SciencePG
Science Publishing Group

A Comparison of Poisson Model and Modified Poisson Model in Modelling Relative Risk of Childhood Diabetes in Kenya

Christine Gacheri Mutuura[1, *], Anthony Kibira Wanjoya[1], Isaiah Njoroge Mwangi[2]

[1]Department of Statistics and Actuarial Sciences, School of Mathematical Sciences, Jomo Kenyatta University of Agriculture and Technology, Juja, Kenya

[2]Centre for Respiratory Disease Research, Kenya Medical Research Institute, Nairobi, Kenya

Email address:
gacherimutuura@gmail.com (C. G. Mutuura)
[*]Corresponding author

Abstract: This study models the relative risk of diabetes, taking obesity and malnutrition as the major risk factors to define exposure, using three different prevalence rates i.e. 3%, 7% and 11% (estimates and projections from various studies). Secondary data consisting of a sample population of 300 children from the Kenya Diabetes Management and Information Centre (DMI), a national central diabetes registry, databases is used. In this research project, the modified Poisson regression approach is used to directly estimate the relative risk of pediatric diabetes in age strata of patients aged between the ages of 0-14years inclusive and for the purpose of model comparison RR estimation is done using Poisson regression which will prove to be less desirable for assessment of risk in this study proving the modified Poisson model gives the best estimates. From the data used in this study it is evident that: exposure (being overweight or underweight) is not a risk factor for diabetes onset in children aged 0-14 years.

Keywords: Type 1 Diabetes (T1D), Relative Risk (RR), Generalized Linear Models (GLMs), Generalized Additive Models (GAMs), Poisson Model and Modified Poisson Model

1. Introduction

Diabetes is a metabolic disorder that is caused by either a lack of or resistance to insulin (A hormone produced in the pancreas and is necessary for the body to efficiently use sugar as energy). Majority of the cases in the country are pegged on physical inactivity and (sedentary lifestyle).

Worldwide 347million people have diabetes, of these 10% have type I of whom 23.14million are children (WHO, 2010). In 2010, 3.4 million people were estimated to have died from diabetes, WHO projects that diabetes will be 7th leading cause of death by 2030. Many countries are documenting higher numbers of newly diagnosed cases of type I diabetes (T1D) particularly in younger children and it is thought to be the result of a combination of genetic and environmental factors.

It is estimated that 1.6 million young people are living with diabetes in Kenya with prevalence rates between 2.7% and 14.7% in rural and urban areas respectively. Incidence of the disease is forecasted to increase in the future due to a change in lifestyle to modern eating habits (junk food) and decline in physical activity. (WDF, 2012)

In this study prevalence rates from previous studies and some estimated projections are used. I look at three different scenarios; a prevalence rate of 3%, 7% and 11% from the DMI rural and urban estimates and the WHO projection for Kenya.

Some of the causes of diabetes among children are:

1) Family history: which alone does not determine the onset of the disease; other environmental factors are needed as triggers.
2) Age – juvenile disease is usually diagnosed during

childhood or young adulthood, though it can occur at any age.

3) Obesity – it is the number one risk of childhood diabetes in children of today.

4) Physical inactivity – in addition to contributing to obesity it also increases the risk for developing diabetes.

5) Malnutrition diabetes also has a proportion of cases in the region.

2. Literature Review

Type I diabetes cases in 2010 in Kenya were estimated at 17,470,000 with an incidence of 0.2 and prevalence of 1.0 per 100,000 persons per year. Age specific incidence rates and prevalence rates are missing in Kenya and therefore these rates had to be estimated from Tanzania rates (Swai *et al*, 1993).

A study carried out on empirical comparison of models that directly estimate the prevalence ratio, (Barros and Hirakata, 2003) came to a conclusion that "the Cox Poisson regression with robust variance and log-binomial regression provide correct and are a better alternative for the analysis of cross-sectional studies with binary outcomes than logistic regression, since the prevalence ratio is more interpretable and easier to communicate to non-specialists than the odds ratio."

Poisson regression has been applied to compare exposed and unexposed cohorts and to evaluate the clinical course of ill subjects. Estimates were obtained for the relative risk between exposed and unexposed individuals and the presence of interactions between the exposure variables and other factors was assessed (Stefano *et al*, 2006).

Majority of studies usually provide incidence rates rather than prevalence rates. Unreliable estimates have to be obtained by extrapolation and are used due to death of published studies (Gyula *et al* 2010). Therefore validity of estimates of numbers of children suffering with T1D is questionable and is therefore treated with considerable caution.

Increase in incidence is very high among the youngest children, though there is a general increase in incidence of childhood onset T1D (Berhan *et al*, 2011). It however stabilizes or even decreases in young adults. 28% of Kenyan children and youth are physically inactive (global WHO guidelines) children from rural Kenya are more physically active than urban children. Generalized additive models (GAM) of Poisson response have been used to investigate trends in incidences (Berhan *et al*, 2011). GAMs are fitted for Poisson family distributions with log link function.

Smoothing functions are allowed in GAM that allow for non-linear modeling of covariates. A non-parametric smoothing function is also used by a penalized regression spline approach. Age and sex interaction are also tested using GAMs.

Zou (2004, 2011) proposed a modified Poisson regression approach to estimate relative risk directly as it is usually the parameter of interest in most epidemiologic and medical studies. Poisson regression model using a sandwich variance estimator has become a viable alternative to logistic regression model for analysis of prospective studies with independent binary outcome. The primary advantage of this approach is that it readily provides covariate-adjusted risk ratios and their associated standard errors.

3. Methods

3.1. Relative Risk

Relative risk is a measure of association between the occurrence of an exposure and probability of disease occurrence. Relative risk is greater than 1 if there is an effect from the exposure. A value less than 1 and greater than 0 is observed for factors inversely related to risk (more of protective factors not exposure) i.e.:

If RR=1, Risk in exposed= Risk in non-exposed (No association between exposure and disease occurrence).

If RR>1, Risk in exposed> Risk in non-exposed (Positive association between exposure and disease occurrence, causal factor).

If RR<1, Risk in exposed< Risk in non-exposed (Negative association between exposure and disease occurrence, protective factor).

3.2. Poisson Regression

Poisson regression analysis is a technique used to describe count data and contingency tables. It is therefore used in epidemiology to investigate occurrence of selected diseases. It is used in place of Cox's model for survival analysis especially where the data is available only in an aggregated form (count data) Poisson is used.

It is used in cases where the dependent variable is a count e.g. events such as number of patients diagnosed with a particular disease and the diagnosis of one patient should be independent i.e. events are independent in the sense that diagnosis of one will not make diagnosis of another more or less likely.

The Poisson model is a regression belonging to the generalized linear models (GLM) class and it uses the logarithm as its link function:

$$g[E(y|\beta,x)] = \beta_0 + \beta_1 x_1 + - - - + \beta_p x_p = \alpha + \beta x_i = \theta' x \qquad (1)$$

Where;

$$\theta' = \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_p \end{pmatrix}$$

$$P(y|\beta,x) = \frac{e^{-\beta}\beta^y}{y!} \qquad (2)$$

Where: $y = 1,2,3,...$

g Is the "link function" which is the log in this case.

Poisson regression allows for modeling of rates even in the presence of groups containing zero observations and linear regression model assumes the same variance for rates within each group under study (homoscedasticity) which is almost impossible in actual cohort data (Kleinbaum et al, 1998).

Poisson regression takes the form in (1) and may be appropriate for rate data. The rate is the number of events occurring in a study area divided by some measure of the study areas exposure.

Therefore,

$$\log(E(y|x)) = \log(\text{exposure}) + \theta'x$$

$$\log(E(y|x)) - \log(\text{exposure}) = \log\left(\frac{E(y|x)}{\text{exposure}}\right) = \theta'x. \quad (3)$$

Poisson mean and variance are equal and in the case where the variance is greater than the mean, it indicates that there is over-dispersion and is solved by using alternative model such as the negative binomial distribution. Poisson also exhibits a problem of excess zeros (zero inflated) whereby in the process of determining how many events occurred, there will be more zeros than Poisson regression predicts.

3.3. Modified Poisson Regression

This is a Poisson regression with a robust error variance to enable direct estimation of a given measure. In this study, the measure is the relative risk based on different prevalence rates. To estimate relative risk directly, modified Poisson regression is appropriate (McNutt *et al*, 2003).

Assuming a 2-by-2 table:

Table 1. Contingency table.

	Y=1 (diabetic)	Y=0 (not diabetic)	Total
X=1 (exposed to lifestyle risk factors)	a	b	$n_1 = a + b$
X=0 (unexposed)	c	d	$n_0 = c + d$
			$n = n_1 + n_0$

Depending on the various ratios of gender and age categories from the data availed from the registry and the prevalence rates projected by DMI (3% for rural and 7% for urban areas and a projection of 11% in future) as well as the exposure and non-exposure ratios obtained from the data that was collected, a and c were estimated appropriately. For the non-diabetics column, the values b and d were assumed to be the remaining portion of the population in the different prevalence categories.

When Poisson regression is used in binomial data relative risk is over-estimated (Zocchetti *et al*, 1995). Using sandwich estimation this is taken care of (Royal, 1986). Assuming the link function in (1), the relative risk is given by:

$$RR = \exp(\beta)$$

Assuming Poisson distribution for Y; log likelihood for (1) is:

$$\log(\alpha, \beta) = C.\sum_{i=1}^{n}[y_i(\alpha + \beta x_i) - \exp(\alpha + \beta x_i)] \quad (4)$$

Where C is a constant.

From the likelihood theory considering equations (1) and (2):

$$\log \lambda(x) = \alpha + \beta x_i$$

The Likelihood is now given as:

$$L(\alpha, \beta) = \prod_{i=1}^{m}\frac{e^{-\lambda(x_i)}\lambda(x_i)^{n_i}}{n_i!} \quad (5)$$

$$\log L = -\sum \lambda(x_i) + \sum n_i \log(\lambda(x_i)) - \sum \log(n_i!) \quad (6)$$

From the contingency table 1, suppose $x's$ take only values 0 or 1:

$$\lambda(0) = e^{\alpha} \text{ And } \lambda(1) = e^{\alpha+\beta}$$

$$\Rightarrow \log L = -\sum e^{\alpha}I_{x_i=0} - \sum e^{\alpha+\beta}I_{x_i=1} + \sum n_i\alpha I_{x_i=0} + \sum n_i(\alpha + \beta)I_{x_i=1} - \sum \log(n_i!) \quad (7)$$

$\log(n_i!)$ Is a constant and does not change with parameter values and is hence not needed in likelihood ratio test since it will cancel out.

Now, differentiating with respect to α and β and setting to 0 to solve the maximum likelihood estimates:

$\hat{\alpha} = \log(\sum\frac{n_iI_{x_i=0}}{m_0})$, Where m_0 is the number of zeros among the $x's$

$$\Rightarrow \exp(\hat{\alpha}) = \sum\frac{n_iI_{x_i=0}}{m_0} = \frac{c}{n_0} \quad (8)$$

$$\hat{\beta} = \log\left(\sum\frac{n_iI_{x_i=1}}{m_1}\right) - \hat{\alpha}$$

$$\hat{\beta} = \log\left(\sum\frac{n_iI_{x_i=1}}{m_1}\right) - \log\sum\frac{n_iI_{x_i=0}}{m_0}$$

$$\hat{\beta} = \log\left(\sum\frac{n_iI_{x_i=1}}{m_1} \div \sum\frac{n_iI_{x_i=0}}{m_0}\right)$$

$$\hat{\beta} = \log\frac{\sum n_iI_{x_i=1}.m_0}{\sum n_iI_{x_i=0}.m_1}$$

m_1 Is the number of 1's among the $x's$

$$\Rightarrow \exp(\hat{\beta}) = \frac{\sum n_iI_{x_i=1}.m_0}{\sum n_iI_{x_i=0}.m_1} = \frac{an_0}{cn_1} \quad (9)$$

I.e. $\exp(\widehat{\alpha}) = \frac{c}{n_0}$ and $\widehat{RR} = \exp(\hat{\beta}) = \frac{an_0}{cn_1}$ from equations

(5) and (6) respectively

$$\text{And } \widehat{var(RR)} = \frac{1}{a} + \frac{1}{c} \qquad (10)$$

Sandwich estimation can be used to estimate the variance of the maximum likelihood estimates (MLE) when the underlying model is incorrect. It yields variances for the MLE that are asymptotically correct. It is also used to avoid overestimating standard errors of parameter estimates.

Now using a sandwich estimator to correct the above variance, (10) it yields;

$$\widehat{var(RR)} = \frac{1}{a} - \frac{1}{n_1} + \frac{1}{c} - \frac{1}{n_0} \qquad (11)$$

This is now consistently estimated.

4. Empirical Results and Presentations

4.1. Poisson Regression Model

The proposed model for this study was the Poisson model with robust variance as suggested by Zou i.e. Modified Poisson model, suitable for fitting epidemiological data especially when estimating the relative risk or odds ratio. For the purpose of model comparison, a poisson regression model was fitted to the data giving the following outputs at 3% prevalence, 7% prevalence and 11% prevalence respectively.

At 3% prevalence the exposure and physical activity variables are the only significant variables in explaining the diabetes status of the child with a p-value< 0.05. However, these two variables appear to have no association with the diabetes status considering the relative risk. The standard

errors associated with these variables are also not robust as will be seen with the modified Poisson regression model. The physical activity and exposure also have a negative correlation with the output variable (diabetes status) from the estimates of their coefficients and this is also supported by the relative risk estimates.

Table 2. 3% Prevalence Poisson regression.

	Estimate	standard err	z-value	p-value
Gender (M)	1.4974	4.1662	0.36	0.719
Gender (F)	1.9363	4.0719	0.48	0.634
Height	−0.0214	0.0394	−0.54	0.588
Weight	0.0589	0.0535	1.10	0.271
Diet	0.1666	0.7604	0.22	0.827
Activity	−2.7330	0.9201	−2.97	0.003
Exposure	−2.0562	0.9774	−2.10	0.035
Age (5-9)	0.3063	1.0354	0.30	0.767
Age (10-14)	0.1195	0.8528	0.14	0.889

Table 3. 7% Prevalence Poisson regression.

	Estimate	standard err	z-value	p-value
Gender (M)	5.0427	2.7002	1.87	0.0618
Gender (F)	5.0354	2.7002	1.86	0.0622
Height	−0.0513	0.0281	−1.83	0.0677
Weight	0.1073	0.0404	2.66	0.0079
Diet	−0.0956	0.5272	−0.18	0.8561
Activity	−2.8023	0.6332	−4.43	$9.6e-06$
Exposure	−2.7303	0.6764	−4.04	$5.4e-05$
Age (5-9)	0.7520	0.7344	1.02	0.3059
Age (10-14)	0.1755	0.6112	0.29	0.7740

Physical activity and exposure are still significant in explaining the diabetes status from their p-values and are negatively correlated to the output variable from the coefficient estimates.

Table 4. 11% Prevalence Poisson regression.

	Estimate	standard err	z-value	p-value	relative risk
Gender (M)	4.1598	2.1576	1.93	0.054	64.06012
Gender (F)	4.2763	2.1158	2.02	0.043	71.97087
Height	−0.0293	0.0209	−1.40	0.160	0.97109
Weight	0.0572	0.0301	1.90	0.057	1.05883
Diet	−0.1288	0.3898	−0.33	0.741	0.87918
Activity	−2.4654	0.4427	−5.57	$2.6e-08$	0.08497
Exposure	−2.1421	0.5222	−4.10	$4.1e-05$	0.11741
Age (5-9)	0.0850	0.5476	0.16	0.877	1.08869
Age (10-14)	−0.2458	0.4772	−0.52	0.606	0.78209

Physical activity and exposure are still significant in explaining the diabetes status from their p-values and are negatively correlated to the output variable from the coefficient estimates. The female gender is also slightly significant in explaining the diabetes status.

4.2. Modified Poisson Regression Model

The following are the modified Poisson regression output at 3% prevalence, 7% prevalence and 11% prevalence respectively.

Table 5. 3% Prevalence modified Poisson.

	Estimate	standard err	Wald's	p-value	relative risk
Gender (M)	1.49743	3.42277	0.191	0.66176	4.47018
Gender (F)	1.93630	3.16642	0.374	0.54086	6.93305
Height	−0.02136	0.02529	0.714	0.39824	0.97886
Weight	0.05892	0.03157	3.482	0.06203	1.06069
Diet	0.16655	0.97566	0.029	0.86446	1.18128
Activity	−2.73297	1.22171	5.004	0.02529	0.06503
Exposure	−2.05620	0.72036	8.148	0.00431	0.12794

	Estimate	standard err	Wald's	p-value	relative risk
Age (5-9)	0.30634	1.01592	0.091	0.76300	1.35844
Age (10-14)	0.11953	0.92488	0.017	0.89717	1.12697

As seen from the above table the exposure, physical activity and weight are seen to be significant in explaining the diabetes status. There appears to be no clear association of the exposure and the physical activity with the diabetes status from the relative risk, although the weight has a detectable positive effect with the diabetes status much as the relative risk is weak.

Here we also see that the gender, age category and diet also have a positive association with the diabetes status. The female gender is once again seen to have a higher relative risk as compared to the male gender implying that females are genetically predisposed to diabetes.

A negative correlation between the height, exposure and physical activity with the output variable is observed and this is also supported by the relative risk estimates for the said variables. The standard errors for the variables are seen to have reduced as compared to those of the Poisson model. This therefore renders this model superior in epidemiological studies of this nature.

Table 6. 7% Prevalence modified Poisson.

	Estimate	standard err	Wald's	p-value	relative risk
Gender (M)	5.04270	2.24995	5.023	0.025010	154.88781
Gender (F)	5.03545	2.31131	4.746	0.029361	153.76865
Height	−0.05131	0.02282	5.055	0.024550	0.94999
Weight	0.10733	0.03195	11.286	0.000781	1.11331
Diet	−0.09558	0.50106	0.036	0.848710	0.90884
Activity	−2.80230	0.67435	17.269	3.25e − 05	0.06067
Exposure	−2.73028	0.48352	31.885	1.64e − 08	0.06520
Age (5-9)	0.75198	0.63695	1.394	0.237767	2.12120
Age (10-14)	0.17542	0.52623	0.111	0.738784	1.19182

As seen from the above table the age categories and diet are seen not to be significant in explaining the diabetes status. There appears to be no clear association of the exposure, the physical activity, the diet and the height with the diabetes status from the relative risk, although the weight and age categories have a detectable positive effect with the

diabetes status much as the relative risk is weak.

A negative correlation between the diet, height, exposure and physical activity with the output variable is observed and this is also supported by the relative risk estimates for the said variables. The standard errors for the variables are seen to have reduced as compared to those of the Poisson model.

Table 7. 11% Prevalence modified Poisson.

	Estimate	standard err	Wald's	p-value	relative risk
Gender (M)	4.15982	1.88037	4.894	0.0270	64.06012
Gender (F)	4.27626	1.86725	5.245	0.0220	71.97087
Height	−0.02933	0.01932	2.304	0.1290	0.97109
Weight	0.05716	0.02795	4.181	0.0409	1.05883
Diet	−0.12876	0.33918	0.144	0.7042	0.87918
Activity	−2.46540	0.44315	30.951	2.65e − 08	0.08497
Exposure	−2.14210	0.44721	22.943	1.67e − 06	0.11741
Age (5-9)	0.08498	0.43433	0.038	0.8449	1.08869
Age (10-14)	−0.24579	0.36952	0.442	0.5066	0.78209

As seen from the above table the exposure, physical activity and weight are seen to be significant in explaining the diabetes status. There appears to be no clear association of the exposure and the physical activity with the diabetes status from the relative risk, although the weight has a detectable positive effect with the diabetes status much as the relative risk is weak and the gender was also noted to have a positive association with the diabetes status.

Here we also see that the gender, age category (5-9) and weight also have a positive association with the diabetes status. The female gender is once again seen to have a higher relative risk as compared to the male gender implying that females are genetically predisposed to diabetes.

A negative correlation between the height, diet, exposure, physical activity and age category (10-14) with the output variable is observed and this is also supported by the relative risk estimates for the said variables. The standard errors for the variables are seen to have reduced as compared to those of the Poisson model.

As can be seen from all the above output and the model comparison tables that follow, the modified Poisson regression gives lesser standard errors hence showing the estimates generated by this model are more precise and accurate as compared to those from the Poisson model above. (The figures in bold indicate significance)

4.3. Model Comparison

Table 8. *3% Prevalence Comparison of Models.*

		Male	Female	Weight	Diet	5-9	10-14
Poisson	Std Err	4.16	4.07	0.05	0.76	1.04	0.85
	RR	4.47	6.93	1.06	1.18	1.36	1.13
Modified Poisson	Std Err	3.42	3.17	0.03	0.98	1.02	0.92
	RR	4.47	6.93	1.06	1.18	1.36	1.13

From the table above it is evident that the modified Poisson regression model is more precise with smaller standard errors and an improved relative risk that is not overestimated. The standard errors of the predictor variables are seen to improve from the Poisson model with the smaller values observed for the modified Poisson. The weight, diet gender and age categories of 5-9 and 10-14 were seen to have a positive association with the diabetes status from their relative risks.

Table 9. *7% Prevalence Model Comparison.*

		Male	Female	Weight	5-9	10-14
Poisson	Std Err	2.70	2.70	0.04	0.73	0.61
	RR	154.9	153.8	1.11	2.12	1.19
Modified Poisson	Std Err	2.25	2.31	0.03	0.64	0.53
	RR	154.9	153.8	1.11	2.12	1.19

Here the gender, weight and age categories 5-9 and 10-14 have a positive association with the diabetes status in children from their relative risks and once again the modified Poisson regression model gives more precise estimates from the standard errors values that are smaller in this model.

Table 10. *11% Prevalence Comparison of Models.*

		Male	Female	Weight	5-9
Poisson	Std Err	2.16	2.12	0.03	0.55
	RR	64.06	71.97	1.06	1.09
Modified Poisson	Std Err	1.89	1.87	0.03	0.43
	RR	64.06	71.97	1.06	1.09

At an 11% prevalence, which is the estimated prevalence rate of diabetes by the year 2030, the gender has a positive association with the diabetes status as well as the weight and 5-9 age category, which have a positive association though very weak. The modified Poisson model gives more precise estimates here as well.

4.4. Confidence Intervals for Modified Poisson Model

The natural log of the sample relative risk is approximately normally distributed and is used to produce the confidence intervals for the relative risk. The null value of the confidence interval for the relative risk is 1. If a 95% confidence interval for the relative risk includes this null value then there is insufficient evidence to conclude that the groups are statistically significantly different. However, if the interval does not include the null value, the finding is statistically significant.

When dealing with a small sample size, as in this study, the confidence is often observed to be very broad and asymmetric (the point estimate of the relative risk does not lie in the exact centre of the confidence interval). As a result the point estimate is imprecise.

Table 11. *95% Confidence intervals at 3% prevalence.*

Variables	Relative risk	Lower relative risk	Upper relative risk
Male	4.470	0.00546	3662.645
Female	6.933	0.01398	3437.067
Height	0.979	0.93152	1.029
Weight	1.061	0.99704	1.128
Diet	1.181	0.17452	7.995
Activity	0.065	0.00593	0.713
Exposure	0.128	0.03118	0.525
Age 5-9	1.358	0.18548	9.949
Age 10-14	1.127	0.18393	6.905

In this output, no variable is observed to have any kind of statistical significance since all the variables that have a positive association with the outcome variable have the null value in the confidence interval. The gender, weight, weight, diet and age category are therefore assumed to have no statistical significance in explaining the diabetes status, though they have a detectable positive effect as explained by the weak relative risk.

Table 12. *95% Confidence Intervals for 7% prevalence.*

Variables	Relative risk	Lower relative risk	Upper relative risk
Male	154.88781	1.88304	1.274e + 04
Female	153.76865	1.65758	1.426e + 04
Height	0.94999	0.90844	1.934e − 01
Weight	1.11331	1.04573	1.185e + 00
Diet	0.90884	0.34040	2.427e + 00
Activity	0.06067	0.01618	2.275e − 01
Exposure	0.06520	0.02527	1.682e − 01
Age 5-9	2.12120	0.60870	7.392e + 00
Age 10-14	1.19182	0.42490	3.343e + 00

At 7% prevalence, the gender and the weight are statistically significant. The height, diet, activity, exposure and age categories have the null value, 1, in the interval and it can therefore be concluded that there is insufficient evidence that these variables are statistically significant in explaining the diabetes status.

Gender and weight at 11% prevalence are statistically significant since their 95% confidence intervals do not include the null value 1 in the interval. Although, the age category 5-9 has a detectable positive effect, the interval

includes the null value and therefore it is considered not to have statistical significance.

Table 13. 95% Confidence intervals for 11% prevalence.

Variables	Relative risk	Lower relative risk	Upper relative risk
Male	64.06012	1.60698	2553.6669
Female	71.97087	1.85246	2796.1842
Height	0.97109	0.93500	1.0086
Weight	1.03883	1.00238	1.1185
Diet	0.87918	0.45224	1.7092
Activity	0.08497	0.03565	0.2025
Exposure	0.11741	0.04887	0.2821
Age 5-9	1.08869	0.46473	2.5504
Age 10-14	0.78209	0.37907	1.6136

5. Conclusion and Recommendations

In this study exposure is not much of a risk factor for diabetes onset since the relative risk of exposure is observed to have a negative association between the exposure variable and the diabetes status, although gender seems to be a major risk factor for diabetes onset since females have a higher risk of having diabetes. Weight is a risk factor for diabetes onset though the relative risk is weak. Modified Poisson model proved to be superior to the Poisson model.

A study that involves family history and other environmental factors as risk factors should be explored to better model diabetes risk such as prenatal influences (i.e. parent's diabetes status, high birth weight, babies born to older women and children with higher birth orders), demographic features such as seasons. Other factors that have been suspected to cause diabetes in the young include: early weaning, other diseases e.g. viral diseases whose treatment may cause failure of the islet cells reducing the pancreas' ability to produce insulin, Lack of vitamin D especially at higher latitudes and other non-specific environmental factors such as the hygiene hypothesis and the accelerator hypothesis. (Gale, 2014). The government or other concerned research organisations should fund more projects on childhood diabetes so as to control the incidence of this disease and also promote more sensitisation campaigns to create awareness on diabetes in children since most people think diabetes is an "Old people disease"

Kenya Demographic Health Surveys (DHS) carried out by KNBS should include diabetes and other non communicable diseases to ensure a detailed national registry with updated figures for prevalence and incidence rates of the diseases.

References

[1] World Health Organisation(2010), www.who.int/diabetes/publications/Definition%20and%20diagnosis%20of%20diabetes_new.pdf

[2] World Diabetes Foundation (2012), Celebrating 10 years of making a difference.

[3] Swai A. B, Lutale J. L, McLarty D. G, (1993), Prospective study of Incidence of Juvenile diabetes mellitus over 10 years in Dar es Salaam Tanzania. BioMedical Science journal, 306(6892): 1570-1572.

[4] Barros J. D and Hirakata V, (2003), an empirical comparison of models that directly estimate the prevalence ratio. Bio Med Central Medical Research Methodology, 3: 21.

[5] Parodi S, Bottarelli E. (2006) Poisson Regression model in Epidemiology, Ann. Fac. Medic. Vet. Al. Parma, XXVI: 25-44.

[6] Gyula S, Paterson C, Dahlquist G, (2010), Diabetes in the Young: A Global Perspective International Diabetes Federation, Diabetes Atlas 4th Edition.

[7] Yonas B, Waernbaum I, Lind T, Mollsten A and Dahlquist G, (2011), Thirty years of Prospective Nationwide Incidence of childhood type 1 diabetes. Diabetes Journal, 60: 577-581.

[8] World Health Organisation (2010), Guidelines, www.who.int/mediacentre/factsheets/fs312/en/, 2013.

[9] Zou, G. (2004), A Modified Poisson Regression Approach to Prospective Studies with Binary Data. American Journal of Epidemiology, 159:702-706.

[10] Kleinbaum, D. G, Kupper, L. L, Muller, K. E and Nizam A, (1998). Applied Regression Analysis and other Multivariate Methods. Third Edition, Brooks/Cole Publishing Company, Duxbury. Press, Pacific Grove (CA).

[11] McNutt, L. A, Wu C, Xue X, (2003), Estimating the Relative Rise in Cohort Studies and Clinical Trials of Common Outcomes. American Journal of epidemiology, 157:940-943.

[12] Zochetti C, Consonni D, Berlazzi P. A, (1995). Estimation of Prevalence Rate and Ratios from Cross-sectional Data. International Journal of Epidemiology, 24:1064-1065.

[13] Royal, R. M, (1986). Model Robust Confidence Intervals using Maximum Likelihood Estimates. International Statistic review, 54: 221-226.

[14] Gale, E. A. M, Environmental factors [www.diapedia.org/type-1-diabetes-mellitus/environmental-factors], 2014 Aug 13; Diapedia 21040851139 rev. no. 45. Available from: http://dx.doi.org/10.14496/dia.21040851139.45.