Foodborne Outbreaks, Product Recalls, and Firm Learning: A Recurrent Event Survival Analysis of U.S. Meat and Poultry Recalls

SHERZOD B. AKHUNDJANOV[†], VERONICA F. POZO[‡], and BRIANA THOMAS[§]

[†] *Corresponding author. Department of Applied Economics, Utah State University, Logan, UT, USA. Address: 4835 Old Main Hill, Logan, UT 84322-4835. E-mail: sherzod.akhundjanov@usu.edu*

[‡] *Department of Applied Economics, Utah State University, Logan, UT, USA.*

[§] *Department of Applied Economics, Utah State University, Logan, UT, USA.*

Abstract

Firms in the food industry may experience more than one contamination incident over time. In a food safety context, increasing the time between foodborne outbreaks represents a key objective for the food industry and public health officials. We evaluate factors that influence time to next recall for food firms and, more importantly, identify the extent of firm learning from previous food safety incidents in the context of increased time to next recall. Using a recurrent event survival analysis framework, we examine repeated recall data of meat/poultry products in the United States for the period 1994-2015. Our findings indicate more diversified firms incur a smaller risk of repeat recall as firm size expands compared to firms producing primarily meat/poultry products. The hazard of a recall incident decreases with the quantity of meat/poultry recalled in the past recall. The risk of a future recall is smaller for firms whose previous recall was of Class I, the most severe class of recalls, compared to those whose past recall was of Class II. Some evidence of firm learning is found, but there is no definitive evidence indicating that a firm's ability to prevent recalls grows with the number of foodborne outbreaks it has experienced.

Keywords: Food safety; Foodborne outbreaks; Food recalls; Recurrent event survival analysis.

Food safety is an issue of shared responsibility across the food supply chain. Public health officials, policy makers, food firms and consumers, all play an important role in minimizing the risk of a foodborne outbreak. Food firms, in particular, invest in the implementation of food safety technologies and protocols to prevent or reduce the occurrence of these incidents (Henson and Reardon, 2005). Despite such efforts, the number of foodborne outbreaks reported in the United States, particularly those related to meat/poultry products, has been on a constant rise. In 2015, the U.S. Department of Agriculture's (USDA) Food Safety and Inspection Service (FSIS), an agency tasked with monitoring the safety of the U.S. supply of meat, poultry, and egg products, reported 150 recalls, which represents almost threefold increase from the number of recalls issued in 2005 (FSIS, 2017). Surprisingly, some firms have experienced more than one food safety incident within a relatively short amount of time, thus raising the question of what factors determine time to next recall for firms producing meat/poultry products, and to what extent these firms learn from previous food safety incidents.

In this study, our main objective is twofold. First, we aim to determine key factors that influence a food firm's risk of recall occurrence at any given time. Towards this end, both firm- and recall-specific factors are evaluated. Previous studies that examined firms' risk of foodborne outbreaks (Salin et al., 2003; Teratanavat et al., 2005) have adopted statistical techniques that are appropriate for modelling the time to a single failure event, which are inefficient or inappropriate in case of repeated recall incidents (Kleinbaum and Klein, 2005; Amorim and Cai, 2015). We implement a recurrent event survival analysis framework, which accommodates the order of recurring events and accounts for intra-firm correlation arising from these events. This method has been commonly used in epidemiology for recurring time-to-event analysis (Kelly and Lim, 2000; Kennedy et al., 2001).

Second, and importantly, we investigate whether food firms that have issued a recall in the past learn from that experience in the context of *increased* time to next recall, and how this learning is influenced by the firm- and recall-specific characteristics. If firms learn from a previous recall incident, they should in theory be able to *lengthen* the amount of time until the next recall event (by, for instance, implementing more effective food quality monitoring systems), thereby increasing the likelihood of their survivorship (i.e., not experiencing a

recall event). As a result, *inter-event* time can plausibly proxy organizational learning as it reflects the extent of efforts taken by the firm upon a food contamination incident to prevent or reduce the occurrence of future incidents.

METHODS

Data

The data used in this study is obtained from the FSIS and corresponds to meat and poultry recalls issued by publicly traded firms in the United States between 1994-2015. Table 1 presents the number of recalls and total quantity of product recalled by 31 publicly traded firms included in our study. Clearly, our data includes many major processors and marketers of meat/poultry.

To ensure adequate sample size in the estimation of each recurrent event and reliable inferences (Kelly and Lim, 2000; Amorim and Cai, 2015), we focus on the first four recall events per firm. Considering more than four events per firm would increase the variance in our models since there are a few firms in the data with five or more recall events. Our sample includes 77 events, of which 31 correspond to first recall, 19 to second recall, 16 to third recall, and 11 to fourth recall.

We consider a firm's "entry" into the study as either January 1, 1994, the date when the recall data started being collected by the FSIS, or the date of the firm's initial public offering (IPO), whichever came later. Firms that went bankrupt or were acquired prior to the end of the study period are considered to have "dropped out" of the study and are right-censored at the date of the firm's acquisition or bankruptcy filing.

The variable of interest in this study, *Time until Recall*, is measured in months and represents the time until a recall event. To acknowledge the order of recurrent recall events, we classify *Time until Recall* by stratum. In particular, stratum 1 represents the time until the first recall event; stratum 2 is the time between first and second recall; stratum 3 is the time between second and third recall; and stratum 4 corresponds to the duration between third and fourth recall event. Figure 1 visualizes the empirical distribution of *Time until Recall* by stratum.

The firm-specific factors considered in the analysis include firm size, level of diversification, and firm age. The information needed to build these variables was obtained from companies' annual reports and 10-K reports. Figure 2 depicts the box plot of *Time until Recall* by *Diversification* and stratum. The recall-specific factors considered in the analysis include recall class and recall size (pounds recalled). Table 2 provides descriptive statistics and further description of study variables.

Statistical Analysis

There are several statistical methods proposed in the literature that can be used to analyze recurrent time-to-event data (Kelly and Lim, 2000; Amorim and Cai, 2015). The choice between these models largely depends on assumptions about the events of interest, as well as the nature of the research question. In what follows, we discuss two such models and elaborate on their suitability for our purposes.

Counting Process Model

The counting process model used in our analysis was developed by Andersen and Gill (1982), hence referred to as the Andersen-Gill or AG model. This approach assumes that recurrent events within subject are conditionally uncorrelated, given the covariates, and are considered identical. When a firm issues a recall, there is close monitoring by the FSIS throughout the process. If at any point during the process it is determined that more product was affected than initially thought, a recall extension will be issued. Thus, if there was a subsisting problem related to the initial recall event, a subsequent recall would not be issued, it would be addressed with a recall extension (FSIS, 2013). Because of this, it is reasonable to treat each recall event within each firm as an independent event.

This model also assumes that covariates are time independent, i.e., variables do not differ depending on whether they are observed for first recall or a fourth recall. This assumption, which is commonly known as the proportional hazard (PH) assumption, is tested using a proportionality test (Grambsch and Therneau, 1994). If one or more covariates did not satisfy the PH assumption, a stratified Cox PH model discussed in the next section would be a more appropriate approach (Kleinbaum and Klein, 2005). Fitting the Andersen-Gill

model allows us to answer the first question of interest: *What factors affect time to next recall (or the rate at which recalls occur) for food firms?*

Let T denote the random variable for duration or event time and the outcome t is time measured in months. The standard Cox PH model (Cox, 1972) is used to carry out the counting process approach. In particular, the hazard function of this model has the functional form of:

$$h(t, \mathbf{x}) = h_0(t) \exp\left(\sum_{i=1}^{p} \beta_i x_i\right) \tag{1}$$

where $\mathbf{x} = (x_1, \ldots, x_p)'$ is a vector of covariates (firm-specific and/or recall-specific factors) and $h_0(t)$ is the baseline hazard function that describes the risk when $\mathbf{x} = \mathbf{0}$.

When using a Cox PH model with recurrent event data, several time intervals on the same subject must be included in the formulation of the likelihood function used to estimate $h(t, \mathbf{x})$. Importantly, unlike in the standard Cox PH model, subjects do not drop out of the risk set after having failed (i.e., experiencing a recall event) or censored. If subjects display multiple failure times (recall events), they remain in the risk set until last interval is completed; that is, their last failure time or censorship. Hence the partial likelihood function (L) used to fit the Cox PH model is formulated as the product of individual likelihoods (L_j, $j = 1, \ldots, J$) for each ordered unique failure time as follows:

$$L = L_1 \times L_2 \times \cdots \times L_J \tag{2}$$

where J is the total number of unique failure times for all subjects (Kleinbaum and Klein, 2005). In our data set, there is 83 unique failure times for all firms from four recall events. For a single failure at the jth ordered failure time $t_{(j)}$, L_j takes the following form:

$$L_j = \text{Prob(failing at time } t_{(j)} | \text{survival up to } t_{(j)}) = \frac{\exp\left(\sum_{i=1}^{p} \beta_i x_{i(j)}\right)}{\sum_{s \in R(t_{(j)})} \exp\left(\sum_{i=1}^{p} \beta_i x_{is(j)}\right)} \tag{3}$$

where $R(t_{(j)})$ is the risk set for time period $t_{(j)}$ and $x_{i(j)}$ is the value of the variable x_i for subject failing at period $t_{(j)}$.

To handle tied survival times, we use the Breslow approximation method (Breslow,

1974). Further, to account for possible correlation among recurrent events on the same subject, robust standard errors are used for model inference as described by Lin and Wei (1989).

Stratified Cox PH Model

The main advantage of the stratified Cox PH model is that we do not have to assume independence between recurrent events of the same subject. Thus, unlike the counting process model, this method conveniently accommodates the order of the events, which in turn allows the effect of covariates to vary from event to event. Estimating this model helps us answer our second question of interest: *Do we see evidence of firm learning after a recall, and if so, how does firm learning differ between various types of firms (as identified by covariates)?*

There are two different versions of the stratified Cox PH model developed by Prentice et al. (1981), hence referred to as PWP models. The first version is called conditional 1 or the PWP Total Time (PWP-TT) model. This model uses time to events from study entry of each subject. The second version, which uses survival from a previous recall, is called conditional 2 or the PWP Gap Time (PWP-GT) model. In the present study, the PWP-GT stratified Cox PH model is used as we are interested in the time to next recall after a previous recall, as opposed to the time to first, second, third, and fourth recall from study entry (Kleinbaum and Klein, 2005). Besides, comparing the performance of different survival models, including AG, PWP-TT, and PWP-GT, Kelly and Lim (2000) conclude that PWP-GT model is "useful for analyzing recurrent event data".

The stratified PWP-GT model can be fit with or without interaction terms between the stratified variables (i.e., dummies for each stratum) and the predictor variables. If the interest is to estimate the overall effects of covariates, while taking account of the order of events, the model without interactions terms can be pursued. However, to determine differential effects of covariates on each stratum, the interaction model needs to be considered.

The stratified PWP-GT model without interaction terms takes the form of:

$$h_g(t, \mathbf{x}) = h_{0g}(t) \exp\left(\sum_{i=1}^{p} \beta_i x_i\right) \tag{4}$$

where $g = 1, 2, 3, 4$ is the stratum of interest. The same set of covariates are included in $\mathbf{x}$ as those for the counting process model. This model is similar to the specification of the Andersen-Gill model but note that in (4) a different baseline hazard is estimated for each stratum, unlike in the Andersen-Gill model.

The stratified PWP-GT model with interaction terms is specified as:

$$h_g(t, \mathbf{x}) = h_{0g}(t) \exp\left(\sum_{i=1}^{p}\left(\beta_i x_i + \sum_{l=2}^{4} \delta_{il} z_l x_i\right)\right) \tag{5}$$

where z_l, for $l = \{2, 3, 4\}$, is a dummy variable for l'th stratum, with stratum 1 dropped out as the base group (Kleinbaum and Klein, 2005). Interacting each covariate with dummies for each stratum allows to check for the differential effects of firm characteristics on each stratum and hence whether learning from previous recall events depends on firm characteristics. For both PWP-GT models with and without interaction terms, we use the Breslow approximation method (Breslow, 1974) to handle tied survival times, and robust standard errors to control for potential within subject correlation.

The choice between the stratified Cox PH model with and without interaction terms can be ascertained by employing a log-likelihood ratio test of the null hypothesis that δ_{il}'s, for $i = \{1, \ldots, p\}$ and $l = \{2, 3, 4\}$, are not different from zero. The corresponding alternative hypothesis is that at least one of the terms is not equal to zero. A failure to reject the null hypothesis indicates that a more parsimonious, no-interaction stratified Cox PH model, is sufficient.

RESULTS

What Factors Affect Time to Next Recall for Food Firms?

Table 3 reports estimation results for the Andersen-Gill counting process model and the PWP-GT stratified Cox PH model. The diagnostic tests in tables 4 and 5 indicate: (i) the PWP-GT model without interaction terms is more plausible than the PWP-GT model with interaction terms for all the specifications in table 3; and (ii) the PH assumption mostly holds for the Andersen-Gill model, except in specification (5) where it does not hold for

Diversification. Therefore, in table 3 we present the results for the Andersen-Gill model and the PWP-GT model without interaction terms.

From specifications (1) and (2) in table 3, it is evident that *Firm Age* is not a statistically significant factor influencing the risk of recall occurrence. In contrast, estimates for *Diversification* are significant at the 5 percent level across both models, and the sign of the parameter estimate matches expectations. In particular, the hazard of recall for firms whose main output is meat and/or poultry product is about 2.5 times the rate of firms that are more diversified. Further, the coefficient of *Firm Size* is statistically significant at the 10 percent level for the Andersen-Gill model. For every one-million-dollar increase in the size of a firm, the log risk of recall increases by 0.005 percentage points.

A natural question is whether the risk of recall associated with the growth of firm size varies by firm's diversification. To answer this question, the joint effect of *Firm Size* and *Diversification* on firms' *Time until Recall* is investigated in specifications (3) and (4) in table 3. When the interaction term for *Firm Size* and *Diversification* is included along with main effects for these variables, only the interaction term comes out to be statistically significant for both the Andersen-Gill and the PWP-GT models. This suggests that what raises the likelihood of a meat/poultry product recall is not being a large firm in general but being a large meat or poultry firm. The parameter estimate on the interaction term indicates that for every one-million-dollar increase in the size of primarily meat producing firms (*Diversification=1*), the log risk of recall increases by an additional 0.0664 (the Andersen-Gill) or 0.0407 (the PWP-GT) percent relative to those for which meat is not the main product (*Diversification=0*). This shows that firms with a more diversified output line incur smaller risk of recall with the expansion of firm size relative to those producing primarily meat products.

Moving on to recall-specific factors, it is crucial to note that since the amount of meat recalled, as well as recall class, for a given recall are not determinants of that recall event but its consequences, it would be erroneous to include these variables directly in the covariates list. Such variables must be lagged (meaning the loss of a first recall event per firm) in order to study their effect on a subsequent recall event. We provide this analysis in specifications (5) and (6) in table 3.

We can observe that the lag of *Class II* is statistically significant at the 5 percent level across the two models. In addition, the lag of *Class I* is significant at the 10 percent level for the Andersen-Gill model, and nearly significant for the PWP-GT model (p-value=0.127). Our results suggest that the hazard of recall for firms that experienced a Class II recall in the past ranges between 7 (the Andresen-Gill) and 13 (the PWP-GT) times the rate of firms that experienced a Class III recall (a reference category) previously. In contrast, the hazard for firms that suffered a Class I recall in the past is approximately 5 times (the Andersen-Gill) the hazard for firms that reported a Class III recall in the past. Comparing risks for firms that experienced Class I and Class II recalls in the past, it is evident that firms that dealt with the latter have almost twice the risk of those that dealt with the former. This suggests that firms seem to learn more after facing a Class I recall, the most severe category, than when they face a Class II recall.

Additionally, the lag of *Recall Size* is statistically significant at the 1 percent level for the PWP-GT model. The parameter estimate for this variable shows that the hazard of a future recall incident decreases with the quantity of meat recalled in the past recall event. Specifically, for every one-million-pounds of meat recalled in the past, the log risk of the next recall for a firm decreases by 3.656 percent, which is also suggestive of potential firm learning.

Do Firms Learn from Previous Food Safety Incidents?

Figure 3 depicts the estimated survival functions for the Andersen-Gill counting process model (panel (a)) and the PWP-GT model (panel (b)). In this context, the estimated survival function conveys the probability of surviving (i.e., no recall event) longer than t periods, i.e., Prob($T > t$). Since the Andersen-Gill model treats recall events within firms as independent, we obtain a single survivorship curve as demonstrated in figure 3(a). The slope of the survival curve is relatively steep between 0-80 months compared to that for >80 months, which indicates the likelihood of survivorship declines rapidly between 0-80 months and significantly slows down thereafter. In other words, firms are generally most prone to issuing a recall within 80 months of operation since either the beginning of business or the last recall event.

To illustrate survivorship curves for each stratum, while acknowledging the order of recurring events, figure 3(b) plots the results from the PWP-GT stratified Cox PH model. It is obvious that the shape of the estimated survival curve for each event stratum is different. Specifically, strata 1 and 2 look somewhat similar, so do strata 3 and 4, but strata 1 and 2 are different from strata 3 and 4. Visually, this is an indication that the Andersen-Gill model, which treats all strata as identical, may not be entirely adequate to model the repeated recall data, and that firms' duration times to next recall behave differently depending on where they fall in time compared to other recall events.

If firms learned from a previous recall event, and were able to *lengthen* the amount of time until next recall, we would generally expect a survivorship curve for the later stratum to be above that of the past stratum, so that the probability of survivorship longer than t periods would be higher for the later stratum. Examining the survival curve for stratum 1 in figure 3(b), we see that all observations in the sample issued a recall by the ≈250th month, when the probability of survivorship converges to zero. From the survivorship curve for stratum 2, we observe that not all firms in the sample issued a second recall event, which is evident from the flat-lining of the estimated survival curve for stratum 2 at probability ≈0.3. Comparing survivorship curves for strata 1 and 2, it is apparent that the curves largely overlap for the first ≈50 months, and after that the survival curve for stratum 2 diverges and lies above that of stratum 1. This implies the likelihood of survivorship (i.e., no recall event) in the first and second recall events appears similar for the first ≈50 months, and after that the probability of survivorship increases in case of the second recall event relative to the first. This prolonged time-to-event in case of the second recall is suggestive of some firm learning after the first recall event.

The survival curve for stratum 3 is similar to that of stratum 2 in that not all firms experienced a third recall event. If firms learned how to avoid a subsequent recall after handling previous two recall events, we would expect a survivorship curve for stratum 3 to be above that of the past strata. From the survivorship curve for stratum 3 in figure 3(b), we find no evidence for that to be the case: the survival time from second to third recall (stratum 3) is lower than that from first to second recall (stratum 2), which is in line with observations from figure 1. Therefore, this result does not support firm learning in this

particular case. Finally, the survival curve for stratum 4 flat-lines at probability ≈0.25 for the same reason explained above. The evidence of firm learning between strata 3 and 4 is more substantial as the survivorship curve for stratum 4 is almost always above that for stratum 3.

Overall, no definitive evidence of learning, as a clear pattern of increasing survivorship after each recall incident, is observed. However, there is some evidence of learning between strata 1 and 2, and particularly, strata 3 and 4.

Decomposing Survivorship by Firm and Recall Characteristics

To decompose survivorship curves for each stratum by measurable covariates, specifically those found to be statistically significant in table 3, the Kaplan-Meier (KM) survivorship curves (Kaplan and Meier, 1958) are estimated. The KM curves non-parametrically estimate the survival function for each stratum as:

$$S(t) = \text{Prob}(T > t) = \prod_{j|t_j \leq t} \frac{n_j - d_j}{n_j} \tag{6}$$

where n_j is number of observations at risk just prior to time t_j and d_j is the number of events (recalls) at time t_j. Time in months is plotted against the estimated probability that a recall event will occur in the next immediate time frame, $\text{Prob}(T > t)$, where T is the random variable for duration and the outcome t is time measured in months.

Figure 4 illustrates the stratified KM survivorship curves decomposed by *Diversification*. This variable describes whether a firm's main output is meat/poultry or if they have more diversified output. As can be seen, the estimated KM curves for firms whose main output is meat/poultry lie below those of diversified firms across all strata, which indicates firms producing meat/poultry as their main output have lower probability of surviving (i.e., not experiencing a recall event) at any given time compared to their more diversified rivals. This observation corroborates our findings from the Andersen-Gill and PWP-GT models in table 3.

In figure 5, we decompose survival functions for strata 2 to 4 by recall class. We can see that the survival functions for firms that experienced a Class I recall in the past are almost

always higher than those for firms that dealt with a Class II recall previously. This implies it takes longer (in months) for a firm to issue a subsequent recall after it faces a Class I recall compared to a Class II recall. This indicates that firms seem to learn and reflect more after experiencing the most severe type of recalls. This observation is in line with our findings from table 3. Given that Class III recalls were observed in the first and second recall events in the data (see table 2), and that there is very few *Class III* observations in our sample, the inferences for Class III recall cannot be made robustly and are thus excluded.

DISCUSSION

In a food safety context, increasing the time between recall events represents a key driver of success to food firms. Although our results suggest that the food industry is targeting its resources appropriately to address those recalls that pose the greatest risks to human health (i.e., Class I), overall firms' ability to prevent recalls does not consistently grow with past recall experience. Hence, public health officials and agencies perhaps need to re-visit their inspection programs and/or design appropriate policies that would incentivize food firms to reflect more effectively on their previous food safety incidents.

References

Amorim, L. D. A. F. and Cai, J. (2015). Modelling recurrent events: A tutorial for analysis in epidemiology. *International Journal of Epidemiology*, 44(1):324–333.

Andersen, P. K. and Gill, R. D. (1982). Cox's regression model for counting processes: A large sample study. *The Annals of Statistics*, 10(4):1100–1120.

Breslow, N. (1974). Covariance analysis of censored survival data. *Biometrics*, 30(1):89–99.

Cox, D. R. (1972). Regression models and life tables. *Journal of the Royal Statistical Society: Series B*, 34(2):187–220.

Fama, E. and French, K. (1992). The cross-section of expected stock returns. *Journal of Finance*, 47(2):427–465.

FSIS (2013). Recall of meat and poultry products. FSIS Directive 8080.1, Washington DC.

FSIS (2017). Recall case archive. Retrieved from: https://www.fsis.usda.gov/wps/portal/fsis/topics/recalls-and-public-health-alerts/recall-case-archive.

Grambsch, P. M. and Therneau, T. M. (1994). Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika*, 81(3):515–526.

Henson, S. and Reardon, T. (2005). Private agri-food standards: Implications for food policy and the agri-food system. *Food Policy*, 30(3):241–253.

Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53(282):457–481.

Kelly, P. J. and Lim, L. L.-Y. (2000). Survival analysis for recurrent event data: An application to childhood infectious diseases. *Statistics in Medicine*, 19(1):13–33.

Kennedy, B., Kasl, S., and Vaccarino, V. (2001). Repeated hospitalizations and self-rated health among the elderly: A multivariate failure time analysis. *American Journal of Epidemiology*, 153(3):232–241.

Kleinbaum, D. G. and Klein, M. (2005). *Survival analysis: A self-learning text*. Springer, New York, 2 edition.

Lin, D. Y. and Wei, L. J. (1989). The robust inference for the Cox proportional hazards model. *Journal of the American Statistical Association*, 84(408):1074–1078.

Prentice, R. L., Williams, B. J., and Peterson, A. V. (1981). On the regression analysis of multivariate failure time data. *Biometrika*, 68(2):373–379.

Salin, V., Hooker, N., and Teratanavat, R. (2003). Survival analysis of U.S. meat and poultry recalls, 1994-2002. Research paper presented at the Second Hawaii International Conference on Statistics and Related Fields.

Teratanavat, R., Salin, V., and Hooker, N. (2005). Recall event timing: Measures of managerial performance in U.S. meat and poultry plants. *Agribusiness*, 21(3):351–373.

Tables

Table 1: Number of recalls and total quantity of meat/poultry recalled (in pounds) by publicly traded firms in the United States between 1994-2015.

Stock	Firm	Recalls	Pounds
AHP	American Home Products Corp.	1	150,000
BOBE	Bob Evans Farms Inc.	1	8,500
CAG	ConAgra Inc.	25	114,724,199
COST	Costco Wholesale Corp.	3	222,123
CPB	Campbell Soup Co.	9	16,342,765
DEG	The Delhaize Group	1	112,230
GIS	General Mills Inc.	2	3,300,000
HAIN	The Hain Celestial Group Inc.	1	983,700
HFI	Hudson Foods Inc.	5	28,313,959
HNZ	Heinz H. J. Co.	3	94,886
HRL	Hormel Foods Corp.	6	234,946
HWKN	Hawkins Inc.	1	529
IBP	Iowa Beef Processors, Inc.	8	1,938,155
K	Kellogg Co.	1	2,790
KFT	Kraft Foods Inc.	5	28,508
KR	Kroger Co.	3	490,131
NSRGY	Nestle SA	13	1,786,393
PPC	Pilgrims Pride Corp.	4	28,806,600
SAFM	Sanderson Farms Inc.	1	10,000
SFD	Smithfield Foods Inc.	13	1,007,821
SJM	Smucker J. M. Co.	1	3,000
SLE	Sara Lee Corp.	13	37,720,286
SVU	Supervalu Inc.	2	962
SYY	Sysco Corp.	1	16,800
TAVI	Thorn Apple Valley Inc.	2	35,009,936
THS	TreeHouse Foods Inc.	3	214,957
TSN	Tyson Foods Inc.	36	4,308,504
UVV	Universal Corp.	1	578,000
WFM	Whole Foods Market Inc.	3	34,834
WIN	Winn Dixie Stores Inc.	1	1,734,002
WMK	Weis Markets Inc.	1	2,852
Total		170	278,182,368

Notes: The data used is obtained from the USDA FSIS Recall Case Archive. The time period of our data is determined by the recall data made available by the FSIS at the time of writing this article.

Table 2: Summary statistics of study variables (by stratum/recall).

Stratum/Recall	Mean	Median	Std. dev.	Min	Max
	Time until Recall (in months)				
1	81.42	58.10	72.68	0.20	256.60
2	64.28	49.20	57.25	1.90	188.70
3	30.91	7.80	52.35	0.01	203.10
4	30.16	20.30	31.36	0.50	121.00
	Firm Size (market capitalization, in million U.S.$)				
1	2,161.59	671.80	3,496.04	58.80	17,166.50
2	3,083.85	1,228.60	3,952.50	11.20	13,680.90
3	4,493.84	1,753.30	5,696.26	11.20	21,392.10
4	4,783.12	3,287.45	5,587.75	321.50	21,392.10
	Diversification (1 if meat is the main product, 0 otherwise)				
1	0.23	0.00	0.43	0.00	1.00
2	0.23	0.00	0.43	0.00	1.00
3	0.32	0.00	0.48	0.00	1.00
4	0.31	0.00	0.48	0.00	1.00
	Firm Age (in years)				
1	76.09	73.81	41.10	4.69	144.96
2	87.76	78.00	46.83	4.84	185.98
3	77.71	64.25	43.90	6.33	150.00
4	77.22	64.75	44.03	11.00	147.00
	Recall Size (in thousand pounds)				
1	1,265.22	40.57	5,097.21	0.46	28,000
2	1,348.17	10.18	6,367.17	0.32	35,000
3	2,080.39	33.19	8,003.84	0.50	35,000
4	168.15	39.88	327.07	1.20	1,300
	Class I (1 if recall is of Class I, 0 otherwise)				
1	0.68	1	0.48	0	1
2	0.61	1	0.50	0	1
3	0.74	1	0.45	0	1
4	0.75	1	0.45	0	1
	Class II (1 if recall is of Class II, 0 otherwise)				
1	0.19	0	0.40	0	1
2	0.29	0	0.46	0	1
3	0.26	0	0.45	0	1
4	0.25	0	0.45	0	1
	Class III (1 if recall is of Class III, 0 otherwise)				
1	0.13	0	0.34	0	1
2	0.10	0	0.30	0	1
3	0.00	0	0.00	0	0
4	0.00	0	0.00	0	0

Notes: *Firm Size* is constructed by multiplying the number of shares outstanding by the stock price quoted 10 days before the recall announcement (Fama and French, 1992), and adjusted against inflation using the Consumer Price Index (CPI) provided by the U.S. Bureau of Labor Statistics. *Firm Age* represents the age of the firm since its establishment. The FSIS classifies food recalls into three broad classes. Class I is the most severe class of recalls, which can cause adverse health consequences or death. Class II has a remote probability of adverse health effects. Class III is the least severe class of recalls and it does not cause any adverse health impacts.

Table 3: Estimated coefficients for the Andersen-Gill and PWP-GT models.

	AG (1)	PWP-GT (2)	AG (3)	PWP-GT (4)	AG (5)	PWP-GT (6)
Firm size	5.065×10^{-5}*	1.620×10^{-5}	4.537×10^{-5}	1.593×10^{-5}	6.469×10^{-5}*	3.542×10^{-5}
	(2.969×10^{-5})	(3.468×10^{-5})	(3.096×10^{-5})	(3.535×10^{-5})	(3.708×10^{-5})	(3.648×10^{-5})
	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]
Diversification	0.903**	0.899**	0.553	0.674	0.982**	1.361***
	(0.364)	(0.357)	(0.414)	(0.416)	(0.411)	(0.358)
	[2.466]	[2.458]	[1.738]	[1.962]	[2.671]	[3.900]
Firm age	-1.175×10^{-3}	-7.391×10^{-4}	-1.120×10^{-3}	-8.292×10^{-4}	-8.865×10^{-4}	-9.066×10^{-5}
	(2.827×10^{-3})	(2.399×10^{-3})	(2.818×10^{-3})	(2.450×10^{-3})	(3.512×10^{-3})	(3.178×10^{-3})
	[0.999]	[0.999]	[0.999]	[0.999]	[0.999]	[1.000]
Firm size×Diversification			6.643×10^{-4}**	4.079×10^{-4}*		
			(2.666×10^{-4})	(2.169×10^{-4})		
			[1.001]	[1.000]		
Lag(Recall size)					-8.876×10^{-6}	-3.656×10^{-5}***
					(1.567×10^{-5})	(1.275×10^{-5})
					[1.000]	[1.000]
Lag(Class 1)					1.617*	1.928
					(0.831)	(1.264)
					[5.039]	[6.879]
Lag(Class 2)					2.003**	2.555**
					(0.891)	(1.291)
					[7.409]	[12.871]
AIC	482.90	382.17	480.78	382.77	192.62	211.73
BIC	490.62	389.89	491.08	393.07	205.58	224.68

Notes: Reported are the estimation results, including $\hat{\beta}_i$, robust-standard error of $\hat{\beta}_i$ (in parenthesis), and $\exp(\hat{\beta}_i)$ (in brackets), for the Anderson-Gill (AG)model and the PWP-GT model without interaction terms. Statistical significance is denoted according to the following convention: *p<0.1; **p<0.05; ***p<0.01. The associated diagnostic tests are provided in tables tables 4 and 5.

Table 4: Log-likelihood ratio test of the PWP-GT model with and without interaction terms.

	$\Delta \log(\mathcal{L})$	df	p-value
Specification (2) in Table 3	6.10	9	0.270
Specification (4) in Table 3	7.49	12	0.176
Specification (6) in Table 3	6.13	10	0.196

Notes: Reported are the log-likelihood ratio ($\Delta \log(\mathcal{L})$) statistics and associated degrees of freedom (df) and p-values from the log-likelihood ratio test of the fits of the PWP-GT model with and without interaction terms. Under the null hypothesis, the PWP-GT model without interaction terms is preferred.

Table 5: Test of the proportional hazard assumption for the Andersen-Gill model.

	χ^2	p-value
Specification (1) in Table 3		
Firm size	0.047	0.83
Diversification	0.096	0.76
Firm age	0.180	0.67
Global	0.215	0.98
Specification (3) in Table 3		
Firm size	0.067	0.80
Diversification	0.621	0.43
Firm age	0.323	0.57
Firm size×Diversification	2.446	0.12
Global	2.579	0.63
Specification (5) in Table 3		
Firm size	0.602	0.438
Diversification	6.927	0.008
Firm age	0.082	0.775
Lag(Meat recalled)	0.773	0.379
Lag(Class 1)	0.835	0.361
Lag(Class 2)	0.789	0.374
Global	8.947	0.177

Notes: Reported are the chi-square (χ^2) statistics and associated p-values from testing the proportional hazards assumption of the Andersen-Gill model fits (Grambsch and Therneau, 1994). Under the null hypothesis, the proportional hazards assumption is satisfied.

Figures

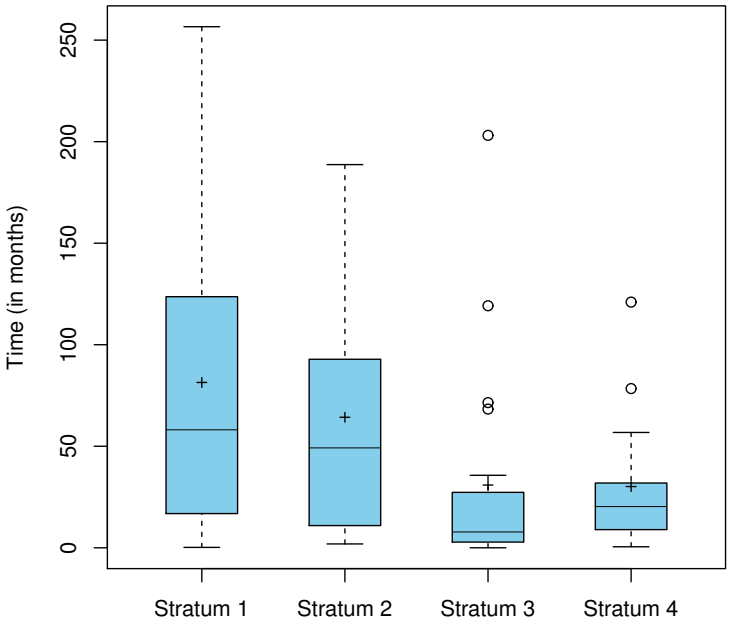


Figure 1: Box plot of time (in months) until recall by stratum.
Notes: Stratum 1 represents the time until the first recall event; stratum 2 is the time between first and second recall; stratum 3 is the time between second and third recall; and stratum 4 is the time between third and fourth recall event. The bottom and the top of a box correspond to the 25th and 75th percentiles, respectively. The horizontal line inside the box represents the median (the 50th percentile), while the small "+" mark denotes the mean. Whiskers indicate variability outside the upper and lower quartiles. Circles denote extreme observations. It takes the longest until recall event for stratum 1, with stratum 2 coming close behind with the median time of about 50 months. For stratum 3 we observe a sizable dip in time relative to stratum 2, with the median time to recall plummeting to 8 months, and then a jump in stratum 4, with the median time of about 20 months.

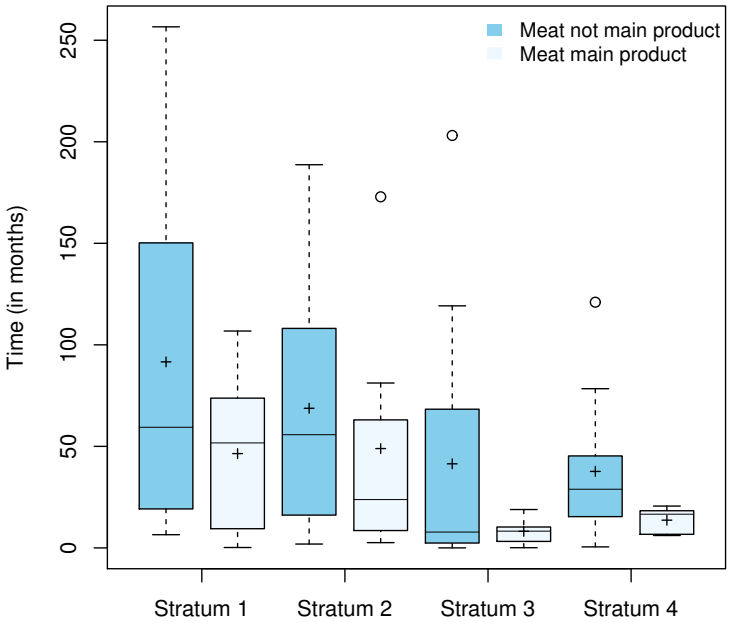
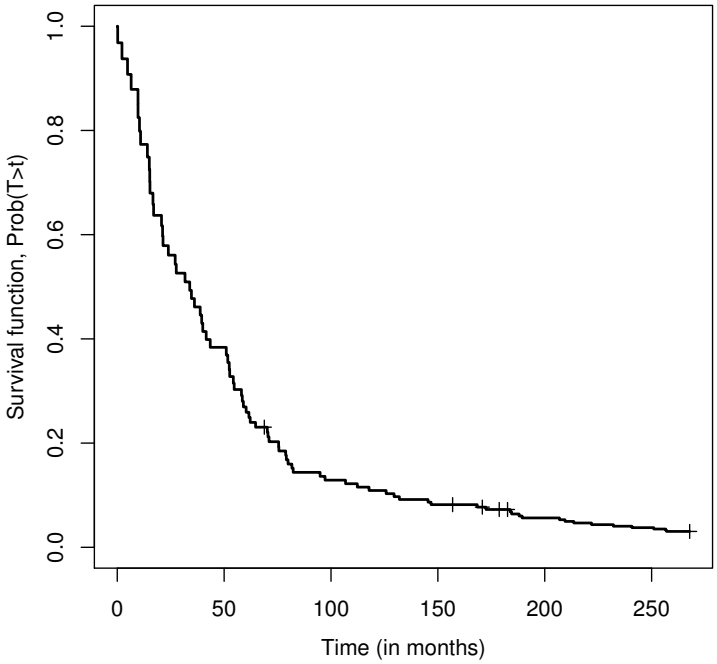
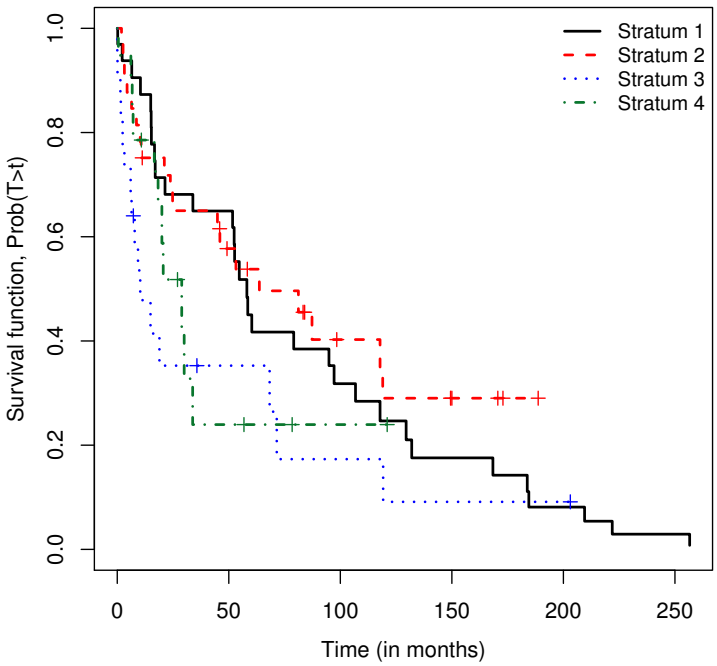


Figure 2: Box plot of time (in months) until recall by diversification and stratum.
Notes: Stratum 1 represents the time until the first recall event; stratum 2 is the time between first and second recall; stratum 3 is the time between second and third recall; and stratum 4 is the time between third and fourth recall event. The bottom and the top of a box correspond to the 25th and 75th percentiles, respectively. The horizontal line inside the box represents the median (the 50th percentile), while the small "+" mark denotes the mean. Whiskers indicate variability outside the upper and lower quartiles. Circles denote extreme observations. Both median and mean times until recall event across different strata are shorter for firms for which meat is the main output relative to those for which meat is not the primary output. This suggests that the risk of foodborne outbreak is greater for firms that handle primarily meat products compared to those whose product line is more diversified.



(a) The Andersen-Gill model fit



(b) The PWP-GT model fit

Figure 3: Estimated survival functions.
Notes: The estimated survival function shows the probability of surviving (i.e., no recall event) longer than t periods, Prob$(T > t)$. Small "+" marks on the plot denote individual firms whose survival times have been right-censored (i.e., they have not demonstrated the event of interest).

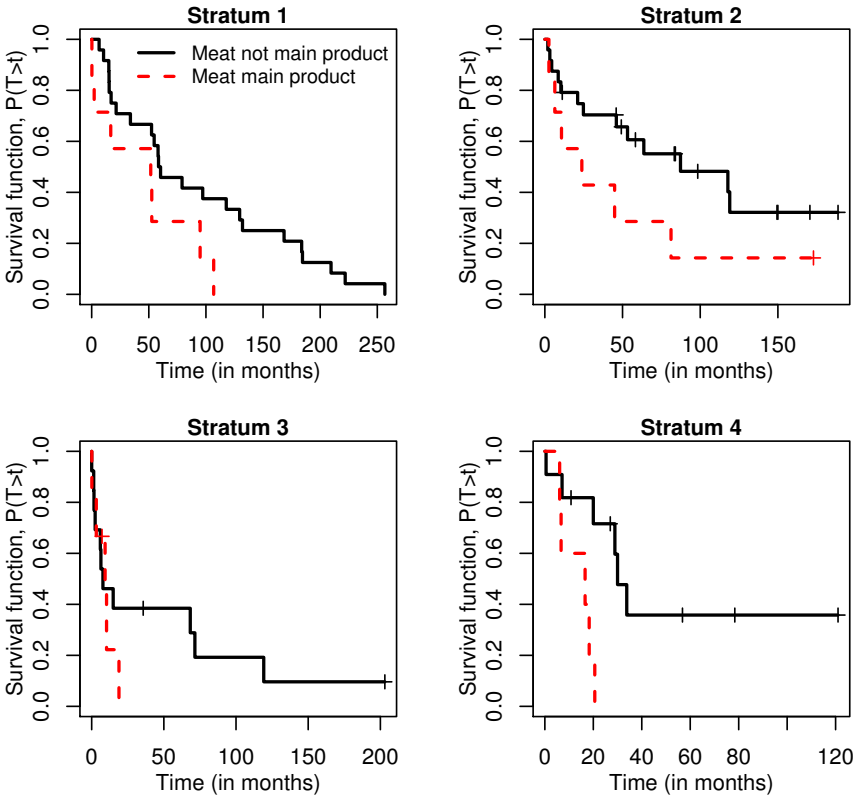


Figure 4: Estimated Kaplan-Meier curves for each stratum by diversification.
Notes: The estimated survival function shows the probability of surviving (i.e., no recall event) longer than t periods, Prob$(T > t)$. Small "+" marks on the plot denote individual firms whose survival times have been right-censored (i.e., they have not demonstrated the event of interest).

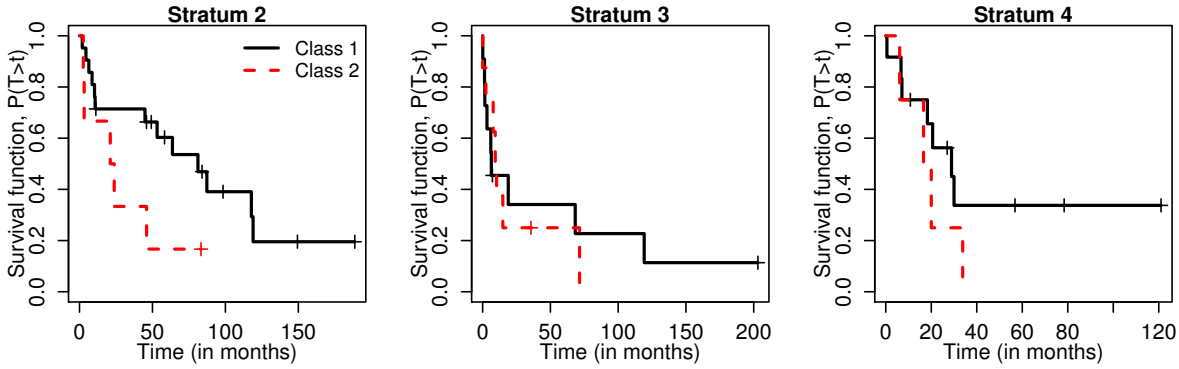


Figure 5: Estimated Kaplan-Meier curves for each stratum by lag of recall class.
Notes: The estimated survival function shows the probability of surviving (i.e., no recall event) longer than t periods, Prob$(T > t)$. Small "+" marks on the plot denote individual firms whose survival times have been right-censored (i.e., they have not demonstrated the event of interest). Stratum 1 is not reported because first recall event per firm (hence stratum 1) is lost when recall-specific factors are lagged.