ACTA MÉDICA PORTUGUESA
A Revista Científica da Ordem dos Médicos

Logistic Regression: Limitations in the Estimation of Measures of Association with Binary Health Outcomes

Regressão Logística: Limitações na Estimação de Medidas de Associação com Desfechos de Saúde Binários

Lara PINHEIRO-GUEDES✉[1,2], Clarisse MARTINHO[2], Maria ROSÁRIO O. MARTINS[3]
Acta Med Port 2024 Oct;37(10):697-705 ▪ https://doi.org/10.20344/amp.21435

ABSTRACT

Introduction: Logistic regression models are frequently used to estimate measures of association between an exposure, health determinant or intervention, and a binary outcome. However, when the outcome is frequent (> 10%), model estimates for relative risks and prevalence ratios might be biased. Despite the availability of several alternatives, many still rely on these models, and a consensus is yet to be reached. We aimed to compare the estimation and goodness-of-fit of logistic, log-binomial and robust Poisson regression models, in cross-sectional studies involving frequent binary outcomes.
Methods: Two cross-sectional studies were conducted. Study 1 was a nationally representative study on the impact of air pollution on mental health. Study 2 was a local study on immigrants' access to urgent healthcare services. Odds ratios (OR) were obtained through logistic regression, and prevalence ratios (PR) through log-binomial and robust Poisson regression models. Confidence intervals (CI), their ranges, and standard-errors (SE) were also computed, along with models' relative goodness-of-fit through Akaike Information Criterion (AIC), when applicable.
Results: In Study 1, the OR (95% CI) was 1.015 (0.970 - 1.063), while the PR (95% CI) obtained through the robust Poisson mode was 1.012 (0.979 - 1.045). The log-binomial regression model did not converge in this study. In Study 2, the OR (95% CI) was 1.584 (1.026 - 2.446), the PR (95% CI) for the log-binomial model was 1.217 (0.978 - 1.515), and 1.130 (1.013 - 1.261) for the robust Poisson model. The 95% CI, their ranges, and the SE of the OR were higher than those of the PR, in both studies. However, in Study 2, the AIC value was lower for the logistic regression model.
Conclusion: The odds ratio overestimated PR with wider 95% CI and higher SE. The overestimation was greater as the outcome of the study became more prevalent, in line with previous studies. In Study 2, the logistic regression was the model with the best fit, illustrating the need to consider multiple criteria when selecting the most appropriate statistical model for each study. Employing logistic regression models by default might lead to misinterpretations. Robust Poisson models are viable alternatives in cross-sectional studies with frequent binary outcomes, avoiding the non-convergence of log-binomial models.

Keywords: Logistic Models; Models, Statistical; Odds Ratio; Outcome Assessment, Health Care; Poisson Distribution

RESUMO

Introdução: A regressão logística é frequentemente utilizada para estimar medidas de associação entre uma exposição, determinante de saúde ou intervenção e um desfecho binário. No entanto, quando o desfecho é frequente (> 10%), estas estimativas podem ser enviesadas. Apesar de existirem modelos estatísticos alternativos, muitos estudos continuam a aplicar modelos de regressão logística indiscriminadamente. O objetivo deste estudo foi comparar as estimativas e o ajuste de modelos de regressão logística, log-binomial e Poisson robustos, em estudos transversais com desfechos binários frequentes.
Métodos: Realizaram-se dois estudos transversais. O Estudo 1 foi um estudo representativo a nível nacional sobre o impacto da poluição atmosférica na saúde mental. O Estudo 2 foi um estudo local sobre o acesso de imigrantes a serviços de urgência. Obtiveram-se *odds ratio* (OR) através de regressões logísticas e razões de prevalência (RP) através de modelos log-binomiais e Poisson robustos. Foram ainda obtidos intervalos de confiança a 95% (IC 95%), suas amplitudes, os erros-padrão (EP) das estimativas e comparados os valores *Akaike Information Criteria* (AIC).
Resultados: No Estudo 1, a OR (IC 95%) foi de 1,015 (0,970 - 1,063) e a RP (IC 95%) obtida através do modelo de Poisson robusto foi de 1,012 (0,979 - 1,045). O modelo de regressão log-binomial não convergiu. No Estudo 2, a OR (IC 95%) foi de 1,584 (1,026 - 2,446), a RP (IC 95%) para o modelo de regressão log-binomial foi de 1,217 (0,978 - 1,515) e para o modelo de Poisson robusto foi de 1,130 (1,013 - 1,261). Os IC 95%, as suas amplitudes e os EP das OR foram superiores ao das RP, em ambos os estudos. No entanto, no Estudo 2, o valor do AIC foi inferior no modelo de regressão logística.
Conclusão: As OR sobrestimaram as RP, com IC 95% mais amplos e EP superiores. A magnitude da sobrestimação foi tanto maior quanto mais prevalente o desfecho em estudo, em linha com estudos prévios. No Estudo 2, a regressão logística foi a que melhor se ajustou aos dados. Este exemplo ilustra a necessidade de avaliar vários critérios para selecionar o modelo estatístico mais apropriado. Os modelos de Poisson robustos são uma alternativa viável em estudos transversais com desfechos binários frequentes e evitam o problema de não convergência dos modelos log-binomiais.

Palavras-chave: Avaliação de Processos e Resultados em Cuidados de Saúde; Distribuição de Poisson; Modelos Estatísticos; Modelos Logísticos; Rácio de Probabilidades

1. Institute of Hygiene and Tropical Medicine. Universidade NOVA de Lisboa. Lisbon. Portugal.
2. Public Health Unit. Unidade Local de Saúde do Tâmega e Sousa. Marco de Canaveses. Portugal.
3. Global Health and Tropical Medicine. Institute of Hygiene and Tropical Medicine. Universidade NOVA de Lisboa. Lisbon. Portugal.
✉ **Autor correspondente:** Lara Pinheiro-Guedes. lpinheiroguedes@gmail.com
Recebido/*Received*: 28/02/2024 - **Aceite/*Accepted*:** 24/06/2024 - **Publicado/*Published*:** 01/10/2024

KEY MESSAGES

- Researchers should avoid defaulting to logistic regression, in studies with frequent binary outcomes (frequency > 10%), as it may lead to misinterpretations of data.
- The Robust Poisson regression is a viable alternative to the logistic regression, in cross-sectional studies with frequent binary outcomes, as it provides less biased estimates for prevalence ratios and avoids the convergence issues often encountered with log-binomial models.
- Careful consideration of the outcome's frequency and appropriate model selection are crucial to ensure accurate estimates.
- The Akaike Information Criterion (AIC) can be used to assess the goodness-of-fit for different models, but a lower AIC does not necessarily mean the model provides the most appropriate estimates for measures of association, as illustrated by the better fit of logistic regression in Study 2, despite its overestimation.

INTRODUCTION

In medical research, logistic regression models are frequently used to estimate measures of association between an exposure, treatment or health determinant and a binary outcome.[1,2] These models are employed mostly in cross-sectional and case-control studies to estimate odds ratios (OR), the odds of disease occurrence between the exposed and unexposed.[1] Logistic regression models are very attractive due to their simplicity and effectiveness.

The equivalent of the relative risk or risk ratio (RR) of prospective studies is the prevalence ratio (PR) in the case of cross-sectional studies. Although under certain conditions OR might allow to infer PR/RR, in many other cases, these inferences should not be made as they introduce bias.[1] If the outcome of the study is rare (≤ 10%), logistic regression models may appropriately estimate PR/RR. However, if the outcome is frequent (> 10%), OR might not be a suitable estimator of PR/RR.[2,3]

In the case of a frequent outcome, OR tend to overestimate PR/RR when their values are greater than one, or to underestimate them when values are lower than one.[2,4] A study reported that 40% of the papers that employed logistic regression models estimated OR which deviated by over 20% from the corresponding RR.[5] Odds ratio is also mistakenly portrayed as PR/RR, which is reflected in the results section of some research papers where 'risk' and 'probability' are wrongly used instead of the correct 'odds' and 'possibility' to refer to OR.[2,3]

While alternative methods to estimate valid adjusted PR/RR in the presence of frequent binary outcomes have been suggested, a consensus on the matter is yet to be reached. Log-binomial,[6-8] and modified Poisson regression models[6,7,9] are the most widely accepted.

The Poisson regression model is the nominal model for count data, being commonly used in cohort studies to estimate RR.[7,9] In cross-sectional studies with binary data, the computed ratios can still be interpreted, but as an approximation to the PR.[7,9] This approach yields correct estimates of the PR directly like log-binomial models, but typically with larger variances and standard errors.[7,9] Modified Poisson regression models, as the robust Poisson model, address this issue. Robust standard errors, determined through the Huber sandwich estimation (Appendix 1: https://www.actamedicaportuguesa.com/revista/index.php/amp/article/view/21435/15510), correct the Poisson model for data overdispersion (when variance is higher than the mean). This robust estimator performs well even when data does not perfectly meet the model's requirement, as in the case of Poisson regression models applied to cross-sectional binary data.[9]

Since methodological consistency is essential to epidemiological and clinical research, and given the potential impact of biased estimates on clinical practice, policymaking, communication and community behaviour, selecting the appropriate statistical models holds paramount importance. This study aimed to compare the estimation and goodness-of-fit of logistic, log-binomial, and robust Poisson regression models, in cross-sectional studies with frequent (> 10%) binary outcomes.

METHODS

We carried out two cross-sectional studies (Study 1 and Study 2) to estimate the measures of association between two different exposures and two different frequent (> 10%) binary outcomes that were adjusted for several variables identified as potential confounders through directed acyclic graphs, elaborated based on a literature review of the associations in the study.

We conducted a comparative analysis of the estimates of the measures of association (OR in the case of logistic regression models, and PR in the case of log-binomial and robust Poisson regression models), in terms of their magnitude and significance. The estimates' confidence intervals, their ranges and standard errors were also compared.

The maximum likelihood estimation method was employed in all the computed models, even for those using complex survey data (Study 1). In this case, we fitted models to complex data incorporating the sampling weights in a designed-based analysis (a statistical analysis of survey data that took the survey design, its stratification and clustering, and the sampling weights into consideration).

As the Akaike Information Criterion (AIC) is a widely used tool in model selection and allows to compare the relative goodness-of-fit of different models, AIC values were obtained for every model in study fitting non-complex data (Study 2) and compared. The optimal model should be the one with the minimum AIC value. Any model yielding an AIC value within two units of the minimum AIC value might also be an appropriate candidate.[10]

All statistical analyses were conducted using Stata®, version 15. The significance level was established at 5%.

Study 1: Association between long-term exposure to ambient air, depression and anxiety

Particulate matter smaller than 10 µm (PM_{10}) accounts for much of the impact of air pollution on health.[11,12] Some studies have assessed the association between long-term exposure to PM_{10} and common mental disorders (CMD), namely depressive and anxiety disorders, but evidence is inconsistent.[13-18] To estimate the association between long-term exposure to PM_{10} and the frequency of probable diagnosis of CMD, a population-based, nationally representative cross-sectional study was conducted, in mainland Portugal.

Long-term exposure to PM_{10} was estimated through one-year average concentrations of PM_{10}, calculated with data from the Portuguese Environment Agency's air quality monitoring stations, and attributed to each individual considering their seven-digit postal codes of residence. The probable diagnosis of CMD was ascertained through the scores obtained in the 5-item Mental Health Inventory (MHI-5) [Appendix 2 – Fig. 2.1 (Appendix 2: https://www.actamedicaportuguesa.com/revista/index.php/amp/article/view/21435/15511)].[19-21] A score ≤ 52 in the MHI-5 represents a situation that implies proper clinical evaluation by a doctor (a "Probable diagnosis of CMD"), while a score > 52 represents a situation that does not imply proper clinical evaluation. The MHI-5 scores and the independent variables' data were obtained from a restricted sample of the participants of the first Portuguese National Health Examination Survey (INSEF).[22] The study participants were the INSEF participants from mainland Portugal who consented on data linkage, had their seven-digit postal codes of residence available in the database, were living within a 30-km radius from a background air quality monitoring station (with available data on PM_{10} atmospheric concentration), like previously detailed,[23-26] and who had answered all the five items of the MHI-5. Data on area-level socioeconomic deprivation, the Portuguese version of the European Deprivation Index, was available online.[27] Individually allocated one-year average temperatures were obtained through a similar methodology to the one applied to estimate exposure to PM_{10}, making use of one-year average temperatures collected from the National Oceanic and Atmospheric Administration data.[23,24,26] Data on area-level walkability was obtained through the weighted sum of residential density, street connectivity, and a land use mix index, for all the parishes of mainland Portugal, and is available at request, in terciles of increasing walkability.

We performed single-level multivariable analyses since, even if we used individual and aggregated variables (at parish level) in our models, the assumptions for carrying out a multilevel analysis were not met [some parishes had just one individual) (models' specifications in Appendix 2 – Table 2.1 (Appendix 2: https://www.actamedicaportuguesa.com/revista/index.php/amp/article/view/21435/15511)]. All estimates were weighted to account for different selection probabilities resulting from the complex sampling design, and to match the population distribution in terms of geographic region, age group and sex, unless specifically stated.

Study 2: Association between immigration status and urgent care use in a paediatric population

Foreign residents have been increasing in Portugal, in the past years.[28] One of the biggest challenges immigrants face in the host countries is to obtain access to healthcare services.[29] The Portuguese National Health Service offers universal and free healthcare services for children up to 18 years old, irrespective of their immigration status.[30] Several studies have already reported the increased urgent care use by migrants, compared to non-migrants,[31-33] but evidence is inconsistent,[31,34-38] and differences exist according to migrants' characteristics and to countries' different healthcare systems.[31,39] To estimate the association between being an immigrant and the urgent care use, in the paediatric population living in Amadora, a population-based, non-representative cross-sectional study was conducted there, in the most densely populated Portuguese municipality, and the second with the highest density of foreign residents.[30]

Data was obtained from the participants of the first wave of the CRIAS (Health Trajectories of Immigrant Children in Amadora) cohort, implemented in Portugal from June 2019 to March 2020. The CRIAS targeted children born in 2015, aged four or five years old, who had records of attending at least one of the primary healthcare centres of Amadora in the two years before the assessment time, also targeting their parents. Self-reported data on the families' demographic and socioeconomic characteristics, migration history, and on the children's lifestyle and health was collected and integrated with data on the children's healthcare use, obtained through electronic health records.[30] From the CRIAS participants, those with missing data on the variables of immigration status or urgent care use were excluded from the present study.

Immigration status was categorized as "immigrant" (born to non-native parents, even if in Portugal) and "non-immigrant" (born in Portugal, to native parents). Urgent care use was categorized as "yes" or "no" according to the children having, or not, at least one visit to the urgent care service of the Hospital Professor Doutor Fernando Fonseca, in 2019.

The AIC values were obtained for all the fitted models [models' specifications in Appendix 3 – Table 3.1 (Appendix 3: https://www.actamedicaportuguesa.com/revista/index.php/amp/article/view/21435/15512)].

Ethical considerations

In this study, no ethical or legal issues of confidentiality were raised, since all the data came from anonymized databases.

The INSEF had its scientific protocol approved by the Ethics Committee for Health of the National Health Institute Doutor Ricardo Jorge (INSA) (Internal Note No. 7/2011), by the National Data Protection Commission (Authorization No. 199/2001), and by the Ethics Commissions of the Northern Portugal Regional Health Administration (Authorization No. 91/2014), the Central Portugal Regional Health Administration (Authorization No. 44/2014), the Lisbon and Tagus Valley Regional Health Administration (Authorization No. 17/2014), the Algarve Regional Health Administration (Authorization No. 2742 of 04/03/2015), and the Health Service of the Autonomous Region of Madeira (Authorization No. 32/2014). All the participants were asked to sign a declaration of informed consent to participate in INSEF, which consisted of accepting to respond to a general health interview, perform a physical examination and donate a blood sample for testing.

The protocol of this study was also approved by the INSA Ethics Committee and by the Institutional Review board of INSEF (INSA-IM60_05/February 2023).

The CRIAS cohort involves human participants and was approved by the Health Ethics Committee of the Lisbon and Tagus Valley Regional Health Administration of, Portugal (001/CES/INV/2019). The parents signed a written information and consent form to participate in the study, including permission to assess data from the child's health centre and hospital medical records.

RESULTS

Study 1

A total of 2398 individuals were included in the study, following the application of the inclusion and exclusion criteria [Appendix 2 – Fig. 2.2 (Appendix 2: https://www.actamedicaportuguesa.com/revista/index.php/amp/article/view/21435/15511)]. Included and excluded individuals were similar regarding most of the analysed characteristics. Differences between these two groups were only found regarding the individual allocated one-year average temperature, and the area-level walkability terciles [Appendix 2 – Table 2.2 (Appendix 2: https://www.actamedicaportuguesa.com/revista/index.php/amp/article/view/21435/15511)].

The study population had higher percentages of female participants (52.6%, 95% CI: 50.2 - 55.1) and individuals belonging to the age groups 35 - 49 years (34.1%, 95% CI: 32.0 - 36.4) and 50 - 64 years (31.5%, 95% CI: 29.4 - 33.7) [Appendix 2 – Table 2.3 (Appendix 2: https://www.actamedicaportuguesa.com/revista/index.php/amp/article/view/21435/15511)]. The median [interquartile range (IQR)] individual allocated one-year average PM_{10} concentration was 18.6 (15.3 - 19.3) μg/m³ [Appendix 2 – Table 2.4 (Appendix 2: https://www.actamedicaportuguesa.com/revista/index.php/amp/article/view/21435/15511)]. A probable diagnosis of DMC was found in 22.7% (95% CI: 20.0 - 25.6) of the study population [Appendix 2 – Table 2.3 (Appendix 2: https://www.actamedicaportuguesa.com/revista/index.php/amp/article/view/21435/15511)].

The adjusted logistic regression model obtained an OR (95% CI) = 1.015 (0.970 - 1.063), and the robust Poisson model a PR (95% CI) = 1.012 (0.979 - 1.045). The adjusted log-binomial regression model did not converge. Adjusted OR overestimated the adjusted PR obtained through the robust Poisson model by 0.003 and presented wider 95% CI (95% CI range: 0.093 *vs* 0.066, respectively) and a higher standard-error (0.022 *vs* 0.016, respectively) (Table 1).

No statistically significant association between long-term exposure to PM_{10} and the frequency of probable CMD diagnosis was observed in any of the models after adjustment.

Study 2

From the CRIAS participants, 410 children were included in the study, since 10 were excluded due to missing data in the variable urgent care use.

Most of the children were male (50.7%), being most of their caregivers (or questionnaire respondents) female (87.6%) with a median (IQR) age of 34 (18 - 75) years old [Appendix 3 – Table 3.2 (Appendix 3: https://www.actamedicaportuguesa.com/revista/index.php/amp/article/view/21435/15512)]. Among the children in study, 50.5% were immigrants [Appendix 3 – Table 3.2 (Appendix 3: https://www.actamedicaportuguesa.com/revista/index.php/amp/article/view/21435/15512)]. Approximately 48.0% of all children used urgent care services (58.4% of them being immigrants) [Appendix 3 – Tables 3.2 and 3.3 (Appendix 3:

Table 1 – Characteristics of logistic, log-binomial and robust Poisson regression models fitted for the estimation of measures of association between the exposure to particulate matter with an aerodynamic diameter equal than or lower to 10 micrometres and the frequency of common mental disorders

	Logistic	Log-binomial	Robust Poisson
Number of individuals included in the adjusted models	2210	2210	2210
Adjusted OR/PR* (95% CI)	1.015 (0.970 - 1.063)	Did not converge	1.012 (0.979 - 1.045)
95% CI range	0.093	Did not converge	0.066
Standard-error of the adjusted estimates	0.022	Did not converge	0.016

OR: odds ratio; PR: prevalence ratio; 95% CI: 95% confidence interval; PM10: particulate matter with an aerodynamic diameter less than or equal to 10 micrometres.
* Adjusted for sex, age groups, education level, employment status, professional occupation, area-level socioeconomic deprivation terciles, individual allocated one-year average temperature, area-level walkability terciles, degree of urbanization (OR for logistic and log-binomial regression models, PR for robust Poisson regression model).

https://www.actamedicaportuguesa.com/revista/index.php/amp/article/view/21435/15512)].

The adjusted logistic regression model obtained an OR (95% CI) = 1.584 (1.026 - 2.446), the adjusted log-binomial regression model a PR (95% CI) = 1.217 (0.978 - 1.515), and the robust Poisson regression model a PR (95% CI) = 1.130 (1.013 - 1.261). The adjusted OR overestimated the adjusted PR obtained through the log-binomial and robust Poisson regression models by 0.367 and 0.454, respectively, presented wider 95% CI (95% CI range: 1.420 *vs* 0.537 and 0.248, respectively) and higher standard-errors (0.351 *vs* 0.136 and 0.063, respectively). However, the AIC value was lower for the logistic regression model than for the log-binomial and robust Poisson models (1.345 *vs* 1.350 and 1.656, respectively) (Table 2).

Only logistic and robust Poisson regression models showed statistically significant associations between immigration status and urgent care service use, in the adjusted models. The log-binomial model, however, did not yield statistically significant associations.

DISCUSSION

This study, based on two distinct real-life epidemiological examples involving outcomes with different prevalence (22.7% in Study 1, and 48.0% in Study 2), contributed to the knowledge about the estimation of measures of association, in cross-sectional studies with frequent outcomes, in different settings. Study 1 assessed a nationally representative sample of 2398 individuals, whereas Study 2 focused on a local sample of 410 individuals. Study 1 estimated the association between a continuous exposure and a binary outcome, adjusting for nine independent variables (including one continuous variable); and Study 2 estimated the association between a binary exposure and a binary outcome, adjusting for three covariates.

In Study 1, the log-binomial regression model did not converge. This likely happened due to the inclusion of two continuous independent variables in the model (one variable which was a potential confounder and the exposure in study).[8,40,41] When in the presence of quantitative variables, the maximum likelihood estimate can lie on the boundary of the parameter space which can lead to model non-convergence as the instantaneous slope of the likelihood may not reach zero at this boundary.[6] This inability to find a solution that fits the model to the data supports authors who advocate for the robust Poisson regression model as the

Table 2 – Characteristics of logistic, log-binomial, and robust Poisson regression models fitted for the estimation of measures of association between immigration status and urgent care use

	Logistic	Log-binomial	Robust Poisson
Number of individuals included in the adjusted models	389	389	389
Adjusted OR/PR* (95% CI)	**1.584 (1.026 - 2.446)**	1.217 (0.978 - 1.515)	**1.130 (1.013 - 1.261)**
95% CI range	**1.420**	0.537	**0.248**
AIC for the adjusted model	1.345	1.350	1.656
Standard-error of the adjusted estimate	0.351	0.136	0.063

OR: odds ratio; PR: prevalence ratio; 95% CI: 95% confidence interval; AIC: Akaike information criteria.
* Adjusted for caregiver's age, education level and professional occupation (OR for logistic and log-binomial regression models, PR for robust Poisson regression model). Results in bold are those statistically significant.

preferred choice to logistic models, when addressing frequent outcomes in cross-sectional studies, instead of log-binomial models.[40,41]

As, in Study 1, the obtained point estimate for the measures of association was higher than 1, the OR overestimated the PR obtained through the robust Poisson regression model. The magnitude of the overestimation was 0.003 units (OR = 1.015 *vs* PR = 1.012). The precision of the OR was lower than the precision of the PR, which presented a narrower 95% CI, with the 95% CI range for the OR being 0.027 units higher than the obtained for the PR through the robust Poisson regression model (0.093 *vs* 0.066, respectively). Consistently, the standard-error was 0.006 units lower for the robust Poisson (robust Poisson: 0.016 *vs* logistic: 0.022, respectively).

In Study 2, the obtained measures of association were also higher than 1, with the consequent overestimation of PR by OR already reported in Study 1. The magnitude of the overestimation was higher when the OR was compared with the PR obtained through the robust Poisson model, than with the PR obtained through the log-binomial model (overestimations of 0.454 and 0.367 units, respectively). This happened because the robust Poisson model obtained a lower value for the PR than the log-binomial model (PR: 1.130 *vs* 1.217, respectively). The robust Poisson model also obtained the most precise estimate (the one with the lowest 95% CI range), followed by the log-binomial and, lastly, by the logistic regression model (95% CI ranges: 0.248, 0.537, and 1.420, respectively), as well as the lowest standard-error for the estimate. However, the AIC values were lower for the logistic regression model, followed by the log-binomial model, and, lastly, by the robust Poisson model, indicating that the logistic was the one better fitting to the data, which was counterintuitive. This might have been because the binomial distribution is ultimately more appropriate to module binary outcomes than the Poisson distribution, and because the mean and the variance of the outcome were different in this study (variance = 0.250, mean = 0.480), affecting the fit of the Poisson model, even if a robust estimator was being used.[7] This example illustrates why relying solely on the AIC value for the selection of the most appropriate statistical model is not adequate.

We also observed that, in Study 2, the log-binomial model could not obtain a statistically significant association, while the logistic and robust Poisson models did. This might have been due to the conjunction of the estimation of a not-so-inflated estimate through the log-binomial model, as through the logistic model, with wider 95% CI for the log-binomial than for the robust Poisson model.

Moreover, in Study 2, the magnitude of the overestimation exceeded the overestimation observed in Study 1 (Study 2: 0.454 and 0.367 units, Study 1: 0.003 units), which was also verified in terms of the 95% CI range. This discrepancy was likely due to the higher prevalence of the outcome in Study 2 compared to Study 1 and illustrates that higher outcome prevalence represents larger OR inflations.[1,3]

In both Study 1 and Study 2, the robust Poisson model exhibited greater precision (narrower 95% CI and lower 95% CI ranges) than log-binomial and logistic regression models. This characteristic is particularly relevant when a study aims to estimate unbiased measures of association (where the expected value is identical to the population parameter being estimated), but not so when the aim is solely to comprehend the general trend or direction of it.[2]

In both examples, we observed that OR obtained through logistic regression models provide biased estimates of PR, in the presence of frequent binary outcomes, and also wider confidence intervals, magnifying the probability of a type II error. Additionally, we verified that the extension of the bias became more pronounced as the prevalence of the outcome increased, consistent with findings from prior studies.[4,6,8,40-44] The challenges associated with estimating measures of association through logistic regression models when there are frequent outcomes in study are not limited to cross-sectional designs. Longitudinal studies and randomized clinical trials might encounter similar estimation issues, and also need to be investigated.[5]

This study reinforces the viability of alternative statistical models and reopens a discussion that has yet to reach a definitive solution. These alternative models are readily accessible within commonly used statistical software packages like SPSS®, Stata® and R Studio®. Nonetheless, they have potential disadvantages that could lead researchers to discard them and stick to the logistic regression models. A disadvantage of the log–binomial regression is that the model might not converge, like in Study 1.[45] Despite robust Poisson regression not showing convergence issues, this model, like the log-binomial, may yield individual predicted probabilities above 1. The robust Poisson regression is indeed a suitable method to obtain valid, unbiased estimates, but not to predict individual probabilities (as in diagnostic or prognostic studies).[3,7,9] Moreover, both models lack reciprocity, underscoring the importance of being mindful of the reference category selected.[4] While other options exist, the log-binomial and the robust Poisson regression models are relatively straightforward to apply, easy to interpret, and offer the ability to control for confounding.[7,9]

Given the curriculum requirements that often compel medical doctors and other health professionals to publish, recognizing the time constraints clinicians face to master research methodologies and statistics, and considering the need for consistency in research, establishing a standardized approach for the selection of the most suitable model to employ in a study would be useful. Several processes have been proposed. One of them involves constructing a contingency table between the exposure and the outcome and comparing the OR and RR/PR obtained through epidemiological formulas of calculation. Relevant disparities between the values would suggest the employment of a model that estimates RR/PR instead of OR (such as the log-binomial and robust Poisson models).[44] However, this methodology is too simplistic, can only be applied in the presence of a categorical exposure, and does not help to choose between log-binomial and robust Poisson models. Another proposal was made by a group of authors who created a flowchart to support researchers, but its interpretation could be complex as it requires some methodological knowledge, and it was not agreed upon consensus.[46]

Strengths and limitations

This study has several strengths. First, it contributed to reigniting the discussion around the validity of the OR to infer RR/PR in the context of frequent binary outcomes, in cross-sectional studies. Second, it used real-life datasets applied to different settings – in terms of topics investigated (the health impact of air pollution, and the access to healthcare services), types of exposure assessed (continuous, and categorical), types of independent variables included (categorical and continuous, or just categorical), and outcome prevalence (22.7%, and 48.0%) – allowing for a comprehensive discussion, pertinent to several different scientific fields and aims. Third, the example studies were conducted according to robust and previously published methodological approaches, allowing us to diminish biases and confounding, and to add findings relevant to the scientific literature.

Some limitations should also be noted. First, a simulation study to calculate the exact relative risk was not performed. Second, the assessed associations did not have substantial effect magnitudes, making the estimates' values closer to one and, across models, to each other (due to reduced variability). Third, logistic regression models were compared against log-binomial and robust Poisson regression models, and not all the possible models. This choice was made not only based on the appropriateness of these models to the studies developed, but also on their popularity and user-friendliness. However, we acknowledge that each model has its own assumptions and limitations

CONCLUSION

Even if both the OR and the RR/PR might be appropriate to understand the direction of a certain measure of association, in a study whose aim is to obtain unbiased estimates, using the OR might lead to misinterpretations. The OR should not be interpreted as RR/PR in cross-sectional studies whose outcomes are not rare. Robust Poisson models could be a viable alternative to logistic regression models, avoiding the non-convergence of log-binomial models.

There is not a standardized approach to the selection process of the most suitable model. The type of outcome variable, the aim of the study (the need or not of unbiased estimates), its intent, the fulfilment of the assumptions for causal inference, and the fulfilment of the assumptions of the available statistical models should guide the decision. Several statistical measures could be used to support the models' selection, including models' estimates, standard-errors, 95% CI, and AIC values. This study emphasizes that no single value should serve as the sole criterion to determine the most appropriate model to employ.

More studies could be conducted to compare other alternatives to logistic regression models (modified Cox, quasi-Poisson, negative binomial, and probit models), not only within cross-sectional, but also in longitudinal studies and clinical trials, provided these studies aim to obtain unbiased estimates. Simulation studies could be useful to fix relative risks. A standard approach to the estimation of measures of association involving frequent binary outcomes would be helpful to guide researchers in the process.

ACKNOWLEDGEMENTS

The authors acknowledge all the support received from the colleagues from the Department of Epidemiology of the National Health Institute Doutor Ricardo Jorge IP (INSA, IP), especially Vânia Gaio, Mafalda Sousa Uva, and also thank Isabel Natário, and Sara Simões Dias.

PREVIOUS AWARDS AND PRESENTATIONS

A poster of this work was presented in the *IV Congresso Nacional dos Médicos de Saúde Pública*.

AUTHOR CONTIBUTIONS

LPG: Conceptualization, methodology, software, formal analysis and investigation, writing – original draft preparation.

CM: Investigation, writing – review & editing.

MROM: Conceptualization, resources, validation, writing – review & editing, supervision.

All authors approved the final version to be published.

PROTECTION OF HUMANS AND ANIMALS

The authors declare that the procedures were followed according to the regulations established by the Clinical Research and Ethics Committee and to the Helsinki Declaration of the World Medical Association updated in 2013.

DATA CONFIDENTIALITY

The authors declare having followed the protocols in use at their working center regarding patients' data publication.

DATA AVAILABILITY

The datasets analysed during the current study are not publicly available due to ethical restrictions. Data is available from the authors upon reasonable request.

COMPETING INTERESTS

The authors have declared that no competing interests exist.

FUNDING SOURCES

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

INSEF was developed as part of a project financed under the Public Health Initiatives Programme ("Improvement of epidemiological health information to support public health decision and management in Portugal. Towards reduced inequalities, improved health and bilateral cooperation") with a 1,500,000 grant from Iceland, Liechtenstein and Norway, the EEA Grants and the Portuguese Government.

CRIAS cohort was financed by the Asylum, Integration and Migration Fund (ref.PT/2018/FAMI/350) under the Multianual Financial Framework 2014/20, by the Portuguese Foundation for Science and Technology (FCT) (RESEARCH4COVID-19-065) and Global Health and Tropical Medicine (GHTM), Institute of Hygiene and Tropical Medicine (IHMT), NOVA University of Lisbon, Portugal (UID/04413/2020). The extension of the cohort study is financed by the Portuguese Foundation for Science and Technology (FCT) (PTDC/SAU-SER/4664/2020).

REFERENCES

1. Hosmer Jr, Lemeshow S, Sturdivant RX. Applied logistic regression. 3rd ed. Hoboken: John Wiley & Sons; 2013.
2. Lash TL, VanderWeele TJ, Haneuse S, Rothman KJ. Modern epidemiology. Philadelphia: Lippincott Williams & Wilkins; 2020.
3. Celentano D, Szklo M. Gordis epidemiology. Amsterdam: Elsevier; 2020.
4. Tamhane AR, Westfall AO, Burkholder GA, Cutter GR. Prevalence odds ratio versus prevalence ratio: choice comes with consequences. Stat Med. 2016;35:5730-5.
5. Knol MJ, Le Cessie S, Algra A, Vandenbroucke JP, Groenwold RH. Overestimation of risk ratios by odds ratios in trials and cohort studies: alternatives to logistic regression. CMAJ. 2012;184:895-9.
6. Petersen MR, Deddens JA. A comparison of two methods for estimating prevalence ratios. BMC Med Res Methodol. 2008;8:9.
7. McCullagh P. Generalized linear models. New York: Routledge; 2019.
8. Barros AJ, Hirakata VN. Alternatives for logistic regression in cross-sectional studies: an empirical comparison of models that directly estimate the prevalence ratio. BMC Med Res Methodol. 2003;3:21.
9. Fleiss JL, Levin B, Paik MC. Statistical methods for rates and proportions. New York: John Wiley & Sons; 2003.
10. Cavanaugh JE, Neath AA. The Akaike information criterion: background, derivation, properties, application, interpretation, and refinements. WIREs Comp Stats. 2019;11:e1460.
11. Manisalidis I, Stavropoulou E, Stavropoulos A, Bezirtzoglou E. Environmental and health impacts of air pollution: a review. Front Public Health. 2020;8:14.
12. World Health Organization. WHO global air quality guidelines: particulate matter (PM2.5 and PM10), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide. 2021.[cited 2023 Aug 06]. Available from: https://iris.who.int/bitstream/handle/10665/345329/9789240034228-eng.pdf?sequence=1&isAllowed=y.
13. Braithwaite I, Zhang S, Kirkbride JB, Osborn DPJ, Hayes JF. Air pollution (particulate matter) exposure and associations with depression, anxiety, bipolar, psychosis and suicide risk: a systematic review and meta-analysis. Environ Health Perspect. 2019;127:126002.
14. Borroni E, Pesatori AC, Bollati V, Buoli M, Carugno M. Air pollution exposure and depression: a comprehensive updated systematic review and meta-analysis. Environ Pollut. 2022;292:118245.
15. Zeng Y, Lin R, Liu L, Liu Y, Li Y. Ambient air pollution exposure and risk of depression: a systematic review and meta-analysis of observational studies. Psychiatry Res. 2019;276:69-78.
16. Liu Q, Wang W, Gu X, Deng F, Wang , Lin H, et al. Association between particulate matter air pollution and risk of depression and suicide: a systematic review and meta-analysis. Environ Sci Pollut Res Int. 2021;28:9029-49.
17. Trushna T, Dhiman V, Raj D, Tiwari RR. Effects of ambient air pollution on psychological stress and anxiety disorder: a systematic review and meta-analysis of epidemiological evidence. Rev Environ Health. 2021;36:501-21.
18. Cuijpers P, Miguel C, Ciharova M, Kumar M, Brander L, Kumar P, et al. Impact of climate events, pollution, and green spaces on mental health: an umbrella review of meta-analyses. Psychol Med. 2023;53:638-53.
19. Stewart AL, Greenfield S, Hays RD, Wells K, Rogers WH, Berry SD, et al. Functional status and well-being of patients with chronic conditions. Results from the Medical Outcomes Study. JAMA. 1989;262:907-13.
20. Stewart AL, Hays RD, Ware JE, Jr. The MOS short-form general health survey. Reliability and validity in a patient population. Med Care. 1988;26:724-35.
21. Berwick DM, Murphy JM, Goldman PA, Ware JE Jr., Barsky AJ, Weinstein MC. Performance of a five-item mental health screening test. Med Care. 1991;29:169-76.
22. Nunes B, Barreto M, Gil AP, Kislaya I, Namorado S, Antunes L, et al. The first Portuguese national health examination survey (2015): design, planning and implementation. J Public Health. 2019;41:511-7.
23. Gaio V, Dias CM. Association between ambient air pollution exposure and biomarkers of cardiovascular risk: link between the first Portuguese health examination survey and the air quality data. [cited 2023 Aug 12]. Available from: https://run.unl.pt/handle/10362/143849.
24. Gaio V, Roquette R, Monteiro A, Ferreira J, Rafael S, Dias CM, et al. Exposure to ambient particulate matter increases blood count parameters with potential to mediate a cardiovascular event: results from a population-based study in Portugal. Air Qual Atmos Hlth. 2021;14:1189-202.
25. Gaio V, Roquette R, Monteiro A, Ferreira J, Matias Dias C, Nunes B. Investigating the association between ambient particulate matter (PM(10)) exposure and blood pressure values: Results from the link between the Portuguese Health Examination Survey and air quality data. Rev Port Cardiol. 2023;42:251-8.
26. Gaio V, Roquette R, Monteiro A, Ferreira J, Lopes D, Dias CM, et al. PM10 exposure interacts with abdominal obesity to increase blood triglycerides: a cross-sectional linkage study. Eur J Public Health. 2022;32:281-8.
27. Ribeiro AI, Launay L, Guillaume E, Launoy G, Barros H. The Portuguese version of the European Deprivation Index: development and association with all-cause mortality. PLoS One. 2018;13:e0208320.
28. Instituto Nacional de Estatística. Censos 2021 dissemination platform (definitive results). [cited 2023 Sep 10]. Available from: https://censos.ine.pt/xportal/xmain?xpgid=censos21_produtos&xpid=CENSOS21&xla

ng=ptdb_censos_2021.html.
29. Buja A, Fusco M, Furlan P, Bertoncello C, Baldovin T, Casale P, et al. Characteristics, processes, management and outcome of accesses to accident and emergency departments by citizenship. Int J Public Health. 2014;59:167-74.
30. Muggli Z, Mertens T, Amado R, Teixeira AL, Vaz D, Pires M, et al. Cohort profile: health trajectories of immigrant children (CRIAS)-a prospective cohort study in the metropolitan area of Lisbon, Portugal. BMJ Open. 2022;12:e061919.
31. Graetz V, Rechel B, Groot W, Norredam M, Pavlova M. Utilization of health care services by migrants in Europe-a systematic literature review. Br Med Bull. 2017;121:5-18.
32. Norredam M, Krasnik A, Moller Sorensen T, Keiding N, Joost Michaelsen J, Sonne Nielsen A. Emergency room utilization in Copenhagen: a comparison of immigrant groups and Danish-born residents. Scand J Public Health. 2004;32:53-9.
33. Ruud SE, Hjortdahl P, Natvig B. Reasons for attending a general emergency outpatient clinic versus a regular general practitioner - a survey among immigrant and native walk-in patients in Oslo, Norway. Scand J Prim Health Care. 2017;35:35-45.
34. Gimeno-Feliu LA, Pastor-Sanz M, Poblador-Plou B, Calderon-Larranaga A, Diaz E, Prados-Torres A. Overuse or underuse? Use of healthcare services among irregular migrants in a north-eastern Spanish region. Int J Equity Health. 2021;20:41.
35. Schoevers MA, Loeffen MJ, van den Muijsenbergh ME, Lagro-Janssen AL. Health care utilisation and problems in accessing health care of female undocumented immigrants in the Netherlands. Int J Public Health. 2010;55:421-8.
36. Gimeno-Feliu LA, Calderon-Larranaga A, Diaz E, Poblador-Plou B, Macipe-Costa R, Prados-Torres A. Global healthcare use by immigrants in Spain according to morbidity burden, area of origin, and length of stay. BMC Public Health. 2016;16:450.
37. Dias S, Gama A, Cortes M, de Sousa B. Healthcare-seeking patterns among immigrants in Portugal. Health Soc Care Community. 2011;19:514-21.
38. Sauzet O, David M, Naghavi B, Borde T, Sehouli J, Razum O. Adequate utilization of emergency services in Germany: is there a differential by migration background? Front Public Health. 2020;8:613250.
39. Tzogiou C, Boes S, Brunner B. What explains the inequalities in health care utilization between immigrants and non-migrants in Switzerland? BMC Public Health. 2021;21:530.
40. Coutinho LM, Scazufca M, Menezes PR. Methods for estimating prevalence ratios in cross-sectional studies. Rev Saude Publica. 2008;42:992-8.
41. Lee J, Tan CS, Chia KS. A practical guide for multivariate analysis of dichotomous outcomes. Ann Acad Med Singap. 2009;38:714-9.
42. Behrens T, Taeger D, Wellmann J, Keil U. Different methods to calculate effect estimates in cross-sectional studies. A comparison between prevalence odds ratio and prevalence ratio. Methods Inf Med. 2004;43:505-9.
43. Martinez BA, Leotti VB, Silva GS, Nunes LN, Machado G, Corbellini LG. Odds ratio or prevalence ratio? An overview of reported statistical methods and appropriateness of interpretations in cross-sectional studies with dichotomous outcomes in veterinary medicine. original research. Front Vet Sci. 2017;4:193.
44. Gnardellis C, Notara V, Papadakaki M, Gialamas V, Chliaoutakis J. Overestimation of relative risk and prevalence ratio: misuse of logistic modeling. Diagnostics. 2022;12:2851.
45. Williamson T, Eliasziw M, Fick GH. Log-binomial models: exploring failed convergence. Emerg Themes Epidemiol. 2013;10:14.
46. Reichenheim ME, Coutinho ES. Measures and models for causal inference in cross-sectional studies: arguments for the appropriateness of the prevalence odds ratio and related logistic regression. BMC Med Res Methodol. 2010;10:66.