THE LANCET HIV

Supplementary appendix

This appendix formed part of the original submission and has been peer reviewed. We post it as supplied by the authors.

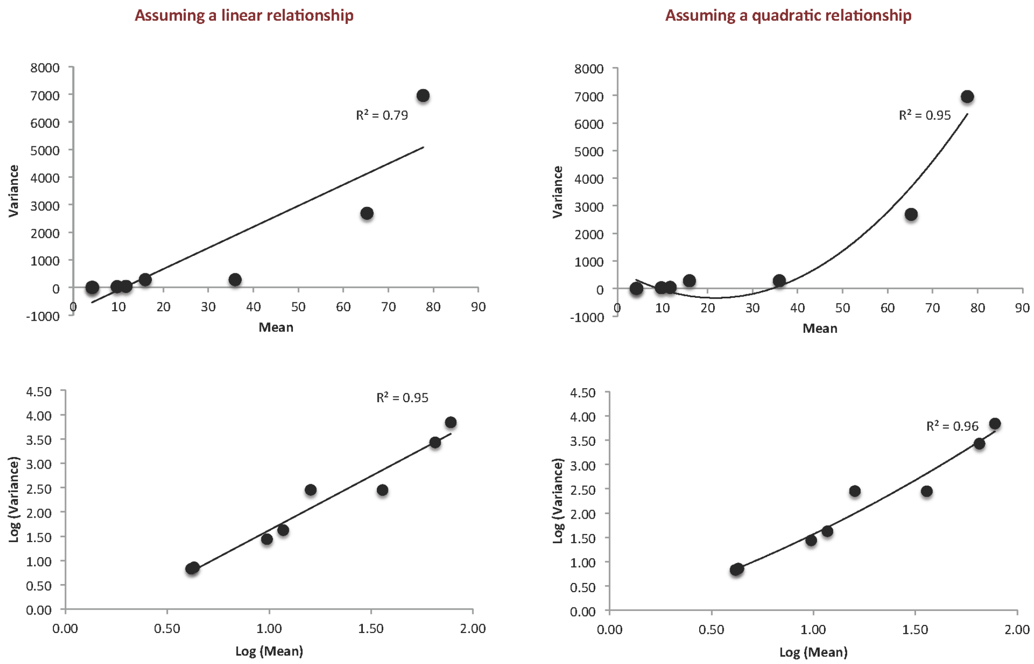
Supplement to: Lima VD, Lourenço L, Yip B, Hogg RS, Phillips P, Montaner JSG. AIDS incidence and AIDS-related mortality in British Columbia, Canada, between 1981 and 2013: a retrospective study. *Lancet HIV* 2015; published online Feb 13. http://dx.doi.org/10.1016/S2352-3018(15)00017-X.

Supplementary Material

Negative Binomial Regression Analysis[1, 2]

We run this model in the statistical software R© version 3.0.2 so that we could make use of the following packages: (1) AER to test for overdispersion and assess whether a Quasi-Poisson or a negative binomial would be the appropriated distribution for our count data; (2) pcsl and lmtest to test the likelihood ratios between the Negative Binomial and Poisson regressions; and (3) MASS to perform the negative binomial regression and obtain robust standard errors for our estimated parameters. Code used can be obtained from the first author in the paper.

1. We assessed whether the relationship between the mean and the variance was linear or quadratic by plotting the mean and the variance for the combined levels of the categorical variables age group and gender.



In summary, the relationship between the mean and the variance seems to be quadratic (based in the curves and the statistic R^2 (larger values are indicative of better fit – maximum is at 1), however, with the small sample sixe for the models, it is hard to really know for sure.

2. The test of overdispersion in the Poisson model was done by comparing the mean to the variance of the counts and by fitting the linear regression $\frac{\left(y_i-\widehat{\mu}_i\right)^2-y_i}{\widehat{\mu}_i}=\alpha\frac{g\left(\widehat{\mu}_i\right)}{\widehat{\mu}_i}+u_i$, where $\widehat{\mu}_i=\exp(x'\widehat{\beta})$ and u_i is the error term. The function $g(.)$ could be $g(\mu)=\mu$ or $g(\mu)=\mu^2$, and in our case we were interested in fitting a negative binomial with overdispersion of the form $g(\mu)=\mu^2$. Note that in this case we were testing the hypothesis: $H_0:\alpha=0$ versus $H_1:\alpha\neq 0$. In this case p-values >0.05 indicated the presence of overdispersion. This test was done using the R library AER function dispersiontest (). For overdispersion, we also examined the ratio between the model's residual deviance and the degrees of freedom and a model without overdispersion would have this ratio equal to one.

3. To assess the goodness of fit of each model, we used different criteria: (1) We compared the likelihood ratio test of the Negative Binomial in relation to the Poisson distribution using two likelihood-based tests (functions lrtest() from the lmtest R library and odTest() from the pcsl R library). In these tests we compared the log-likelihoods of a negative binomial regression model and a Poisson regression model – in this case p-values <0.05 indicate that the Negative Binomial distribution provides a better fit to the data; (2) We examined the estimate and standard error (SE) for the dispersion parameter to assess its magnitude; and (3) We plotted the Pearson residuals by the model predicted values, and we also tested the significance of the sum of the Pearson

residuals. We used the chi-square test built into R – in this case p-values ≥0.05 indicated the appropriateness of the model in fitting the data.

Next, we present the tables with the diagnostic/test information, described in items 2 and 3 above, for each of the fitted models in Table 2 in the manuscript. For the calendar periods 1981-1996 and 2004-2013, we fitted a model with both age and gender as covariates. Based on the assumptions check for the models in periods 1997-1999 and 2000-2003, we had to fit separated models for gender and age to fulfill the requirements of fitting a Negative Binomial regression. We believe that the small sample size in these groups might have influenced the model's diagnostic statistics and tests.

Diagnostic Statistics/Tests	**Period**	**Period**
	1981 -1996	**2004 - 2013**
Mean of Outcome	38.74	22.88
Variance of Outcome	4688.54	499.39
Model with age and gender as covariates		
Overdispersion tests		
Dispersion parameter - Negative Binomial	0.97 (SE 0.15)	6.96 (SE 1.62)
Residual deviance - Poisson	2879.90	313.60
Residual deviance - Negative Binomial	82.70	79.72
Degrees of freedom	72.00	72.00
Deviance/degrees of freedom - Poisson	40.00	4.36
Deviance/degrees of freedom - Negative Binomial	1.15	1.11
Hypothesis test for overdispersion (p-value)	0.9456	0.1260
Goodness of fit tests		
AIC Poisson	3212.5	670.98
AIC Negative Binomial	624.42	534.26
Hypothesis test for the appropriateness of the Negative Binomial versus a Poisson distribution (p-value)	0.0000	0.0000
Likelihood test comparing the Poisson and the Negative Binomial distributions (p-value)	0.0000	0.0000
Hypothesis test, based on residuals, for the model fit (p-value) - Poisson	0.0000	0.0000
Hypothesis test, based on residuals, for the model fit (p-value) - Negative Binomial	0.9618	0.4173

Diagnostic Statistics/Tests	Period	Period
	1997 - 1999	2000 - 2003
Mean of Outcome	27.71	24.09
Variance of Outcome	884.81	491.54
Model with age and gender as covariates		
Overdispersion tests		
Dispersion parameter - Negative Binomial	31.23 (SE 20.80)	20.38 (SE 19.70)
Residual deviance - Poisson	61.57	53.49
Residual deviance - Negative Binomial	37.01	35.37
Degrees of freedom	26.00	18.00
Deviance/degrees of freedom - Poisson	2.37	2.97
Deviance/degrees of freedom - Negative Binomial	1.42	1.97
Hypothesis test for overdispersion (p-value)	0.0245	0.0129
Goodness of fit tests		
AIC Poisson	210.92	167.76
AIC Negative Binomial	204.89	167.12
Hypothesis test for the appropriateness of the Negative Binomial versus a Poisson distribution (p-value)	0.0023	0.0519
Likelihood test comparing the Poisson and the Negative Binomial distributions (p-value)	0.0046	0.1038
Hypothesis test, based on residuals, for the model fit (p-value) - Poisson	0.0000	0.0000
Hypothesis test, based on residuals, for the model fit (p-value) - Negative Binomial	0.0768	0.0048
Model with only gender as a covariate		
Overdispersion tests		
Dispersion parameter - Negative Binomial	2.27 (SE 0.65)	1.23 (SE 0.31)
Residual deviance - Poisson	358.83	593.61
Residual deviance - Negative Binomial	32.79	34.07
Degrees of freedom	29.00	27.00
Deviance/degrees of freedom - Poisson	12.37	21.99
Deviance/degrees of freedom - Negative Binomial	2.65	1.55
Hypothesis test for overdispersion (p-value)	0.0000	0.0000
Goodness of fit tests		
AIC Poisson	502.18	740.96
AIC Negative Binomial	244.34	267.18
Hypothesis test for the appropriateness of the Negative Binomial versus a Poisson distribution (p-value)	0.0000	0.0000
Likelihood test comparing the Poisson and the Negative Binomial distributions (p-value)	0.0000	0.0000
Hypothesis test, based on residuals, for the model fit (p-value) - Poisson	0.0000	0.0000
Hypothesis test, based on residuals, for the model fit (p-value) - Negative Binomial	0.3200	0.7685
Model with only age as a covariate		
Overdispersion tests		
Dispersion parameter - Negative Binomial	2.73 (SE 0.92)	1.80 (SE 0.55)
Residual deviance - Poisson	195.28	298.70
Residual deviance - Negative Binomial	23.86	24.26
Degrees of freedom	21.00	19.00
Deviance/degrees of freedom - Poisson	9.30	15.72
Deviance/degrees of freedom - Negative Binomial	2.57	1.54
Hypothesis test for overdispersion (p-value)	0.0000	0.0000
Goodness of fit tests		
AIC Poisson	303.55	410.96
AIC Negative Binomial	179.5	192.48
Hypothesis test for the appropriateness of the Negative Binomial versus a Poisson distribution (p-value)	0.0000	0.0000
Likelihood test comparing the Poisson and the Negative Binomial distributions (p-value)	0.0000	0.0000
Hypothesis test, based on residuals, for the model fit (p-value) - Poisson	0.0000	0.0000
Hypothesis test, based on residuals, for the model fit (p-value) - Negative Binomial	0.7633	0.6967

Generalized Additive Models[3, 4]

We run this model in the statistical software R© version 3.0.2 so that we could make use of the following packages: (1) mgcv, function gam (), for fitting the generalized additive model and test the models' goodness of fit through diagnostic plots commonly using when fitting a generalized liner model; and (2) function anova() to perform an analysis of deviance comparing each of the fitted models for each outcome.

In brief, we used a cubic regression spline to smooth the trend of the number of ADIs and the proportion of deaths due to an ADI by calendar year. In this case calendar year is the explanatory variable being smoothed. It is important to mention that we decided to transform the outcome number of ADIs using the logarithm function to obtain a better fit for the model and we assumed a Gaussian distribution for this outcome. For the outcome proportion of deaths due to an ADI, we modeled it using the number of deaths due to an ADI as outcome and as an offset the overall number of deaths for each calendar year. In this case we also transformed the outcome using the logarithm function and assumed a Gaussian distribution.

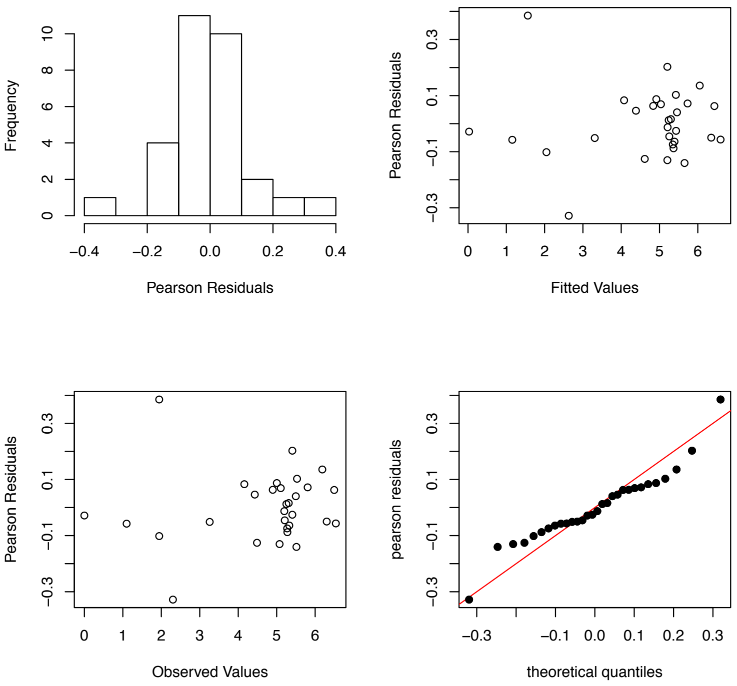
Additionally, different models were fitted based on the number of knots used to smooth the trend in both outcomes when estimation was via restricted maximum likelihood. We also used a built-in smoothing parameter estimation method "GCV.Cp" to estimate the number of knots in each of the models. To assess the goodness of fit of each model, R outputs the adjusted R^2, the percentage of the deviance explained, the generalized cross validation (GCV) score (lower values produce better fitted models), and a test to check the appropriateness of the number of knots in the model.

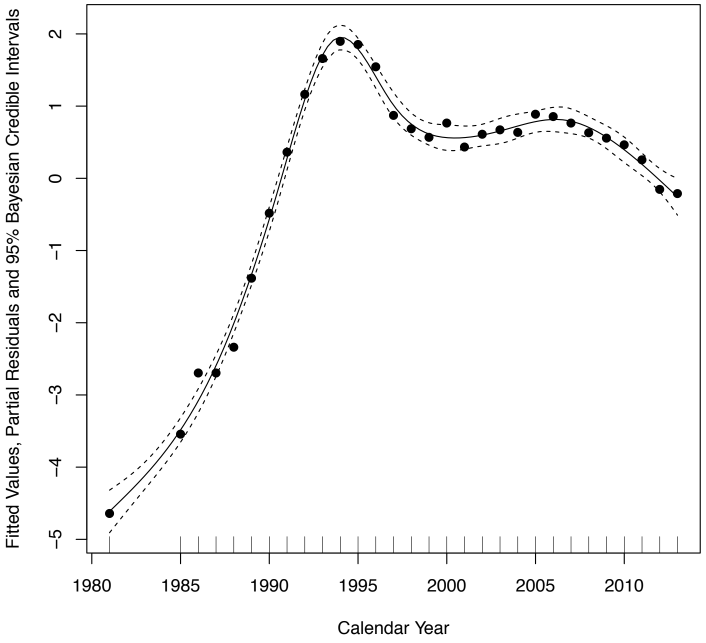
Next, we present the table with the diagnostic statistics for the final model of each outcome as outlined above.

Coefficients/Diagnostic Statistics	**Number of ADIs**	**Proportion of deaths due to an ADI**
Model Parameters		
Significance of spline term for calendar year (p-value)	0.0000	0.0000
Goodness of fit statistics		
R^2	0.9910	0.9730
Deviance explained	99.40%	98.10%
Generalized cross validation (GCV) score	0.0333	0.0019
Sum of squared Pearson residuals	0.0000	0.0000

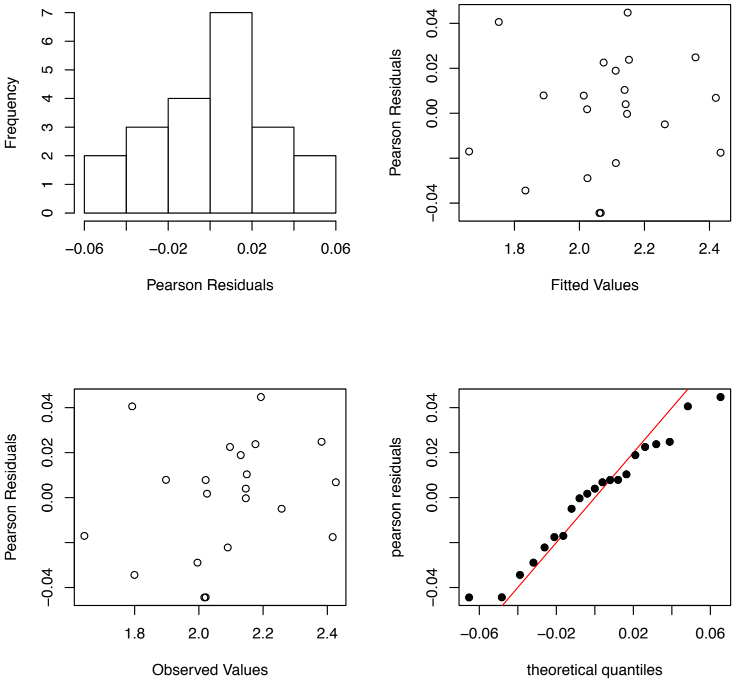
The diagnostic plots for each of the models are presented next. In summary, these plots indicate that we obtained an excellent fit for each of our outcomes.

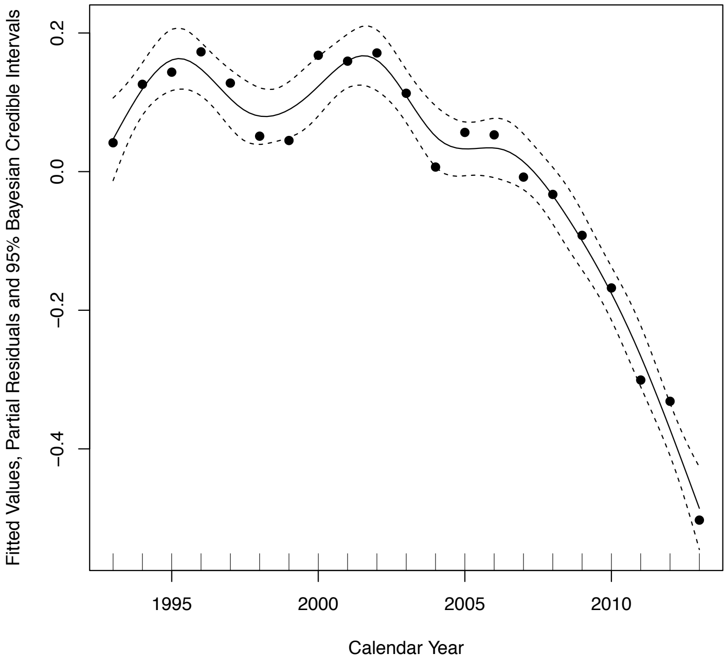
Model Diagnosis Plots for the Number of ADIs





Model Diagnosis for the Proportion of Deaths due to an ADI





References

1. Cameron AC, Trivedi PK. Regression analysis of count data. UK: Cambridge University Press; 1998.
2. Cameron AC, Trivedi PK. Microeconometrics: Methods and Applications. UK: Cambridge University Press; 2005.
3. Clark M. Generalized Additive Models: Getting Started with Additive Models in R 2013. Available from: http://www3.nd.edu/~mclark19/learn/GAMS.pdf.
4. Wood SN. Generalized Additive Models. An Introduction with R. US: Chapman & Hall/CRC; 2006.

	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	Total
All	2	0	0	0	3	7	7	10	26	64	150	334	547	696	665	491	253	207	187	225	163	192	203	197	252	244	223	197	183	167	136	90	84	6205
Men	0	0	0	0	3	7	6	9	25	60	146	315	518	659	616	441	215	181	160	199	131	159	174	169	197	201	190	150	143	126	108	73	62	5443
Women	1	0	0	0	0	0	1	1	1	4	3	17	27	32	46	45	33	25	23	23	29	32	29	27	55	43	33	45	37	39	26	16	22	715
<30	1	0	0	0	2	6	2	6	4	17	26	61	55	62	66	45	25	20	13	16	15	14	12	14	13	22	7	14	14	10	12	3	2	579
30–39	0	0	0	0	1	0	3	3	14	37	80	164	257	365	299	247	135	80	82	94	69	68	83	58	84	66	56	50	42	40	31	9	18	2535
40–49	0	0	0	0	0	1	0	1	8	10	35	89	187	192	218	132	59	83	63	69	43	69	56	89	96	102	112	81	74	59	54	41	29	2052
≥50	0	0	0	0	0	0	2	0	0	0	8	18	46	72	79	62	29	23	25	44	33	40	52	35	59	54	48	50	51	56	37	36	35	994

Table: Number of ADIs each year

Data on sex and age group not available for all ADIs. ADI=AIDS-defining illness.

	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	Total
Not caused by AIDS	0	0	0	0	0	0	10	39	65	119	147	173	75	44	31	11	20	38	42	9	11	7	28	56	55	51	72	72	97	102	133	145	174	1826
Caused by AIDS	0	0	0	0	0	0	0	0	0	0	0	1	181	267	261	241	125	99	104	156	140	141	140	105	150	135	123	106	105	79	63	62	44	2828
% caused by AIDS												1%	71%	86%	89%	96%	86%	72%	71%	95%	93%	95%	83%	65%	73%	73%	63%	60%	52%	44%	32%	30%	20%	

Table: Number and proportion of deaths caused by AIDS