

Dependency of the Distribution of Salmonella-Specific Antibodies SP-Ratios on Weight and Sampling Time

Isaac Akpor Adjei^{1,*}, Md. Rezaul Karim², Rachid Muleia³, Peter Jouck⁴

¹Department of Mathematics, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana

²Department of Statistics, Jahangirnagar University, Dhaka, Bangladesh

³Department of Mathematics and Informatics, Faculty of Science, Eduardo Mondlane University, Maputo, Mozambique

⁴FPS Health, Food Chain Safety and Environment, Brussels, Belgium

Email address:

isaac.adjei@gmail.com (I. A. Adjei)

*Corresponding author

To cite this article:

Isaac Akpor Adjei, Md. Rezaul Karim, Rachid Muleia, Peter Jouck. Dependency of the Distribution of Salmonella-Specific Antibodies SP-Ratios on Weight and Sampling Time. *American Journal of Theoretical and Applied Statistics*. Vol. 5, No. 3, 2016, pp. 87-93.

doi: 10.11648/j.ajtas.20160503.12

Received: March 28, 2016; **Accepted:** April 14, 2016; **Published:** April 26, 2016

Abstract: Salmonella is one of the major sources of toxin-infection in humans worldwide and is due to transmission of pathogens via pork. This paper aims at investigating the effect of season and animal weight on SP-ratio. Pigs were sampled from different herds and SP-ratios were measured and categorized into two different groups. Depending on the categorization of the response and whether or not clustering is taken into account, different binary logistic and multicategory logit models were considered. Without taking clustering into account, ordinary logistic regression, adjacent logit, continuation-ratio logit and proportional odds model were fitted. GEE and GLMM were considered to correct for the herd-effect. Among the multicategory logit models, the proportional odds model is preferred, since it did not reject the assumption of common slopes. However, regarding the goodness-of-fit test, this model did not adequately fit the data. Both GEE and GLMM have their advantages, depending on the specific focus and question of interest. In all models, the interaction between weight and season was not significant. Weight was found significant, while season was insignificant in all models. As it was expected, weight as indicator for age was found to have a significant effect on SP-ratios.

Keywords: Salmonella, SP-ratios, Logit Models, GEE, GLMM, Herd-Effect

1. Introduction

A foodborne disease is when a person eats food containing harmful microorganisms, which then grow in the intestinal tract and cause illness. Some bacteria, all viruses, and all parasites cause foodborne illness via infection. The foodborne bacteria that cause infection are: *Salmonella spp.*, *Listeria monocytogenes*, *Campylobacter jejuni*, *Vibrio parahaemolyticus*, *Vibrio vulnificus*, and *Yersinia enterocolitica* [1]. *Salmonella* is an important foodborne pathogen worldwide due to its associated public health concerns and estimated cost [2].

Salmonella is one of the major sources of toxin-infection in humans worldwide. The incidence of human *salmonellosis* has considerably increased over the past 20 years and this can

be largely attributed to epidemics of *Salmonella enteritidis* phage type 4 in poultry in numerous countries. *Salmonella* infection is usually caused by eating raw or undercooked meat, poultry, eggs or egg products. The incubation period ranges from several hours to two days. Possible signs and symptoms include: nausea, vomiting, abdominal cramps, diarrhea, fever, chills, headache, blood in stool. It is a major concern in most industrialized countries having a significant economic impact [3]. Global estimations vary between 14 and 120 cases of *Salmonellosis* per 100,000 people [4]. The majority of the cases of *Salmonellosis* is due to *Salmonella Enteritidis* and *Salmonella Typhimurium* infections, which comprised almost 80% of the total number of *Salmonella* infections in Belgium in 2005 [5]. Of the reported *Salmonella* infections 26% of these are likely to be due to transmission

of the pathogen via pork. Although pork is less associated with foodborne illness than other meat sources, it remains significant due to its large consumption in a variety of products. Pork is the most consumed meat in the world [6]. *Salmonella enterica* infections are transmitted not only by animal-derived foods but also by vegetables, fruits, and other plant products [7]. In 2009, *Salmonella* was the most commonly reported bacteriological agent of human foodborne disease in the USA, causing approx. 44% of confirmed foodborne bacterial infections [8]. Foodborne pathogens are a major contributor to human illnesses, hospitalizations, and deaths each year. The Centers for Disease Control and Prevention (CDC) estimates that 47.8 million illnesses and 3000 deaths are caused by foodborne pathogens each year [9].

The main objective of this paper is to examine the dependency of the SP-ratios on the estimated animal weight and the time of recording (season). In section 2, description of the Salmonella data is provided, followed by the methodology in section 3. Next, results of the study are presented in section 4. Finally, main findings and conclusions are described, together with some suggestions for further research.

2. Salmonella Data

Blood samples from 1402 pigs were taken and the dataset which consist of the herd identification number, weight of the pig and sampling time were recorded. Pigs were sampled from different herds so the observations were assumed to be independent.

Depending on the size of a herd, a number of samples of pigs from each professional meat producing herd of different weight categories were collected each 3 to 4 months per year. Blood samples were taken and analyzed using an indirect enzyme-linked immunosorbent assay (ELISA) in order to measure sample-to-positive ratios (SP-ratios). Normally, SP-ratio ranges between 0 and 4, but higher values can be observed. In this aspect, SP-ratio is positively skewed. In addition, confounding factors such as seasonal effects and animal age are known with higher expected SP-ratios during the summer and for older animals [10].

$$\text{logit}[\Pr(y_{1i} = 1)] = \alpha + \beta_1 \text{Spring}_i + \beta_2 \text{Summer}_i + \beta_3 \text{Autumn}_i + \beta_4 \text{Weight}_i \quad (1)$$

In order to check if the interaction between season and weight on SP-ratio is significant, a likelihood ratio test was conducted. Since the nature of the data were clustered in herds, generalized estimating equations (GEE) for the marginal model and generalized linear mixed model (GLMM) by taking a herd-specific intercept, were both considered. GEE is a multivariate extension of quasi-likelihood and was used to estimate the average SP-ratio over all pigs, by assuming a working correlation structure. The independent and exchangeable working correlation structures were considered in the GEE estimation. However, even when

3. Methodology

3.1. Variables

All variables in the dataset were considered as categorical. SP-values are recoded in two different settings. The first takes the SP-values as a binary response:

$$y_{1i} = \begin{cases} 1 & \text{if } SP\text{-ratio} > 0.5 \text{ Salmonella positive} \\ 0 & \text{if } SP\text{-ratio} \leq 0.5 \text{ Salmonella negative} \end{cases}$$

As the infection status of an animal with SP-ratio between 0.5 and 1 is doubtful, a middle category was considered leading to the second setting for trinomial response:

$$y_{2i} = \begin{cases} 1 & \text{if } SP\text{-ratio} \leq 0.5 \text{ Salmonella negative} \\ 2 & \text{if } 0.5 < SP\text{-ratio} \leq 1 \text{ inconclusive} \\ 3 & \text{if } SP\text{-ratio} > 1 \text{ Salmonella positive} \end{cases}$$

The time variable “days” was categorized into seasons, with winter as reference category:

$$y_{3i} = \begin{cases} 1, & \text{if } 0 \leq \text{days} \leq 90 \text{ for Winter} \\ 2, & \text{if } 91 \leq \text{days} \leq 181 \text{ for Spring} \\ 3, & \text{if } 182 \leq \text{days} \leq 273 \text{ for Summer} \\ 4, & \text{if } 274 \leq \text{days} \leq 365 \text{ for Autumn} \end{cases}$$

Because of the natural ordering of the weight variable, it was treated as an ordinal categorical covariate (< 40kg, 40 - 59kg, 60 - 80kg and > 80kg) with < 40kg as reference category. The scores for weight variable were calculated by considering mid-value of the range in each class interval. Mid-values 20, 50, 70 and 100 for the last category were considered. The choice of mid-point 100 was based on the fact that the maximum weight for pigs at slaughter was about 110 to 120 kg. Finally, all pigs belonging to a specific herd were indicated by a herd ID.

3.2. Logit Regression Models

Without taking the clustered nature of the data into account, an ordinary logistic regression model for probability of infection $\Pr(y_{1i} = 1)$ with season and weight as explanatory variables was fitted:

the working correlation structure is misspecified, the parameter estimates will still be valid [11]. To define the most appropriate working assumption for the data, the model-based and empirical standard errors were compared for the two working correlation structures. The working assumption, where the standard errors of both empirical and model-based estimates are most close to each other, is considered as being more plausible [11]. In both GEE and GLMM the interaction between season and weight on SP-ratio were tested. Quasi-likelihood with beta-binomial-type variance or inflated binomial variance was not considered

since the covariates were at the level of the pig and not at the level of the herd.

3.3. Multicategory Logit Models

In this section, three types of ordinal multicategorical logit models were considered. Since the baseline logits model treat the response in a nominal way, it was not considered. For these models, the assumption of common slopes was tested in order to possibly simplify the interpretations and to reduce the number of parameters. Loglikelihood ratio (LR) test for

common slopes was performed for the adjacent and continuation-ratio logit models and also score test for proportional odds assumption for the proportional odds model was conducted. The AICs and BICs for these logit models were also obtained.

3.3.1. Adjacent Logit Model

The adjacent logit model for ordinal response was used for comparing one category to adjacent category. The adjacent-categories logit model ($j = 1; 2$) for the data is of the form:

$$\log\left(\frac{\pi_{j+1}}{\pi_j}\right) = \alpha_j + \beta_{1j}Spring + \beta_{2j}Summer + \beta_{3j}Autumn + \beta_{4j}Weight \quad (2)$$

The adjacent-categories logits, like the baseline-category logits, determine the logits for all pairs of adjacent response categories.

3.3.2. Continuation-Ratio Logit Model

This method forms logits for ordered response categories in a sequential manner. A continuation-ratio logit model contrasts each category with a grouping of categories from higher levels of the response scale:

$$\log\left(\frac{\pi_j}{\pi_{j+1} + \dots + \pi_J}\right) = \alpha_j + \beta_{1j}Spring + \beta_{2j}Summer + \beta_{3j}Autumn \quad (3)$$

Continuation-ratio logit models are useful when a sequential mechanism determines the response outcome, in the sense that an observation must potentially occur in category j before it can occur in a higher category [12].

3.3.3. Proportional Odds Model

The proportional (cumulative) odds model is similar to the continuation-ratio logit model, with the advantage that it contrasts each category with a grouping of categories from lower and higher levels of the response scale. In other words, the proportional odds model uses the entire response scale in forming each logit. The model fitted is of the form

$$\log\left(\frac{\pi_{j+1} + \dots + \pi_J}{\pi_1 + \dots + \pi_j}\right) = \alpha_j + \beta_{1j}Spring + \beta_{2j}Summer + \beta_{3j}Autumn \quad (4)$$

For making easy interpretations in terms of odds ratios, we only considered the logit link function instead of probit or loglog link.

3.4. Multicategory Logit Models Accounting for Clustering

To be able to study the effect of weight and season on the multicategorical version of SP ratio and accounting for the correlation within cluster, GEE and GLMM were fitted. The GEE technique and GLMM considering random intercept were done on model (4). GEE model was only fitted using the independent working correlation structure. Other correlation structures such as exchangeable and Autoregressive of order p were not considered due to the fact that SAS only allows for independent working correlation structure. Additionally, GEE was fitted using cumulative logit link. The logit link was used due to its simplicity in the interpretation of the results. Link functions such as probit and cloglog were not considered due to limitation in selection criteria, i.e., GEE models are not based on the full likelihood, so it does not provide AIC that can help in selecting the best model and inability to perform

likelihood ratio test.

With regards to GLMM, to account for the heterogeneity across clusters, herd was considered as a random component. In fitting GLMM, only logit link is considered. In both GEE and GLMM the interaction between weight and season was tested. SAS version 9.4 was used for statistical analysis.

4. Results

4.1. Exploratory Data Analysis

From the random subset of 1402 pigs, it is observed that the SP-ratios range from 0.005 to 3.440, with an average of 0.333 and a standard deviation of 0.472. After dichotomization of the SP-ratios, only 245 pigs (17.48%) have SP-ratio larger than 0.5 and the remaining 1157 pigs (82.52%) have SP-ratio less than or equal to 0.5. Next, only 123 pigs (8.77%) have SP ratio larger than 1 SP-ratio while 122 pigs (8.70%) have a SP-ratio in between 0.5 and 1.

The summary statistics of SP-ratio with respect to weight and season are presented in Table 1. In general an increase in

SP-ratio is noticed when the weight of the pigs increases. Furthermore, the average SP-ratio is highest and smallest during winter and spring seasons respectively.

Table 1. Summary Statistics of SP-ratio by Weight and Season.

Weight	N	Mean	Std Dev	Season	N	Mean	Std Dev
<40 kg	86	0.303	0.624	Winter	307	0.401	0.505
40-59 kg	434	0.244	0.393	Spring	318	0.3	0.425
60-80 kg	474	0.336	0.473	Summer	418	0.325	0.488
>80 kg	408	0.429	0.493	Autumn	359	0.312	0.459

4.2. Logistic Regression Models

With and without taking the herd-effect into account, the interaction between season and weight was found not to be significant. For instance, in the simple logistic regression case, it was not significant (p -value equal to 0.6436). Therefore, the interaction term was ignored and the model was refitted without interaction.

In Table 2, all estimates together with their standard errors are provided for each of the logistic models. The binomial ML model was considered as the best fitted model when clustering is not taken into account. For the clustered setting, GEE with independence working correlation structure and GLMM are both good fits depending on the specific interest: population-averaged or herd-specific interpretation. GEE with exchangeable correlation assumption is not preferred because the discrepancy between the standard errors for model-based and empirical based estimates is larger compared to the GEE analysis with independence working

correlation structure.

It is observed that the estimates are the same for binomial ML and GEE with independence correlation structure. However, the standard errors are larger for GEE compared to ordinary logistic regression and GLMM. In general it does not really affect the estimation of the parameters, because they are close to each other. Additionally, it is observed that the sign of the estimate for autumn parameter is different when taking herd-effect into account in GLMM and GEE (Exch) compared to Binomial ML and GEE under independence. Moreover, as shown in Table A1, season is not significant. Also, weight was highly significant in the simple logistic model (p -value < 0.0001), significant in GEE assuming independence (p -value = 0.0299) and in GLMM (p -value = 0.0056). It is worth mentioning that weight is not significant in GEE with exchangeable working correlation structure whereas it is significant in all other logit models. However, since this model does not adequately fit the data, its results should be interpreted with caution.

Table 2. Parameter Estimates (Standard Errors) for the Logistic Regression Models.

Type of Logit Model	Binomial ML	GEE (indep)	GEE (Exch)	GLMM
Parameter				
Intercept	-2.466 (0.291)	-2.466 (0.700)	-2.149 (0.840)	-2.529 (0.390)
Spring	-0.344 (0.208)	-0.344 (0.353)	-0.260 (0.383)	-0.293 (0.243)
Summer	-0.375 (0.193)	-0.375 (0.424)	-0.151 (0.424)	-0.170 (0.222)
Autumn	-0.321 (0.200)	-0.321 (0.363)	0.057 (0.369)	0.097 (0.235)
Weight	0.016 (0.003)	0.016 (0.007)	0.010 (0.009)	0.011 (0.004)
Overdispersion	None	None	$\hat{\rho} = 0.142$	$\hat{\sigma} = 1.461$

In Table 2 above it can be seen that the parameter estimate for weight is positive in each model, which indicates that the probability of pigs having SP-ratio larger than 0.5, is larger for those pigs having high weight. Although the effect of season is not significant, it was observed that the estimated parameters for the season categories were negative. This indicates that the probability of pigs having SP-ratio larger than 0.5 is high when measured in the winter season, compared to the other seasons. The variance for the random herd-effect in GLMM is estimated to be 1.461. This corresponds to an approximate intra-herd correlation of 0.3075.

4.3. Multicategory Logit Models

From Table A2 in appendix, it is observed that the AIC and the Likelihood ratio test indicate that the Adjacent and

Continuation ratio models with different slopes are preferable. For the proportional odds model AIC and Likelihood ratio test evidently indicate that a model with common slope is best. Moreover, from Table A4 in the appendix, it is seen that for all multicategory logit model not taking clustering into account weight has a significant effect and season is found not to be significant.

In Table 4 below the parameter estimates indicate similar effect for both adjacent and continuation ratio logit model, that is, the sign for the parameter estimates for both models are the same. Furthermore, from these two models, it is seen that pigs with high weight tend to have high odds of the risk for developing the infection. In addition to these, considering the proportional odds model with common slopes in Table 3, similar effect of the covariates on the SP-ratio is observed. Figure A1 in appendix clearly shows that for pigs with SP-

ratio in the second and the third level, there is a high probability of developing the infection with higher weight.

Table 3. Parameter Estimates and Standard Errors for Proportional Odds Model with Common Slopes.

Parameter	Estimate	Std. Error	Z-value	P-value
Intercept 1	-2.4491	0.2891	8.4710	< 0.0001
Intercept 2	-3.2532	0.2974	10.9380	< 0.0001
Spring	-0.3189	0.2068	1.5420	0.1230
Summer	-0.3375	0.1914	1.7630	0.0778
Autumn	-0.3330	0.2001	1.6640	0.0962
Weight	0.0158	0.0032	-4.9360	< 0.0001

Table 4. Parameter Estimates (Standard Errors) for Multicategorical Logit Models with different Slopes.

Multicategorical Ordinal Logit Model		
Parameter	Adjacent	Continuation Ratio
Intercept 1	-3.3366 (0.3995)	-2.4603 (0.2892)
Intercept 2	0.3443 (0.5282)	0.2651 (0.5308)
Spring 1	-0.4338 (0.2790)	-0.3437 (0.2083)
Spring 2	0.1819 (0.3719)	0.1869 (0.3722)
Summer 1	-0.6287 (0.2679)	-0.3737 (0.1930)
Summer 2	0.4790 (0.3467)	0.4920 (0.3469)
Autumn 1	-0.2382 (0.2570)	-0.3209 (0.2002)
Autumn 2	-0.1787 (0.3588)	-0.1532 (0.3591)
Weight 1	0.0193 (0.0044)	0.0162 (0.0032)
Weight 2	-0.0060 (0.0058)	-0.0051 (0.0059)

Table 5. Parameter Estimates (Standard Errors) for the GEE and GLMM using cumulative logit link.

Multicategorical Ordinal Logit Model		
Parameter	GEE	GLMM
Intercept 1	-3.2533 (0.7310)	-3.5100 (0.4045)
Intercept 2	-2.4491 (0.6966)	-2.5990 (0.3967)
Spring	-0.3189 (0.3549)	-0.2592 (0.2420)
Summer	-0.3375 (0.4299)	-0.1245 (0.2222)
Autumn	-0.3330 (0.3571)	0.0478 (0.2341)
Weight	0.0158 (0.0073)	0.0104 (0.0042)
Overdispersion	None	$\hat{\sigma} = 1.574$

4.4. Multicategory Logit Models Accounting for Clustering

Based on the results presented in Appendix Table A6, it is clear that the AIC for GLMM is smaller under the cumulative probit link (1496.45) though there is not a large difference

with the AIC for the model fitted under cumulative logit link (1496.76), showing that there is not a lot of improvement shifting from logit to probit link. Therefore, a model fitted under cumulative logit is preferred. Logit link is preferred due to its simplicity in the interpretation of the results. The GEE model was also fitted under cumulative logit link.

From Table A4 it can be seen that for both GEE and GLMM only weight has a significant effect on SP-ratio. Although the season has an insignificant effect, it was considered in both models. It was observed that for one unit increase in weight the log odds for high SP-ratio increases by 0.0158 units (see Table 3), and conditioning on the herd, the estimated log odds will increase by 0.01031 units. Furthermore, it is observed that the estimates and standard errors under GLMM were smaller compared to GEE. It is also observed that there is a difference in sign for the parameter related to the effect of autumn, that is the parameter estimate in GEE is negative and in GLMM is positive. The variance for the random herd-effect in GLMM is estimated to be 1.574, which corresponds to an approximate intra-herd correlation of 0.3236.

5. Conclusions

Salmonella is one of the major sources of toxic-infection in humans worldwide and mainly transmitted via pork. This paper aimed at investigating the effect of season and animal weight on SP-ratio. Blood samples from pigs were collected from different herds and SP-ratio was measured and categorized under two different settings. One setting with binomial response and the other as trinomial response. Weight of pigs was treated in an ordinal way, while seasons were on a nominal scale. To achieve the objective, different logit models were fitted with and without taking the herd-effect into account. Depending on the categorization of the response, binary logistic and multicategory logit models were considered.

For all models the interaction between weight and season was not significant on SP-ratio. Weight was found to be significant, while season was insignificant in all the models. This is in contrast with the literature where SP-ratios were expected to be larger during the summer [10]. For weight it was observed that SP-ratio increases when the weight of the pig increases. In the case where SP-ratio was dichotomized and taking clustering into account, GEE and GLMM were fitted. Depending on the interest either on the whole population of the pigs or on the level of a particular pig, both models were found to fit adequately well. Among the multicategory logit models, the proportional odds model is preferred, since it did not reject the assumption of common slopes, which gives a model with a reduced number of parameters. However, regarding the goodness-of-fit test, these models did not fit well. GEE and GLMM were again fitted to account for clustering in the multicategorical case.

In summary, based on all analyses, SP-ratio tends to be high for pigs having a high weight and also season has no effect on the SP-ratio.

Appendix

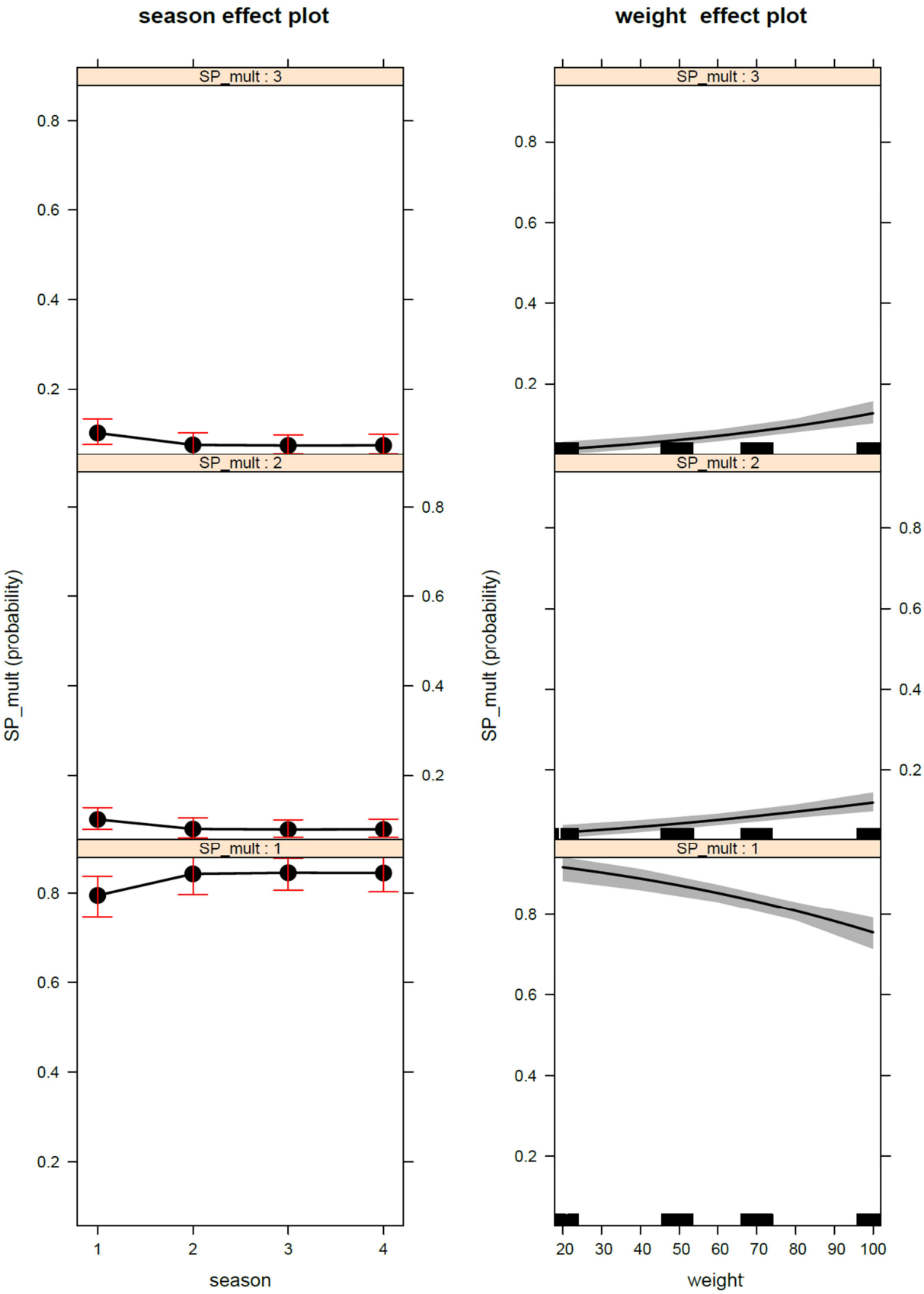


Figure A1. Predicted probabilities using proportional odds model.

Table A1. Analysis of the Effect of Season and Weight for the Logit Regression Models.

Effect	P-value			
	Binomial ML	GEE (Indep)	GEE (exch)	GLMM
Season	0.1919	0.7593	0.7650	0.4189
Weight	< 0.0001	0.0299	0.2345	0.0056

Table A2. The AIC and BIC values of Multicategorical Logistic Models with likelihood ratio test for common slopes.

	Common slopes		Different slopes		Likelihood-ratio
Model	AIC	BIC	AIC	BIC	P-value
Adjacent	1624.465	1655.939	1619.278	1671.735	0.0104
Continuation ratio	1629.53	1661.004	1619.476	1671.932	0.0012
Proportional odds	1617.81	1649.284	1620.059	1672.516	0.2185

Table A3. Score Test for the Proportional Odds Assumption.

Chi-Square	df	P-value
6.24	4	0.1817

Table A4. Analysis of the Effect of Season and Weight for different Multicategory Logistic Models.

Effect	P-value				
	Adjacent	Continuation-Ratio	Proportional Odds	GEE	GLMM
Season	0.2174	0.2227	0.2434	0.777	0.6922
Weight	<0.0001	<0.0001	<0.0001	0.0318	0.0187

Table A5. Deviance and Pearson Goodness-of-Fit Statistics.

Criterion	Value	df	Value/DF	P-value
Deviance	47.2379	26	1.8168	0.0066
Pearson	58.1642	26	2.2371	0.0003

Table A6. AIC values for GLMM (Multicategorical response) for different Link Function Cumulative Link.

Logit	Probit	Cloglog	loglog
1496.76	1496.45	1497.26	1499.89

References

- [1] Chin, J. (2000). Control of communicable diseases manual (Vol. 17). Washington, DC: American Public Health Association.
- [2] Hoelzer, K., Moreno Switt, A. I., & Wiedmann, M. (2011). Animal contact as a source of human non-typhoidal salmonellosis. *Vet res*, 42(1), 34.
- [3] Mead, P. S., Slutsker, L., Dietz, V. McCaig, L. F., Bresee, J. S., Shapiro, C., Griffin, P. M. and Tauxe, R. V. (1999). Food-related illness and death in the United States. *Emerging Infectious Diseases*, 5(5), 607-625.
- [4] Thorns, C. J. (2000). Bacterial food-borne zoonoses. Review Scientific et Technique (International Office of Epizootics), 19(5), 226-239.
- [5] NRRS (2005). National Reference Centre for Salmonella and Shigella: Annual report. Retrieved from http://www.iph.fgov.be/bacterio/iframes/rapports/2004/Salm_2004_NL_cover.pdf.
- [6] Delgado CL, Rosegrant MW, Meijer S. 2001. Livestock to 2020: the revolution continues. Auckland, New Zealand: Intl. Trade Research Consortium (IATRC).
- [7] Baer, Arica A. and Miller, Michael J. and Dilger, Anna C. (2013). Pathogens of Interest to the Pork Industry: A Review of Research on Interventions to Assure Food Safety. *Comprehensive Reviews in Food Science and Food Safety*, 12(2), 183-217.
- [8] CDC (Centers for Disease Control and Prevention): Preliminary Food Net data on the incidence of infection with pathogens transmitted commonly through food - 10 states. *MMWR Morb Mortal Wkly Rep* 2009, 59(14):418-422.
- [9] CDC. 2011. CDC estimates of foodborne illness in the United States. Center for Disease Control. Available from: <http://www.cdc.gov/foodborneburden/2011-foodborne-estimates.html>.
- [10] Bollaerts, K., Aerts, M., Ribbens, S., Van der Stede Y., Boone, I. and Mintiens, K. (2008). Identification of Salmonella high risk pig farms in Belgium using semi-parametric quantile regression. *Journal of Royal Statistical Society (series A)* 171(2), 449-464.
- [11] Molenberghs, G. and Verbeke, G. (2005). Models for discrete longitudinal data. New York: Springer.
- [12] Gerhard Tutz (1991). Sequential models in categorical regression. *Computational Statistics & Data Analysis*, 11(3), 275-295.