution and a plot of the inverse values of the square roots. **d**, The mathematical formulas for the hyperbolic shape. Note that the OR is not logarithmic, and only the length of the error bar is logarithmic (see the upper left position of panel d). We generate binomial distribution data because normal distribution could be approximated by binominal distribution. Around the center of distribution (within ±2SD of the mean), the parabola was a good fit for the points calculated from the simulated data. Also, from around the over ±2SD to ±3SD range, the parabola is positioned below the points. Although a catenary curve (dashed curve) is more likely suitable for fit, the number of patients who fall into the categories at both ends is slightly larger than the expected number. Usually, data placed over the range of ±3 SD from the distribution center are summarized as “greater than” or “less than.” For example, in the case of 10 hours is the center of the distribution, the time category might be following sets such as “less than”

medRxiv preprint doi: https://doi.org/10.1101/2022.06.06.22275944; this version posted March 28, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

5 hours, 6-8 hours, 9-11 hours, 12-14 hours, and “more than” 15 hours. The value of 1 over root N (1/sqrt(N)) becomes smaller. So, using a parabola seemed to be based on the reality of the data handling method in the medical and biological fields.

medRxiv preprint doi: https://doi.org/10.1101/2022.06.06.22275944; this version posted March 28, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Fig. 2

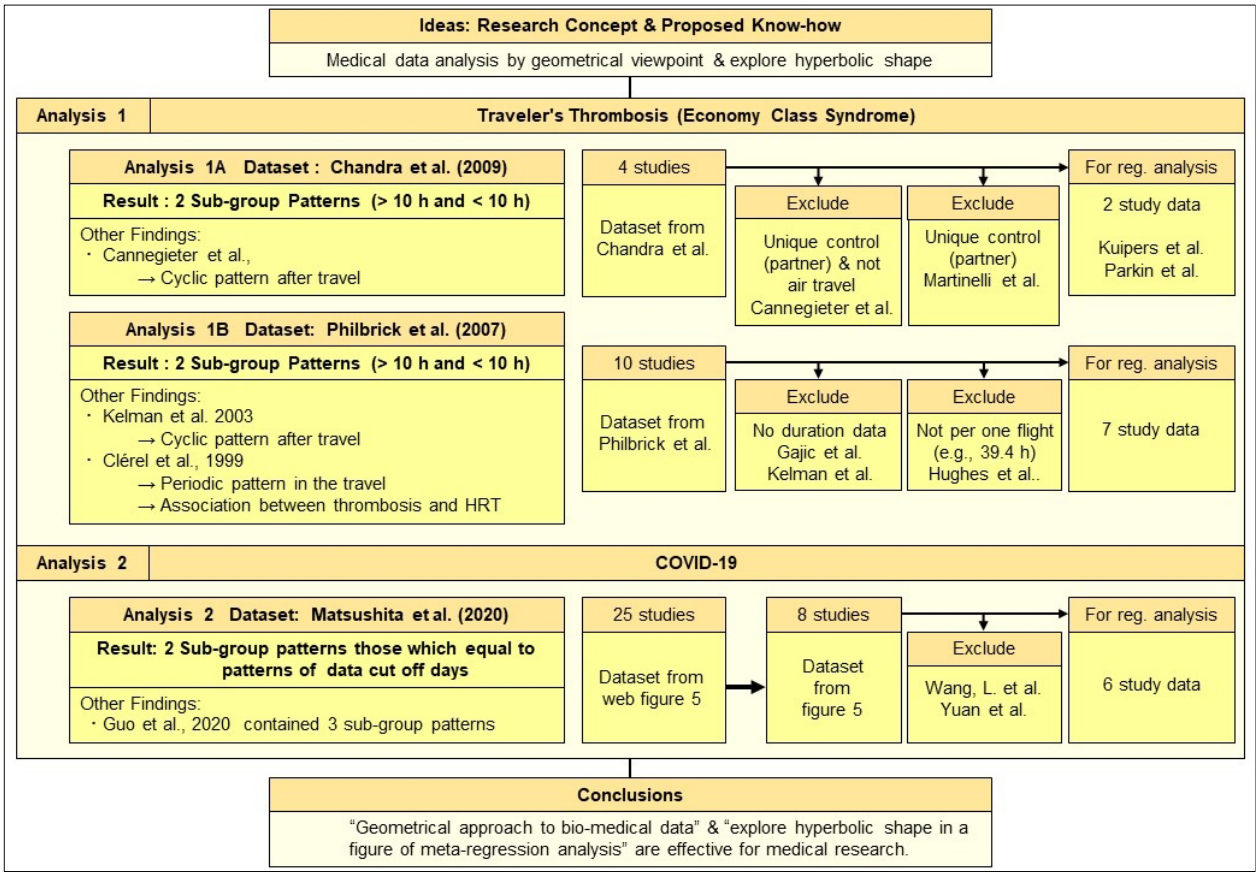

Fig. 2. Overall framework of this study.

We applied our concepts to travelers' thrombosis and COVID-19. These were similar in research history (see **Supporting discussion 1: history of traveler's thrombosis** & **Supporting discussion 2: research situations of COVID-19-related thrombosis**). In the case of traveler's thrombosis, we analyzed a dataset collected by Chandra et al. and Philbrick et al. (*13*, *16*). In the case of COVID-19, we analyzed a dataset collected by Matsushita et al. (*16*). Flow charts show data accept or reject flows. To avoid duplicating data, Matsushita et al. selected only eight studies data (25 studies included in initial web Figure 5) (*34*).

medRxiv preprint doi: https://doi.org/10.1101/2022.06.06.22275944; this version posted March 28, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Fig. 3

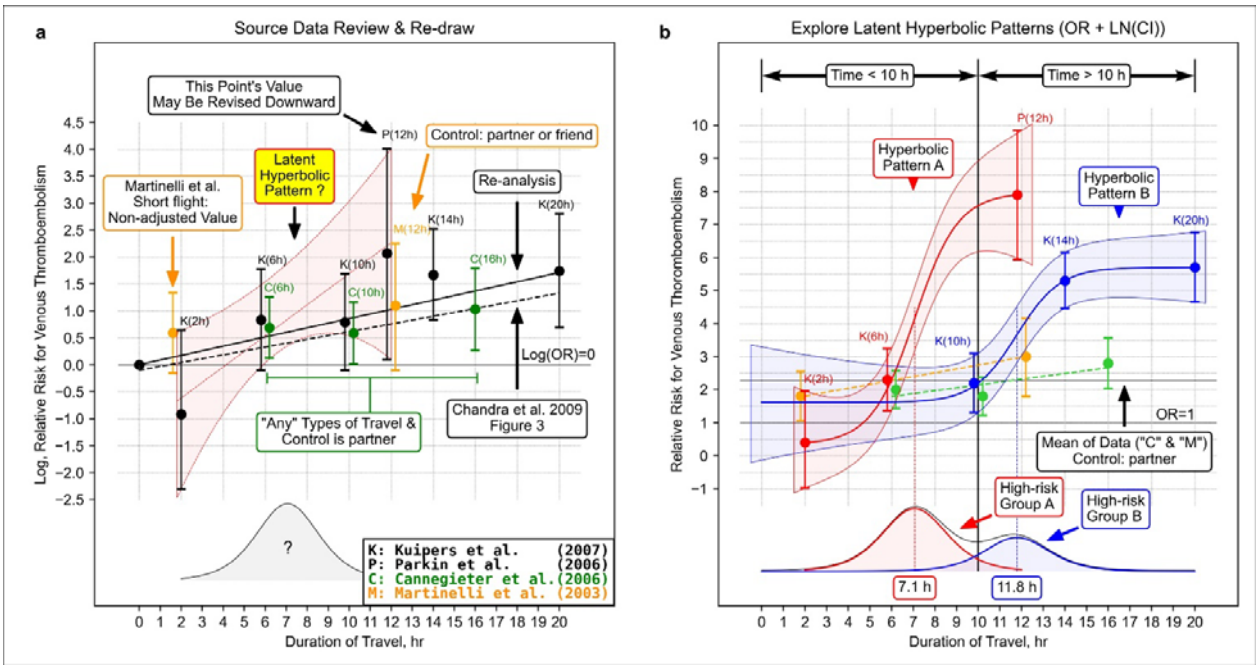

Fig. 3. Re-analysis based on the proposed ideas.

a, Data review and meta-regression analysis using the same method as Chandra et al. Ann Intern Med. 2009; 151(3):180-190. Figure 3 (*16*). We added information from original studies and our hypothesis to the previously published form, such as latent distribution. This figure is regarded as a re-drawn figure from Chandra D, Parisini E, Mozaffarian D. Meta-analysis: travel and risk for venous thromboembolism. Ann Intern Med. 2009 Aug 4;151(3):180-90. doi: 10.7326/0003-4819-151-3-200908040-00129. Epub 2009 Jul 6. © 2009 American College of Physicians. Adapted with permission. The original figure has been shown on the American College of Physicians website, which links to PubMedR (https://pubmed.ncbi.nlm.nih.gov/19581633/). **b**, grouping the data and hyperbolic shape fitting. The dots represent the odds ratio with error bars of the confidence interval are the adjusted odds ratio described in the four studies cited by Chandra et al. Since Cannegieter et al. did not describe values of confidence intervals, we calculated the values from the case-count data. We also calculated the odds ratio at the time point of short travel in the research reported by Martinelli et al. Chandra et al. did not display it might be due to the no adjusted odds ratio initially described. Our result of a meta-regression analysis was slightly different from the straight line drawn by Chandra et al. might be due to the software used by Chandra et al. (STATA) being different from the one we used (R and R package “metafor”) or they might conduct a meta-regression analysis reversing the front head and the front side of the cross-tabulation. It was impossible to fit S-shaped curves for the points excluded from the search for hyperbolic shapes (Martinelli et al. & Cannegieter et al.) due to the small number of data or the form of the point sequence. However, if we

medRxiv preprint doi: https://doi.org/10.1101/2022.06.06.22275944; this version posted March 28, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

manually fit S-curves, the two inflection points would appear close to the positions of the hyperbolic shapes' inflection points.

medRxiv preprint doi: https://doi.org/10.1101/2022.06.06.22275944; this version posted March 28, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Fig. 4

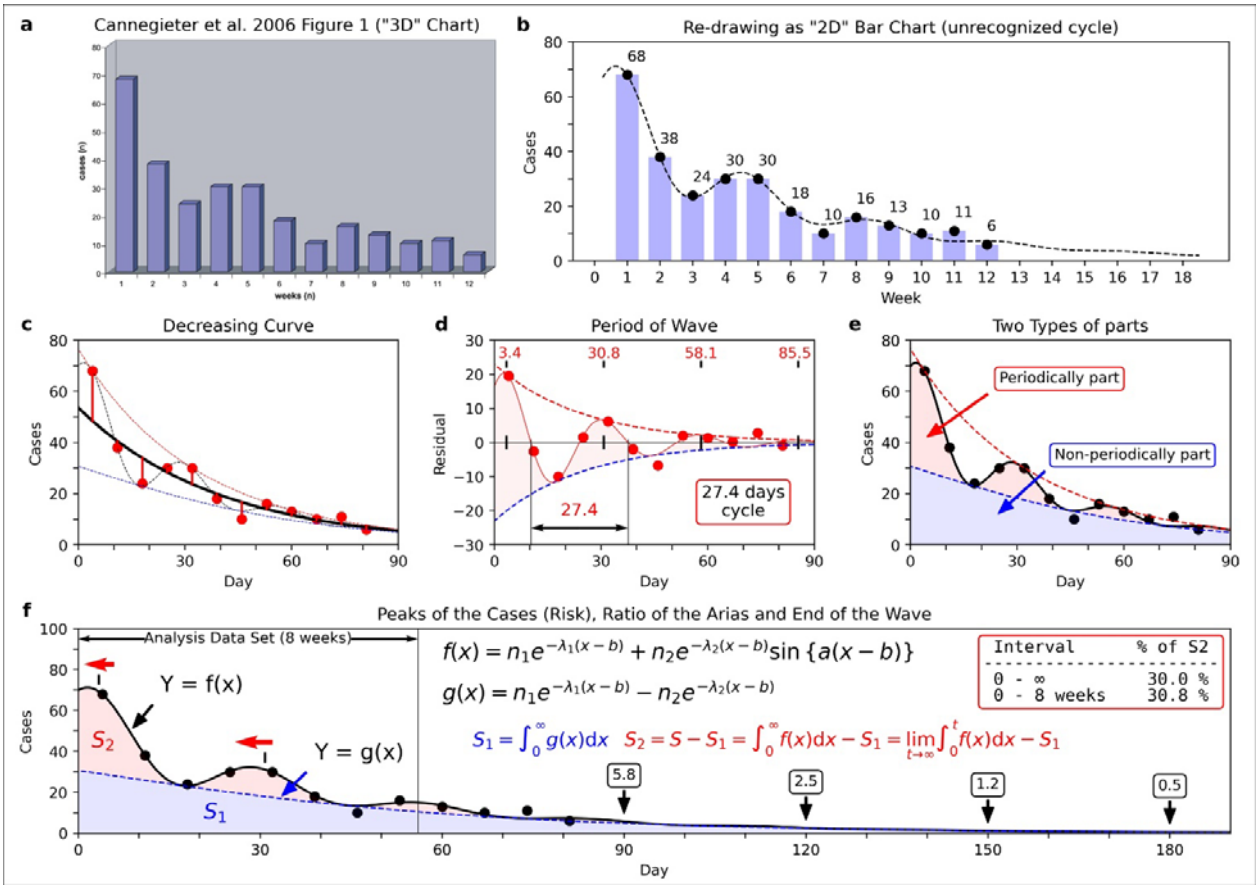

Fig. 4. Unrecognized cyclic pattern of thrombosis onset after travel (*2*).

a, A bar graph showing the relationship between weeks after travel and the number of thrombosis onset in a figure reported by Cannegieter et al. (*2*). **b**, Re-expressing as a two-dimensional bar graph avoiding the three-dimensional representation. **c**, Applying a damped wave by non-linear regression analysis. **d**, Extraction of damped wave part by subtracting the monotonic decrease function. **e**, Dividing into 2 sub-parts areas by the envelopes of the damped wave that touch the lower parts of the wave. **f**, Enlarged and added explanation of the ratio of the area. Cannegieter et al. (*2*) used the data up to week eight. The ratio of sub-parts 2 to the total number of patients (ratio of S2 to the area of S1 + S2) was 30.0%. In the integration for the infinite interval, it was 30.8%. Peak shifts of the wave (peak position of the wave shifted from the original to the other) appeared when comparing panels d and f due to putting by regression curve located in the center (c + d). Panel a was re-used from Cannegieter et al. PLoS Med. 2006; 3(8):e307. Figure 1.

https://www.ncbi.nlm.nih.gov/labs/pmc/articles/PMC1551914/figure/pmed-0030307-g001/ Copyright © 2006

medRxiv preprint doi: https://doi.org/10.1101/2022.06.06.22275944; this version posted March 28, 2023. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Cannegieter et al. Creative Commons Attribution License. In 2006, the Creative Commons Attribution 2.0 Generic, License (CC BY 2.0) was available. https://creativecommons.org/licenses/by/2.0/

medRxiv preprint doi: https://doi.org/10.1101/2022.06.06.22275944; this version posted March 28, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Fig. 5

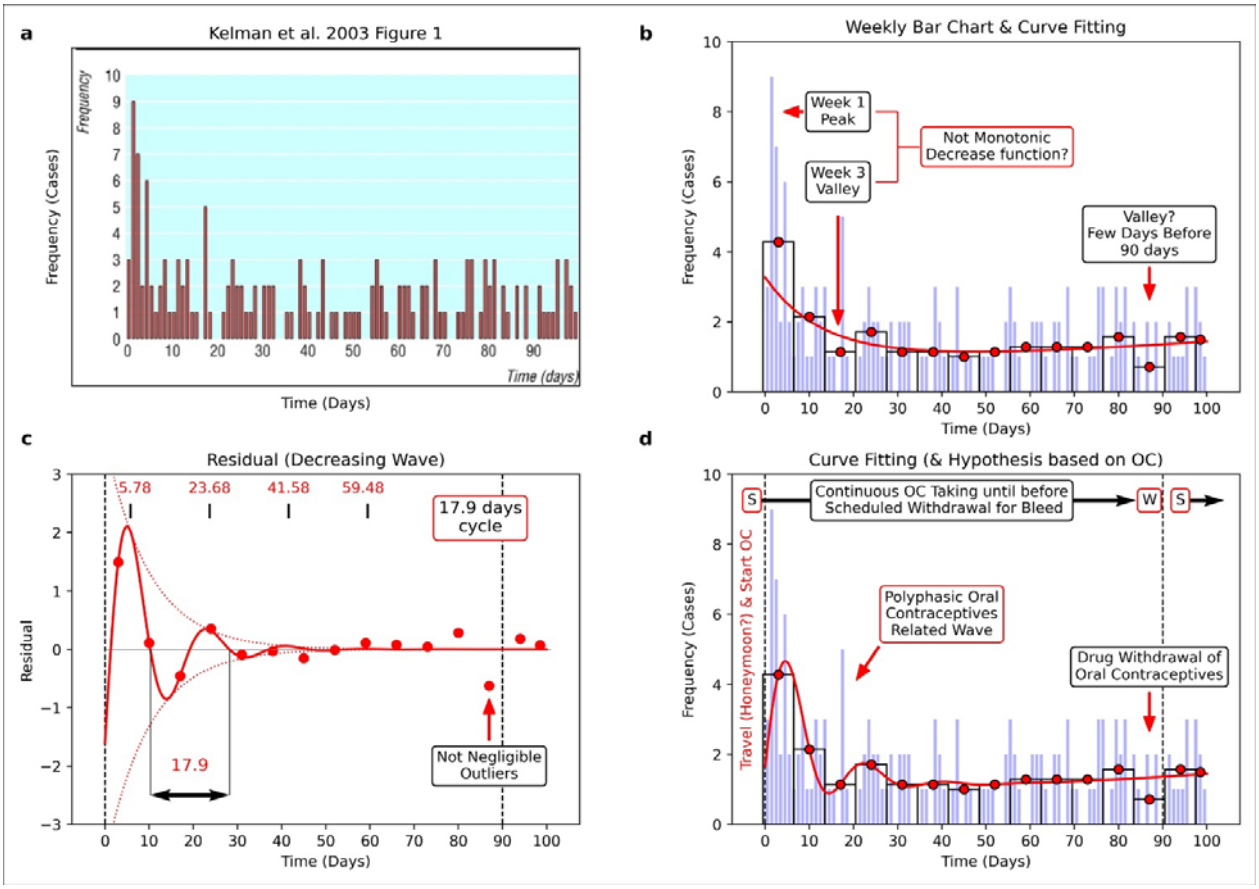

Fig. 5. Cyclic pattern of thrombosis onset appeared in a figure by Kelman et al. (*3*).

a, A thrombosis onset distribution reported by Kelman et al. (*3*). **b**, The regression curve is located in the center of all average points (the points express the average of thrombosis onset during seven days). **c**, Curve fitting to the residual data of the regression curve in panel b. **d**, Application of damped wave function and adding interpretation of the results assuming some women started taking oral contraceptives (OC) in the timing of travel. The observed cyclic pattern was relatively unclear than Cannegieter et al.(*2*) **(see Fig. 4)**. Noteworthy, the thrombosis onset downed in the days before 90 days, and it seems to be the scheduled withdrawal period of OC use. This figure was re-used and re-drawn from Kelman et al. BMJ. 2003; 327(7423):1072. Figure 1 https://www.bmj.com/content/327/7423/1072#F1 Copyright © 2003 BMJ Publishing Group Ltd. All rights reserved. The BMJ permission team thankfully confirms this figure adaptation. Also, we obtained permission to re-use.

medRxiv preprint doi: https://doi.org/10.1101/2022.06.06.22275944; this version posted March 28, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Fig. 6

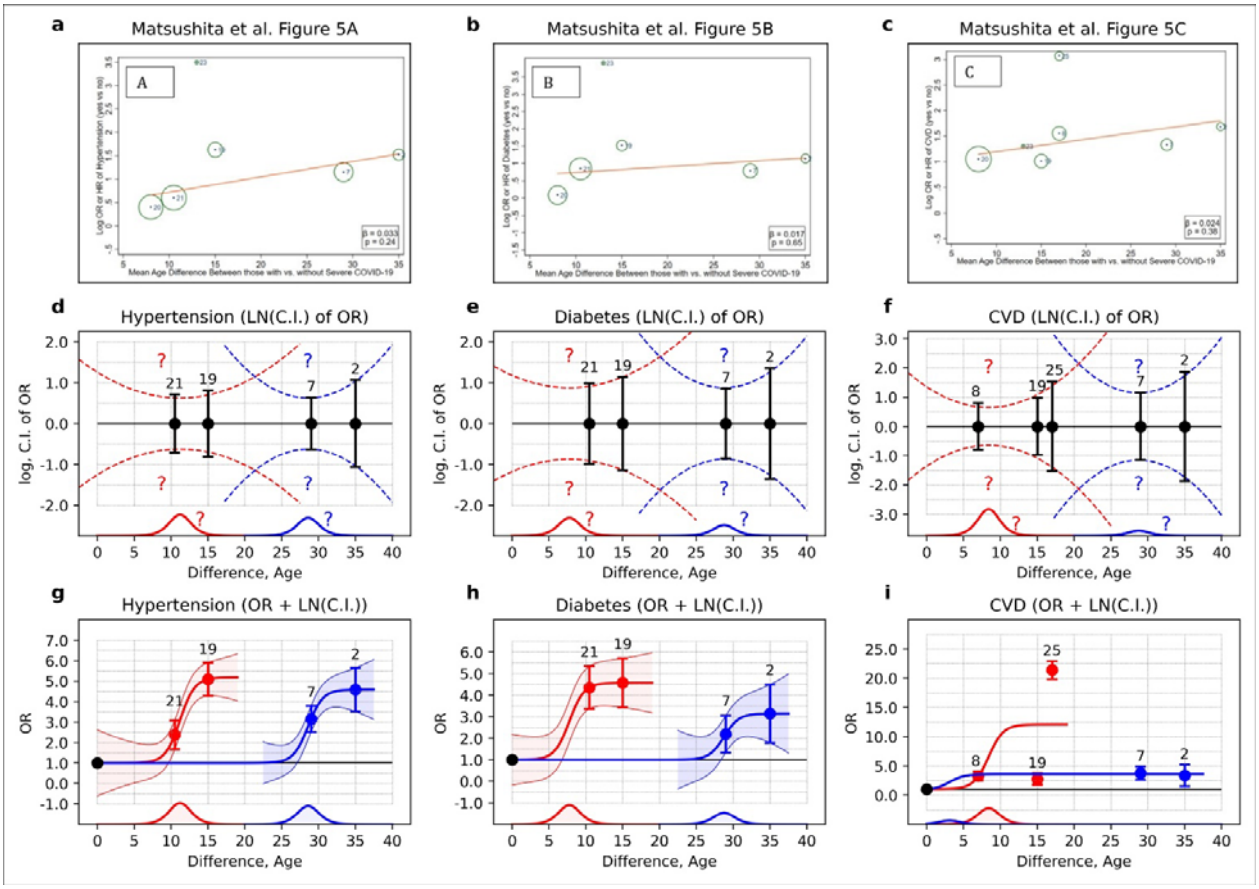

Fig. 6. Hyperbolic shapes found in the figure reported by Matsushita et al. (*34*).

a, A figure reported by Matsushita et al. shows the relationship between the severe and non-severe groups (*34*), which shows age difference on the horizontal axis and odds ratio (OR) or hazard ratio of hypertension on the vertical axis. To evaluate potential confounding for relative risk by age, Matsushita et al. conducted meta-regression analyses based on the assumption that there was the possibility of confounding by age in the case that the study with a larger age difference has a higher relative risk (*34*). **b**, Diabetes. **c**, Cardiovascular disease (CVD). **d-f**, Comparisons of error bars, which show 95% confidence interval (C.I.) s. It corresponds to the upper figure. **g-i**, Hyperbolic patterns were fitted to the OR and the 95% confidence limit of the OR. In the panel i, a hyperbolic shape could not be fitted due to the considerable data variation, likely due to the inconsistency of the term "CVD" (see **Supporting information: Method details**). The numbers marked at each point are the same as the numbers shown in the original figure. The sources of each data are shown in Methods. This figure was re-used from Matsushita et al. Glob Heart. 2020; 15(1):64. Figure 5.

https://www.ncbi.nlm.nih.gov/labs/pmc/articles/PMC7546112/figure/F5/

medRxiv preprint doi: https://doi.org/10.1101/2022.06.06.22275944; this version posted March 28, 2023. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Copyright © 2020 The Authors. Creative Commons Attribution 4.0 International License (CC-BY 4.0)

https://creativecommons.org/licenses/by/4.0/

medRxiv preprint doi: https://doi.org/10.1101/2022.06.06.22275944; this version posted March 28, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Fig. 7

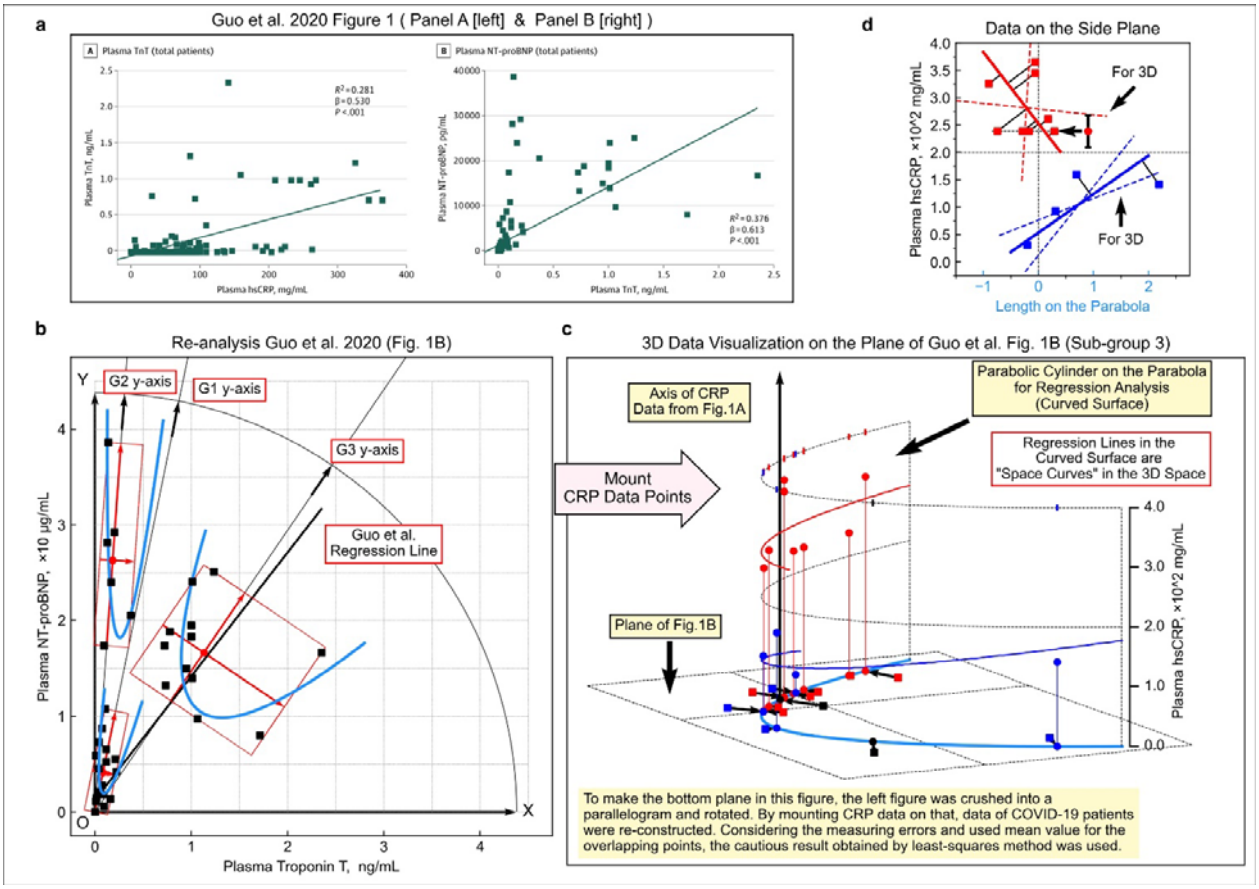

Fig. 7. Parabola shape patterns in the figure reported by Guo et al. (*4*).

a, Relationship between high sensitive C-reactive protein (hsCRP) and cardiac troponin T (TnT) in COVID-19 patients (left) and the relationship between cardiac troponin T (TnT) and N-terminal pro-brain natriuretic peptide (NT-proBNP) (right). **b**, Data points from the right side of panel a and fitting parabola. c, Three-dimensional data visualization was constructed by mounting the value of hsCRP onto panel b. **d**, Linear regression analysis on the side surface of the parabolic cylinder in panel **c**. The points were replaced with their average if the values could not be determined due to overlapping (the points indicated by the left-pointing arrow and the error bar, which are the average and the range of values, respectively). Panel a was re-used from Guo T et al. JAMA Cardiol. 2020; 5(7):811-818. Figure 1

https://www.ncbi.nlm.nih.gov/labs/pmc/articles/PMC7101506/figure/hoi200026f1/

Copyright © 2020 Guo T et al. JAMA Cardiology. Creative Commons Attribution License (CC-BY).

https://creativecommons.org/licenses/by/4.0/

medRxiv preprint doi: https://doi.org/10.1101/2022.06.06.22275944; this version posted March 28, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Table 1

Table 1. Changing female percentage from four studies reported in the 1970s to 2000s.

Author	Year	Total N	Male		Female		Larger one
			Male N	%	Female N	%	
Symington & Stack (***31***)	1977	n=8	n=5	62.5%	n=3	37.5%	Male
Cruickshank et al. (***32***)	1988	n=6	n=5	83.3%	n=1	16.7%	Male
Clérel & Caillard (***26***)	1999	n=64	n=15	23.4%	n=49	76.6%	Female
Parkin et al. (***25***)	2006	n=11	n=4	36.4%	n=7	63.6%	Female