Journal of
Infectious Diseases and Epidemiology

RESEARCH ARTICLE

Statistical Behavior of Hospital Admissions for Respiratory Diseases by Probability Distribution Functions

Amaury de Souza [ID][1*], *Flavio Aristone* [ID][1], *Widinei A Fernandes*[1], *Zaccheus Olaofe* [ID][2], *Marcel Carvalho Abreu* [ID][3], *José Francisco de Oliveira Júnio* [ID][4], *Guilherme Cavazzana* [ID][5] and *Cícero Manoel dos Santos* [ID][6]

[1]*Federal University of Mato Grosso do Sul, CP 549, Campo Grande, MS, 79070-900, Brazil*

[2]*University of Cape Town, Rondebosch, Western Cape, South Africa*

[3]*Universidade Federal Rural do Rio de Janeiro, Seropédica, Rio de Janeiro, Brazil*

[4]*Institute of Atmospheric Sciences, Universidade Federal de Alagoas, Maceió, Brazil*

[5]*Faculdade de Mato Grosso do Sul, Universidade Católica Dom Bosco, Brazil*

[6]*Federal University of Para, Altamira, Brazil*

*Corresponding authors: *Amaury de Souza, Federal University of Mato Grosso do Sul, CP 549, Campo Grande, MS, 79070-900, Brazil*

Abstract

Climate change has a high impact on health and morbidity/mortality in respiratory system diseases and remains poorly investigated in probability distribution modeling. The objective of this study was to analyze the adjustments of Burr (Bu), Inv Gausian 3P (IG3P), Lognormal (LN), Pert (Pe), Rayleigh 2P (Ra 2P) and Weibull 3P (W3P) distributions of the historical series of hospitalizations for respiratory diseases (total hospital admissions) for the period from 2004 to 2018, in Campo Grande, MS. For the data series, the shape and scale parameters of the distributions were determined to verify the quality of fit of the observation data, the Goodness-of-Fit Tests (GOF): Kolmogorov-Smirnov Test, Anderson -Darling Test, Chi-square Test tests were used to verify an optimal estimate for the hospitalization data hospital.

All PDFs are able to describe well the characteristics of hospitalizations. The results presented (total admissions), (summer) show that the functions Weibull 3P (W3P) and Inv Gausian 3P (IG3P); (fall) show that the functions Burr and Weibull 3P (W3P); (winter) shows that the Burr and Inv Gausian 3P (IG3P) and (spring) functions for lognormal and Rayleigh (2P) functions provided the best observed fit for hospital admissions.

Keywords

Hospital admission, Pneumonia, Modeling, Probability, Child and adults

Introduction

Several studies of adjustment of probability density distribution or probability estimates using theoretical models of probability in relation to a historical series of data have been developed, highlighting the benefits in the planning of activities that minimize the risks, among which can be cited: precipitation [1-18], air temperature [19-22], solar radiation [20], concentration of pollutant gases [23,24], for the historical series of hospital admissions for respiratory diseases there are no published works with this methodology.

The use of probability density functions is directly linked to the nature of the data to which they relate. Some have good estimation capacity for small numbers of data, others require a large number of observations. Due to the number of parameters of your equation, some can take different forms, being framed in a greater number of situations, that is, they are more flexible. Since respecting the aspect of data representativeness, the estimates of its parameters for a given region can be established as general purpose, without prejudice to the precision in the estimation of probability [4].

Climate change has become one of the most seri-

de Souza et al. J Infect Dis Epidemiol 2019, 5:098

• Page 1 of 12 •

ous environmental concerns for urban areas in recent decades. Several epidemiological studies in recent years have reported associations between high levels of climatic changes and increased rates of death and hospitalization for respiratory and cardiovascular diseases [25-31]. Some epidemiological studies show that air pollution affects human health, even concentrations of air pollutants are below the air quality standards [32-34].

Respiratory diseases and related mortality have been increasingly associated with exposure to climate change. Sensitive and vulnerable groups, such as pregnant women, children, the elderly, and those who already suffer from respiratory illnesses and other serious diseases, or from low-income groups, are especially affected by climatic variation. Studies have shown that the number of respiratory diseases in children and the elderly increases due to the higher concentrations of air pollution [35-41]. According to these studies, children are more susceptible because they need twice the amount of air inhaled by adults, and the elderly are more affected because of their weakened immune and respiratory systems and have been exposed to a large amount of air pollution in all your life.

In this study, we focused on determining the best statistical model that describes the number of hospitalizations for the city of Campo Grande.

Methodology

Data on the number of hospitalizations for respiratory diseases were used, referring to the period from 2004 to 2018 of Campo Grande-MS, whose geographic coordinates are: 20° 27S; 54° 37W; 530 m and an estimated population of 850,000 inhabitants.

Health data

For the correlation of the meteorological data with the aggravation of respiratory diseases, data of hospital admission were collected, together with the health departments of the Department of Informatics of the SUS - Unified Health System (DATASUS).

The data available came from the Hospital Information System of SUS (SIH/SUS), managed by the Ministry of Health, through the Health Care Secretariat, in conjunction with the State Health Secretariats and the Municipal Health Department, and processed DATASUS, of the Executive Secretariat of the Ministry of Health.

The hospital units, participants of the SUS (public or private parties), send the hospitalization information, made by the AIH-Hospital admission authorization, to the municipal managers (full management) or state (the others). This information is processed in DATASUS, generating the credits related to the services provided and forming a valuable database, which contains a large part of hospital admissions, carried out in Brazil.

It should be noted that the SIH/SUS collects variables related to hospitalizations: Identification and qualification of the patient, procedures, examinations and medical acts performed, diagnosis, reasons for discharge, amounts due, etc. Through the Internet, DATASUS provides the main information for tabulation on the databases of the SIH/SUS.

Study area

This is an ecological study of time series. This type of design is characterized by studying groups of individuals, generally by geographic regions. In the case of this work the site studied is the city of Campo Grande-MS.

The study population was made up of people living in the city of Campo Grande who were hospitalized due to diseases of the respiratory system from 2004 to 2018. We analyzed all hospitalizations with diagnosis of respiratory diseases from all hospitals, to the Unified Health System (SUS). These data are records of the Hospital Hospitalization Authorizations (AIH) of public and private hospitals and that serve the portion of the population that does not have private health plans, private or funded by companies (http://www.datasus.gov.br). The information in the database is: The number of the Taxpayer's Registry (CGC) of the hospital, the city where it is located, the age of the patient, sex, cause of hospitalization, procedure performed, patient's postal code, hospitalization, date of discharge or death, days of UTI stay, among other information. Among the information in the database were selected for this study the date of hospitalization, the diagnosis, the age of the patient.

Statistical Analysis

In this study, a descriptive analysis of the variables was performed, we used the Burr (Bu), Inv Gausian 3P (IG3P), Lognormal (LN), Pert (Pe), Rayleigh 2P (Ra 2P) and Weibull 3P (W3P) functions to model hospital admission data in Campo Grande. Performance indicators are calculated by comparing observed values to predicted values. The observed values are the classified values of the observation data, while the predicted values are the values obtained from the adjusted distribution.

Probability Distributions

In this study, the efficacy of six probability distributions of a component is evaluated. We use the parametric probability distribution functions (pdfs) of a component because our data have a unimodal distribution. These six models were selected among other models of a component because of their successful applications according to the literature.

Goodness-of-Fit Tests (GOF)

The GOF is used to determine the best model among the distributions tested in the characteristic of hospital-

izations for respiratory diseases. The goodness of fit test is performed to test the following hypothesis:

H0: The amount of monthly hospital admission data does not follow

H1: The amount of monthly hospital admission data does not follow a distribution

A couple of goodness-of-fit test have been conducted such as Kolmogorov-Smirnov test, Anderson-Darling test along with the chi-square test at significance level (α = 0.05) for choosing the best probability distribution [42].

Kolmogorov-smirnov test

The Kolmogorov-Smirnov test [43] is used to decide if a sample comes from a population with a specific distribution.

The Kolmogorov-Smirnov (K-S) test is based on the empirical distribution function (ECDF). Given N ordered data points Y_1, Y_2, ..., Y_N, the ECDF is defined as

$$E_N = \frac{n(i)}{N} \qquad (1)$$

Where n(i) is the number of points less than Y_i and the Y_i are ordered from smallest to largest value. This is a step function that increases by 1/N at the value of each ordered data point.

The Kolmogorov-Smirnov test statistic (D) is defined as

$$D = max_{1 \leq i << N}[F(Y_i)] - \frac{i-1}{N}, \frac{i}{N} - F(Y_i) \qquad (2)$$

Where F is the theoretical cumulative distribution of the distribution being tested which must be a continuous distribution (i.e., no discrete distributions such as the binomial or Poisson), and it must be fully specified (i.e., the location, scale, and shape parameters cannot be estimated from the data).

The hypothesis regarding the distributional form is rejected if the test statistic, D, is greater than the critical value obtained from a table.

Anderson-darling test

The Anderson-Darling test (Stephens, 1974) is used to test if a sample of data comes from a population with a specific distribution. It is a modification of the Kolmogorov-Smirnov (K-S) test and gives more weight to the tails than does the K-S test. The K-S test is distribution free in the sense that the critical values do not depend on the specific distribution being tested. The Anderson-Darling test makes use of the specific distribution in calculating critical values. This has the advantage of allowing a more sensitive test and the disadvantage that critical values must be calculated for each distribution. Currently, tables of critical values are available for the normal, lognormal, exponential, Weibull, extreme value type I, and logistic distributions.

The Anderson-Darling test statistic (A) is defined as

$$A^2 = -N - \frac{1}{N}\sum_{i=1}^{N}(2i-1)\big[lnF(X_i) + \ln(1 - F(X_{N-i+1}))\big] \qquad (3)$$

F is the cumulative distribution function of the specified distribution. Note that the Y_i are the ordered data. The critical values for the Anderson-Darling test are dependent on the specific distribution that is being tested. Tabulated values and formulas have been published [44] for a few specific distributions (normal, lognormal, exponential, Weibull, logistic, extreme value type I). The test is a one-sided test and the hypothesis that the distribution is of a specific form is rejected if the test statistic, A, is greater than the critical value.

Chi-square test

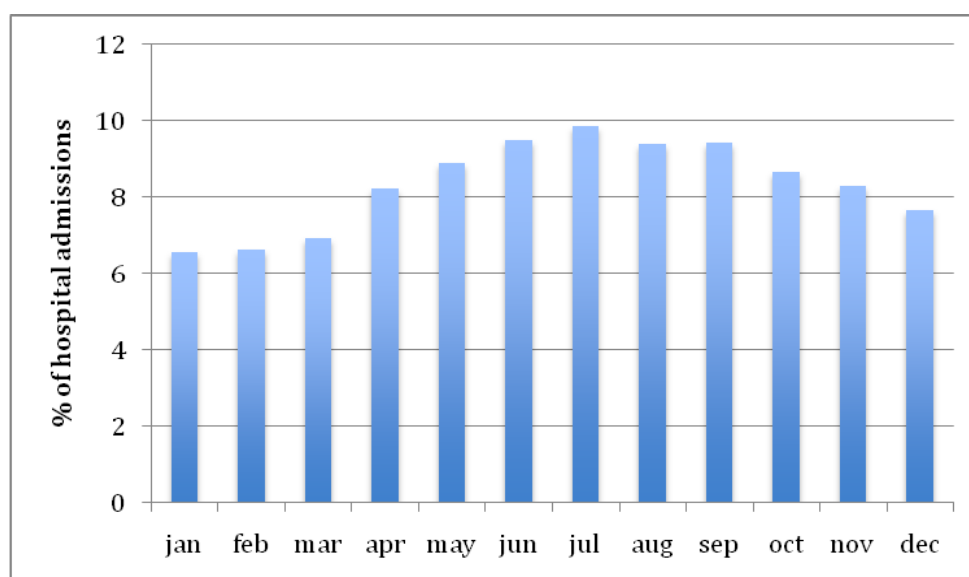The Chi-square test χ^2 assumes that the number of

Figure 1: Monthly percentage of respiratory disease morbidity (DAR) in the years 2004 to 2018.

observations is large enough so that the chi-square distribution provides a good approximation as the distribution of test statistic. The Chi-squared statistic is defined as

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)}{E_i} \qquad (4)$$

Where O_i is observed frequency, E_i is expected frequency, 'i' is number observations (1, 2, ……k), calculated by $E_i = F(X_2) - F(X_1)$, F is the CDF of the probability distribution being tested.

The observed number of observation (k) in interval 'i' is computed from equation given below

$$k = 1 + \log_2 n \qquad (5)$$

Where n is sample size.

This equation is for continuous sample data only and is used to determine if a sample comes from a population with a specific distribution [42].

Result and Discussion

Descriptions of hospitalizations of respiratory diseases

Figure 1 illustrates a typical pattern of hospital admissions (morbidity) for respiratory diseases, the average of the months of the years 2004 to 2018, as an example of a typical pattern.

During the study period (January 1, 2004 to December 31, 2018) the number of hospitalizations for respiratory diseases was 63,316, with an average of 4221 hospitalizations per month, with a maximum number of hospitalizations from May to October (Table 1).

Figure 1 shows the behavior of mean monthly admissions for respiratory diseases. According to the data, a seasonal pattern was observed between periods of rainfall, drought and transition, especially in the months (April, May, June, July, August and September), where the peak of hospitalizations corresponding to the dry season, low precipitation, relative humidity and minimum temperatures.

It is essential to recognize that social and economic factors play a significant role in predicting the change in the risk of infectious diseases caused by climate change [45-47]. Some populations and regions are more vulnerable to high risks because of

their inability to respond effectively to the tensions and challenges posed by climate change [48-50]. Vulnerability levels are partly a function of the programs and measures implemented to reduce the burden of climate-sensitive health determinants and outcomes, and partly to the success of traditional public health practices, including access to safe and better drinking water, sanitation and biosafety and surveillance programs to identify and respond to outbreaks of infectious diseases [48,51-53]. The vulnerability of a society to climate change induced by the risk to the health of infectious diseases is related to its social development. Many infectious diseases often occur in developing countries after tropical cyclones, but are rare in developed countries [54].

The vulnerability of a society to climate change induced by the risk to the health of infectious diseases is even more related to the public health system and to the existing infrastructure. Developing countries tend to be more sensitive to a high health risk posed by climate change because of the lack of resources and capacities for their public health system to effectively respond to the various challenges. Vulnerability to changes in the risks of infectious diseases can be reduced by appropriate adaptation measures. Adaptation can be effective in addressing the challenges posed by climate change. It is important to emphasize that the success of a proactive adaptation depends to a large extent on the correct prediction of the change from the health risk scenario to infectious diseases.

Adaptation measures may be informed by better weather forecasts, including prediction of extreme weather events and weather hazards. By developing an early warning system based on accurate weather forecasts, a society can better prepare for the health risks related to climate change.

Souza and Santos [55] calculated hospitalizations attributed to heat and cold, defined as temperatures above and below optimal temperature and for moderate and extreme temperatures, defined using cutoff point the 2.5th and 97.5th percentiles. They analyzed 148, 849 admissions in several periods. In total, 6.62% (95% IC-6.53-6.82) was due to the non-optimal temperature. The percentile minimum morbidity temperature ranged from approximately the 60th percentile. The temperatures attributable to hospitalization were

Table 1: Descriptive statistics of hospital admissions.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Mean	541	550	570	670	749	794	825	776	790	712	679	630
Stdev	56	55	58	73	88	68	106	94	74	74	51	57
Median	288	297	288	347	375	394	396	393	401	355	354	326
Minimum	127	177	188	251	224	269	253	284	254	282	276	246
Maimum	341	354	375	490	506	536	676	681	514	558	452	418
Skewness	-2	-1	0	0	0	0	1	2	0	1	0	0

de Souza et al. J Infect Dis Epidemiol 2019, 5:098

• Page 4 of 12 •

caused by cold (6.38%, 95% IC 6.04-6.58) rather than by heat (0.39%, 0.28-0.42).

Little evidence is available on the association between ozone exposure and health in Campo Grande, Brazil [56] examined the effects of surface ozone concentrations (O$_3$) on respiratory morbidity in Campo Grande, Brazil. A Poisson time series model was used to examine the effects of O$_3$ on hospital admissions, while controlling for seasonality, long-term trend, temperature, and relative humidity. A nonlinear distributed delay function was used for O$_3$, temperature and relative humidity. We examined the effects of O$_3$ in different age groups (0-4 years, 5-60 years and > 60 years). The relationship between ozone and respiratory morbidity was not linear, with a threshold of 13 ppb (less than 25% of the ozone distribution percentage). The relative risk of hospital admission at the 75th percentile of the O$_3$ distribution compared to 25% of the O$_3$ distribution percentile. The effect of O$_3$ on respiratory morbidity was delayed by two days and lasted 4 days for all age groups, except for people between 5 and 60 years. Children and the elderly were much more vulnerable to ozone pollution than people aged 5 to 60. The study suggests that ozone pollution has negative impacts on respiratory diseases in Campo Grande, Brazil. Children and the elderly were susceptible to exposure to O$_3$. These findings should be used to develop policies to protect people from O$_3$ pollution.

When comparing this pattern with the burned indexes, it is observed an increase in respiratory attendance due to the fires, between August, September and October that compose the dry period of the year and intensification of the fires and higher concentration of fires [55-57].

The use of different methods may lead to inconsistent or even contradictory results. For example, in a study investigating the climatic effects of respiratory diseases in the city of Campo Grande, temperature accounts for a large part of the variance in a multivariate Poisson regression model [55]. However, using the same data source, but based on an integrated autoregressive mobile time series model, another study reported no significant association of temperature and hospital admissions with respiratory diseases. We must realize that a better understanding of the mechanisms of interactions requires the integration of methods with detailed monitoring. However, this has been rare in existing literature.

Among the diseases cataloged in the International Coding of Diseases (ICD 9a and 10a revisions) as respiratory diseases (460-496 and J00-J99 respectively), it was verified that the highest daily averages of hospitalizations were due to influenza and pneumonia (480-487 and J10-J18) representing 52.3% of all admissions. Second, chronic diseases of the lower airways, such as chronic bronchitis, simple and mu-

cus-purulent bronchitis, emphysema, asthma, malaise asthmaticus and bronchiectasis (490-496 and J40-J47) with 19.3%. Third, with 11.3% and a daily average of 5.2, there were the other upper airway diseases (470-478 and J30-J39), such as allergic and vasomotor rhinitis, rhinitis, chronic nasopharyngitis and pharyngitis, chronic sinusitis, nasal polyp, other disorders of the nose and paranasal sinuses, chronic diseases of the tonsils and adenoids, chronic laryngitis and laryngotracheitis. Other acute lower respiratory infections (466 and J20-J22), such as acute bronchitis and bronchitis, which had a daily mean of 2.8 and accounted for 6% of all hospitalizations. Finally, acute upper respiratory infections (460-465 and J00-J06) accounted for 4.3% of hospital admissions, with a daily average of 2.0. These diseases were as follows: Acute nasopharyngitis (common cold), sinusitis, pharyngitis, tonsillitis, laryngitis, obstructive laryngitis, epiglottitis and tracheitis. The other diseases of the respiratory system did not reach 2.0 hospitalizations and had a very low percentage - except for the grouping of unspecified diseases (508 and J95-J99), which represented 4.9% of hospitalizations and a daily average of 2.2. However, this grouping, as its name implies, represents unspecified diseases and, therefore, it is not possible to know the actual cause of hospitalizations.

Probability distributions and their parameter estimation

The parameters of the estimates of the tested distributions are presented in Table 2, these parameters are obtained 'using the Matlab software. Figure 2 shows, respectively, the histogram of the hospital admissions of hospitalizations of the years 2004 to 2018, adjusted by the six probability density functions studied and their cumulative frequency adjusted by the six functions of cumulative distribution.

Total hospital admissions (DAR)

Table 2, Table 3, Figure 2.

Hospital admission for summer

Table 4, Table 5, Figure 3.

Hospital admission for autumn

Table 6, Table 7, Figure 4.

Hospital admission for winter

Table 8, Table 9, Figure 5.

Hospital admission for Spring

Table 10, Table 11, Figure 6.

The statistical parameters for assessing the suitability of the analyzed PDFs are presented in Table 3, Table 5, Table 7 and Table 9. All the PDFs are able to describe well the characteristics of the hospital-

Table 2: Estimated parameters for the distributions studied.

Distribution	Parameters		
Burr	k = 428.17	α = 8.046	β = 820.65
Inv. Gaussian	λ = 1.219 × 10^{-7}	μ = 3258.1	γ = 2894.2
Lognormal	σ = 0.1492	μ = 5.8848	
Pert	m = 369.21	a = 279	b = 440.0
Rayleigh (2P)	σ = 82.126	γ = 260.17	
Weibull (3P)	α = 2.4349	β = 134.25	γ = 244.73

Table 3: Criteria for quality adjustment of the historical series of total hospital stays (2004-2018) for respiratory diseases, for the six distribution models Probability using different fit quality tests.

#	Distribution	Kolmogorov Smirnov		Anderson Darling		Chi-Squared	
		Statistic	Rank	Statistic	Rank	Statistic	Rank
1	Burr	0.2121	3	4.1693	5	N/A	
2	Inv. Gaussian	0.16038	2	0.24219	1	0.23349	1
3	Lognormal	0.27767	4	0.81603	2	2.4778	2
4	Pert	0.33265	6	4.4897	6	4.0	4
5	Rayleigh (2P)	0.31562	5	1.0689	3	3.6964	3
6	Weibull (3P)	0.12357	1	3.8219	4	N/A	

Table 4: Estimated parameters for the distributions studied.

Distribution	Parameters		
Burr	k = 578.76	α = 6.1518	β = 820.65
Inv. Gaussian	λ = 6.019 × 10^8	μ = 12555	γ = -12272
Lognormal	σ = 0.22941	μ = 5.6214	
Pert	m = 298.93	a = 101.6	b = 375.0
Rayleigh (2P)	σ = 121.04	γ = 121.49	
Weibull (3P)	α = 19.98	β = 940.55	γ = -632.40

Table 5: Criteria for quality adjustment of the historical series of hospital stays for the summer (2004-2018) for respiratory diseases, for the six distribution models Probability using different fit quality tests.

#	Distribution	Kolmogorov Smirnov		Anderson Darling		Chi-Squared	
		Statistic	Rank	Statistic	Rank	Statistic	Rank
1	Burr	0.06679	1	0.36242	2	1.6316	1
2	Inv. Gaussian	0.10144	4	0.69138	3	2.5386	4
3	Lognormal	0.14763	5	1.5347	4	5.6829	5
4	Pert	0.09632	3	1.8946	5	2.0647	3
5	Rayleigh (2P)	0.2385	6	3.2002	6	19.807	6
6	Weibull (3P)	0.07028	2	0.23247	1	1.7043	2

Table 6: Estimated parameters for the distributions studied.

Distribution	Parameters		
Burr	k = 578.76	α = 6.1518	β = 858.03
Inv. Gaussian	λ = 6.019 × 10^8	μ = 12555	γ = -12272
Lognormal	σ = 0.22941	μ = 5.6214	
Pert	m = 298.93	a = 101.6	b = 375.0
Rayleigh (2P)	σ = 121.04	γ = 121.49	
Weibull (3P)	α = 19.98	β = 940.55	γ = -632.40

izations. The results presented in Table 3 (total hospitalizations) and Table 5 (summer) clearly show that the functions weibull 3P (W3P) and Inv Gausian 3P (IG3P), provided the best adjustment of hospital admissions observed. The results presented in Table 7 (autumn) clearly show that the Burr and Weibull 3P

de Souza et al. J Infect Dis Epidemiol 2019, 5:098

• Page 6 of 12 •

Table 7: Criteria for quality adjustment of the historical series of hospital stays for the autumn (2004-2018) for respiratory diseases, for the six distribution models Probability using different fit quality tests.

#	Distribution	Kolmogorov Smirnov		Anderson Darling		Chi-Squared	
		Statistic	Rank	Statistic	Rank	Statistic	Rank
1	Burr	0.09225	1	0.34523	1	1.1533	4
2	Inv. Gaussian	0.9304	2	0.38907	2	2.2031	5
3	Lognormal	0.11172	5	0.54699	5	0.85595	3
4	Pert	0.14483	6	4.7137	6	10.067	6
5	Rayleigh (2P)	0.11042	4	0.5268	4	0.63362	1
6	Weibull (3P)	0.09938	3	0.39328	3	0.6789	2

Table 8: Estimated parameters for the distributions studied.

Distribution	Parameters		
Burr	k = 1.0239	α = 7.8603	β = 394.71
Inv. Gaussian	λ = 3653.4	μ = 315.87	γ = 87.305
Lognormal	σ = 0.22268	μ = 5.9743	
Pert	m = 371.27	a = 253	b = 681.0
Rayleigh (2P)	σ = 138.29	γ = 231.02	
Weibull (3P)	α = 1.8149	β = 184.29	γ = 239.0

Table 9: Criteria for quality adjustment of the historical series of hospital stays for the **Winter** (2004-2018) for respiratory diseases, for the six distribution models Probability using different fit quality tests.

#	Distribution	Kolmogorov Smirnov		Anderson Darling		Chi-Squared	
		Statistic	Rank	Statistic	Rank	Statistic	Rank
1	Burr	0.06727	2	0.1944	1	0.98542	2
2	Inv. Gaussian	0.08153	5	0.25861	3	1.32242	5
3	Lognormal	0.06784	3	0.2251	2	0.77493	1
4	Pert	0.07553	4	3.3263	6	1.0	3
5	Rayleigh (2P)	0.6724	1	0.30086	4	1.278	4
6	Weibull (3P)	0.08364	6	0.345	5	2.9276	6

Table 10: Estimated parameters for the distributions studied.

Distribution	Parameters		
Burr	k = 1.0559	α = 9.5207	β = 343.54
Inv. Gaussian	λ = 2392.6	μ = 215.42	γ = 131.07
Lognormal	σ = 0.17914	μ = 5.8316	
Pert	m = 318.73	a = 246	b = 558.0
Rayleigh (2P)	σ = 94.51	γ = 229.05	
Weibull (3P)	α = 1.7257	β = 121.63	γ = 237.76

Table 11: Criteria for quality adjustment of the historical series of hospital stays for the **Spring** (2004-2018) for respiratory diseases, for the six distribution models Probability using different fit quality tests.

#	Distribution	Kolmogorov Smirnov		Anderson Darling		Chi-Squared	
		Statistic	Rank	Statistic	Rank	Statistic	Rank
1	Burr	0.09711	3	0.4032	3	0.50776	1
2	Inv. Gaussian	0.11332	5	0.4303	4	0.98539	3
3	Lognormal	0.09179	1	0.36903	1	0.52826	2
4	Pert	0.09762	4	27314	6	1.4667	6
5	Rayleigh (2P)	0.09242	2	0.40188	2	1.0569	4
6	Weibull (3P)	0.11462	6	0.47038	5	1.2634	5

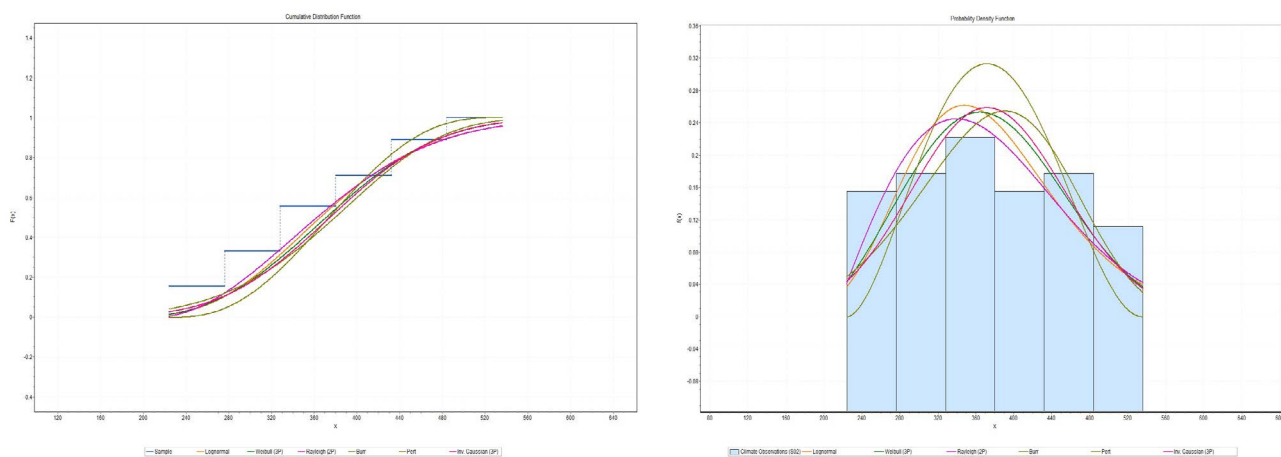de Souza et al. J Infect Dis Epidemiol 2019, 5:098

• Page 7 of 12 •

Figure 4: Graphs cdf (left) and pdf (right) of the distributions obtained for the monthly averages of hospital admissions for **autumn** (2004-2018).
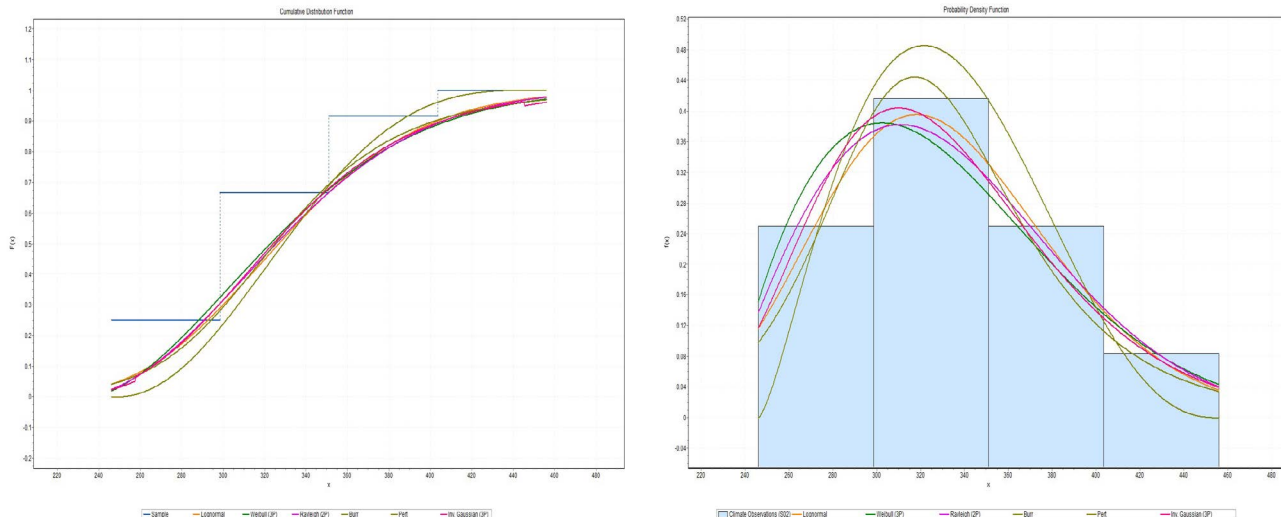

Figure 2: Graphs cdf (left) and pdf (right) of the distributions obtained for the monthly averages of hospital admissions for the years (2004-2018).
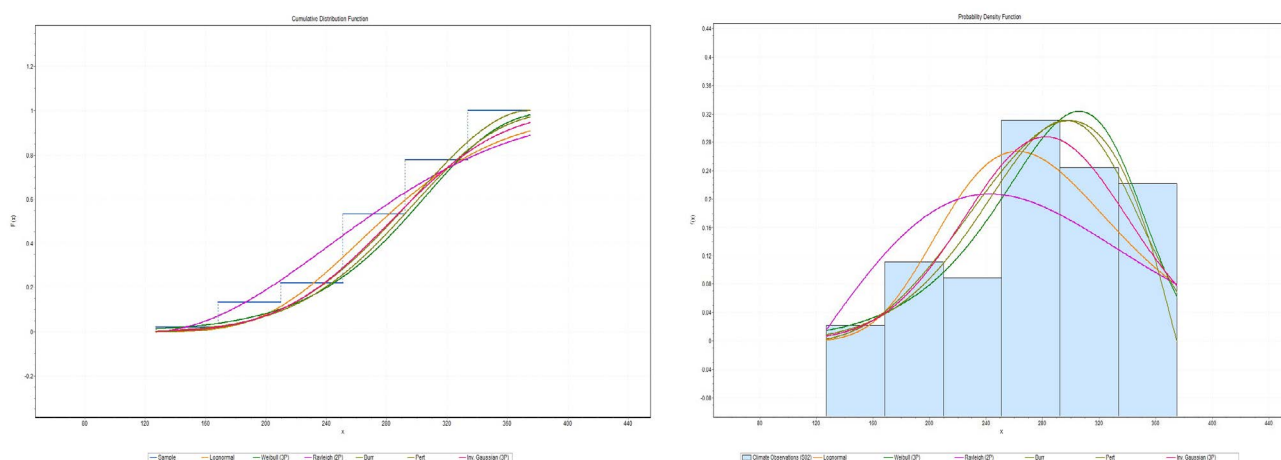

Figure 3: Graphs cdf (left) and pdf (right) of the distributions obtained for the monthly averages of hospital admissions for **summer** (2004-2015).

(W3P) functions provided the best hospital adjustment observed, Table 9 (winter) clearly shows that the Burr and Inv Gausian 3P (IG3P) functions, provided the best adjustment for hospital admissions observed. Table 9 shows that the Rayleigh (2P) and Burr functions provided the best hospital adjustment

de Souza et al. J Infect Dis Epidemiol 2019, 5:098

• Page 8 of 12 •

Figure 5: Graphs cdf (left) and pdf (right) of the distributions obtained for the monthly averages of hospital admissions for **Winter** (2004-2018).

Figure 6: Graphs cdf (left) and pdf (right) of the distributions obtained for the monthly averages of hospital admissions for **Spring** (2004-2018).
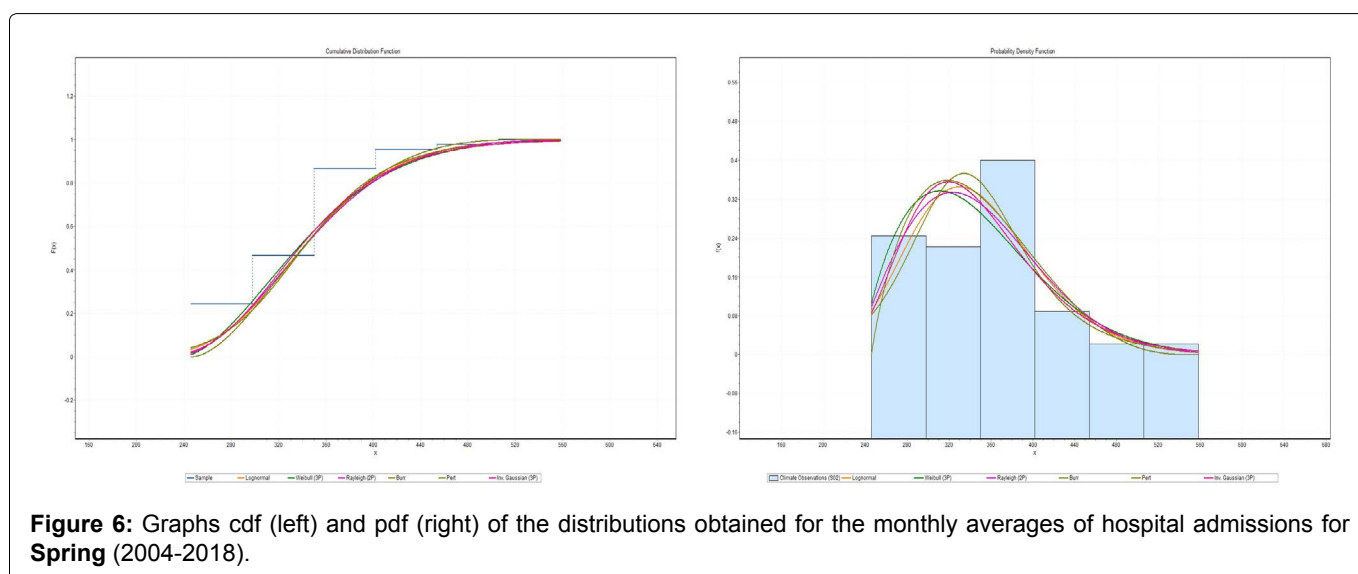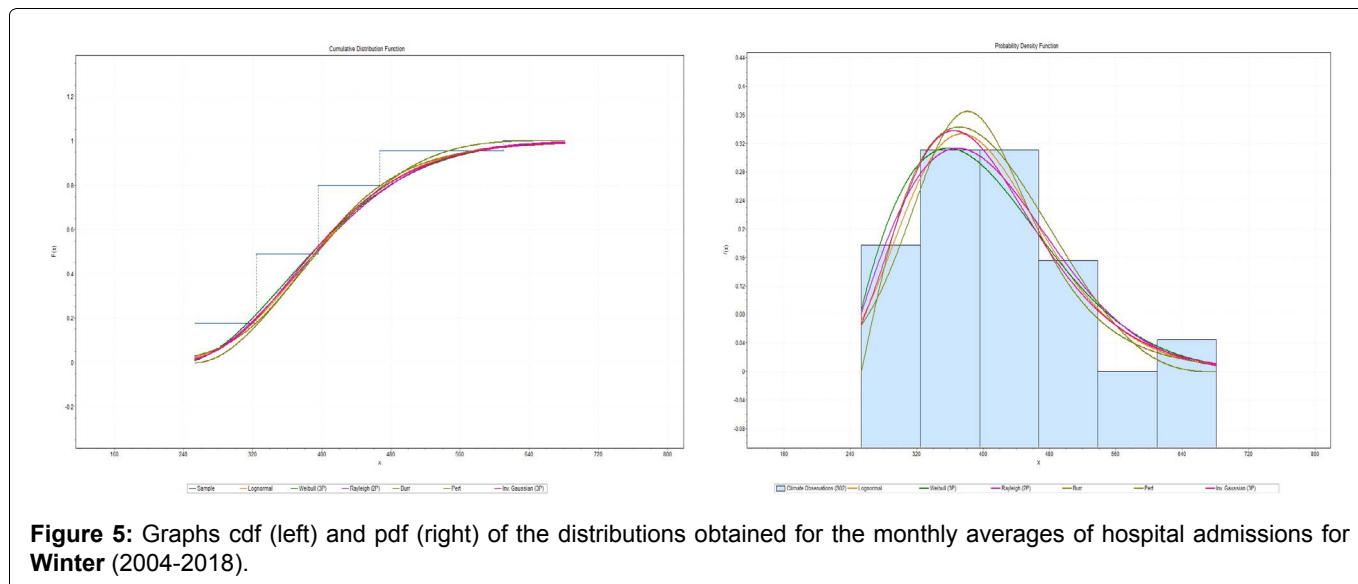
observed, Table 11 (spring) for the lognormal and Rayleigh (2P) functions, provided the best adjustment of the hospital admissions observed.

In Figure 2, Figure 3, Figure 4, Figure 5 and Figure 6, we present the time series of hospital stays for the years 2004 to 2018. There is no clear variation or trend around an average value. To extract more information about the behavior of the series, Table 1 shows the calculations of mean, standard deviation, variance, asymmetry for the mean of the studied years. It is quite evident the lack of homogeneity in the data of each time series, because in all cases, it is observed that the standard deviation presents average amplitude and a very high variance. These facts are corroborated by the fact that these measurements are highly sensitive to atmospheric conditions, sudden physical changes at measurement sites, and other changes that make the series very heterogeneous and unpredictable in the long run. With regard to asymmetry, it is possible to notice a situation of asymmetry in all cases, with right or positive displacement.

The adjustment of the six functions of probability

density was through the Kolmogorov-Smirnov test, Anderson Darling and Chi Squared. Table 2, Table 3, Table 4, Table 5, Table 6, Table 7, Table 8, Table 9, Table 10 and Table 11 presents the test results for the probability density functions that best fit the data as well as the parameter values of the models. In this table it is possible to compare the adjustment is approved according to K-S criteria. Approved functions are highlighted in the table. It is also possible to observe that all the adjustments approved by the tests presented very low values for the maximum quadratic error, corroborating, therefore, adjustments as being the best ones for the respective series.

Finally, the Figures shows the comparison of the adjustments of the best probability density functions with the respective original series (observed). By means of these figures it is possible to note again the trend with positive displacement of both the fitted models and the observed data. In addition, the small differences between the observed profile and the adjusted functions are more clearly demonstrated, because in this reconstruction, different from what happens with the accu-

de Souza et al. J Infect Dis Epidemiol 2019, 5:098

• Page 9 of 12 •

mulated frequency, the errors are no longer attenuated by the accumulation of frequencies, so that discrete discrepancies become, if clearer. Such discrepancies do not invalidate the results, since the representativeness presented by the models is larger than expected given the heterogeneity of the series.

Conclusion

The results show that all temporary series studied are able to describe hospitalizations well. The results (total admissions), (summer) show the functions Weibull 3P (W3P) and Inv Gausian 3P (IG3P); fall functions Burr and Weibull 3P (W3P); (winter) shows how Burr and Inv Gausian 3P (IG3P) and (spring) functions for lognormal and Rayleigh (2P) functions provide the best observed fit for hospital admissions.

Through the study, it can be concluded that all time series studied presented positive displacement asymmetry. Regarding the adjustment of the best probability distribution model, it was found that for each season a set of PDFs is satisfactorily adjusted.

Authors' Contributions

All authors performed the data analysis, applied the model, created the figures and prepared the original draft, analyzed the raw data and revised the original draft, helped to collect the raw data from the monitoring station.

Highlights

Climate change impacts human infectious disease via pathogen, host and transmission.

Go beyond empirical observation of association between climate and health effect.

Improve prediction of associated shifts in infectious diseases at various scales.

The health impacts may be controlled through adopting certain proactive measures.

Establish local early warning system for health effect of predicated climate change.

Database Statement/Availability of Data

The meteorological database is public domain and is available at: Center for Monitoring Weather, Climate and Water Resources of Mato Grosso do Sul (Cemtec/MS), an agency linked to the State Secretariat of Environment, Economic Development, Production and Family Agriculture (Semagro), http://www.cemtec.ms.gov.br/laudos-meteorologicos/.

The ozone pollutant database belongs to the physics institute of the federal university of mato grosso do sul and may be requested from Prof Dr Amaury de Souza, e-mail: amaury.souza@ufms.br.

Funding

References

1. Ávila AMH, Berlato MA, Silva JB, Fontana DC (1996) Probabilidade de ocorrência de precipitação pluvial mensal igual ou maior que a evapotranspiração potencial para a estação de crescimento das culturas de primavera-verão no Estado do Rio Grande do Sul. Pesquisa Agropecuária Gaúcha 2: 149-154.

2. Castro R (1996) Distribuição probabilística de precipitação na região de Botucatu - SP. Universidade Estadual Paulista, Botucatu.

3. Botelho VA, Morais AR (1999) Estimativas dos parâmetros da distribuição gama de dados pluviométricos do município de Lavras, Estado de Minas Gerais. Ciências e Agrotecnologia 23: 697-706.

4. Catalunha MJ, Sediyama GC, Leal BG, Soares CPB, Ribeiro A (2002) Aplicação de cinco funções densidade de probabilidade a séries de precipitação pluvial no Estado de Minas Gerais. Revista Brasileira de Agrometeorologia 10: 153-162.

5. Dourado Neto D, Assis JP, Timm LC, Manfron PA, Sparovek G, et al. (2005) Ajuste de modelos de distribuição de probabilidade a séries históricas de precipitação pluvial diária em Piracicaba-SP. Revista Brasileira de Agrometeorologia 13: 273-283.

6. Junqueira Júnior JA, Gomes NM, Mello CR de, Silva AM (2007) Precipitação provável para a região de Madre de Deus, Alto Rio Grande: Modelos de probabilidades e valores característicos. Ciência e Agrotecnologia 31: 842-850.

7. Sampaio SC, Queiroz MMF, Frigo E, Longo AJ, Suszek M (2007) Estimativa e distribuição de precipitações decendiais para o Estado do Paraná. Irriga 12: 38-53.

8. Silva JC, Heldwein AB, Martins FB, Trentin G, Grimm EL (2007) Análise de distribuição de chuva para Santa Maria, RS. Revista Brasileira de Engenharia Agrícola e Ambiental 11: 67-72.

9. Lima JSS, Silva SA, Oliveira RB, Cecílio RA, Xavier AC (2008) Variabilidade temporal da precipitação mensal em Alegre - ES. Revista

10. Martin TN, Dourado Neto D, Vieira Junior PA, Manfron PA (2008) Homogeneidade espaço-temporal e modelos de distribuição para a precipitação pluvial no estado de São Paulo. Revista Ceres 55: 476-481.

11. Oliveira LFC, Antonini JCA, Fioreze AP, Silva MAS (2008) Métodos de estimativa de precipitação máxima para o Estado de Goiás. Revista Brasileira de Engenharia Agrícola e Ambiental 12: 620-625.

12. Ávila LF, Mello CR, Viola MR (2009) Mapeamento da precipitação mínima provável para o sul de Minas Gerais. Revista Brasileira de Engenharia Agrícola e Ambiental 13: 906-915.

13. Mello CR, Silva AM (2009) Modelagem estatística da precipitação mensal e anual e no período seco para o estado de Minas Gerais. Engenharia Agrícola e Ambiental 13: 68-74.

de Souza et al. J Infect Dis Epidemiol 2019, 5:098

• Page 10 of 12 •

14. Moreira PSP, Dallacort R, Magalhães RA, Inoue MH, Stieler MC, et al. (2010) Distribuição e probabilidade de ocorrência de chuvas no município de nova Maringá-MT. Revista de Ciências Agro-Ambientais 8: 9-20.

15. Dallacort R, Martins JA, Inoue MH, Freitas PSL, Coletti AJ (2011) Distribuição das chuvas no município de Tangará da Serra, médio norte do Estado de Mato Grosso, Brasil. Acta Scientiarum Agronomy 33: 193-200.

16. Hartmann M, Moala FA, Mendonça MA (2011) Estudo das precipitações máximas anuais em presidente prudente. Revista Brasileira de Meteorologia 26: 561-568.

17. Rodrigues JA, Silva APCM, Santos Filho J (2014) Uso de distribuições de probabilidade na modelagem de intensidade de secas ocorridas em Laranjeiras do Sul, PR. Revista Brasileira de Biometria 32: 570-583.

18. Kist A, Virgens Filho JS, Jorim Sousa das (2015) Análise probabilística da distribuição de dados diários de chuva no estado do Paraná. Revista Ambiente Água 10: 172-181.

19. Assis JP, Sousa RP, Batista BDO, Linhares PCF (2018) Probabilidade de chuva em Piracicaba, SP, através da distribuição densidade de probabilidade Gama. Revista Brasileira de Geografia Física 11: 814-825.

20. Assis JP, Neto DD, Manfron PA, Martin TN, Sparovek G, et al. (2004) Ajuste de séries históricas de temperatura e radiação solar global diária às funções densidade de probabilidade normal e log-normal, em Piracicaba, SP. Revista Brasileira de Agrometeorologia 12: 113-121.

21. Sansigolo CA (2008) Distribuições de extremos de precipitação diária, temperatura máxima e mínima e velocidade do vento em Piracicaba, SP (1917-2006). Revista Brasileira de Meteorologia 23: 341-346.

22. Araújo EM, Silva IN, Oliveira JB, Cavalcante Junior EG, Almeida BM (2010) Aplicação de seis distribuições de probabilidade a séries de temperatura máxima em Iguatú-Ceará. Revista Ciência Agronômica 41: 36-45.

23. Souza Amaury de, Oliveira Soetânia S. de, Aristone Flavio, Olaofe Zaccheus, Kodicherla Shiva Prashanth Kumar, et al. (2018) Modeling of the function of the ozone concentration distribution of surface to urban areas. European Chemical Bulletin 7: 98-105.

24. Souza Amaury de, Olaofe Zaccheus, Kodicherla Shiva Prashanth Kumar, Ikefuti Priscilla, Nobrega Luciana, et al. (2018) Probability distributions assessment for modeling gas concentration in Campo Grande, MS, Brazil. European Chemical Bulletin 6: 569.

25. Samet JM, Zeger SL, Dominici F, Curriero F, Coursac I, et al. (2000) The national morbidity, mortality, and air pollution study. Part II: Morbidity and mortality from air pollution in the United States. Res Rep Health Eff Inst 94: 5-70.

26. Atkinson RW, Anderson HR, Sunyer J, Ayres J, Baccini M, et al. (2001) Acute effects of particulate air pollution on respiratory admissions: Results from APHEA 2 project. Am J Respir Crit Care Med 164: 1860-1866.

27. Katsouyanni K, Touloumi G, Samoli E, Gryparis A, Le Tertre A, et al. (2001) Confounding and effect modification in the short-term effects of ambient particles on total mortality: Results from 29 European cities within the APHEA2 Project. Epidemiology 12: 521-531.

28. Dominici F, Peng RD, Bell ML, Pham L, McDermott A, et al. (2006) Fine particulate air pollution and hospital admission for cardiovascular and respiratory diseases. J Am Stat Assoc 295: 1127-1134.

29. Samoli E, Peng R, Ramsay T, Pipikou M, Touloumi G, et al. (2008) Acute effects of ambient particulate matter on mortality in Europe and North America: Results from the APHENA study. Environ Health Perspect 116: 1480-1486.

30. Mahiyuddin WRW, Sahani M, Aripin R, Latif MT, Thach T-Q, et al. (2013) Short-term effects of daily air pollution on mortality. Atmos Environ 65: 69-79.

31. Li Y, Henze DK, Jack D, Henderson BH, Kinney PL (2016) Assessing public health burden associated with exposure to ambient black carbon in the United States. Sci Total Environ 539: 515-525.

32. Souza A, Fernandes WA, Pavao H, Lastoria G, Albrez EA (2012) Potential impacts of climate variability on respiratory morbidity in children, infants, and adults. J Bras Pneumol 38: 708-715.

33. Peters A, Dockery DW, Muller JE, Mittleman MA (2001) Increased particulate air pollution and the triggering of myocardial infarction. Circulation 103: 2810-2815.

34. Peng RD, Dominici F, Louis TA (2006) Model choice in time series studies of air pollution and mortality. J R Stat Soc Ser 169: 179-203.

35. Souza Amaury de, Andrade Fernanda A, Oguntunde Pelumi E, Arsic Milica, Santos Débora Aparecida da Silva (2018) Climate indicators and the impact on morbidity and mortality of acute respiratory infections. Advanced Studies In Medical Sciences 6: 5-20.

36. Souza Amaury de, Aristone Flavio, Andrade Fernanda A, Pavao Hamilton G, Fernandes Widinei A, et al. (2018) The effects of climate on hospitalization for respiratory diseases by age group. Advanced Studies In Medical Sciences 6: 21-35.

37. Souza Amaury de, Santos Débora Aparecida da Silva, Ikefuti Priscilla Venâncio (2017) Association between climate variables, pollutants, aerosols and hospitalizations due to asthma. O Mundo Da Saúde 41: 359-367.

38. Santos DAS, Azevedo Pedro Vieira de, Olinda Ricardo Alves de, Santos Carlos Antonio Costa dos, Souza Amaury de, et al. (2017) A relação das variáveis climáticas na prevalência de infecção respiratória aguda em crianças menores de dois anos em Rondonópolis-MT, Brasil. Ciencia & Saude Coletiva 22: 3711-3722.

39. Braga ALF, Conceiçao GMS, Pereira LAA, Kishi HS, Pereira JCR, et al. (1999) Air pollution and pediatric respiratory hospital admissions in Sao Paulo, Brazil. J Occup Environ Med 1: 95-102.

40. Braga AL, Saldiva PH, Pereira LA, Menezes JJ, Conceiçao GM, et al. (2001) Health effects of air pollution exposure on children and adolescents in Sao Paulo, Brazil. Pediatr Pulmonol 31: 106-113.

41. Viegi G, Maio S, Simoni M, Baldacci S, Annesi-Maesano I (2009) The epidemiological link between ageing and respiratory diseases. European respiratory monograph: Respiratory diseases in the elderly. Eur Respir Soc 43: 1-17.

42. Sharma MA, Singh JB (2010) Use of probability distribution in rainfall analysis. NY Sci J 3: 40-49.

43. Chakravarti, Laha, Roy (1967) Hand book methods of applied statistics. John Wiley and Sons, 1: 11-27.

44. Stephens MA (1974) EDF Statistics for goodness of fit and some comparisons. Journal of the American Statistical Association 69: 730-737.

45. Semenza JC, Menne B (2009) Climate change and infectious diseases in Europe. Lancet Infect Dis 9: 365-375.

de Souza et al. J Infect Dis Epidemiol 2019, 5:098

• Page 11 of 12 •

46. Wu XX, Tian HY, Zhou S, Chen LF, Xu B (2014) Impact of global change on transmission of human infectious diseases. Sci China Earth Sci 57: 189-203.

47. Xu B, Jin ZY, Jiang ZB, Guo JP, Timberlake M (2016) Climatological and geographical impacts on global pandemic of influenza A (H1N1). Environment International 86: 14-23.

48. Li R, Jiang Z, Xu B (2014) Global spatiotemporal and genetic footprint of the H5N1 avian influenza vírus. Int J Health Geogr 13: 14.

49. Wang Y, Jiang Z, Jin Z, Tan H, Xu B (2013) Risk factors for infectious diseases in backyard poultry farms in the Poyang lake area, China. PLoS One 8: e67366.

50. Wei T, Yang S, Moore JC, Shi PJ, Cui XF, et al. (2012) Developed and developing world responsibilities for historical climate change and CO2 mitigation. Proc Natl Acad Sci 109: 12911-12915.

51. Bai L, Cirendunzhu, Woodward A, Dawa, Xiraoruodeng, et al. (2014) Temperature and mortality on the roof of the world: A time-series analysis in three Tibetan counties. Sci Total Environ 485: 41-48.

52. U Confalonieri, B Menne, R Akhtar, KL Ebi, M Hauengue, et al. (2007) Human health. ML Parry, OF Canziani, JP Palutikof, PJ van der Linden, CE Hanson, Human Health Climate Change 2007: Impacts, adaptation and vulnerability. Intergovernmental Panel on Climate Change, Cambridge, UK.

53. Jiang Q, Zhou J, Jiang Z, Xu B (2014) Identifying risk factors of avian infectious disease at household level in Poyang Lake region, China. Prev Vet Med 116: 151-160.

54. Guill CK, Shandera WX (2001) The effects of Hurricane Mitch on a community in northern Honduras. Pre hosp Disaster Med 16: 166-171.

55. Souza, A, Santos DAS (2016) Temperatura como um fator de risco para internações hospitalares em 70 cidades do ms. Ciência E Natura 38: 1512.

56. Souza A, Guo Y, Pavão HG, Fernandes WA (2014) Effects of air pollution on disease respiratory: Structures lag. Health 6: 1333-1339.

57. Souza A, Ozonur D (2019) Statistical behavior of O_3, O_x, NO, NO_2, and NO_x in urban environment. Ozone-Science & Engineering 1: 1-13.

de Souza et al. J Infect Dis Epidemiol 2019, 5:098

• Page 12 of 12 •