BMJ Open

Study of depression influencing factors with zero-inflated regression models in a large-scale population survey

Tao Xu,[1] Guangjin Zhu,[2] Shaomei Han[1]

ABSTRACT

Objectives The number of depression symptoms can be considered as count data in order to get complete and accurate analyses findings in studies of depression. This study aims to compare the goodness of fit of four count outcomes models by a large survey sample to identify the optimum model for a risk factor study of the number of depression symptoms.

Methods 15 820 subjects, aged 10 to 80 years old, who were not suffering from serious chronic diseases and had not run a high fever in the past 15 days, agreed to take part in this survey; 15 462 subjects completed all the survey scales. The number of depression symptoms was the sum of the 'positive' responses of seven depression questions. Four count outcomes models and a logistic model were constructed to identify the optimum model of the number of depression symptoms.

Results The mean number of depression symptoms was 1.37±1.55. The over-dispersion test statistic O was 308.011. The alpha dispersion parameter was 0.475 (95% CI 0.443 to 0.508), which was significantly larger than 0. The Vuong test statistic Z was 6.782 and the P value was <0.001, which showed that there were too many zero counts to be accounted for with traditional negative binomial distribution. The zero-inflated negative binomial (ZINB) model had the largest log likelihood and smallest AIC and BIC, suggesting best goodness of fit. In addition, predictive probabilities for many counts in the ZINB model fitted the observed counts best.

Conclusions All fitting test statistics and the predictive probability curve produced the same findings that the ZINB model was the best model for fitting the number of depression symptoms, assessing both the presence or absence of depression and its severity.

Strengths and limitations of this study

► This study explored methods of constructing Poisson, NB, ZIP and ZINB models for the number of depression symptoms and compared the fitting goodness of four count outcomes regression models.
► The alpha dispersion parameter, O test and Vuong test showed over-dispersion and excessive zeroes for the number of depression symptoms.
► The likelihood ratio statistic and predictive probability curve suggested that the ZINB model was the best model for fitting the number of depression symptoms.
► The ZINB model could provide more accurate information about the risk factors for depression status than the logistic model and other count outcomes models.
► Categorising of the count data would lead to loss of some useful information; therefore, logistic regression was not the appropriate model for count outcomes study.

[1]Department of Epidemiology and Statistics, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences & School of Basic Medicine, Peking Union Medical College, Beijing, China
[2]Department of Physiopathology, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences & School of Basic Medicine, Peking Union Medical College, Beijing, China

Correspondence to
Professor Shaomei Han;
hansm1@vip.sina.com

INTRODUCTION

In statistics, count data are a type of data in which the observations can take only the non-negative integer values {0, 1, 2, 3, … }, and where these integers arise from counting rather than ranking.[1] Count data are commonly encountered in medical studies, such as the number of depression symptoms, dental caries, adverse events of clinical trials, physical activity days, etc. During statistical treatment, they are usually considered as continuous outcomes or transferred to dichotomous data. However, being treated as continuous data, count data are often extremely concrete and do not follow normal distribution. Therefore, arithmetic mean and standard deviation are not applicable statistics, and linear regression is therefore not an appropriate analytical method due to skewed distribution and over-dispersion. Moreover, count data are different from dichotomous data, in that the observations can take only two values, usually represented by 0 and 1. The categorisation of count data to be used in crude rate and logistic regression will lead to loss of information. Furthermore count data are different from ordinal data, which may also consist of integers, but where the individual values fall on an arbitrary scale and only the relative ranking is important. Hence, treating count data as a continuous variable in linear regression or dichotomous variable in logistic regression models is likely to bias the results.[2 3]

In view of these limitations, Poisson regression or negative binomial (NB) regression

are commonly used to model count outcomes assuming Poisson distribution or negative binomial distribution are applicable distributions. But probability of zeroes based on Poisson distribution or negative binomial distribution cannot account for excess zero counts. Neglect of excess zeroes will bias the estimation of parameters.[4] Zero-inflated (ZI) regression models consider the raw dataset as a mixture of an all-zero subset and another subset following Poisson distribution or negative binomial distribution. A ZI model has been the best model so far to solve this issue in relation to excess zeroes.[5–15]

Depression is usually assessed with some scales, which refer to the number and severity of depression symptoms. In general, participants will be categorised into two or several categories based on their positive depression symptom items. Prevalence rate and logistic regression are used to study the incidence intensity and risk factors of depression.[16–20] These traditional analysis methods are vulnerable to loss of information because every depressive subject may have different numbers of depressive symptoms, resulting in an inability to assess the severity of depression. This study aims to compare the goodness of fit of several count outcomes models—Poisson model, NB model, zero-inflated Poisson (ZIP) model and zero-inflated negative binomial (ZINB) model—by a large-scale cross-sectional sample to identify the optimum model of depression symptom study.

METHODS
Sample and participants
The sample was part of a large-scale population survey about Chinese subjects' physiological and psychological constants, supported by the basic performance key project by the Ministry of Science and Technology of the People's Republic of China. This survey was conducted in Yunnan Province, southwest of China, and the two-stage cluster sampling method was used. First, two cities were sampled, then several communities and villages were randomly selected in the cities. In these selected communities and villages, all eligible people were referred to as our survey subjects who were aged 10 to 80 years old, were not suffering from serious chronic diseases, and had not run a high fever in the past 15 days. All subjects signed informed consent forms. The study was approved by the review board of the Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences (ethics approval number 005–2008).

Depression assessment
Trained medical professionals carried out the survey and interviews. Before the survey, they were trained about the depression assessment scale. The depression assessment scale was designed based on the Composite International Diagnostic Interview Short Form for Major Depression (CIDI-SFMD).[21 22] Subjects were asked if there was ever a time when they felt sad, blue or depressed for 2 weeks or more in a row during the past 12 months. Seven questions were asked about whether they had lost interest in things, felt tired or low energy, gained or lost weight, had more trouble falling asleep or concentrating than usual, thought a lot about death and had a feeling of worthlessness. The number of depression symptoms was the sum of 'positive' responses of these seven depression questions (range 0–7), which was the main outcome measure.

Potential risk factors of depression symptoms in the models included age, sex, hypertension status, occupation, tobacco smoking, alcohol consumption, nationality, marital status, obesity, stress at work or home, and positive or negative life events. Negative events included loss of job, retirement, loss of crop/business failure, household break-in, marital separation/divorce, other major intra-family conflict, major personal injury or illness, violence, death of a spouse, death/major illness of another close family member or other major stress. Positive events were wedding of a family member, new job or birth in the family.

Analysis methods
Poisson regression, NB regression, ZIP model and ZINB model were constructed and their goodness-of-fit were compared. These four models were useful for count outcomes. ZI models were first introduced by Lambert to account for excess zero counts.[23] Cheung mentioned that ZI regression models can be interpreted as reckoning a two-step disease regression.[4] At the beginning subjects are not at risk, so they have zero counts. The influence of covariates may move them into the at-risk population and the outcomes follow a Poisson or NB regression distribution. A covariate may or may not have the same impact on the outcome distribution in the two steps.[4]

ZI models are two-part models, consisting of both binary and count model sections in order to account for excess zero counts.[24] The ZIP model refers to raw dataset as a mixture including an all-zero subset and a subset following Poisson distribution.[23] The ZIP model supposes that:

$$p(y_i) = \begin{cases} \pi_i + (1-\pi_i)e^{-\mu} & y_i = 0 \quad \text{logit section} \\ (1-\pi_i)\dfrac{e^{-\mu}\mu^{y_i}}{y_i!} & y_i \geq 1 \quad \text{poisson section} \end{cases}$$

At the same time, the ZINB model refers to raw dataset as a mixture including an all-zero subset and a subset following NB distribution.[24] The probability density function of the ZINB model is:

$$p(y_i) = \begin{cases} \pi_i + (1-\pi_i)\left(\dfrac{1}{1+\alpha\mu}\right)^{\alpha^{-1}} & y_i = 0 \quad \text{logit section} \\ (1-\pi_i)\dfrac{\Gamma(y_i+\alpha^{-1})(\alpha\mu)^{y_i}}{y_i!\Gamma(\alpha^{-1})(1+\alpha\mu)^{y_i+\alpha^{-1}}} & y_i \geq 1 \quad \text{NB section} \end{cases}$$

Ln and logit link functions were used for parameters μ and π_i. $\ln(\mu) = B_i\beta$; $\text{logit}(\pi_i) = \ln[\pi_i/1-\pi_i] = G_i\gamma$. In the logit section, the explanations of regression coefficients are similar to those in logistic regression. In the

Table 1	Characteristics of respondents		
Characteristics		**n**	**%**
Sex	Male	6601	42.70
	Female	8861	57.31
Occupation	Mental	10133	65.53
	Physical	5329	34.47
Nationalities	Han	10527	68.08
	Yi	3855	24.93
	Others	1080	6.98
Marital status	Married	5089	35.92
	Single	8825	62.29
	Widowed or divorced	254	1.79
Hypertension	Yes	2453	15.86
	No	13009	84.14
Obesity	Yes	667	4.31
	No	14795	95.69
Tobacco smoking	Yes	1953	12.63
	No	13509	87.27
Alcohol drinking	Yes	1706	11.03
	No	13756	88.97
Stress	Yes	1361	8.80
	No	14102	91.20
Positive events	Yes	644	4.17
	No	14818	95.83
Negative events	Yes	3274	21.17
	No	12188	78.83

Table 2	Proportions and predictive probabilities of each counts (%)				
Count	**Observed**	**Poisson**	**NB**	**ZIP**	**ZINB**
0	39.28	28.10	36.89	39.22	39.04
1	23.74	33.19	28.02	20.63	22.67
2	15.23	21.61	16.40	18.67	17.64
3	10.38	10.42	8.83	11.75	10.58
4	6.33	4.24	4.62	5.84	5.46
5	3.21	1.58	2.41	2.47	2.58
6	1.40	0.56	1.27	0.94	1.15
7	0.43	0.20	0.68	0.33	0.50

NB, negative binomial; ZINB, zero-inflated negaitve binomial; ZIP, zero-inflated poisson.

life events were reported by 4.17% and 21.17% of respondents, respectively.

The second column of table 2 presents the observed distribution of the number of depression symptoms. Among the total of 15462 respondents, 39.28% reported no depression symptoms. The larger number of symptoms means the lower proportion of respondents. The mean number of depression symptoms was 1.37±1.55. The variance was greater than the mean. The over-dispersion test statistic O was 308.011 and the P value was <0.001. Furthermore, the alpha dispersion parameter was 0.475 and 95% CI of α was 0.443 to 0.508, which was significantly larger than 0. So the number of depression symptoms was over-dispersed. The Vuong test statistic Z was 6.782, and the P value was <0.001, which suggested that there were too many zero counts to be accounted for with traditional negative binomial distribution. Table 3 demonstrates the fitting goodness of four regression models. ZINB model had the largest log likelihood and the smallest AIC and BIC, suggesting best goodness of fit. The Poisson regression model fitted the data worst.

The last four columns of table 2 presented the predictive probabilities for each count in four regression models. Figure 1 shows the predictive probabilities distribution curve of four regression models and the observed proportions. From table 2 and figure 1, it was clear that the Poisson regression model fitted worst, in which the predictive probability of each count was significantly different from the observed proportion. The NB

Poisson or NB sections, the explanations are the same as in the traditional Poisson or NB regression models.

In this study, SAS version 9.2 was used for the regression model construction. The alpha dispersion parameter and O test were used to identify the over-dispersion.[25] The Vuong test was conducted to judge whether there were excessive zero counts.[24 26] The fitting goodness of regression models were determined by the predictive probability curve of each count, and the likelihood ratio test statistics: log-likelihood, AIC (Akaike information criterion) and BIC (Bayesian information criterion). A logistic regression model was also conducted.

RESULTS

A total of 15820 subjects agreed to take part in this survey, of which 15462 subjects completed all the survey scales. The response rate was 97.7%. The mean age was 26.7±16.7 years (range 10– 80 years). Other characteristics of the sample are shown in table 1: 57.31% of respondents were female; about a quarter of respondents were Yi nationalities; 8.8% of respondents felt psychological stress at work or home; and both positive and negative

Table 3	Fitting goodness statistics of regression models		
Models	**Log likelihood**	**AIC**	**BIC**
Poisson	−25012	50053	50168
NB	−24128	48289	48411
ZIP	−23912	47884	48113
ZINB	−23854	47771	48008

AIC, akaike's information criterion; BIC, bayesian information criterion; NB, negative binomial; ZINB, zero-inflated negaitve binomial; ZIP, zero-inflated poisson.
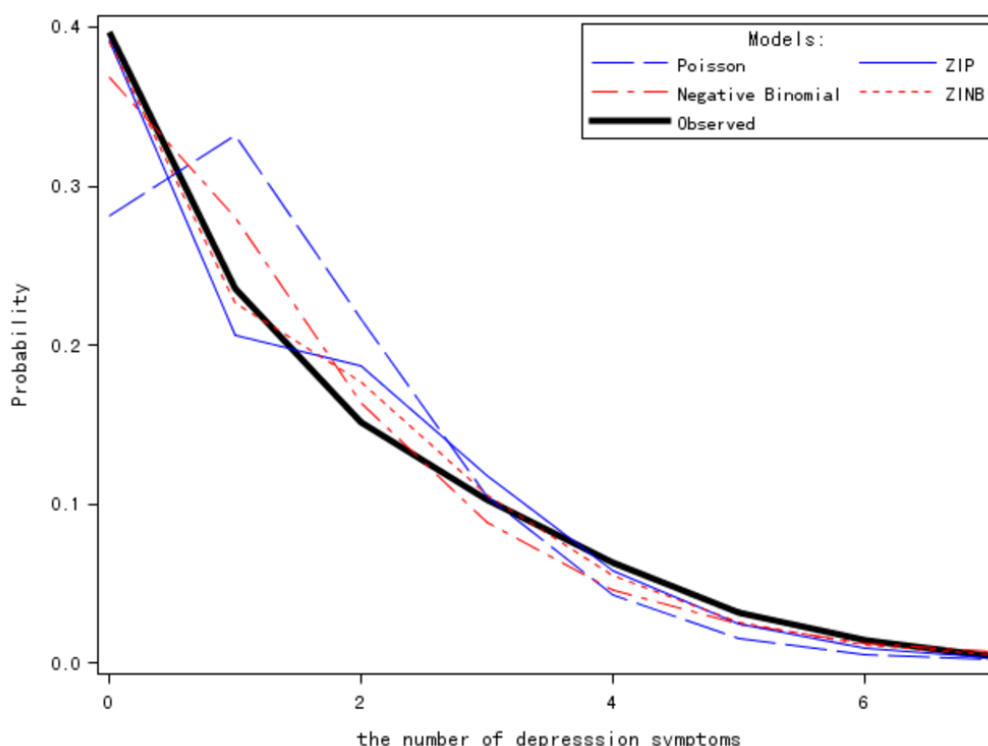
Figure 1 Predictive probabilities of the four models. ZINB, zero-inflated negative binomial model; ZIP, zero-inflated Poisson model.

model was a little better than the Poisson model. The ZIP and ZINB models fit the data better and the predictive probability for zero count of the two ZI models was very close to the observed proportion, especially in the ZINB model. With the exception that the predictive probability of 2 was a little larger than the observe count, the probabilities for the other counts in the ZINB model fitted the observed counts very well. However, the ZIP model predicted fewer 1s and more 2s and 3s.

Based on the alpha dispersion parameter, over-dispersion O test, Vuong test, fitting goodness statistic and the predictive probabilities for counts, the ZINB model was the optimum model for fitting the number of depression symptoms.

Regression coefficients of the ZINB model are shown in table 4. The logit section on the left side of the table is only for zero count. It was clear that sex, occupation, alcohol drinker, Yi nationality, single status, stress, and positive or negative events were risk factors for whether any depression symptoms were encountered or not. Female respondents, mental labourers, alcohol drinkers, Yi nationality, single status, respondents suffering from stress, and respondents with positive or negative events were more at risk for depression. The negative binomial section on the right side showed that age, sex, occupation, alcohol drinker, stress and negative events had a significant effect on the severity of depression. Female respondents, mental labourers, alcohol drinker, single status, and respondents suffering from stress or negative events reported more symptoms

of depression. However, older individuals had a smaller number of depression symptoms.

Table 5 shows the logistic regression model coefficients for risk factors of depression. Female, younger age, mental labourers, alcohol drinkers, Yi nationality, single, widowed or divorced, obesity, stress, and positive or negative events were associated with increased odds of reporting one or more depression symptoms.

DISCUSSION

This study explored methods of constructing Poisson, NB, ZIP and ZINB models for the number of depression symptoms and compared the goodness of fit of four count outcome regression models.

Over-dispersion and terrible skewed distribution reduced the utility of linear regression for count outcomes. Traditional Poisson regression and negative binomial regression were the common models for count outcomes. However, strict limitation of variance equalling the mean resulted in it being very difficult for over-dispersed count data to follow a Poisson distribution. With the error item of gamma function, NB distribution allows for the over dispersion. But excessive zero counts had a bad effect on the Poisson regression and NB regression models. ZI models were introduced for resolving excessive zeroes. ZI models provide assessment of the risk factors of depression severity and not just the presence or absence of depression, because ZI models can model depression in a continuum instead of the dichotomous outcome. In

Table 4 ZINB regression coefficients for the number of depression symptoms

	Logit section				Negative binomial section			
	β	Z	P	95% CI for β	β	Z	P	95% CI for β
Age	−0.003	−0.850	0.396	−0.010 to 0.004	−0.004	−4.100	<0.0001	−0.007 to −0.002
Sex (female)	−0.256	−3.150	0.002	−0.415 to −0.097	0.131	6.220	<0.0001	0.090 to 0.173
Hypertension	−0.085	−0.810	0.419	−0.292 to 0.121	−0.002	−0.070	0.944	−0.068 to 0.063
Mental labourers	−0.804	−6.530	<0.001	−1.045 to −0.563	0.094	2.830	0.005	0.029 to 0.159
Smoker	0.158	1.380	0.169	−0.067 to 0.384	0.017	0.450	0.652	−0.056 to 0.090
Alcohol drinker	−0.377	−3.060	0.002	−0.619 to −0.136	0.087	2.510	0.012	0.019 to 0.155
Yi nationality	0.698	9.010	<0.001	0.547 to 0.850	0.016	0.680	0.497	−0.031 to 0.063
Other race	0.007	0.040	0.969	−0.329 to 0.342	0.012	0.330	0.744	−0.059 to 0.082
Widowed or divorced	−0.642	−1.820	0.068	−1.332 to 0.048	0.017	0.210	0.833	−0.144 to 0.179
Single	−0.445	−4.370	<0.001	−0.644 to −0.245	−0.005	−0.190	0.852	−0.057 to 0.047
Obesity	−0.150	−0.880	0.379	−0.483 to 0.184	0.078	1.620	0.106	−0.017 to 0.174
Stress	−0.997	−5.820	<0.001	−1.333 to −0.661	0.472	18.620	<0.001	0.422 to 0.522
Positive events	−2.179	−2.010	0.045	−4.306 to −0.053	−0.040	−0.910	0.364	−0.127 to 0.047
Negative events	−0.449	−4.460	<0.001	−0.646 to −0.252	0.250	11.990	<0.001	0.209 to 0.290
Intercept	−0.052	−0.250	0.806	−0.462 to 0.359	0.260	4.270	<0.001	0.141 to 0.380

ZINB, zero-inflated negaitve binomial.

particular, ZINB models can resolve both over dispersion and excessive zeroes in the same time.

In this study, the *O* test, Vuong test, AIC and BIC statistic and predictive probability curve indicated that the ZINB model was the best model for the number of depression symptoms with about 40% zero counts. The ZIP model fitted the data worse than the ZINB model perhaps because the over dispersion of the number of depression symptoms restricted the utility of the ZIP model. Many studies reported similar results that the ZINB model was the best model for count outcomes.[8 12–15] However, in a physical

activity study and another depressive symptoms study, ZIP was considered a better model than the ZINB model.[27 28]

In the ZINB model, the influence of risk factors on depression can be assessed by two aspects: whether or not there is depression, and the severity of depression. The logistic regression model reported different risk factors for depression from the ZINB model, especially for obesity and widowed or divorced status. In the ZINB model, obesity and widowed or divorced status were not found to have a strong effect on depression symptoms, although P values approached 0.05 (P=0.106, P=0.068). In

Table 5 Logistic regression coefficients for depression

	β	Wald	P value	OR	95% CI for OR
Age	−0.005	9.501	0.002	0.995	0.992 to 0.998
Sex (female)	0.286	75.085	<0.001	1.331	1.248 to 1.420
Hypertension	0.027	0.320	0.572	1.027	0.937 to 1.126
Mental labourers	0.476	87.311	<0.001	1.610	1.457 to 1.779
Smoker	−0.066	1.381	0.240	0.936	0.839 to 1.045
Alcohol drinker	0.293	27.855	<0.001	1.340	1.202 to 1.494
Yi nationality	−0.291	64.958	<0.001	0.748	0.697 to 0.803
Other race	0.000	0.000	0.997	1.000	0.893 to 1.120
Widowed or divorced	0.310	6.627	0.010	1.363	1.077 to 1.726
Single	0.184	16.295	<0.001	1.202	1.099 to 1.314
Obesity	0.191	6.522	0.011	1.211	1.045 to 1.402
Stress	1.246	573.265	<0.001	3.476	3.139 to 3.849
Positive events	0.314	18.732	<0.001	1.369	1.187 to 1.578
Negative events	0.575	251.969	<0.001	1.777	1.656 to 1.908

addition, age influenced only the severity of depression and not whether depression was present, and positive life events had the opposite influence. Categorising of the count data would lead to loss of some useful information so that logistic regression was not the appropriate model for count outcomes study.

Several limitations are worth noting. First, the Poisson or NB distributional assumption has no upper limit for the counts. However, in the medical field, the outcome variables might always have a specific upper limit for the counts. In this study, the number of depression symptoms ranged from 0 to 7, which was not spread widely. This might be a reason for the poor goodness of fit for traditional Poisson regression and NB regression models. Second, some potential risk factors were correlated with each other, such as stress and positive or negative life events, but the correlation was not too strong to have a disruptive influence on the results of this study.

Despite these limitations, we can conclude that all fitting test statistics and predictive probability curves produced the same finding that the ZINB model was the best model for fitting the number of depression symptoms, not only assessing the presence or absence of depression but also assessing the severity of depression.

REFERENCES

1. Cameron AC, Trivedi PK. *Regression analysis of count data*. London: Cambridge University Press, 1998:1–3.
2. Hall DB. Zero-inflated Poisson and binomial regression with random effects: a case study. *Biometrics* 2000;56:1030–9.
3. Agresti A. *An introduction to categorical data analysis*. New York: Chichester Wiley, 1996:74–83.
4. Cheung YB. Zero-inflated models for regression analysis of count data: a study of growth and development. *Stat Med* 2002;21:1461–9.
5. Solinas G, Campus G, Maida C, et al. What statistical method should be used to evaluate risk factors associated with dmfs index? Evidence from the National Pathfinder Survey of 4-year-old Italian children. *Community Dent Oral Epidemiol* 2009;37:539–46.
6. Kipnis V, Midthune D, Buckman DW, et al. Modeling data with excess zeros and measurement error: application to evaluating relationships between episodically consumed foods and health outcomes. *Biometrics* 2009;65:1003–10.
7. Wang H, Heitjan DF. Modeling heaping in self-reported cigarette counts. *Stat Med* 2008;27:3789–804.
8. Zaninotto P, Falaschetti E. Comparison of methods for modelling a count outcome with excess zeros: application to Activities of Daily Living (ADL-s). *J Epidemiol Community Health* 2011;65:205–10.
9. Lord D, Washington SP, Ivan JN. Poisson, Poisson-gamma and zero-inflated regression models of motor vehicle crashes: balancing statistical fit and theory. *Accid Anal Prev* 2005;37:35–46.
10. Sheu ML, Hu TW, Keeler TE, et al. The effect of a major cigarette price change on smoking behavior in California: a zero-inflated negative binomial model. *Health Econ* 2004;13:781–91.
11. Denwood MJ, Stear MJ, Matthews L, et al. The distribution of the pathogenic nematode Nematodirus battus in lambs is zero-inflated. *Parasitology* 2008;135:1225–35.
12. Lewsey JD, Thomson WM. The utility of the zero-inflated Poisson and zero-inflated negative binomial models: a case study of cross-sectional and longitudinal DMF data examining the effect of socio-economic status. *Community Dent Oral Epidemiol* 2004;32:183–9.
13. Javali SB, Pandit PV. Using zero inflated models to analyze dental caries with many zeroes. *Indian J Dent Res* 2010;21:480–5.
14. Rose CE, Martin SW, Wannemuehler KA, et al. On the use of zero-inflated and hurdle models for modeling vaccine adverse event count data. *J Biopharm Stat* 2006;16:463–81.
15. Akram K, Pedersen-Bjergaard U, Carstensen B, et al. Frequency and risk factors of severe hypoglycaemia in insulin-treated type 2 diabetes: a cross-sectional survey. *Diabet Med* 2006;23:750–6.
16. Müller H, Rehberger P, Günther C, et al. Determinants of disability, quality of life and depression in dermatological patients with systemic scleroderma. *Br J Dermatol* 2012;166:343–53.
17. Xie RH, Yang J, Liao S, et al. Prenatal family support, postnatal family support and postpartum depression. *Aust N Z J Obstet Gynaecol* 2010;50:340–5.
18. Wiersma JE, Hovens JG, van Oppen P, et al. The importance of childhood trauma and childhood life events for chronicity of depression in adults. *J Clin Psychiatry* 2009;70:983–9.
19. Maly RC, Liu Y, Leake B, et al. Treatment-related symptoms among underserved women with breast cancer: the impact of physician-patient communication. *Breast Cancer Res Treat* 2010;119:707–16.
20. Aragonès E, Piñol JL, Labad A, et al. Prevalence and determinants of depressive disorders in primary care practice in Spain. *Int J Psychiatry Med* 2004;34:21–35.
21. Patten SB. Performance of the Composite International Diagnostic Interview Short Form for major depression in community and clinical samples. *Chronic Dis Can* 1997;18:109–12.
22. Patten SB, Brandon-Christie J, Devji J, et al. Performance of the composite international diagnostic interview short form for major depression in a community sample. *Chronic Dis Can* 2000;21:68–72.
23. Lambert D. Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics* 1992;34:1–14.
24. Joseph M H. *Negative Binomial Regression*. London: Cambridge Univ Press, 2007:173–7.
25. Böhning D, Dietz E, Schlattmann P. Zero-inflated count models and their applications in public health and social science. In: Rost J, Langeheine R, eds. *Applications of latent trait and latent class models in the Social Sciences. 1st edn*. Germany: Munster Waxmann, 1997:333–44.
26. Vuong QH. Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica* 1989;57:307–33.
27. Slymen DJ, Ayala GX, Arredondo EM, et al. A demonstration of modeling count data with an application to physical activity. *Epidemiol Perspect Innov* 2006;3:3.
28. Bandiera FC, Arheart KL, Caban-Martinez AJ, et al. Secondhand smoke exposure and depressive symptoms. *Psychosom Med* 2010;72:68–72.