Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Increased risk of COVID-19 from walking dogs? Most likely, a spurious finding

ARTICLE INFO

Keywords
Multiple hypotheses testing
P-value
False discovery rate (FDR)
Health geography
SARS-CoV-2

ABSTRACT

We analyse the paper “The spread of SARS-CoV-2 in Spain: Hygiene habits, sociodemographic profile, mobility patterns and comorbidities” authored by Rodríguez-Barranco et al. (2021), published in Environmental Research, vol.192, January 2021. The study was carried out under challenging conditions and provides original data of great value for exploratory purposes. Nevertheless, we found that the authors have not considered the potential effect of the multiple hypothesis testing carried out until obtaining the final model on the increased occurrence of false discoveries by mere chance. After adjusting the results provided in the paper for the effects of multiple testing, we conclude that only one of the five factors cited as statistically significant and relevant in the article, living with someone who has suffered from COVID-19, remained significantly related to the relative prevalence of COVID-19. Therefore, the prominent role given in the analysed work to walking the dog as one of the main transmission routes of COVID-19 probably does not correspond to an actual effect. Instead, until replicated by other studies, it should be considered a spurious discovery.

1. Introduction

We have read with great interest the paper recently published in Environmental Research, entitled “The spread of SARS-CoV-2 in Spain: Hygiene habits, sociodemographic profile, mobility patterns and comorbidities” (Rodríguez-Barranco et al., 2021).

This article attracted much attention when it was first published online because no one imagined that walking the dog - one of the few outdoor activities that had been allowed during the harshest phase of confinement - could carry so much risk.

Above all, this is an important study carried out under challenging conditions and provides original data of great value. Nevertheless, we think that some of their achievements may have been oversized, as some critical limitations of the study were not explicitly recognised.

1.1. Case study summary

The general aim of the study (Rodríguez-Barranco et al., 2021) was to identify the significant risk factors for contracting COVID-19 disease. The survey was based on 41 questions after a reworking of the 53 initially included in the questionnaire delivered as supplementary info. They tested differences in the estimated prevalence of COVID-19 between categories of the variables studied and found that six out of the 41 used single predictors were significantly related to differences in COVID-19 prevalence. Namely, living with a household member who has suffered COVID-19, smoking, disinfecting purchased products upon arrival home, using public transportation, walking with pets, and working on-site at the workplace during confinement.

Additionally, the authors fitted a multivariate logistic regression model, using backwards variable selection on a selected subset of predictors, including those best correlated with the risk of COVID-19. Other predictors having a p = 0.20 or lower in the bivariate test were also included. In the following step, predictors having a p-value above 0.1 in the Wald test were discarded. The final fitted model retained five predictors (see Table 5 from Rodríguez-Barranco et al., 2021). Namely, walking the pet, living with a COVID-19 patient, applying a disinfectant product on the products purchased from the market upon arrival home, using home delivery service, and working on-site at the workplace (or travelling to the workplace).

In the discussion section, the authors highlight a significant relationship between the estimated prevalence of the COVID-19 and having walked a dog during the period of confinement. Specifically, walking the pet increases the risk of contracting COVID-19 by 78%, according to the odds ratio (1.78 in Table 5).

They concluded that “These results point to living with dogs as a potent risk factor for COVID-19 infection” and “The results of this study demonstrate that living with dogs, working on-site, purchasing essential commodities by using home delivery service, and especially, living with a COVID-19 patient, have been the main routes of transmission of SARS-CoV-2 during the most restrictive period of confinement in Spain”.

1.2. Methodological concerns

We have some general concerns about the work mentioned above and, particularly, about its most outstanding discovery, which gives walking the dog as one of the main transmission routes of SARS-CoV-2 during confinement in Spain.

As a starting hypothesis, we consider that the significant role of walking the dog in COVID-19 contagions found in Rodríguez-Barranco et al. (2021) is a spurious finding derived from methodological flaws.

2. Methods

We first analysed Rodríguez-Barranco et al. (2021) to address the so-called “multiple testing problem” to test our starting hypothesis. This is a relevant issue both in environmental and public health...
research (García, 2003; Patel and Ioannidis, 2014; Gutiérrez-Hernández and García, 2021). It is particularly critical in exploratory studies in which many statistical tests are performed on many variables measured in the same data set (Goodman, 1998; Young et al., 2009). When multiple testing results are evaluated individually (i.e. on a per-test basis), the chance of false positives increases with the number of tests performed. In a table of 45 simultaneous tests, the probability of obtaining at least one spurious result rises to near 90%, and the expected number of false positives is two.

To account for the multiplicity problem, we controlled the False Discovery Rate (FDR) Benjamini and Hochberg (1995), that is, the expected proportion of false discoveries among all rejected null hypotheses (see García, 2004 for further details and algorithms). It is a powerful statistical technique that allows the researcher to control the increased chance of false discoveries derived from multiple testing while avoiding the large power losses associated with the Bonferroni-like procedures. Its use has been strongly recommended to handle the multiple testing problem in epidemiology and public health studies (Glickman et al., 2014).

Additionally, we analysed other methodological issues mainly related to data collection and analysis and interpreting results, which also influenced the conclusions reached by the authors in the studied paper.

3. Results

3.1. Methodological concerns related to results generation

There are some concerns regarding how the research was conceived and planned and how the data were gathered.

3.1.1. Objectives of the study

According to Rodríguez-Barranco et al. (2021), the final aim was “to clarify the main causes of transmission of the COVID-19 disease during confinement and some risk factors, in order to detect the critical points of exposure to the virus and thus reduce its spread”.

Since no explicit working hypothesis was stated, it can be assumed that, in practice, the study’s aim was searching for one or more predictors, among those derived from the survey carried out, that related to the estimated COVID-19 prevalence in a statistically significant way.

3.1.2. Data gathering and variable building

A second concern is related to how the data were collected and codified. Although it must be pointed out that it would not be reasonable to demand entirely usual standards in the challenging conditions in which the work was carried out, some potentially relevant issues are summarised in Supplementary Table 1.

3.2. Methodological concerns related to data analysis

3.2.1. Sequential model fitting and multiple testing

Rodríguez-Barranco et al. (2021) first carried out a bivariate association study, which made it possible to select a subset of 19 predictors most closely related to the estimated prevalence of the COVID-19 disease. In turn, these predictors were used to fit a logistic regression model, whose coefficients were interpreted in terms of adjusted odds ratio (OR). This model, summarised in Table 5 from Rodríguez-Barranco et al. (2021), allows for a direct quantitative interpretation and discussion of the adjusted relative influence of each predictor in the outcome variable. In addition, the results of the multivariate analysis inform on the statistical significance of each contribution, based on the value of its coefficient (estimated OR), the width of the confidence interval (precision), and the p-value (statistical significance) associated with each predictor coefficient.

It seems clear that, in the analysed paper, both the bivariate and the multivariate analysis contribute sequentially to achieve the final optimal explanatory model. In the first step, the 43-test bivariate analysis contributed to selecting the predictors to reach the modelling phase. According to the p-values obtained in the bivariate analysis, 24 non-promising predictors (p > 0.20) were discarded. The remaining 19 predictors were used in the modelling phase. According to the modelling described in the text (Section 2.4), one could think that the whole 19 selected predictors were first fitted in a full model (i.e., 19 tested coefficients) and then discarded according to a fixed threshold (p > 0.10). Nevertheless, Table 5 from Rodríguez-Barranco et al. (2021) shows that a “backward variable selection procedure” was used to fit the multivariate risk logistic regression model. Since the departure set had 19 predictors, a full backwards procedure would involve performing over a hundred statistical tests, which would significantly increase the potential number of false positives over the one-step complete model-discarding option. We will consider the latter option, which involves a minimum of 19 statistical tests.

3.2.2. Discard of untested predictors

One first concern is related to the pre-select predictors’ method by looking at the p-values of the bivariate analysis and a fixed p-value threshold. Probably, an in-depth analysis of the whole predictor set, looking at the redundancies and the meaning of each group of predictors, could have led to a more convincing preselection of predictors. This raises doubts about whether a meaningful predictor could have been excluded from the subsequent modelling process due to applying a criterion not well-justified in the preselection step. For example, smoking appeared as the second-best related variable to COVID-19 estimated prevalence (p = 0.003), but disappeared with no explanation in the modelling phase (see supplementary Table 1).

3.2.3. The multiple testing problem

A primary concern in the whole data analysis process is that the authors have not even considered the possibility that some significant relationship between the selected predictors and the increased risk of contracting COVID-19 could be a false positive derived from the substantial multiple testing involved in the whole process of model building.

To test a potential spurious relationship between the COVID-19 estimated prevalence and walking the dog, we applied the above-cited Benjamini and Hochberg (1995) procedure by calculating corrected significance thresholds to control the false discovery rate at the 5% level. Any uncorrected p-value equal to or less than the FDR-corrected threshold is considered significant after controlling for the increased probability of spurious finding associated with multiple testing.

The 43 p-values resulting from the bivariate exploratory analysis (Tables 1–4 from Rodríguez-Barranco et al., 2021) were directly analysed. In multivariate analysis, there is some additional “hidden multiplicity” derived from predictor selection (Forstmeier and Schielzeth, 2011; Young and Karr, 2011; Tsujiya, 2014). Since, in the bivariate results shown in the mentioned tables, 19 predictors have a p-value lower than 0.20 (preselection criterion), we assume that at least 19 tests were performed before discarding predictors and getting the definitive 5-predictor model (Table 5 from Rodríguez-Barranco et al., 2021). To evaluate the effects of this “hidden multiplicity”, we used the approach for partially unknown p-values implemented in the function p.adjust() (R Core Team, 2020).

The calculated threshold values for controlling the FDR at the 5% level in the bivariate (n = 43) and multivariate (n = 19) analyses were 0.001 and 0.00049 (a proxy to ensure the declared p < 0.001, even after rounding up), respectively. Comparing the p-values included in Tables 1 through 4 of the analysed paper with the calculated FDR-corrected significance threshold leads to the conclusion that only the predictor “living with a COVID-19 patient” is significantly related to the variable of interest after correcting for multiplicity. The same conclusion is reached when the p-values of Table 5 from Rodríguez-Barranco et al. (2021) are compared to the corresponding corrected significance.
threshold. Therefore, only the predictor “living with a COVID-19 patient” would contribute to explaining the estimated prevalence of the COVID-19 disease in a significant way.

4. Discussion

Our analysis of the Rodríguez-Barranco et al. (2021) paper has revealed different issues related to the objectives pursued, the methodologies applied, and interpreting the results.

4.1. Low support for causality

The ultimate aim of the analysed work was “to clarify the main causes of transmission of the COVID-19 disease during confinement”. Although hypothesis testing is a helpful tool informing on the likelihood of observing the results obtained under the null hypothesis, it cannot directly reveal causes nor determine the scientific relevance of a “statistically significant result” (i.e., a result that has a low probability of occurring when the null hypothesis is true).

Causal inference is multifactorial using observational/non-experimental evidence. A single hypothesis test results fall far short of the threshold for establishing a causal association between dog walking and COVID-19 disease, as proposed by Rodríguez-Barranco et al. (2021). According to the classic epidemiological criteria for causality (Hill, 1965), it is necessary to have a clear temporal relationship (i.e. exposure precedes outcome), reproducibility of results across studies in different populations, the strength of association (since a weak association is more likely to be biased or confounded than a stronger one) and biological plausibility. It is more plausible that interactions with a more significant number of people confer a greater risk than dogs serve as a mechanical vector of disease.

On the other hand, in observational studies oriented to explore relationships among many variables, without having a well-established working hypothesis to be tested, the probability of getting extreme results under the null hypothesis (i.e., significant results) is much higher than expected in well-focused planned research, because of the effects derived of data dredging. In these circumstances, it is necessary to interpret the discoveries correctly in the light of the previous knowledge.

In the work that we analyse, the direct evaluation of the p-values, effect sizes and precision data provided by the authors (Table 5 from Rodríguez-Barranco et al., 2021) lead to the conclusion that there is one highly significant relationship worthy of attention between the variable of interest (COVID-19) and “living with a COVID-19 patient”. The remaining variables included in the final adjusted model show a weak relationship with the variable of interest. In particular, in the relationship’s case with walking the pet that stands out as the most relevant achievement, neither the p-value (0.028), nor the value of the lower end of the interval of trust (1.03), point to an especially close relationship that can be labelled as “particularly relevant” from a statistical viewpoint. With an isolated result, without precedent (not replicated), and whose biological foundations are not well-established, it is more reasonable to present it as a mere sign or suggestion and stress that it requires confirmation.

4.2. Multiple testing and inflated type I error

In the analysed paper, the authors partially carry out a widespread practice, consisting of, from a data set with a high number of variables measured in the same sample, being first split it into groups and testing differences, then submitting it to bivariate analysis and test associations and, finally, performing a sequential process for building the model that culminates the work. There is nothing to object to this practice, except to point out that, in some instances, chance may be capitalised throughout the entire process and contribute to spurious results and conclusions.

For example, extensive table testing, including uncorrected multiple comparisons between several or many groups, large correlation/association matrices, and stepwise modelling with many predictors, are frequent sources of spurious findings (Rice, 1989; García, 2004). In particular, model selection is a primary source of hidden multiplicity that, in most cases, goes unnoticed (Mundry and Nunn, 2009; Forstmeier and Schielzeth, 2011; Young and Karr, 2011). Tsujiya (2014) has shown that using only random predictors, over 50% of the logistic models fitted included at least two significant but spurious predictors.

Therefore, it is very convenient to use some procedure, such as controlling the false discovery rate (Benjamini and Hochberg, 1995), that reports which result of multiple testing may have been generated by mere chance, as recommended by Glickman et al. (2014) for health studies. Similarly, the p-value is used to know the probability of a false positive when performing a single test. Both should be understood as useful information to consider when evaluating the research results.

Ignoring that multiple testing - mainly when data dredging is involved - can generate spurious findings with severe consequences for the reproducibility and credibility of the scientific research (Elsner, 2018) and is a major error demanding a solution. Nevertheless, it has nothing to do with discarding or ignoring rigorous research results because they involve multiple testing or non-significant p-values (Goldberg and Silbergeld, 2011; Rothman, 2014).

In the paper we analysed, multiplicity analysis confirms that several of the various effects reported as significant, particularly the one related to walking the pet, have a high chance of being spurious and should not have been presented as a demonstrated cause of COVID-19 transmission.

5. Conclusions

Rodríguez-Barranco et al. (2021) carried out an important study under very challenging conditions and provided original data of great value. However, they should have considered the effects of multiplicity on the significance of their results, at least as a potential limitation of their study. As a result of the above, the importance of dogs as possible transmitters of COVID-19 has likely been overestimated.

Finally, we want to spotlight the importance of highlighting the provisional nature of the results of the exploratory research, especially with unconfirmed shocking results, avoiding the use of terms such as discovered, demonstrated or proven.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We acknowledge the contributions of Dr Cristina Aponte and three referees who greatly help us to improve the ms.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.envres.2021.111600.

References

Benjamin, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. Roy. Stat. Soc. B 57, 89–230. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.

Elsner, D.A., 2018. Reproducibility of science: fraud, impact factors and carelessmess. J. Mol. Cell. Cardiol. 114, 364–368. https://doi.org/10.1016/j.yjmcc.2017.10.009.

Forstmeier, W., Schielzeth, H., 2011. Cryptic multiple hypotheses testing in linear models: overestimated effect sizes and the winner’s curse. Behav. Ecol. Sociobiol. 65, 47–55. https://doi.org/10.1007/s00265-010-1038-5.

García, L.V., 2004. Escaping the Bonferroni iron claw in ecological studies. Oikos 105, 657–663. https://doi.org/10.1111/j.0030-1299.2004.13046.x.
García, L.V., 2003. Controlling the false discovery rate in ecological research. Trends Ecol. Evol. 18, 553–554. https://doi.org/10.1016/j.tree.2003.08.011.

Glickman, M.E., Rao, S.R., Schultz, M.R., 2014. False discovery rate control is a recommended alternative to Bonferroni-type adjustments in health studies. J. Clin. Epidemiol. 67, 850–857. https://doi.org/10.1016/j.jclinepi.2014.03.012.

Goldberg, M., Silberfeld, E., 2011. On multiple comparisons and on the design and interpretation of epidemiological studies of many associations. Environ. Res. 111, 1007–1009. https://doi.org/10.1016/j.envres.2011.08.010.

Goodman, S.N., 1998. Multiple comparisons, explained. Am. J. Epidemiol. 147, 807–812. https://doi.org/10.1093/aje/147.8.807.

Gutiérrez-Hernández, O., García, L.V., 2021. On the usefulness of the bioclimatic correlative models of SARS-CoV-2. Environ. Res. 195, 110818. https://doi.org/10.1016/j.envres.2021.110818.

Hill, A.B., 1965. The environment and disease: association or causation? J. R. Soc. Med. 58, 295–300. https://doi.org/10.1177/003591576505800503.

Mundry, R., Nunn, C.L., 2009. Stepwise model fitting and statistical inference: turning noise into signal pollution. Am. Nat. 173, 119–123. https://doi.org/10.1086/593303.

Patel, C.J., Ioannidis, J.P.A., 2014. Placing epidemiological results in the context of multiplicity and typical correlations of exposures. J. Epidemiol. Community Health 68, 1096–1100. https://doi.org/10.1136/jech-2014-204195.

R Core Team, 2020. R: a language and environment for statistical computing. R Found. Stat. Comput. https://doi.org/10.1007/978-3-540-74686-7.

Rice, W.R., 1989. Analyzing tables of statistical tests. Evolution 43, 223–225. https://doi.org/10.1111/1558-5646.1989.tb00420.x.

Rodríguez-Barranco, M., Rivas-García, L., Quiles, J.L., Redondo-Sánchez, D., Aranda-Ramírez, P., Llopis-González, J., Sánchez Pérez, M.J., Sánchez-González, C., 2021. The spread of SARS-CoV-2 in Spain: Hygiene habits, sociodemographic profile, mobility patterns and comorbidities. Environ. Res. 192, 110223. https://doi.org/10.1016/j.envres.2020.110223.

Rothman, K.J., 2014. Six persistent research misconceptions. J. Gen. Intern. Med. 29, 1060–1064. https://doi.org/10.1007/s11606-013-2755-z.

Tsuchiya, T., 2014. Risk of performing multiple logistic regression analysis without considering multiplicity: an overview for clinicians and practitioners. J. Med. Stat. Inf. 2, 7. https://doi.org/10.7243/2053-7662-2-7.

Young, S.S., Bang, H., Oktay, K., 2009. Cereal-induced gender selection? Most likely a multiple testing false positive. Proc. Biol. Sci. 276, 1211–1212. https://doi.org/10.1098/rspb.2008.1405.

Young, S.S., Karr, A., 2011. Deming, data and observational studies. Significance 8, 116–120. https://doi.org/10.1111/j.1740-9713.2011.00506.x.

Oliver Gutiérrez-Hernández a, Luis V. García b

a Department of Geography, University of Málaga, Bulevar Louis Pasteur 27, 29010, Málaga, Spain

b Institute of Natural Resources and Agrobiology of Seville (IRNAS), Spanish National Research Council (CSIC), Av. Reina Mercedes 10, 41012, Seville, Spain

* Corresponding author.

E-mail addresses: olivergh@uma.es (O. Gutiérrez-Hernández), lv.garcia@csic.es (L.V. García).