

# Using the Zero Inflated Negative Binomial Model to assess Factors Associated with Occurrence of Childhood Pneumonia

Oyindamola B. Yusuf<sup>1</sup>, Onoja M. Akpa<sup>1\*</sup>, Kehinde P. Akinpelu<sup>2</sup>, Abass O. Gbolahan<sup>3</sup>

<sup>1</sup>Department of Epidemiology and Medical Statistics, Faculty of Public Health, College of Medicine, University of Ibadan, Nigeria

<sup>2</sup>Infectious Disease Institute, College of Medicine, University of Ibadan, Ibadan, Nigeria

<sup>3</sup>Department of Planning, Research and Statistics, Oyo State Ministry of Health, Nigeria

**Abstract** The present study was intended to model the number of cases of childhood pneumonia using zero inflated negative binomial (ZINB) regression which accounts for both excess zeros and over dispersion. Data on childhood pneumonia was obtained from the Integrated Disease Surveillance and Response (IDSR) of the State Ministry of Health for the period: 2010 - 2014. Over-dispersion was ascertained using the mean and variance. Incidence per year was determined. The ZINB model was fitted, and incidence rate ratios with their 95% CI were determined. Goodness of fit was assessed using the Vuong z-statistic criteria. There was an initial rise and a subsequent decline in the incidence of childhood pneumonia cases; 7 per 1000 in 2012 and 0.5 per 1000 in 2014. The risk of childhood pneumonia was lowest in 2011 (IRR = 0.24, 95% CI: 0.15, 0.38) with a slight increase in 2014 (IRR= 0.17, 95% CI: 0.09, 0.32). The Vuong z-statistic was -7.952 (i.e.  $V < -1.96$ ). The zero inflated negative binomial model yielded optimal results in modeling the number of cases of childhood pneumonia when there is an indication of over dispersion and excess zeros.

**Keywords** Over-dispersion, Childhood Pneumonia, Zero inflated negative binomial regression

## 1. Introduction

In the analysis of count data, there are situations where there are an excessive number of individuals with a count of zero. The poisson model generally known for count data may not be appropriate because it assumes that the conditional variance of the dependent variable is equal to the conditional mean and so it will give a poor fit. Zero inflated models have been developed to accommodate for this scenario. The zero inflated poisson model (ZIP) allows for overdispersion but does not cater for these excess zeros. However, the zero inflated negative binomial model (ZINB) fits better because it allows for overdispersion and also excessive zeros.

The rationale behind zero-inflated count modelling is that real-life data are usually prone to over-dispersion and excess zeros and they adjust well to data from a particular mixture of two populations; one that has only zero counts and another in which the counts are the realizations of a discrete distribution [1]. The ZINB model deals with both over-dispersed and zero inflated data by assuming there are

two distinct data generating processes. The result of a Bernoulli trial is used to determine which of the two processes is used. The ZINB regression model has been applied to human microbiota sequence data with random effects. The results revealed that the simulation study used indicated that the method of estimation gave unbiased results for both fixed effects and random effects [2]. The theoretical distribution of ZIP and ZINB has been extensively reviewed and reported by several authors [2-4].

Pneumonia being the most common clinical manifestation of pneumococcal infection remains a childhood disease in developing countries such as Nigeria, with high morbidity rates [5] Pneumonia deaths is also the commonest cause of under-5 mortality in Nigeria. The disease morbidity is usually presented as count data such as number of cases with the disease. The methods of analysis of such count data may however pose a problem when extra zeros in the count dataset remain unaccounted for. This may subsequently result in biased parameter estimates, low power, loss of information to detect a covariate effect and ultimately wrong inferences. Hence, the use of ZINB models to analyze such data which helps to overcome the problems of underestimation of parameters, account for over-dispersion and improve inferences. Therefore we determined factors associated with number of cases of Pneumonia in south west Nigeria.

\* Corresponding author:

onojamatthew@gmail.com (Onoja M. Akpa)

Published online at <http://journal.sapub.org/ijps>

Copyright © 2018 Scientific & Academic Publishing. All Rights Reserved



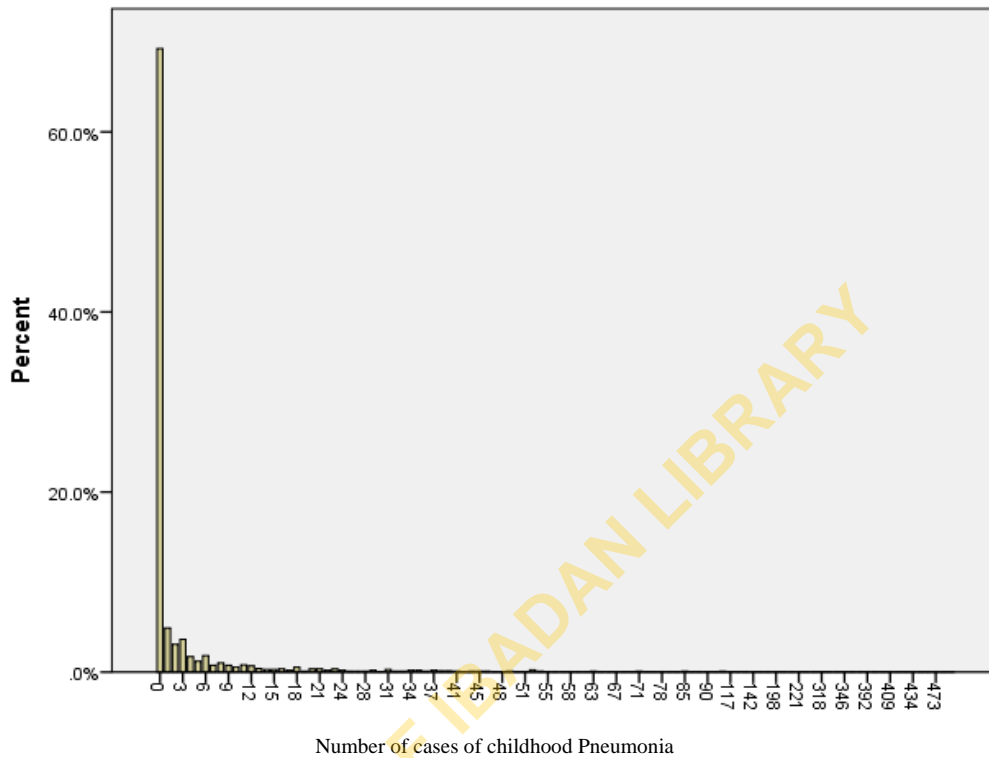
Intercept	0.502	0.274	0.102	0.733	<0.001	0.597	1.155	0.358	3.721	0.809
2010 (Year 1)*										
2011 (Year 2)	0.236	0.238	0.150	0.379	<0.001	0.048	1.537	1.399	1.689	<0.001
2012 (Year 3)	0.238	0.218	0.136	0.347	<0.001	0.046	2.255	2.060	2.467	<0.001
2013 (Year 4)	0.249	0.209	0.128	0.341	<0.001	0.049	2.303	2.092	2.535	<0.001
2014 (Year 5)	0.321	0.173	0.092	0.324	<0.001	0.070	0.915	0.798	1.049	0.205
April *										
August	0.202	1.549	1.043	2.302	0.030	0.038	0.843	0.782	0.908	<0.001
Dec	0.257	0.872	0.527	1.443	0.595	0.042	1.114	1.026	1.210	0.010
Feb	0.205	0.769	0.515	1.150	0.200	0.038	0.824	0.765	0.887	<0.001
Jan	0.201	0.874	0.589	1.296	0.501	0.039	0.731	0.677	0.789	<0.001
July	0.207	1.751	1.167	2.627	0.007	0.034	1.289	1.206	1.378	<0.001
June	0.197	1.456	0.990	2.143	0.056	0.035	1.212	1.131	1.298	<0.001
March	0.198	1.009	0.684	1.487	0.963	0.039	0.894	0.828	0.965	0.004
May	0.207	1.155	0.770	1.733	0.485	0.036	1.016	0.947	1.090	0.655
Nov	0.237	1.428	0.897	2.272	0.133	0.041	1.131	1.044	1.226	0.003
Oct	0.275	1.366	0.797	2.342	0.257	0.048	0.937	0.853	1.030	0.173
Sept	0.225	1.467	0.944	2.280	0.088	0.040	1.003	0.927	1.085	0.935
Afijio *										
Akinyele	0.548	3.258	1.113	9.536	<0.001	0.620	1.262	0.374	4.255	0.706
Atiba	0.590	3.979	1.252	12.647	<0.001	0.610	3.068	0.928	10.141	0.066
Atisbo	0.537	5.496	1.918	15.745	<0.001	0.610	1.931	0.584	6.383	0.281
Egbeda	0.532	4.195	1.479	11.902	<0.001	0.601	9.915	3.053	32.200	<0.001
Ibarapa central	0.663	4.473	1.22	16.403	<0.001	0.687	0.666	0.173	2.559	0.554
Ibadan North	0.532	1.430	0.504	4.058	0.101	0.605	2.547	0.778	8.338	0.122
Ibadan North East	0.516	4.455	1.620	12.248	<0.001	0.600	3.706	1.143	12.013	<0.001
Ibadan North West	0.531	2.557	0.903	7.241	0.201	0.600	4.958	1.530	16.071	<0.001
Ibadan South East	0.511	3.600	1.322	9.802	<0.001	0.600	6.417	1.980	20.801	0.002
Ibadan South West	0.523	1.554	0.558	4.332	0.201	0.600	8.697	2.683	28.191	<0.001
Ibarapa East	0.547	2.323	0.795	6.788	0.111	0.603	4.509	1.383	14.700	0.013
Ibarapa North	0.814	3.083	0.625	15.202	0.101	0.652	2.878	0.802	10.329	0.105
Ido	0.505	2.757	1.024	7.417	<0.001	0.599	5.063	4.254	44.525	<0.001
Irepo	0.500	1.319	0.495	3.515	0.511	0.601	3.6	1.109	11.693	0.033
Iseyin	0.527	5.238	1.897	14.968	<0.001	0.609	1.556	0.472	5.133	0.468
Itesiwaju	0.554	5.317	1.795	15.75	<0.001	0.621	1.451	0.429	4.900	0.549
Iwajowa	0.54	3.877	1.345	11.172	<0.001	0.609	2.524	0.765	8.328	0.128
Kajola	0.516	1.025	0.373	2.819	0.101	0.603	3.669	1.125	11.964	0.031
Lagelu	0.522	1.311	0.471	3.648	0.111	0.604	2.743	0.840	8.961	0.095
Ogbomoso South	0.551	1.812	0.401	3.480	0.201	0.614	1.989	0.597	6.629	0.262
Ogbomoso North	0.516	2.408	0.876	6.622	0.131	0.601	5.280	1.626	17.149	0.006
OgoOluwa	0.616	4.599	1.375	15.384	<0.001	0.663	0.765	0.209	2.805	0.686
Olorunsogo	0.519	1.349	0.488	3.729	0.181	0.604	3.287	1.006	10.739	0.049
Oluyole	0.536	5.578	1.951	15.952	<0.001	0.601	9.816	3.022	31.879	<0.001
Onaara	1.613	7.308	0.310	172.514	0.217	1.155	0.392	0.041	3.773	0.418
Orelope	0.647	5.692	1.601	20.229	<0.001	0.615	6.309	1.890	21.061	0.003
Orire	1.122	6.404	0.710	57.75	0.301	0.630	1.031	0.300	3.546	<0.001
Oyo East	0.537	2.398	0.837	6.872	0.101	0.609	1.974	0.598	6.512	0.264
Oyo West	0.513	2.392	0.875	6.537	0.331	0.601	4.581	1.411	14.879	0.011
Saki East	0.504	4.047	1.507	10.868	<0.001	0.599	1.031	0.319	3.337	<0.001
Saki West	0.504	1.379	0.514	3.706	0.271	0.598	6.573	2.036	21.223	<0.001
Surulere	0.550	1.184	0.403	3.480	0.121	0.629	1.122	0.327	3.849	0.855
Inf.Intercept	2.521	0.129	0.001	18.177	0.418	0.634	1.262	0.364	4.374	<0.001

Inf.2010 (Year 1) *										
Inf.2011 (Year 2)	0.340	0.019	0.010	0.037	<0.001	0.229	0.055	0.035	0.087	<0.001
Inf.2012 (Year 3)	0.408	0.016	0.007	0.037	<0.001	0.226	0.066	0.042	0.103	<0.001
Inf.2013 (Year 4)	0.282	0.067	0.039	0.117	<0.001	0.224	0.142	0.092	0.220	<0.001
Inf.2014 (Year 5)	0.293	0.746	0.420	1.325	0.318	0.264	0.993	0.592	1.666	0.978
Inf.April *										
Inf.August	0.390	0.779	0.363	1.673	0.521	0.288	0.612	0.348	1.076	0.088
Inf.Dec	0.426	4.973	2.158	11.461	<0.001	0.314	3.515	1.899	6.504	<0.001
Inf.Feb	0.458	0.419	0.171	1.028	0.058	0.292	0.729	0.412	1.293	0.280
Inf.Jan	0.430	0.454	0.196	1.055	0.066	0.292	0.647	0.365	1.147	0.136
Inf.July	0.391	1.062	0.493	2.285	0.879	0.286	0.873	0.498	1.529	0.635
Inf.June	0.389	0.792	0.370	1.698	0.549	0.285	0.742	0.424	1.296	0.295
Inf.March	0.435	0.393	0.168	0.922	0.032	0.289	0.627	0.356	1.105	0.105
Inf.May	0.404	1.154	0.523	2.547	0.722	0.289	1.129	0.641	1.989	0.676
Inf.Nov	0.412	4.023	1.794	9.021	<0.001	0.303	2.557	1.412	4.631	0.002
Inf.Oct	0.435	7.591	3.236	17.807	<0.001	0.325	4.379	2.316	8.281	<0.001
Inf.Sept	0.401	1.958	0.892	4.297	0.094	0.291	1.349	0.762	2.385	0.303
Inf.Afijio *										
Inf.Akinyele	2.561	4.191	0.028	634.324	0.083	0.676	0.177	0.047	0.667	0.010
Inf.Atiba	2.552	3.789	0.025	563.361	0.113	0.685	0.502	0.131	1.921	0.314
Inf.Atisbo	2.549	4.983	0.034	736.596	0.071	0.663	0.179	0.049	0.655	0.009
Inf.Egbeda	2.536	1.446	0.010	208.421	0.334	0.657	0.194	0.053	0.702	0.012
Inf.Ibarapa central	2.589	6.726	0.042	1075.391	0.223	0.745	0.486	0.113	2.092	0.332
Inf.Ibadan North	2.54	5.551	0.038	806.255	0.064	0.658	0.135	0.037	0.491	0.002
Inf.Ibadan North East	2.533	1.817	0.013	260.26	0.070	0.652	0.105	0.029	0.377	<0.001
Inf.Ibadan North West	2.535	1.366	0.009	196.488	0.066	0.656	0.174	0.048	0.631	0.008
Inf.Ibadan South East	0.532	1.062	0.374	3.012	0.109	0.653	0.064	0.018	0.231	<0.001
Inf.Ibadan South West	0.534	7.057	2.478	20.098	0.051	0.654	0.142	0.039	0.511	0.003
Inf.Ibarapa East	1.541	4.203	0.205	86.173	0.062	0.661	0.240	0.066	0.877	0.031
Inf.Ibarapa North	1.631	4.869	0.199	119.076	0.074	0.835	2.214	0.431	11.377	0.341
Inf.Ido	0.529	5.202	1.844	14.671	0.029	0.655	0.043	0.012	0.156	<0.001
Inf.Irepo	1.524	2.083	0.105	41.307	0.061	0.615	0.156	0.047	0.520	0.002
Inf.Iseyin	0.541	1.236	0.428	3.569	0.097	0.663	0.092	0.025	0.338	<0.001
Inf.Itesiwaju	0.608	6.746	2.049	22.213	0.034	0.668	0.219	0.059	0.810	0.023
Inf.Iwajowa	0.572	4.158	1.355	12.758	0.035	0.663	0.233	0.064	0.856	0.028
Inf.Kajola	2.545	1.063	0.007	155.898	0.111	0.653	0.104	0.029	0.375	<0.001
Inf.Lagelu	0.538	3.622	1.262	10.397	0.021	0.655	0.113	0.031	0.407	<0.001
Inf.Ogbomoso South	0.590	3.622	1.140	11.512	0.018	0.667	0.255	0.069	0.941	0.040
Inf.Ogbomoso North	0.538	3.010	1.049	8.641	0.006	0.652	0.095	0.026	0.341	<0.001
Inf.OgoOluwa	0.588	3.725	1.176	11.793	0.040	0.720	0.372	0.091	1.524	0.169
Inf.Olorunsogo	0.569	2.012	0.660	6.136	0.150	0.654	0.127	0.035	0.457	0.002
Inf.Oluyole	0.537	4.225	1.475	12.104	0.032	0.657	0.194	0.053	0.702	0.012
Inf.Onaara	0.044	3.248	2.98	3.540	0.007	1.377	3.785	0.255	56.256	0.334
Inf.Orelope	1.514	2.588	0.133	50.322	0.006	0.733	0.998	0.237	4.198	0.100
Inf.Orire	1.739	1.175	0.039	35.497	0.301	1.167	7.207	0.732	70.974	0.091
Inf.Oyo East	1.545	2.221	0.108	45.888	0.059	0.661	0.163	0.045	0.594	0.006
Inf.Oyo West	1.532	2.604	0.129	52.443	0.118	0.653	0.064	0.018	0.230	<0.001
Inf.Saki East	1.529	1.543	0.077	30.903	0.174	0.657	0.035	0.010	0.128	<0.001
Inf.Saki West	1.528	1.820	0.088	34.949	0.159	0.656	0.039	0.011	0.142	<0.001
Inf.Surulere	1.680	4.166	0.155	112.146	0.365	0.696	0.199	0.051	0.780	0.021
Dispersion	1.025									

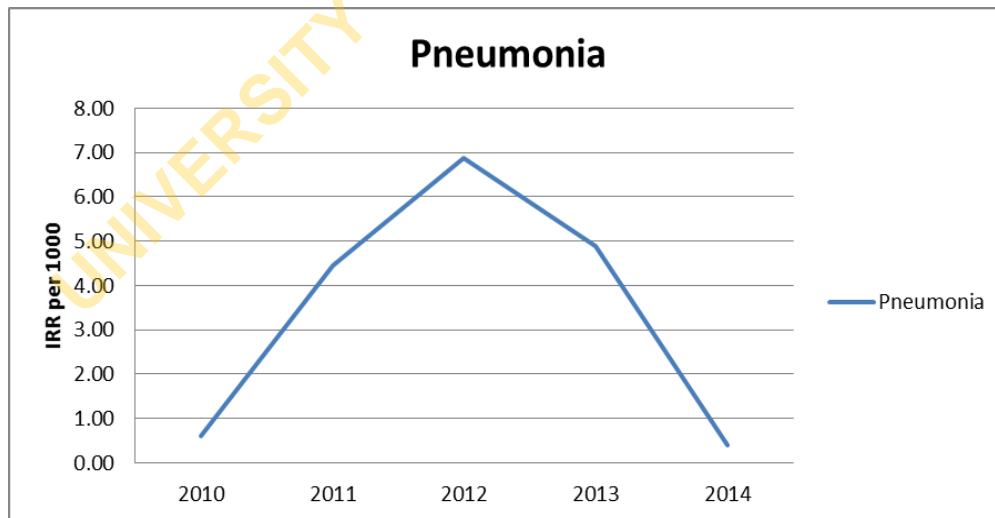
\*Reference category

**Table 2.** Descriptive statistics of number of cases of childhood pneumonia in 2010 to 2014

	N	Mean	Variance	Min	Max	No of zero cases (%)	No of non-zero cases
No of Pneumonia cases	1980	8.69	1865.203	0	544	1371(69.0%)	609(31.0%)



**Figure 1.** Distribution of Childhood Pneumonia cases in 2010 to 2014



**Figure 2.** Pattern of childhood pneumonia 2010 to 2014

**Table 3.** Comparison Test for childhood Pneumonia models (ZIP and ZINB)

Model comparison for Childhood Pneumonia			
Model 2: Zero-Inflated Negative Binomial			
Model 1: Zero-inflated Poisson		Vuong z-statistic	P-value
	Vuong statistic	-7.952249	<0.001

Test-statistic is asymptotically distributed  $N(0, 1)$  under the null hypothesis

## 5. Discussion

Unlike categorical data whose distribution is almost always known, count data is different and may sometimes have excess zeros. The childhood pneumonia cases showed excess zeroes and over-dispersion (the variance was greater than the mean). The zero inflated negative binomial regression model (ZINB) was preferred since it accommodates over-dispersion in the outcome data. Several studies have used the ZINB [1] and [6].

The annual incidence of childhood pneumonia showed a sustained decrease from 2010 onwards with the greatest decrease in year 2014. This reduction in incidence of pneumonia appears to coincide with the measures put in place to combat pneumonia both locally and globally. These measures include the annual observance of World Pneumonia Day and the introduction of immunization against HIBb, pneumococcus, measles and whooping cough (pertussis) [7].

Having established that ZINB regression model was best suited to determine the effect of month on the disease occurrence in Oyo state, months of reporting also had significant effect on the incidence of new cases of childhood pneumonia. Previous studies have shown that pneumonia illness is most common in the winter months and the greatest incidence during the rainy season [8, 9]. The geographical locations were also predictors of the occurrence of childhood pneumonia. In Nigeria, several environmental factors that could predispose children to pneumonia vary across geographic locations. The variation may be occasioned by seasonal and or occupational differences. A previous study on acute respiratory infections among un-der-five children in Indian slums [10] has identified the role of hazardous environment in the incidence of pneumonia while the impact of seasonal variation in pneumonia infection was noted by [11] who highlighted the increase in episodes typically occurring during the rainy season.

The years of reporting, months and the local government areas were found to be predictors for non-cases of pneumonia. The non-cases of pneumonia in the years, months and local government areas may be explained by the increased measures taken in Nigeria to improving access to vaccination in recent years.

## 6. Conclusions

The incidence of childhood pneumonia had a sustained decrease from 2010 onward and was highest in 2014. In view of observed sustained decrease in the incidence of childhood pneumonia, the government should be encouraged to improve on the intervention measures that have brought about this degree of success so far attained. The zero inflated negative binomial regression has been shown to adequately model such count data having excess zeros and over-dispersion and hence should be considered

as the statistical method of choice for analyzing such data.

## ACKNOWLEDGEMENTS

The authors sincerely thank the Department of Planning, Research and Statistics, Oyo State Ministry of Health for providing the dataset used for the study.

## REFERENCES

- [1] Greene, W.H., (1994). Accounting for excess zeros and sample selection in Poisson and negative binomial regression models, working paper, Department of Economics, Stern School of business, New York University, New York.
- [2] Fang, R. (2008). Zero-inflated negative binomial (ZINB) regression model for over-dispersed count data with excess zeros and repeated measures, an application to Human Microbiota sequence data. Bachelor of Medicine, North China Coal Mining Medical College. pp 1.
- [3] Yau, K., Wang, K., & Lee, A., (2003). Zero-inflated negative binomial mixed regression modeling of over-dispersed count data with extra zeros. *Biometrical Journal*, 45(4), 437-452.
- [4] Ridout, M., Hinde, J., & Demetrio, C., (2001). A score test for testing a zero-inflated Poisson regression model against zero-inflated negative binomial alternatives. *Biometrics*, 57.1: 219-223.
- [5] Akinpelu, K.P., Yusuf, O.B., Akpa, O.M., Gbolahan, A.O., 2016. Zero Inflated Regression Models with Application to Malaria Surveillance Data. *International Journal of Statistics and Applications*, 6(4), 223-234.
- [6] Abhulimhen-Iyoha, B.I., and Okolo, A.A. (2012). Morbidity and mortality of childhood illnesses at the emergency paediatric unit of the University of Benin Teaching Hospital, Benin City. *Niger Journal Paed*; 39(2), 71-74.
- [7] Hinde, J., and Demetrio, C., (1998). Over-dispersion: models and estimation. *Computational Statistics and Data Analysis*, 27, 151-170.
- [8] Erchick Daniel, Chizoba Wonodi, and Lois Privor-Dumm, 2011. Global coalition against child pneumonia. <http://www.jhsph.edu/research/centers-and-institutes/ivac/projects/nigeria/World-Pneumonia-Day-Nigeria-2011-Report.pdf>.
- [9] Paynter, Stuart, Robert S Ware, Philip Weinstein, Gail Williams, Peter D. Sly, (2010). Childhood pneumonia: a neglected, climate-sensitive disease. *The Lancet* 376:1804.
- [10] Dowell, S.F., Whitney, C.G., Wright, C., Rose, C.E., Schuchat, A., (2003). Seasonal Patterns of Invasive Pneumococcal Disease *Emerging Infectious Diseases*, 9(5), 574-579.
- [11] Singh, A., Singh M.N., (2014). Diarrhoea and acute respiratory infections among under-five children in slums: Evidence from India. *PeerJ PrePrint*, available at: <http://dx.doi.org/10.728, 208v1>.

- [12] Akanbi, M.O., Ukoli, C.O., Erhabor, G.E., Akanbi, F.O., Gordon, S.B., (2009). The burden of respiratory disease in Nigeria. *Mera: African Journal of Respiratory Medicine*, 10 -17.

UNIVERSITY OF IBADAN LIBRARY