

Springer Proceedings in Mathematics & Statistics

Nicola Salvati

Cira Perna

Stefano Marchetti

Raymond Chambers *Editors*

Studies in Theoretical and Applied Statistics

SIS 2021, Pisa, Italy, June 21–25



**Springer Proceedings in Mathematics &
Statistics**

Volume 406

This book series features volumes composed of selected contributions from workshops and conferences in all areas of current research in mathematics and statistics, including data science, operations research and optimization. In addition to an overall evaluation of the interest, scientific quality, and timeliness of each proposal at the hands of the publisher, individual contributions are all refereed to the high quality standards of leading journals in the field. Thus, this series provides the research community with well-edited, authoritative reports on developments in the most exciting areas of mathematical and statistical research today.

Nicola Salvati · Cira Perna · Stefano Marchetti ·
Raymond Chambers
Editors

Studies in Theoretical and Applied Statistics

SIS 2021, Pisa, Italy, June 21–25



Editors

Nicola Salvati
Department of Economics and Management
University of Pisa
Pisa, Italy

Cira Perna
Department of Economics and Statistics
University of Salerno
Fisciano, Salerno, Italy

Stefano Marchetti
Department of Economics and Management
University of Pisa
Pisa, Italy

Raymond Chambers
School of Mathematics and Applied
Statistics
University of Wollongong
Wollongong, NSW, Australia

ISSN 2194-1009 ISSN 2194-1017 (electronic)
Springer Proceedings in Mathematics & Statistics
ISBN 978-3-031-16608-2 ISBN 978-3-031-16609-9 (eBook)
<https://doi.org/10.1007/978-3-031-16609-9>

Mathematics Subject Classification: 62-06

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2022, **corrected publication 2023**

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

This book gathers selected peer-reviewed papers presented during the 50th Scientific Meeting of the Italian Statistical Society (SIS2021). Due to the Covid-19 pandemic, which limited the mobility of the staff of many universities and research centres, SIS2021 was conducted remotely from the 21st to the 25th of June 2021.

This biennial conference is a traditional meeting for promoting interactions among national and international researchers in statistics, demography, and applied statistics in Italy. The aim of the conference is to bring together national and foreign researchers and practitioners to discuss recent developments in theoretical and applied statistics as well as in demography and statistics for the social sciences.

The Scientific Program Committee and the Organizing Committee of SIS2021 put together a balanced and stimulating program which was of great interest to all participants.

The conference program included 4 plenary sessions, 15 specialized sessions, 20 solicited sessions, 37 contributed sessions, and the poster exhibition. The meeting also hosted three Satellite Events on ‘Measuring uncertainty in key official economic statistics’, ‘Covid-19: the urgent call for a unified statistical and demographic challenge’, and ‘Evento SIS-PLS Statistica in classe: verso un insegnamento laboratoriale’. There were 323 submissions accepted by the Scientific Program Committee, including 128 that were presented at invited plenary, specialized and solicited sessions, and 195 that were submitted as contributed papers for oral presentation and for the poster sessions.

This book of selected papers from those presented at SIS2021 covers a wide variety of subjects and provides an overview of the current state of Italian scientific research in theoretical and applied statistics. The papers contained in this book cover areas that include Bayesian models, survey methods, time series models, spatial models, finance models, clustering methods, and new methods and applications to Covid-19.

The Scientific Program Committee, the Organizing Committee, and many volunteers contributed to the organization of SIS2021 and to the refereeing of the papers included in this book. Our heartfelt thanks go to all of them. A special thank you

goes to Francesco Schirripa Spagnolo for his continuous assistance and support in the organization of the conference and in the editing of this book.

Wishing you a productive and stimulating reading experience.

Pisa, Italy

Salerno, Italy

Pisa, Italy

Wollongong, Australia

Nicola Salvati

Cira Perna

Stefano Marchetti

Raymond Chambers

Contents

A Composite Index of Economic Well-Being for the European Union Countries	1
Andrea Cutillo, Matteo Mazziotta, and Adriano Pareto	
A Dynamic Power Prior for Bayesian Non-inferiority Trials	15
Fulvio De Santis and Stefania Gubbiotti	
A Graphical Approach for the Selection of the Number of Clusters in the Spectral Clustering Algorithm	31
Cinzia Di Nuzzo and Salvatore Ingrassia	
A Higher-Order PLS-SEM Approach to Evaluate Football Players' Performance	45
Mattia Cefis and Maurizio Carpita	
A Latent Markov Approach for Clustering Contracting Authorities over Time Using Public Procurement Red Flags	57
Simone Del Sarto, Paolo Coppola, and Matteo Troia	
A Multiplex Network Approach for Analyzing University Students' Mobility Flows	75
Ilaria Primerano, Francesco Santelli, and Cristian Usala	
A Statistical Model for Predicting Child Language Acquisition: Unfolding Qualitative Grammatical Development by Using Logistic Regression Model	91
Andrea Briglia, Massimo Mucciardi, and Giovanni Pirrotta	
Adaptive COVID-19 Screening of a Subpopulation	105
Fulvio Di Stefano and Mauro Gasparini	
Alternative Probability Weighting Functions in Behavioral Portfolio Selection	117
Diana Barro, Marco Corazza, and Martina Nardon	

Bayesian Quantile Estimation in Deconvolution	135
Catia Scricciolo	
Can the Compositional Nature of Compositional Data Be Ignored by Using Deep Learning Approaches?	151
Matthias Templ	
Citizen Data and Citizen Science: A Challenge for Official Statistics	167
Monica Pratesi	
Detecting States of Distress in Financial Markets: The Case of the Italian Sovereign Debt	175
Maria Flora and Roberto Renò	
Forecasting Combination of Hierarchical Time Series: A Novel Method with an Application to CoVid-19	185
Livio Fenga	
Frequency Domain Clustering: An Application to Time Series with Time-Varying Parameters	219
Raffaele Mattera and Germana Scepti	
Heterogeneous Income Dynamics: Unemployment Consequences in Germany and the US	239
Raffaele Grotti	
How Much Do Knowledge About and Attitude Toward Mobile Phone Use Affect Behavior While Driving? An Empirical Study Using a Structural Equation Model	263
Carlo Cavicchia, Pasquale Sarnacchiaro, and Paolo Montuori	
Impact of Emergency Online Classes on Students' Motivation and Engagement in University During the Covid-19 Pandemic: A Study Case	281
Isabella Morlini	
Local Heterogeneities in Population Growth and Decline. A Spatial Analysis of Italian Municipalities	297
Federico Benassi, Annalisa Busetta, Gerardo Gallo, and Manuela Stranges	
Model Predictivity Assessment: Incremental Test-Set Selection and Accuracy Evaluation	315
Elias Fekhari, Bertrand Iooss, Joseph Muré, Luc Pronzato, and Maria-João Rendas	
Multiple Measures Realized GARCH Models	349
Antonio Naimoli and Giuseppe Storti	

Multiversal Methods in Observational Studies: The Case of COVID-19 369
 Venera Tomaselli, Giulio Giacomo Cantone, and Vincenzo Miracula

Neural Network for the Statistical Process Control of HVAC Systems in Passenger Rail Vehicles 393
 Fiorenzo Ambrosino, Giuseppe Giannini, Antonio Lepore, Biagio Palumbo, and Gianluca Sposito

On the Use of the Matrix-Variate Tail-Inflated Normal Distribution for Parsimonious Mixture Modeling 407
 Salvatore D. Tomarchio, Antonio Punzo, and Luca Bagnato

Population Size Estimation by Repeated Identifications of Units. A Bayesian Semi-parametric Mixture Model Approach 425
 Tiziana Tuoto, Davide Di Cecco, and Andrea Tancredi

Spatial Interdependence of Mycorrhizal Nuclear Size in Confocal Microscopy 435
 Ivan Sciascia, Andrea Crosino, Gennaro Carotenuto, and Andrea Genre

Spatially Balanced Indirect Sampling to Estimate the Coverage of the Agricultural Census 449
 Federica Piersimoni, Francesco Pantalone, and Roberto Benedetti

The Assessment of Environmental and Income Inequalities 463
 Michele Costa

The Italian Social Mood on Economy Index During the Covid-19 Crisis 475
 A. Righi, E. Catanese, L. Valentino, and D. Zardetto

The Rating of Journals and the Research Outcomes in Statistical Sciences in Italian Universities 487
 Maria Maddalena Barbieri, Francesca Bassi, Antonio Iripino, and Rosanna Verde

Trusted Smart Surveys: Architectural and Methodological Challenges Related to New Data Sources 513
 Mauro Bruno, Francesca Inglese, and Giuseppina Ruocco

Web Surveys: Profiles of Respondents to the Italian Population Census 531
 Elena Grimaccia, Alessia Naccarato, and Gerardo Gallo

Publisher Correction to: Forecasting Combination of Hierarchical Time Series: A Novel Method with an Application to CoVid-19 C1
 Livio Fenga

Multiversal Methods in Observational Studies: The Case of COVID-19



Venera Tomaselli, Giulio Giacomo Cantone, and Vincenzo Miracula

Abstract In the present study, 13 covariates have been selected as potentially associated with 3 metrics of the spread of COVID-19 in 20 European countries. Robustness of the linear correlations between 10 of the 13 covariates as main regressors and the 3 COVID-19 metrics as dependent variables have been tested through a methodology for sensitivity analysis that falls under the name of “Multiverse”. Under this methodology, thousands of alternative estimates are generated by a single hypothesis of regression. The capacity of identification of a robust causal claim for the 10 variables has been measured through 3 indicators over a Janus Confusion Matrix, which is a confusion matrix that assumes the likelihood to observe a True claim as the ratio between the absolute difference of estimates with a different sign and the total of estimates. This methodology provides the opportunity to evaluate the outcomes of a shift from the common level of significance $\alpha = .05$ to the alternative $\alpha = .005$. According to the results of the study, in the dataset the benefits of the shifts come at a very high cost in terms of false negatives.

Keywords Multiverse analysis · Model mis-specification · *p*-hacking, significance level, COVID-19

V. Tomaselli (✉)

Department of Political and Social Sciences, University of Catania, 8, Vittorio Emanuele II, 95131 Catania, Italy
e-mail: venera.tomaselli@unict.it

G. G. Cantone · V. Miracula

Department of Physics and Astronomy “E. Majorana”, University of Catania, 64, S. Sofia, 95123 Catania, Italy
e-mail: giulio.cantone@phd.unict.it

V. Miracula

e-mail: vincenzo.miracula@phd.unict.it

1 Introduction

In hypothesis testing, the probability of randomly drawing, in the set of theoretical circumstances described under the null hypothesis, a value as extreme as the value empirically observed, is referred as the p -value. Historically, a p -value smaller than the level of $\alpha = 0.05$ signalled a *statistically significant* the result of the test, which means that the presented evidence is not compatible with the null hypothesis. In this case the researcher has a justification to reject the null hypothesis [87].

Over time, the status of p -values in scientific research reached a situation of paradox: null hypothesis statistical testing (usually with a significance level $\alpha < 0.05$) is the most taught and the most used method for scientific inference, and at the same time, by its broad use, it started to be regarded as the main culprit for the lack of reliability in science [30, 55, 67, 82]. Skepticism and self-criticism towards standard scientific practices raised after the mediatic emergence of the so called ‘replicability crisis’ that invested psychological research [2, 10, 21, 52, 58, 85], health sciences [4, 34, 36], and is a concern for other fields [2, 6, 20, 61, 80, 91].

While there are methodological reasons why a null hypothesis statistical testing could lead into a lack of not replicated scientific claims, the focus is on the noxious practice of ‘ p -hacking’ [33, 54, 77]. It consists in collecting a large amount of model specifications of a scientific claim in order to randomly see a significant result popping up by chance, then not reporting the exact number of attempts before reaching a $p < \alpha$.

Many scholars [35, 38, 48, 83, 88] saw an opportunity for a reflection about the possibilities for an advancement beyond the rule the *status quo* of hypothesis testing with $\alpha < 0.05$. A widely discussed proposal is to lower conventional α into 0.005 [5, 37, 42]. This proposal would not require to upgrade introductory classes in Statistics towards Bayesian or other methodologies. However, the raise in the false negative rate could prevent researcher to pursue highly innovative hypotheses. Also, to lower the α is virtually useless against the most extensive forms of p -hacking, e.g., on massive datasets with hundreds of variables (Big Data).

The issue of p -hacking inspired a different approach that falls under the name of multiversal methodology or ‘multiverse-style methods’ [18]. This methodology aims at observing a multivariate population (usually consisting of p -values and estimates $\hat{\beta}_x$) throughout mapping every single reasonable model specification of a causal relationship between a dependent variable y and a set of regressors (x, Z) . This population is referred as ‘the multiverse’ of a study.

The multiverse tells something about the sensitivity of specific hypotheses $x \rightarrow y$ to their alternative specifications or *vibrations* [65]: multiversal statistics are informative regarding the practical possibility to p -hack the hypothesis or to incur false positives due mis-specification.

In the present study, the multiversal methods are discussed to estimate coefficients of multiple specifications of regressive models in Sect. 2. In Sect. 3, population, mobility, pollution, and public health variables in European countries and COVID-

19 data are selected for the analysis. In Sect. 4, the multiverse models are employed to data analysis. Lastly, in Sect. 5, multiverse models provide evidences about the COVID-19 pandemic spread and its effects on the health-care systems.

2 Theory: Multiversal Methods

All the multiversal methods are based on the estimation of coefficients of a large number of regressive models organised as a ‘family’. The technical premises are the same of Extreme Bound Analysis (EBA) [29, 46, 71] but research questions are broader. The theoretical connection between the number of different attempts to ‘make a model work’ and the robustness of its scientific claim was made explicit in [77] and [28], which popularised the concept of ‘Researcher’s degree of freedom’ with the metaphor of the “Garden of Forking Paths”, a literary invention of novelist Borges.

A multiverse of specifications can be analysed through plotting:

- The Multiverse grid [32, 81] is a multidimensional array with all the specifications represented by their p -values, clustered in the grid space by the divergences in the ‘Garden of Forking Paths’. Significant p -values are highlighted. This tool is impractical for a high number of specifications and is usually uninformative on the estimate.
- The Vibration-of-Effect (VoE) plot [65], a cartesian representation where
 - in the x-axis are represented the estimates of a standardised regressor x in the specifications of the multiverse,
 - in the y-axis are represented the logarithms of $\log_{10}(p)$ associated to the null-hypothesis of coefficient equal to 0, multiplied *per* -1 .

VoE is used to display the sensitivity of a causal relationship between regressor and dependent as the p -value decreases. At the same time, it allows to detect the so-called “Janus effect”, i.e., the sign discordance between estimates in the same family of specifications.

- The p -curve [7, 81] is a representation of the probability density associated to multiversal p -values: the more the density in p -curve is right-skewed towards lower p -values, the more the regressor is validated.
- The Specification Curve [78], which allows to compare different families of specifications (i.e., aggregations of micro-variants of a single causal hypothesis) and to associate these families both to an interval of p -values and to an interval of estimates.

Worth to mention in the family of multiversal methods is the Computational Framework for Multimodel Analysis [56, 92], which is an alternative to Machine Learning for model selection in a multivariate context.

An interesting application of the multiverse analysis is in [72]. In that paper the three authors had *divergent* results regarding the determinants of political behaviour of U.S. representatives. Instead of defending own theses to the bitter, the three shared their methodological designs (e.g., how to collect data), chose to collaborate to investigate the robustness of the divergent hypotheses through a multiverse of specifications (in the jargon of [65], they ‘vibrated’ them), and finally converged into a unique set of scientific claims.

Other applications of multiverse analysis in empirical research are in [16, 50, 59, 70].

A software to perform Specification Curve Analysis and, more in general, to generate a multiverse from a dataset is *spectr* [51]¹.

However, a properly unified multiversal methodology is still in development. The most theoretical contribution to the topic is by Del Giudice and Gangestad [18]. While the authors highlight both promising features and pitfalls of “The Multiverse”, their main concern is with the phase of analytical choice in order to differentiate (*vibrate*) the specification of a hypothesis into families of model specifications. The argument provided by Del Giudice and Gangestad follows the more known scientific controversy regarding the introduction of a collider variable as a control in a regression model [48].

In the context of multiverse analysis, the controversy could be simplified to only a question: what is a reasonable vibration for a hypothesis?

Del Giudice and Gangestad [18] discuss about the covariates’ selection as a basic issue in the literature. According to Simonsohn et al. [79] the covariates are linked to the chance to provide different answers to different research hypotheses. On the contrary, Patel et al. [65] demonstrate that the VoE emerges only with robustness analyses involving selected and alternative covariates. These controversial claims show that the lacking of agreement about clear and accurate guidelines does not allow to increase the potential of multiverse methods in data analysis.

2.1 What Is a Specification?

A model specification of the causal relationship $x \rightarrow y$:

$$y = f(x, Z) \tag{1}$$

is a member of the family of regressive models formalised as:

$$y_{k_y} \rightsquigarrow F_{k_f}(\overset{\leftrightarrow}{x}_{k_x}, \overset{\leftrightarrow}{Z}_{k_z}) + \epsilon \tag{2}$$

where \leftrightarrow indicates an operationalization, i.e., a decision to represent a theoretical concept (e.g., a statistical population) through a full identified object (e.g., an empiric

¹ Suggested tutorial: https://dcosme.github.io/specification-curves/SCA_tutorial_inferential.

sample). In Eq. 1, Z is a set of control covariates that the researchers necessary deem for the correct causal inference. In Eq. 2, F represents the set of equivalent functions that link the joint predictors x and $z \in Z$ to the outcomes y . k are indices for single operationalizations of the respective constructs [78].

An interesting propriety of the operationalization is that it does not only involve decisions about what to measure but also about *how* to measure it. Indeed, a core element of the methodology in Simonsohn, Simmons, and Nelson [78] is that they stress the importance of recoding the same observations through different scales of measurements.

In the literature, five elements of a specification are often reported:

1. The **Subsetting** of observations: here the decision regards mostly the inclusion of outliers or other peculiar clusters of observations. However, as a general rule, finding a reasonable criterion to split the dataset into subsets should help to assert the sensitivity of the relationship [3].
2. The **Regressors** (x, Z): this operationalization can be split into two different decisions:
 - a. one regards the controversy about full inclusion of all the $n!$ combinations of the n covariates as regressors x (and/or controls) or to make a ‘reasonable selection’. Anyway, already in the operationalization of the Subset there is an implicit decision regarding what variables to observe. In [65] the decision comes after a literature review, so it is only natural that the authors include all the covariates in the multiverse both as x and Z . The same approach could have not been feasible for the goal of convergence into a unique set of claims in [72] or for correct identification of the causal model [18]. Multi-model analysis [56, 92] is an interesting method to solve this controversy, since it includes all combinations to begin but then it tunes the multiverse model (“Multimodel”) by removing the comparatively worse specifications. However, if the goal of the multiverse is exploratory and not conclusive, a full inclusion could be more useful than risky.

The criterion to exclude a z variable from the possibility to work as a x in the multiverse may be disconnected to any scientific evaluation and be more practical. For example, being older could decrease bone mass but the researchers could have no practical interest in just ‘revert people age’, while being interested in asserting dietary advice to contrast reduction of bone mass due aging. In this case the exclusion of variable ‘age’ from x does not mean that age does not control the impact of observed diet in the multiverse but that the statistics (p -value, estimates) of ‘age’ as a regressor are not reported among the multiversal statistics, since they are not of research’s interest.

- b. the second decision regards how to measure the conceptual dimension implied in the hypothesis, for example, by adoption of proxies.²

² Think about the deep metrological differences between Richter and Mercalli scales in measurement of earthquake magnitude.

3. The **Dependent** y : again, there are two approaches. In a sense, it is true that two different dependents provide ‘two answers for two different questions’, hence they generate two different multiverses [18]; but at the same time it also makes sense the adoption of different proxies of the same response variable if the question truly regards sensitivity of the analysis to the *adoption of a proxy*, which is a legitimate research interest that can be explored through multiversal methods.
4. The **Type** of regressions: the main issue in processing the same hypothesis under different types of regression is that even after standardisation of the variables, estimates are not always comparable if not forcing some functional form for x . In binomial regression *vs.* linear regression, to keep comparability one have to force y to assume values in the unitary interval and then estimate the coefficient on $\log(x)$ and at the same time avoid the logarithmic transformation on z : there is an addition of variety of vibration in a sense but also a negation of variety in another one. However, if the research is focused on p -values and not on estimates, it is definitely worth to compute multiversal statistics for more than one type of regression.
5. The **Functional Form** (f) has analogue issues to the type of regression. It is already mentioned that sometimes it depends by the decision regarding vibration of Regression Type. It is worth to remember that the function takes as many arguments as the n of covariates, hence for any $n > 3$ the size of the whole spectrum of many alternative functional forms could be impractical to compute. For $n < 3$, the impact of an exotic change in the functional form could seriously make impossible the interpretation of the coefficients. In absence of reasons to do so, having degrees of freedom regarding f design could be an error.

Decisions on Subsetting and Regression Type impact on all of $\overset{\leftrightarrow}{y}_{k_y}$, $\overset{\leftrightarrow}{F}_{k_F}$, $\overset{\leftrightarrow}{x}_{k_x}$, and $\overset{\leftrightarrow}{Z}_{k_Z}$. Each alternative decision about an element of the specification increases the number of specifications in the multiverse.

2.2 Janus Effect

Given a null value β_0 for coefficient and a value of α of statistical significance, the ‘Janus effect’ [62, 65] is the presence in the multiverse of regressor x of vibrated estimates $\hat{\beta}_x > \beta_0$ and $\hat{\beta}_x < \beta_0$, both such that $p(\beta) < \alpha^3$.

The interpretation of the presence of Janus effect is worrisome: it means that, given α , it is possible for a researcher to claim both positive or negative association between two observational variables in a population. Just by mapping a multiverse and specifying an *ad hoc* model, a desired causal claim can be p -hacked. The implications of Janus effect for clinical research are broadly discussed in [65] and [62].

³ Janus was the Roman god of gates and was always represented with two faces pointing towards opposite directions, hence the name of the effect.

Table 1 Janus matrix of a x -multiverse

	$\beta_x < \beta_0$	$\beta_x > \beta_0$
$p(\beta_x) < \alpha$	A	B
$p(\beta_x) \geq \alpha$	C	D

Table 2 Janus confusion matrix

	One-faced	Two-faced
$p(\beta_x) < \alpha$	$ A - B $	$(A + B) - A - B $
$p(\beta_x) \geq \alpha$	$ C - D $	$(C + D) - A - B $

Given the dependency of Janus effect from α and β_0 is not surprising to see proposals to lower α or to shift the research on estimation of intervals for coefficients instead of looking for significant effects.

In [65] Janus effect is treated mostly as something that is there or is not, however in a multiverse made of many specifications, the magnitude of Janus effect can be observed through counting how many significant specifications hold $\beta_x > \beta_0$, and how many $\beta_x < \beta_0$.

More in general, all the estimates $\hat{\beta}_x$ in a x -multiverse can be represented through a tetrachoric matrix (Table 1):

The two dimensions in Table 1 do not share the same proprieties, though: $p(\beta_x) < \alpha$ does signal a desired condition, $p(\beta_x) \geq \alpha$ does not. The same cannot be said by comparing $\beta_x < \beta_0$ and $\beta_x > \beta_0$. The desirable outcome is to maximise into 1 the ratio:

$$\frac{|A - B|}{A + B} \tag{3}$$

which can be interpreted as the fraction of the significant results leading towards a supposedly *true* direction of the coefficient.

One can catch here the analogy of the statistical measure of Precision $\frac{TruePositives}{Positives}$. However, Precision alone does not account for sensitivity of the test to false negatives, so is usually paired to Recall $\frac{TruePositives}{True}$.

The whole Table 1 can be remapped as a Confusion Matrix (Table 2):

So, if $Precision \simeq \frac{|A-B|}{A+B}$, then,

$$Recall \simeq \frac{|A - B|}{|A - B| + |C - D|} = \frac{|A - B|}{|(A + C) - (B + D)|} \tag{4}$$

3 Materials

3.1 *Why Coronavirus*

COVID-19 emerged in Wuhan (China) in late 2019. In two years, the virus has spread to more than 200 countries worldwide. The outbreak was declared a pandemic by the World Health Organisation on February 22, 2020, and hundreds of millions of COVID-19 cases have been reported, causing millions of deaths [90].

Given the uniqueness of the virus and its biological characteristics, it has been able to spread so rapidly that COVID-19 has become a public health problem [49]. The rapid growth in infection rates in each country has had a severe impact on the capability of health-care services to tackle the pandemic. As discussed in [66], many policies were implemented to reduce the deaths due to the spread of the virus, to limit the growth in the number of infected people, and overall to empower the health-care services. In addition, population control strategies have been implemented, too.

The most monitored and analysed variables associated with infection risk [69] to study the COVID-19 spread have been demographic characteristics [23], passenger mobility [12, 57], air pollution [17, 19], and comorbidity [8, 9, 26, 63].

The rapid spread of COVID-19 has triggered an uncommon increase in research activities leading to an extensive production of several observational studies. However, most of the studies do not go beyond modeling the relationship between COVID-19 and some variables (e.g., air pollution). As a consequence, a data analysis based only on single relationships could be limited and provide misleading results.

In particular, the case of the COVID-19 outbreak is showing the critical role of information dissemination which can strongly influence people's behaviour and alter the effectiveness of countermeasures implemented by governments [27, 45].

It is common for researchers to explore several analytical alternatives [11, 41, 68, 74], i.e., to look for a significant combination in order report only it [5, 48, 84]. Multiple approaches are capable of drawing causal findings from observational data as shown in [7, 53]. In all of the approaches, substantial uncertainty remains about the best model to apply. A multiverse of possible alternatives (or other forms of robustness checks) needs in order to explore how the findings would differ if different assumptions were been adopted [36].

In the present study, a multiverse modeling approach is proposed to process COVID-19 pandemic data. Multiverse analysis is proposed as a suitable method to analyse data when uncertainty could lead to mis-specification of relationships among variables [73, 74].

3.2 Selection of Covariates

One of the challenges in epidemiology is that epidemics happen within societies, and societies are very complex phenomena with a lot of features being relevant for epidemiological models.

Following the example of [62, 65], and [14], the present study aims at selecting covariates that have been broadly discussed in the epidemiological literature of COVID-19 and are available through National Public Health Departments.

The dataset of the present study is made by 16 covariates (see, Table 3). Of the 16, 3 variables are entered as dependents in the models:

1. the count of hospitalised patients with COVID-19
2. the count of hospitalised in intensive care (ICU) with COVID-19
3. the reported count of cases of COVID-19, in the countries.

The observed values of these 13 variables are summary statistics of epidemiological dimensions observed in 20 European countries. They are counted in 4 different time intervals plus the cumulative count from the start of the first interval to the end of the fourth, so, for 20 countries, the total amount of observations in the dataset is $(4 * 20) + 20 = 100$. The other 13 population variables are not collected along the 4 time intervals but they are updated at 2019 and 2020, so their values are fixed across time.

All the variables and the counts are normalised to the population of each country and then standardised. Furthermore, four conceptual dimensions are detected as shown in Table 3:

Details about the selected variables on the basis of the more updated research findings in the literature are below described.

COVID-19 Data

Metrics on the spread of Covid over the four phases (Sect. 4) are those collected by World Health Organisation⁴ and the European Center of Disease and Control (ECDC)⁵, which collect case data submitted by national governments. Where possible, they aim to report confirmed cases.

The main difference among the three cumulative metrics of COVID-19 in their impact over health-care systems, is that the count of reported cases is likely to be biased by many sources of under-reporting while hospitalised patients, both in ICU and not, are unlikely to be asymptotically biased [89]. The variable *Hosp* refers to people who have contracted COVID-19 and need hospitalization, both in the ICU or in other hospital departments and the variable *ICU* includes only patients in ICU departments. Both the variables deal with the two conditions (infected and in need of hospitalisation) as if two events are causally independent. These are also referred as “patients-*with*-COVID-19” and are distinct from “patients-*for*-COVID-19”, which

⁴ <https://www.who.int/>.

⁵ <https://www.ecdc.europa.eu/en>.

Table 3 Variables in multiverse model

Dimension	Description of variable	Labels	Source	Year
COVID-19	N. Confirmed Cases	Cases	National Public Health Dept.	
COVID-19	N. Intensive Care Unit	ICU	National Public Health Dept.	
COVID-19	N. Hospitalised	Hosp	National Public Health Dept.	
Demography	Urban Population Index	UrbanPop	EuroStat	2020
Demography	N. over 65+ years	Over65	EuroStat	2020
Demography	Population Density Index	PopDensity	EuroStat	2020
Health	% Cardiovascular risk	Cardio	EuroStat	2020
Health	% Diabetes prevalence	Diabetes	EuroStat	2020
Health	% Smokers	Smoking	EuroStat	2020
Health	% Obeses	Obesity	EuroStat	2020
Health	% High blood pressure	HiPressure	EuroStat	2020
Pollution	PM2.5 Index	PM2.5	EuroStat	2019
Pollution	PM10 Index	PM10	EuroStat	2019
Pollution	CO2 Index	CO2	EuroStat	2019
Mobility	N. aeroportual passengers	AirPass	EuroStat	2019
Mobility	N. train passengers	TrainPass	EuroStat	2019

is the case when COVID-19 induces hospitalisation. *ICU* and *Hosp* are two mutual proxies.

Sources of under-reporting of COVID-19 cases are due to both delays in reporting cases and prevalence of asymptomatic infected people. In particular, the sources claim that suspected cases are not reported. Another issues dealing with different countries and institutions is that the delay in updating the number of confirmed cases is never consistent among cases. This is due to differences in times of reporting a new tested case and its inclusion in national statistics.

In general, especially in the first phase (see, Table 4), the number of confirmed cases is underestimated. Nevertheless this outcome could be a *vulnus* for scientific research about the spread of the virus, the divergence between *Cases* and the two variables *ICU* and *Hosp* is useful to shown the capability of multiversal methods in regressive analysis.

Demography

Epidemics spread over populations. All the variables in the dataset are weighted to the total population of the countries, but three variables are selected in particular to summarises demographic characteristics of the country. These are: the ratio of people aged over 65 at 2020, the index of urbanisation and the density of population.

Table 4 Epidemic phases

Phase	Start	End	What happened
1st	30/12/2020	30/03/2020	WHO reported evidence of transmission from symptomatic, pre-symptomatic, and asymptomatic infected people with COVID-19
2nd	01/04/2020	23/11/2020	The UK authorities reported a variant of SARS-CoV-2 to the WHO
3rd	24/11/2020	01/01/2021	Pfizer/Biontech vaccine was the first to receive emergency use validation from WHO for efficacy against COVID-19
4th	02/01/2021	30/03/2021	End of data collection

Most official data sources report more severe impacts of COVID-19 on the elderly [47], probably due both to an inherent weakness of their immune system and to the coexistence of other chronic diseases. According to [66], the age is a very important predictor of severe COVID-19. The risk of severe outcomes increases sharply by age, even after controlling for other potential confounding factors, including sex and pre-existing disease conditions. Age is also an important confounder in the associations between some underlying conditions and severe COVID-19 outcomes.

The index of urbanisation is the proportion of people living in a urban center over total population of the country [25]. It measures the demographic current phenomenon of population mobility from rural to urban areas. The high concentration of people and activities in urban areas makes vulnerable the populations to the exposure to COVID-19 infection due to the large amount of social networking [76].

In addition, many studies [1, 31] provide evidences on the correlation between density and the spread of the pandemic. The population density, due to economic and social reasons, affects the spread of COVID-19 infection and the incidence of the cases in the territorial areas.

Comorbidity

Comorbidity is the presence of two or more conditions occurring in a patient, either at the same time or successively. Population with multiple co-existing illness conditions are widespread [22]. This awareness has led to a growing interest among practitioners and researchers in assessing the impact of comorbidity on mortality, health-related quality of life, and efficiency of health care systems.

In the present study, population variables observed as potentially associated to patients infected with COVID-19 are selected: diabetes [26], obesity [8, 9], smoking [23], hypertension, and cardiovascular risk [64].

However, data related to above population variables are not collected from medical records of COVID-19 hospitalised patients but they are epidemiological surveillance data [40] on risk factors for the public health due to the onset or the complications of diseases.

Passengers Mobility

According to [12], mobility data are often used to correlate population mobility and the spread of an infection.

Based on the current literature, in countries where mobility is high, the number of people infected with COVID-19 is higher [11]. Likely there is a positive association between a high airport mobility in a country and a high spread of COVID-19 infection and, as a consequence, the number of both ICU and other hospital admissions increases [60].

In addition to the number of airport passengers, to capture mobility between countries and within national borders, also the number of train passengers are taken into account due both lacking of airports in some territorial areas and regular use of train for mobility.

Air Pollution

According to [19, 66], there are differences in the association between the spread of COVID-19 and the concentration levels of particulate matter (PM10 and PM2.5), and CO2. Long-term previous exposure to air pollution could be an important mediator of deaths by COVID-19 in Europe. In addition, the latest estimates made by the European Environment Agency (EEA) [24] show that the exposure to particulate matter has a strong impact on health [57, 63].

Air pollution has been postulated to affect the viability and transport of viral particles in the air. Long-term exposure could increase the risk of infection by altering the immune system [44]. Particulate matter (PM) is able to deeply enter into the respiratory tract and increase the risk of respiratory diseases [17]. Exposure to PM2.5 is positively associated with COVID-19 infection and with severity of the disease [15, 75].

Furthermore, for the purposes of the present analysis, the choice to *vibrate* the y helps to simulate the sensitivity of the linear regressive model specified on observational data to a non linear epidemic phenomenon as COVID-19 pandemic [43].

In order to define the timings of observation of COVID-19 pandemic, in Table 4 are shown 4 time phases:

Data are collected in 20 European countries: Austria, Belgium, Bulgaria, Croatia, Cyprus, Czechia, Denmark, Estonia, Finland, France, Germany, Ireland, Italy, Luxembourg, Netherlands, Portugal, Slovenia, Spain, Sweden, United Kingdom.

All of countries provide summary statistics of the variables through to National Health Departments and Eurostat.⁶

⁶ UK provides demographic data to Eurostat being a EU member until 2021. Other countries (e.g., Poland) are excluded from the dataset due to missing values.

Table 5 Correlation coefficients among variables

	Cases	ICU	Hosp
Cases	1	-0.4	-0.4
ICU	-0.4	1	0.91
Hosp	-0.4	0.91	1
UrbanPop	0.03	0.11	0.07
Over65	-0.09	0.01	0.06
PopDensity	0.12	0.10	0.05
Cardio	-0.03	-0.01	0.11
Diabetes	-0.02	-0.07	-0.12
Smoking	0.03	0.01	0.07
Obesity	-0.03	-0.11	-0.14
HiPressure	-0.04	-0.06	-0.01
PM2.5	0.15	-0.07	-0.06
PM10	0.14	-0.07	-0.06
CO2	0.10	-0.03	-0.04
AirPass	-0.09	-0.03	-0.09
TrainPass	0.01	-0.00	-0.01

4 Methods: Multiverse Models

The values of correlation coefficients (Table 5) show that in the dataset the cases of COVID-19 are unaligned with the two variables of hospital data. Only *ICU* and *Hosp* are strongly correlated (0.91), since *ICU* is obviously a proportion of *Hosp*, hence the high correlation.

All the values of Bravais-Pearson’s correlation coefficients between the 3 *y* and the other 13 covariates are much weaker than the correlations among the 3 *y*. This is the reason for *vibrating* the regressive models into multiverses in order to explore how these fixed observations are congruent with causal relationships $x \rightarrow y$ [43].

To generate the multiverse, the covariates assume the role of *x* or ‘first regressor’ in the linear model:

$$y \sim \beta_0 + \beta_x(x) + B_{K_x}(K_n \subseteq Z_x) + \epsilon \tag{5}$$

where Z_x is the set of all possible additions of all the *other non-x*, *z*-covariates as control, e.g., $Z_x : (K_1 = \emptyset, K_2 = z_1, K_3 = z_1 + z_2, K_4 = z_1 + z_2 + z_3, \dots, K_n = z_2, K_n = z_2 + z_3, \dots, K_{max(n)-1} = z_{11} + z_{12}, K_{max(n)} = z_{12})$ and B_{K_x} is the vector of coefficients associated to K_n .

$\hat{\beta}_x$ and $p(x)$ are recorded for each vibration. However, while all the non-*y* covariates are members of each Z_x , not all the covariates assume the role of ‘first regressor’ *x*. Indeed, the variables of the Demographic dimension (*UrbanPop*, *Over65*, and

PopDensity) are excluded from this role and assume only the role of z , hence their p -values and β_x are not recorded in the result.

The reason for this exclusion is twofold: the first is the fact that variables as age, density and urban density must have an impact in epidemics does not generate any kind of scientific controversy, or at least it does not ask the same scientific questions as variables as the rate of obesity or railway mobility; the second reason of exclusion is that other variables have space of intervention for a specific, clear-cut public policy (for example, an obesity reduction program), while the claim of a causal relationship between old age and an epidemic would still require other circumstantial assessments of feasibility that cannot be captured by a linear model.

No other functional form nor different scale of measurement for the variables are modeled for the multiverse with the only exception of PM10 and PM2.5, which can be considered mutual proxies since their value of Pearson’s correlation $r \sim 0.99$. In this sense the adopted methodology is more akin to EBA and VoE than to Multiverse Analysis or Specification Curve. However, these differences regard more the goal of the research than the technical procedure.

4.1 Measuring the Outcome of a Shift in α

In the Sect. 2 is mentioned the proposal to shift the conventional α from 0.05 to 0.005. The multiverse offers an opportunity to evaluate the consequence of this change. The main expected result by the shift is to reduce ambiguity in causal interpretation of estimates of a regressive model. However, this result comes at the cost of a higher degree of false negatives.

The goodness of α at making much harder to reach ambiguous results (see, Sect. 2) can be measured by the pseudo-Precision in (3). To ponder this metric to the sensitivity to false negatives, pseudo-Recall (4) is measured too and the two measures are compounded into a indicator J_α as harmonic mean of the two ones (6):

$$J_\alpha = \frac{2}{\frac{A+B}{|A-B|} + \frac{|(A+C)-(B+D)|}{|A-B|}} = \frac{2 A - B|}{A + B + |A - B + C - D|} \tag{6}$$

which is an analogue of the F_1 Score. This metric has been criticised by Chicco and Jurman [13], so the Phi correlation (ϕ_α) and tetrachoric correlation ($r_{tet,\alpha}$) on the Janus Confusion Matrix (Table 2) are provided as alternative metrics [39, 86].

Finally, to estimate the impact of the shift from $\alpha = 0.05$ to $\alpha = 0.005$, the net differences in J , ϕ , and r_{tet} are measured:

$$\begin{aligned} \Delta(J) &= J_{\alpha=0.005} - J_{\alpha=0.05} \\ \Delta(\phi) &= \phi_{\alpha=0.005} - \phi_{\alpha=0.05} \\ \Delta(r_{tet}) &= r_{tet,\alpha=0.005} - r_{tet,\alpha=0.05} \end{aligned} \tag{7}$$

5 Results

For each y , ten linear models with a different first regressor x have been vibrated through all the possible combinations of 12 controls (plus the ‘no control’ case). The observations regarding y are expanded by splitting the counts among 4 time phases (see, Table 4).

This generates 409,600 specifications for each y split in 10 groups of 40,960 specifications for each first regressor, multiplying *per* 3 y , the total is of 30 groups. Multiverse statistics summarise information regarding these 30 groups. The total amount of observed specifications in the study is $409,600 * 3 = 1,228,800$.

The p -curves of the 10 multiversal linear models are plotted as density curves in Figs. 1, 2, and 3.

In Fig. 1 is shown an effect between PM2.5, PM10, Air Passenger, and Cases.

The Figs. 2 and 3 illustrate that Cardiovascular risk and Smoking are very important predictors for ICU cases and Hosp cases. However, for ICU cases, Air Passengers show a significant relationship similar to Fig. 1. Instead, Obesity plays an important role in *Hosp*.

Multivariate statistics for the multiverse $x \rightarrow Cases$ are displayed in Table 6, for $x \rightarrow ICU$ in Table 7, and for $x \rightarrow Hosp$ in Table 8.

For models on COVID-19 cases (Table 6), lowering α leads into an average negative impact in the indicators of evaluation, with the exception of PM10. However, looking at the VoE plot of PM10 and PM2.5 (Fig. 4), this result is mostly determined by imbalance among classes in the Janus Confusion Matrix of PM10. Indeed, once results are paired with those of the proxy of PM2.5, significant results for $\alpha < 0.005$ are paradoxically opposite. Given also the small effect size of the estimates, these are spurious occurrences determined by overfitting of the specifications.

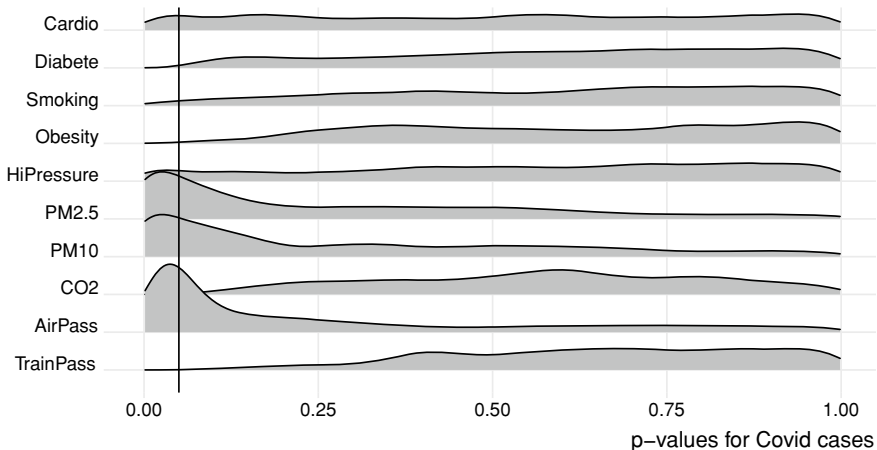


Fig. 1 Density of p -values in the multiverse $x \rightarrow Cases$

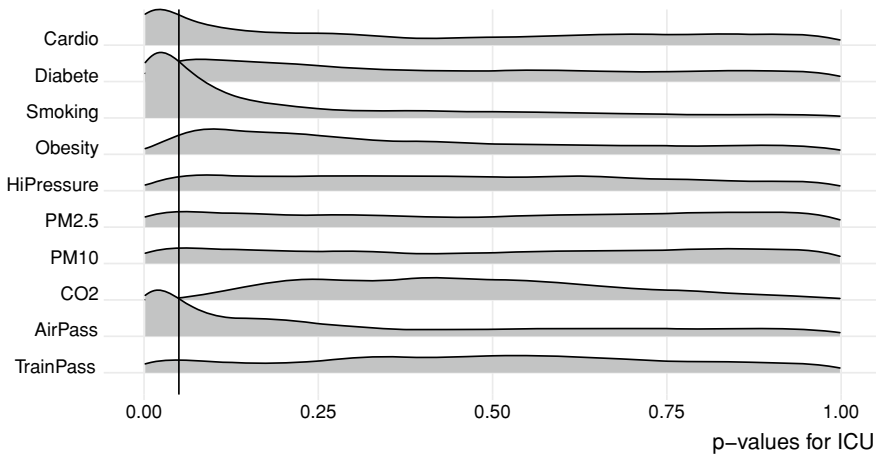


Fig. 2 Density of p -values in the multiverse $x \rightarrow ICU$

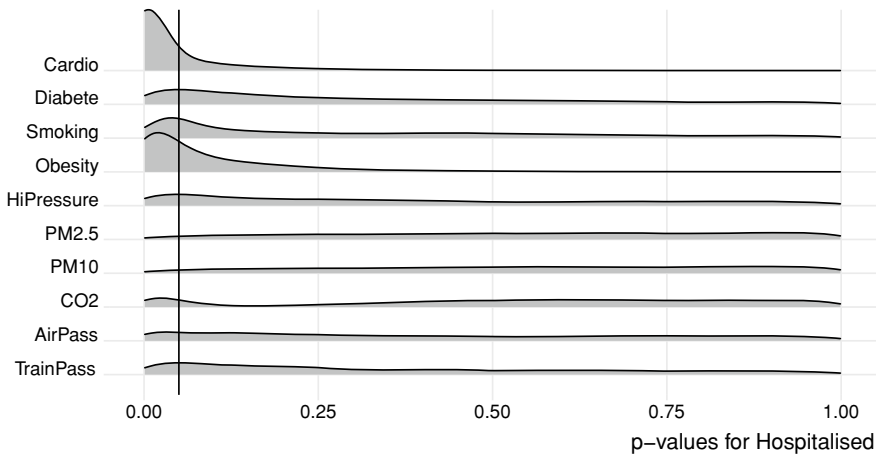


Fig. 3 Density of p -values in the multiverse $x \rightarrow Hosp$

In Fig. 4 is noteworthy the presence of a mirror Janus effect around β_0 . In this case this result is reached by controlling x through its proxy, which is an easy avoidable statistical fallacy.

For models $ICU \sim x$ (Table 7), the analysis confirms that lowering the α is not helpful to highlight good scientific findings and in many cases it could be detrimental.

For example, one can consider the VoE for $ICU \sim Diabetes$, as shown in Fig. 5.

In Fig. 5 there is a clear-cut case of Janus effect that can be common in observational studies. Since *Diabetes*, unlike PM, is relatively robust to ‘bad controls’, the estimate span is much more narrow even if all the variables are standardised.

Table 6 Summary statistics for the multiverse $Cases \sim x$

x	r_ρ	$\%(p < 0.05)$	Med. (p)	$\bar{\beta}$	$\Delta(J)$	$\Delta(\phi)$	$\Delta(r_{ret})$
Cardio	-0.03	0.06	0.52	-0.06	-0.14	-0.11	-0.52
Diabetes	-0.02	0.00	0.63	0.06	-0.01	-0.05	-0.41
Smoking	0.03	0.01	0.63	0.06	-0.04	-0.06	-0.29
Obesity	-0.03	0.00	0.61	-0.05	-0.00	-0.03	-0.60
HiPressure	-0.04	0.05	0.59	-0.01	-0.30	-0.30	-0.17
PM2.5	0.15	0.25	0.19	-1.14	-0.26	-0.10	-0.10
PM10	0.14	0.22	0.21	-0.93	0.33	0.50	0.69
CO2	0.10	0.00	0.58	0.06	-0.00	-0.01	-0.20
AirPass	-0.09	0.31	0.12	-0.24	-0.51	-0.22	-0.54
TrainPass	0.01	0.00	0.67	0.02	0.00	0.00	0.00

Table 7 Summary statistics for the multiverse $ICU \sim x$

x	r_ρ	$\%(p < 0.05)$	Med. (p)	$\bar{\beta}$	$\Delta(J)$	$\Delta(\phi)$	$\Delta(r_{ret})$
Cardio	-0.01	0.21	0.31	0.03	-0.22	-0.02	0.30
Diabetes	-0.07	0.07	0.38	-0.07	-0.66	-0.65	-0.69
Smoking	0.01	0.38	0.09	0.05	-0.39	-0.08	-0.19
Obesity	-0.11	0.04	0.33	-0.12	-0.16	-0.22	-0.87
HiPressure	-0.06	0.04	0.45	-0.04	-0.27	-0.04	0.26
PM2.5	-0.07	0.07	0.52	-0.16	0.36	0.27	0.43
PM10	-0.07	0.07	0.51	-0.33	-0.11	0.02	0.25
CO2	-0.03	0.00	0.44	-0.03	-0.01	-0.05	-0.51
AirPass	-0.03	0.27	0.20	-0.03	-0.31	-0.16	-0.10
TrainPass	-0.00	0.06	0.49	-0.05	-0.16	-0.14	-0.22

Table 8 Summary statistics for the multiverse $Hosp \sim x$

x	r_ρ	$\%p < 0.05$	Med. (p)	$\bar{\beta}$	$\Delta(J)$	$\Delta(\phi)$	$\Delta(r_{ret})$
Cardio	0.11	0.76	0.00	0.16	-0.20	0.00	-0.17
Diabetes	-0.12	0.15	0.23	-0.15	-0.28	-0.14	-0.26
Smoking	0.07	0.21	0.24	0.08	-0.37	-0.14	-0.44
Obesity	-0.14	0.49	0.05	-0.17	-0.43	-0.06	-0.21
HiPressure	-0.01	0.12	0.33	-0.08	-0.28	-0.10	0.20
PM2.5	-0.06	0.03	0.56	0.10	-0.06	0.02	0.58
PM10	-0.06	0.03	0.56	-0.01	-0.13	-0.07	0.36
CO2	-0.04	0.10	0.58	-0.04	-0.26	-0.21	-0.33
AirPass	-0.09	0.10	0.39	-0.07	-0.21	-0.13	-0.16
TrainPass	-0.01	0.12	0.30	-0.08	-0.23	-0.11	-0.30

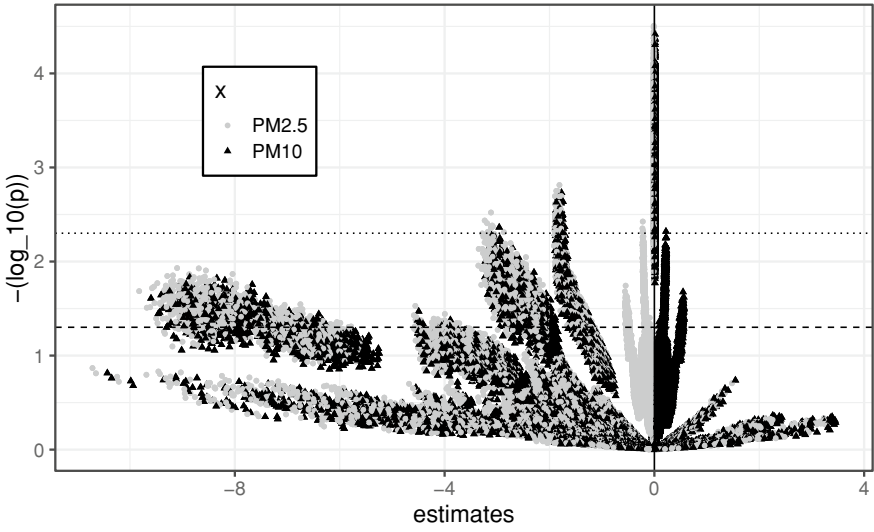


Fig. 4 VoE plot of Covid cases ~ Levels of Particulate Matter

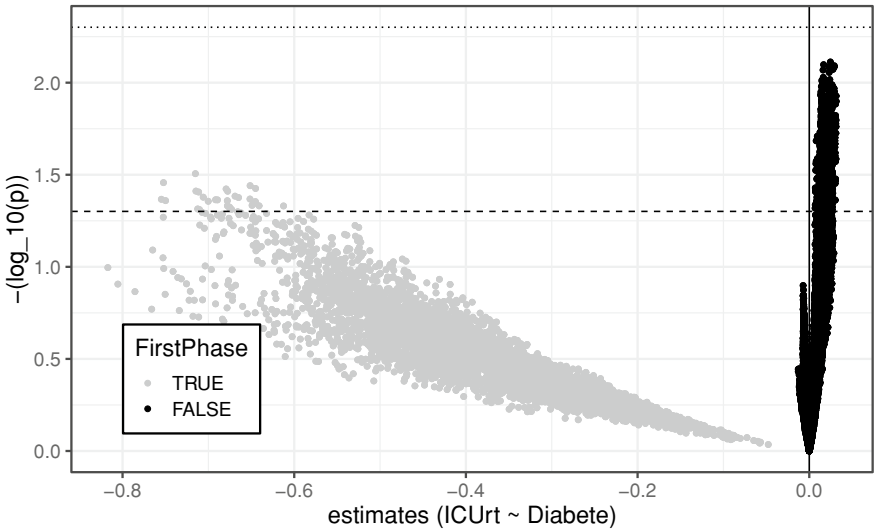


Fig. 5 VoE plot of ICU patients with Covid ~ % of people with diabetes in the country

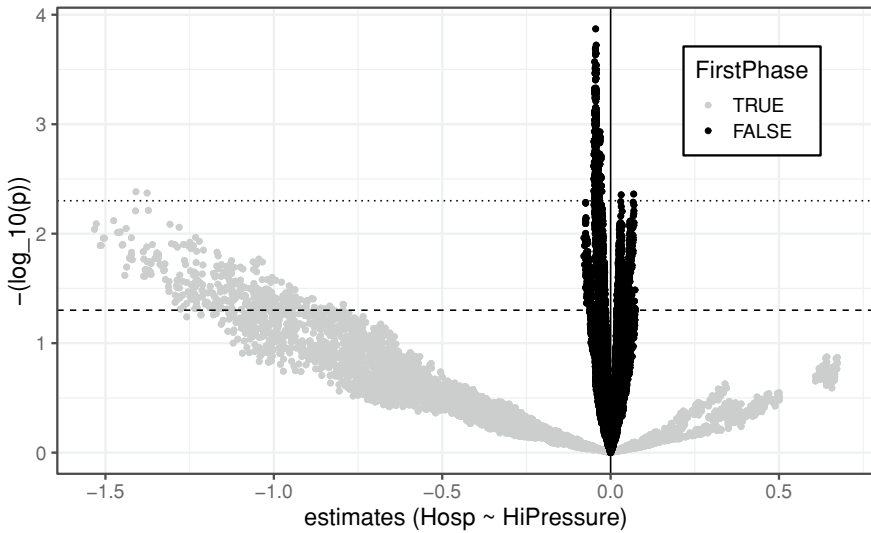


Fig. 6 VoE plot of *Hospitalised* with COVID-19 ~ % of people with high pressure in the country

However, all the estimates of specifications, when run on the data available before April 1st, 2020 (see, Table 4) could lead to think that people affected by diabetes would be, in a way or another, less prone to ICU by COVID-19 (given also a negative linear correlation between the two variables). Indeed, this is only an effect of a hasty analysis that would lead into a scientific controversy:

- $\alpha = 0.05$ is already good ‘gatekeeper’ since only a minority of specifications can be *p*-hacked into a significant result.
- Time reveals that if there is a causal relationship between diabetes and harsh COVID-19, it is extremely weak and positive, even if $\hat{\beta}$ of the multiverse is negative, so it would be easy to *p*-hack.

Since $\alpha = 0.05$ is a good filter while $\alpha = 0.005$ misses the opportunity to actually ‘say something about’ (no specification is significant under such level), the evaluation indicators penalise $\alpha = 0.005$ so much.

For the multiverse of *Hosp*, *HiPressure* is the most ambiguous case (Table 8, Fig. 6). For $\alpha = 0.05$ this is a clear-cut case of Janus effect. Considering the whole dataset of 100 observations, lowering the level of significance to $\alpha = 0.005$ would point towards a unique interpretation of the relationship: it is negative and it is weak. Considering only the first Phase, by $\alpha = 0.05$ while the inference of the sign would have been correct, the size of the negative effect would be overestimated.

The result would have been very hard to *p*-hack under a level of significance $\alpha = 0.005$. Indicators for Janus effect do not detect significant differences in estimates within the same class, they do only the relation between estimates $\hat{\beta}$ and β_0 . However the tetrachoric correlation identifies in this case the potential benefit of adoption of $\alpha = 0.005$.

6 Discussion

In this study 3 indicators of Janus effect have been employed to evaluate the consequence of a shift towards the level of significance $\alpha = 0.005$ in terms of risk of mis-specification of a small-entity causal relationship in observational small samples. The impact has been estimated on a multiverse which is generated by an observational sample with high volatility in the lagged y dependent but fixed x .

$\alpha = 0.005$ does a good job of ‘gatekeeping’ from the possibility of p -hacking a desired outcome but this is a tautology if compared to all of the cases where there is a nearly unambiguous relation (see, Fig. 5) and the relation would be lost as a False Negative setting such a low value of α .

It could still be argued that the impact of False Negatives can be weighted by the low estimates and $\alpha = 0.005$ would not miss stronger effects. If this is the case, then is true that null hypothesis testing provides only a limited contribution to the epidemiological research on observational data and, as a consequence, methodologies focused on effect estimation and not on the statistical significance of the relationship, should be adopted.

The main limitation of the study regards the adoption of J index, ϕ index, and tetracoric correlations as indicators of the overall impact of α on the trustworthiness of claims from observational data. α is there to decrease false positives in scientific studies, so it is consequential that it is penalised by measures that evaluate the rate of false negatives.

The limitation of the Janus Confusion Matrix is that it is insensitive to the actual rate of $+$ and $-$ estimates in the multiverse but only to their absolute divergence, i.e. there could be cases where $+$ is dominant per $p(\beta_x) \geq \alpha$ and $-$ is dominant per $p(\beta_x) < \alpha$.

The above Fig. 5 highlights one of these cases. In these scenarios, the shift of α is actually effective at isolating a True scientific claim but the measures of the study could not capture it. However, occurrences like these are rare and better and more specific analytical choices in generation of the multiverse should totally avoid them [18].⁷

References

1. Almagro, M., Orane-Hutchinson, A.: JUE insight: the determinants of the differential exposure to COVID-19 in New York city and their evolution over time. *J. Urban Econ.* (2020). <https://doi.org/10.1016/j.jue.2020.103293>
2. Anderson, S.F., Maxwell, S.E.: Addressing the “replication crisis”: using original studies to design replication studies with appropriate statistical power. *Multivar. Behav. Res.* **52**(3), 305–324 (2017)

⁷ The paper is also part of research line on vulnerability and risk management of the project *GRIDAVI* Risk Management, Decision Uncertainties and Social Vulnerabilities by the University Research Incentive Plan 2020/2022 called PIACERI.

3. Athey, S., Imbens, G.: A measure of robustness to misspecification. *Am. Econ. Rev.* **105**(5), 476–80 (2015)
4. Begley, C.G., Ellis, L.M.: Raise standards for preclinical cancer research. *Nature* **483**(7391), 531–533 (2012)
5. Benjamin, D.J., et al.: Redefine statistical significance. *Nat. Hum. Behav.* **2**(1), 6–10 (2018)
6. Bossuyt, P.M.: Laboratory measurement’s contribution to the replication and application crisis in clinical research. *Clin. Chem.* **65**(12), 1479–1480 (2019)
7. Bruns, S.B., Ioannidis, J.P.: P-curve and p-hacking in observational research. *PLoS ONE* **11**(2), e0149144 (2016)
8. Busetto, L., et al.: Obesity and COVID-19: an Italian snapshot. *Obesity* **28**(9), 1600–1605 (2020)
9. Caci, G., et al.: COVID-19 and obesity: dangerous liaisons. *J. Clin. Med.* **9**(8), 2511 (2020)
10. Camerer, C.F., et al.: Evaluating the replicability of social science experiments in *Nature* and *Science* between 2010 and 2015. *Nat. Hum. Behav.* **2**(9), 637–644 (2018)
11. Chang, S., et al.: Mobility network models of COVID-19 explain inequities and inform reopening. *Nature* **589**(7840), 82–87 (2021)
12. Charaudeau, S., et al.: Commuter mobility and the spread of infectious diseases: application to influenza in France. *PLoS ONE* **9**(1), e83002 (2014)
13. Chicco, D., Jurman, G.: The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genom.* **21**(1), 1–13 (2020)
14. Chu, L., et al.: Vibration of effects in epidemiologic studies of alcohol consumption and breast cancer risk. *Int. J. Epidemiol.* **49**(2), 608–618 (2020)
15. Conticini, E., et al.: Can atmospheric pollution be considered a co-factor in extremely high level of SARS-CoV-2 lethality in Northern Italy? *Environ. Pollut.* **261**,(2020). <https://doi.org/10.1016/j.envpol.2020.114465>
16. Cosme, D., et al.: Multivariate neural signatures for health neuroscience: assessing spontaneous regulation during food choice. *Soc. Cogn. Affect. Neurosci.* **15**(10), 1120–1134 (2020)
17. Cui, Y., et al.: Air pollution and case fatality of SARS in the People’s Republic of China: an ecologic study. *Environ. Health* **2**(1), 1–5 (2003)
18. Del Giudice, M., Gangestad, S.W.: A traveler’s guide to the multiverse: promises, pitfalls, and a framework for the evaluation of analytic decisions. *Adv. Methods Pract. Psychol. Sci.* **4**(1) (2021). <https://doi.org/10.1177/2515245920954925>
19. Donzelli, G., et al.: Relations between air quality and COVID-19 lockdown measures in Valencia, Spain. *Int. J. Environ. Res. Public Health* **18**(5), 2296 (2021)
20. Duvendack, M., et al.: What is meant by “replication” and why does it encounter resistance in economics? *Am. Econ. Rev.* **107**(5), 46–51 (2017)
21. Earp, B.D., Trafimow, D.: Replication, falsification, and the crisis of confidence in social psychology. *Front. Psychol.* (2015). <https://doi.org/10.3389/fpsyg.2015.00621>
22. Ebinger, J.E., et al.: Pre-existing traits associated with Covid-19 illness severity. *PLoS ONE* **15**(7), e0236240 (2020)
23. Espejo-Paeres, C., Núñez-Gil, I.J., Estrada, V., et al.: Impact of smoking on COVID-19 outcomes: a HOPE Registry subanalysis. *BMJ Nutr. Prev. Health* **4**, (2021). <https://doi.org/10.1136/bmjnp-2021-000269>
24. European Environmental Agency: Air Quality in Europe—2020 report, 9/2020, EEA Report (2020)
25. Eurostat: Eurostat Regional Yearbook, Edition 2020 (2020)
26. Fadini, G.P., et al.: Newly-diagnosed diabetes and admission hyperglycemia predict COVID-19 severity by aggravating respiratory deterioration. *Diabetes Res. Clin. Pract.* **168** (2020). <https://doi.org/10.1016/j.diabres.2020.108374>
27. Gauchat, G.: Politicization of science in the public sphere: a study of public trust in the United States, 1974 to 2010. *Am. Sociol. Rev.* **77**(2), 167–187 (2012)
28. Gelman, A., Loken, E.: The statistical crisis in science. Data-dependent analysis—a “garden of forking paths”—explains why many statistically significant comparisons don’t hold up. *Am. Sci.* **102**(6), 460–466 (2014)

29. Granger, C.W., Uhlig, H.F.: Reasonable extreme-bounds analysis. *J. Econometrics* **44**(1–2), 159–170 (1990)
30. Halsey, L.G., et al.: The fickle P value generates irreproducible results. *Nat. Methods* **12**(3), 179–185 (2015)
31. Hamidi, S., et al.: Does density aggravate the COVID-19 pandemic? Early findings and lessons for planners. *J. Am. Plann. Assoc.* **86**(4), 495–509 (2020)
32. Harder, J.A.: The multiverse of methods: extending the multiverse analysis to address data-collection decisions. *Perspect. Psychol. Sci.* **15**(5), 1158–1177 (2020)
33. Head, M.L., et al.: The extent and consequences of p-hacking in science. *PLoS Biol.* **13**(3) (2015). <https://doi.org/10.1371/journal.pbio.1002106>
34. Hicks, D.J.: Open science, the replication crisis, and environmental public health. *Accountability Res.* 1–29 (2021). <https://doi.org/10.1080/08989621.2021.1962713>
35. Imbens, G.W.: Statistical significance, p-values, and the reporting of uncertainty. *J. Econ. Perspect.* **35**(3), 157–74 (2021)
36. Ioannidis, J.P., et al.: Increasing value and reducing waste in research design, conduct, and analysis. *The Lancet* **383**(9912), 166–175 (2014)
37. Ioannidis, J.P.: The proposal to lower P value thresholds to 0.005. *JAMA* **319**(14), 1429–1430 (2018)
38. Ioannidis, J.P.: What have we (not) learnt from millions of scientific papers with P values? *Am. Stat.* **73**(sup1), 20–25 (2019)
39. Islam, T.U., Rizwan, M.: Comparison of correlation measures for nominal data. *Commun. Stat. Simul. Comput.* 1–20 (2020). <https://doi.org/10.1080/03610918.2020.1869984>
40. ISS EpiCentro: I dati per l'Italia. La Sorveglianza Passi d'Argento (2020). <https://www.epicentro.iss.it/passi-argento>
41. Jewell, N.P., et al.: Predictive mathematical models of the COVID-19 pandemic: underlying principles and value of projections. *JAMA* **323**(19), 1893–1894 (2020)
42. Johnson, V.E.: Revised standards for statistical evidence. *Proc. Natl. Acad. Sci.* **110**(48), 19313–19317 (2013)
43. Klau, S., et al.: Sampling uncertainty versus method uncertainty: a general framework with applications to omics biomarker selection. *Biometrical J.* **62**(3), 670–687 (2020)
44. Kogevinas, M., et al.: Ambient air pollution in relation to SARS-CoV-2 infection, antibody response, and COVID-19 disease: a cohort study in Catalonia, Spain (COVICAT study). *Environ. Health Perspect.* **129**(11) (2021). <https://doi.org/10.1289/EHP9726>
45. Kreps, S.E., Kriner, D.L.: Model uncertainty, political contestation, and public trust in science: evidence from the COVID-19 pandemic. *Sci. Adv.* **6**(43) (2020). <https://doi.org/10.1126/sciadv.abd4563>
46. Leamer, E.E.: Sensitivity analyses would help. *Am. Econ. Rev.* **75**(3), 308–313 (1985)
47. Lee, Y.J.: The impact of the COVID-19 pandemic on vulnerable older adults in the United States. *J. Gerontol. Soc. Work* **63**(6–7), 559–564 (2020)
48. Leek, J.T., Peng, R.D.: Statistics: P values are just the tip of the iceberg. *Nat. News* **520**(7549), 612 (2015)
49. Mao, Z., et al.: Investigating the self-reported health status of domestic and overseas Chinese populations during the COVID-19 pandemic. *Int. J. Environ. Res. Public Health* **18**(6), 3043 (2021)
50. Masur, P.K.: Understanding the effects of conceptual and analytical choices on ‘finding’ the privacy paradox: a specification curve analysis of large-scale survey data. *Inf. Commun. Soc.* 1–19 (2021). <https://doi.org/10.1080/1369118X.2021.1963460>
51. Masur, P.K., Scharkow, M.: *specr: Conducting and Visualizing Specification Curve Analyses*. R Package (2020)
52. Maxwell, S.E., et al.: Is psychology suffering from a replication crisis? What does “failure to replicate” really mean? *Am. Psychol.* **70**(6), 487–498 (2015)
53. Mayo, D.G., Spanos, A.: Methodology in practice: statistical misspecification testing. *Philos. Sci.* **71**(5), 1007–1025 (2004)

54. Mazzola, J.J., Deuling, J.K.: Forgetting what we learned as graduate students: HARKing and selective outcome reporting in I-O journal articles. *Ind. Organ. Psychol.* **6**(3), 279–284 (2013)
55. McShane, B.B., et al.: Abandon statistical significance. *Am. Stat.* **73**(sup1), 235–245 (2019)
56. Muñoz, J., Young, C.: We ran 9 billion regressions: eliminating false positives through computational model robustness. *Sociol. Methodol.* **48**(1), 1–33 (2018)
57. Nižetić, S.: Impact of coronavirus (COVID-19) pandemic on air transport mobility, energy, and environment: a case study. *Int. J. Energy Res.* **44**(13), 10953–10961 (2020)
58. Open Science Collaboration: Estimating the reproducibility of psychological science. *Science* **349**(6251) (2015). <https://doi.org/10.1126/science.aac4716>
59. Orben, A., Przybylski, A.K.: The association between adolescent well-being and digital technology use. *Nat. Hum. Behav.* **3**(2), 173–182 (2019)
60. Oztig, L.I., Askin, O.E.: Human mobility and coronavirus disease 2019 (COVID-19): a negative binomial regression analysis. *Public Health* **185**, 364–367 (2020). <https://doi.org/10.1016/j.puhe.2020.07.002>
61. Page, L., et al.: The replication crisis, the rise of new research practices and what it means for experimental economics. *J. Econ. Sci. Assoc.* 1–16 (2021). <https://doi.org/10.1007/s40881-021-00107-7>
62. Palpacuer, C., et al.: Vibration of effects from diverse inclusion/exclusion criteria and analytical choices: 9216 different ways to perform an indirect comparison meta-analysis. *BMC Med.* **17**(1), 1–13 (2019)
63. Pansini, R., Fornacca, D.: COVID-19 higher mortality in Chinese Regions with chronic exposure to lower air quality. *Front. Public Health* **8**, (2021). <https://doi.org/10.3389/fpubh.2020.597753>
64. Parohan, M., et al.: Risk factors for mortality in patients with Coronavirus disease 2019 (COVID-19) infection: a systematic review and meta-analysis of observational studies. *The Aging Male* **23**(5), 1416–1424 (2020)
65. Patel, C.J., et al.: Assessment of vibration of effects due to model specification can demonstrate the instability of observational associations. *J. Clin. Epidemiol.* **68**(9), 1046–1058 (2015)
66. Perone, G.: The determinants of COVID-19 case fatality rate (CFR) in the Italian regions and provinces: an analysis of environmental, demographic, and healthcare factors. *Sci. Total Environ.* **760**, (2021). <https://doi.org/10.1016/j.scitotenv.2020.142523>
67. Pike, H.: Statistical significance should be abandoned, say scientists. *BMJ* **364**, (2019). <https://doi.org/10.1136/bmj.11374>
68. Pluchino, A., et al.: A novel methodology for epidemic risk assessment of COVID-19 outbreak. *Sci. Rep.* **11**(1), 1–20 (2021)
69. Qiu, F., et al.: Impacts of cigarette smoking on immune responsiveness: up and down or upside down? *Oncotarget* **8**(1), 268–284 (2017)
70. Rohrer, J.M., et al.: Probing birth-order effects on narrow traits using specification-curve analysis. *Psychol. Sci.* **28**(12), 1821–1832 (2017)
71. Sala-i-Martin, X.: I just ran four million regressions. *Am. Econ. Rev.* **87**(2), 178–183 (1997)
72. Saraceno, J., et al.: Reevaluating the substantive representation of lesbian, gay, and bisexual Americans: a multiverse analysis. *J. Politics* **83**(4), 1837–1843 (2021)
73. Schmeiser, H., et al.: The risk of model misspecification and its impact on solvency measurement in the insurance sector. *J. Risk Finance* **13**(4), 285–308 (2012)
74. Seitshiro, M.B., Mashele, H.P.: Quantification of model risk that is caused by model misspecification. *J. Appl. Stat.* 1–21 (2020). <https://doi.org/10.1080/02664763.2020.1849055>
75. Setti, L., et al.: Potential role of particulate matter in the spreading of COVID-19 in Northern Italy: first observational study based on initial epidemic diffusion. *BMJ Open* **10**(9), e039338 (2020)
76. Sharifi, A., Khavarian-Garmsir, A.R.: The COVID-19 pandemic: impacts on cities and major lessons for urban planning, design, and management. *Sci. Total Environ.* (2020). <https://doi.org/10.1016/j.scitotenv.2020.142391>
77. Simmons, J.P., et al.: False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychol. Sci.* **22**(11), 1359–1366 (2011)

78. Simonsohn, U., et al.: Specification curve analysis. *Nat. Hum. Behav.* **4**(11), 1208–1214 (2020)
79. Simonsohn, U., et al.: Specification curve: Descriptive and inferential statistics on all reasonable specifications. Available at:SSRN (2019). <https://doi.org/10.2139/ssrn.2694998>
80. Sönning, L., Werner, V.: The replication crisis, scientific revolutions, and linguistics. *Linguistics* **59**(5), 1179–1206 (2021)
81. Steegen, S., et al.: Increasing transparency through a multiverse analysis. *Perspect. Psychol. Sci.* **11**(5), 702–712 (2016)
82. Taroni, F., et al.: Statistical hypothesis testing and common misinterpretations: should we abandon p-value in forensic science applications? *Forensic Sci. Int.* **259**, e32–e36 (2016)
83. Trafimow, D.: Five nonobvious changes in editorial practice for editors and reviewers to consider when evaluating submissions in a post $p < 0.05$ universe. *Am. Stat.* **73**(sup1), 340–345 (2019)
84. Trafimow, D., et al.: Manipulating the alpha level cannot cure significance testing. *Front. Psychol.* **9**, (2018). <https://doi.org/10.3389/fpsyg.2018.00699>
85. Vanpaemel, W., et al.: Are we wasting a good crisis? The availability of psychological research data after the storm. *Collabra* **1**(1), 1–5 (2015)
86. Warrens, M.J.: On association coefficients for 2×2 tables and properties that do not depend on the marginal distributions. *Psychometrika* **73**(4), 777 (2008)
87. Wasserstein, R.L., Lazar, N.A.: The ASA statement on p-values: context, process, and purpose. *Am. Stat.* **70**(2), 129–133 (2016)
88. Wasserstein, R.L. et al.: Moving to a world beyond “ $p < 0.05$ ”. *Am. Stat.* **73**(sup. 1), 1–19 (2019)
89. Wei, E.K., et al.: Nine lessons learned from the COVID-19 pandemic for improving hospital care and health care delivery. *JAMA Intern. Med.* **181**(9), 1161–1163 (2021)
90. World Health Organization: Coronavirus Disease 2019 (COVID-19) (2021). Available at: <https://apps.who.int/iris/handle/10665/331475>
91. Yaffe, J.: From the editor—do we have a replication crisis in social work research? *J. Soc. Work Educ.* **55**(1), 1–4 (2019)
92. Young, C., Holsteen, K.: Model uncertainty and robustness: a computational framework for multimodel analysis. *Sociol. Methods Res.* **46**(1), 3–40 (2017)