

Longitudinal analysis of foodborne disease outbreaks in Saudi Arabia

Trends and demographic patterns (2017-2023)

Hissah A. Alzahrani, PhD.

ABSTRACT

الأهداف: تساعد دراسة تحليل بيانات الأمراض المنقولة بالغذاء في المملكة العربية السعودية بين عامي 2017م و 2023م باستخدام نماذج تحليل البيانات الطولية.

المنهجية: أُجريت دراسة تحليل البيانات الطولية لسبع سنوات على بيانات فاشيات الأمراض المنقولة بالأغذية من جميع مناطق المملكة العربية السعودية، 13 منطقة. فحصت هذه الدراسة 5 نتائج: عدد المصابين بالأمراض المنقولة بالأغذية، وعدد الأمراض المنقولة بالأغذية، وكذلك عدد المصابين مصنفة حسب الجنس والجنسية والفئات العمرية. واستخدم في التحليل أسلوب نماذج GEE و GLMM بناء على توزيع ذي الحدين السالب.

النتائج: كشفت النتائج عن انخفاض كبير في أعداد المصابين بالأمراض المنقولة بالأغذية، وكذلك عدد الأمراض المنقولة بالأغذية، من عام 2017 إلى عام 2020، تليها زيادة إلى عام 2023. لم يلاحظ أي فرق جوهري في أعداد المصابين بين الذكور والإناث. أيضا يتضح أن أعداد المصابين بالأمراض المنقولة بالأغذية من السعوديين أعلى من غير السعوديين والفئات العمرية 5-19 و 20-49 عامًا هما الأعلى عدداً من المصابين بالأمراض المنقولة بالأغذية.

الخلاصة: وضحت هذه الدراسة بيانات الأمراض المنقولة بالغذاء في المملكة العربية السعودية قبل و بعد سنة 2020، واستنتجت زيادة أعداد الإصابات بالأمراض المنقولة بالغذاء بعد أزمة كوفيد 19 والضرورة لبذل جهود مستهدفة لتقليل الأعداد وحالات الإصابة واتخاذ إجراءات مستقبلية.

Objectives: To explore the long-term patterns of foodborne disease outbreaks (FBDOs) in Saudi Arabia between 2017-2023 using longitudinal data analysis models.

Methods: A 7-year longitudinal study was carried out on FBDO data from all 13 regions of Saudi Arabia. This study examined 5 outcomes: the number of FBDOs, FBDO cases, and FBDO cases classified by gender, nationality, and age groups. The analysis used generalized estimating equations and generalized linear mixed models based on a negative binomial distribution.

Results: The findings revealed a significant decrease in FBDO counts and cases from 2017-2020, followed by an increase until 2023. Public sources consistently reported higher numbers of FBDOs and cases

than home sources. No substantial difference was observed in FBDO cases between men and women. Saudi nationals had higher FBDO case counts than non-Saudi residents. The age groups 5-19 years and 20-49 years exhibited the highest FBDO case numbers.

Conclusion: The observed patterns, particularly the decline toward 2020 and the subsequent increase, underscore the necessity for adaptable public health strategies to mitigate the heightened incidence following the COVID-19 pandemic in Saudi Arabia and may inform future prevention and control measures.

Keywords: foodborne, longitudinal, GEE, GLMM, outbreaks

*Saudi Med J 2025; Vol. 46 (5): 507-514
doi: 10.15537/smj.2025.46.5.20241010*

From the Department of Mathematics, Faculty of Applied Sciences, Umm Al-Qura University, Makkah, Kingdom of Saudi Arabia.

Received 1st December 2024. Accepted 8th April 2025.

*Address correspondence and reprint request to: Dr. Hissah A. Alzahrani, Mathematics Department, Faculty of Applied Sciences, Umm Al-Qura University, Makkah, Kingdom of Saudi Arabia. E-mail: hahzahrani@uqu.edu.sa
ORCID ID: <https://orcid.org/0000-0002-8146-2375>*

Studying foodborne disease outbreaks (FBDOs) in Saudi Arabia is crucial for comprehending the epidemiology of foodborne illnesses and developing effective prevention and control measures. Previous studies have reported statistical analyses of specific regions of FBDOs in Saudi Arabia; however, limited information is available on the longitudinal trends of these outbreaks over several years.^{1,2} Comprehensive data on the longitudinal trends of FBDOs in Saudi Arabia are lacking, which limits our understanding of

the epidemiology of foodborne illnesses, particularly before and after the COVID-19 pandemic. This study aimed to address this knowledge gap by comprehensively analyzing the longitudinal trends of FBDOs in Saudi Arabia from 2017-2023.

Longitudinal data analysis is a statistical technique used to study the data collected from the same subjects over time. The subject of this study was the Saudi Arabian region, and data were collected annually from 2017-2023. Longitudinal data analysis has gained significant attention in recent years owing to its ability to capture changes and patterns over time. Deciphering the dynamics of longitudinal data is crucial in various fields, including healthcare, social sciences, and economics as it allows researchers to identify trends and make informed decisions. This study was focused on applying the main statistical methods for longitudinal data analysis: generalized estimating equations (GEE) and generalized linear mixed models (GLMM). The 2 available methods, GEE and GLMM, are popular for analyzing longitudinal data. The GEE is often preferred when interested in population-level trends and when handling missing data or imbalanced repeated measures. In contrast, GLMM is preferred when the emphasis is on subject-specific predictions or when variability among subjects is of particular interest. The appropriate method for FBDO datasets is GEE, because the main goal is the population mean rather than the subject-specific mean. However, GLMM is more effective with a smaller sample size; therefore, this study applied both.^{3,4} This research intended to investigate the longitudinal trends of FBDOs in Saudi Arabia from 2017-2023 and identify any regional variations in these trends using the GEE and GLMM methods. This study hypothesized a significant increase in the number of FBDOs in Saudi Arabia after the COVID-19 pandemic and a variation in the incidence of these outbreaks across different demographic variables such as gender, nationality, and age groups.

Methods. This longitudinal study was carried out on FBDOs in Saudi Arabia. This work comprised counts of FBDOs from 2017-2023 for all regions in Saudi Arabia (13 regions). The annual counts of FBDOs and their cases are announced by the Ministry of Health based on various factors such as age group, gender,

source of infection (public or home), and nationality (Saudi or non-Saudi) in each region.⁵ Missing data is not a concern because the statistics are completely announced. Longitudinal data modeling was carried out for the counts of FBDOs (model 1) and total cases of FBDOs (model 2). The other 3 models (models 3-5) were FBDO cases based on gender, nationality, and age groups, as shown in the flow chart **Figure 1**.

In this longitudinal study, the subjects were the regions of Saudi Arabia. The outcome was the count, and 5 major outcomes were modeled over the years, with every outcome in a separate model. Initially, the FBDO outcomes in each region of Saudi Arabia were documented. Subsequently, the counts of FBDO cases for each region in the country and the counts based on gender (male or female), nationality (Saudi Arabia or non-Saudi Arabia), and age groups (<1, 1-4, 5-19, 20-49, and 50+ years) were recorded. Each model had its covariate, the categorical variable, and all outcomes were obtained 7 times from 2017-2023 (**Figure 1**). The year variable was also a nominal variable from 2017-2023 and was considered nominal for comparing the primary outcomes over the years, with 2017 as the reference year.

Longitudinal analysis of count data is common in epidemiological studies. This type of analysis assumes only non-negative integer values using Poisson distribution. The Poisson random variable Y has the feature $E(Y)=V(Y)$.⁶ Overdispersion is also a common problem in count data, where $E(Y)<V(Y)$. Negative binomial (NB) is the appropriate choice for modeling the count data, and the random variable follows the mean: $E(Y)$ and $V(Y)=E(Y)+\alpha(E(Y))^2$, where α controls overdispersion.^{7,8} Moreover, the longitudinal dataset model is a special regression model, a type of regression for correlated datasets.⁹⁻¹¹

In the regular regression model, a primary assumption of independence among the dataset subjects is carried out. This assumption is broken in the longitudinal or repeated-measures dataset, where readings are obtained repeatedly from the same subject. In this study, FBDO counts in each region of Saudi Arabia had at least 7 readings from 2017-2023. Regression analysis of correlated datasets accounted for this problem and estimated the valid regression parameters. Subsequently, the longitudinal dataset of FBDO counts from 2017-2023 was analyzed using correlated regression models. The 2 main methodologies used for longitudinal data analysis are marginal and conditional models. Starting with the conditional model, the GLMM is the common model with a random effect part

Disclosure. Author has no conflict of interests, and the work was not supported or funded by any drug company.

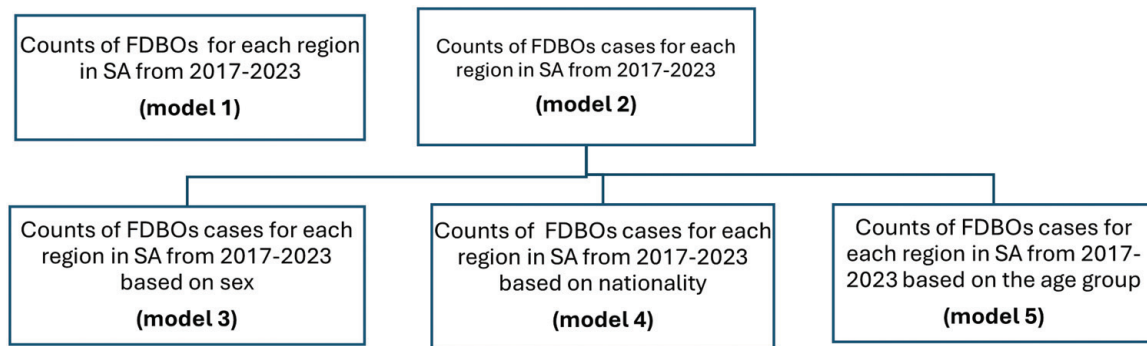


Figure 1 - The flow chart of the datasets and models are used in the study. FDBOs: foodborne disease outbreaks, SA: Saudi Arabia

in the regression equation that describes the behavior of the outcome for the same subject, where a region in Saudi Arabia is a subject. This method was focused on interpreting the subject or region. This random effect part accounted for the dependency among the readings of the same subject, and the appropriate link function for the mean was the log function used in the model.¹⁰ A total of subjects followed up for t repeated measures and p covariates were considered using a single random intercept representing the differences among the individuals. The regression equation for the count outcome was the differences among the individuals, as follows:

$$\text{Log}(E(Y_{it})) = b_0 + b_1 X_{1it} + b_2 X_{2it} + \dots + b_p X_{pit} + Z_i$$

Where $b_0, b_1, b_2, \dots, b_p$ are the regression coefficients and Z_i is the random effect assumed to be normally distributed with a zero mean and estimated variance from the dataset.^{12,13} The parameters were determined using the maximum likelihood computed using iterative methods. The correlation structure can be assumed to be of different types; in this study, it was assumed to be exchangeable. Model selection was based on the Akaike information criterion, and the smallest was the best.¹⁴

In contrast, the strategy and interpretation differ in the marginal model, but the parameter estimations are close. A GEE was used as the marginal model. The GEE focuses on estimating population-averaged parameters, allowing researchers to account for correlations within subjects, while providing robust standard errors. The GEE is a simple method for analyzing longitudinal data without assuming a particular distribution for the data. This method requires only a link function through the mean; in this case, it is a log-link function appropriate for count data and a suitable correlation structure. The model is robust even when the correlation structure is misspecified.¹⁰

The GEE model is as follows:

$$\text{Log}(E(Y_{it})) = b_0 + b_1 X_{1it} + b_2 X_{2it} + \dots + b_p X_{pit}$$

Where $b_0, b_1, b_2, \dots, b_p$ are the regression coefficients. The “working” correlation matrixes are estimated to adopt the correlation among the subjects’ readings using the iterative method of Newton Raphson. In addition, the correlation structure can be assumed to be of different types. In this case, it was assumed to be exchangeable. Model selection was carried out based on the criteria of Quasi-likelihood (QAIC); the smallest is the best. This method is beneficial when the interest lies in understanding the average effect of covariates across all clusters (all ratings for one region), rather than subject cluster effects.^{10,11,15}

There are certain major differences in the methods proposed to analyze FBDOs. The GLMM suggests that the correlation among the repeated measures comes from the same subject. The shared random effect adopts this problem, followed by inference and covariate effects based on the subject. In the GEE approach, covariate effects are based on population rather than the subject mean. This method may be less effective with smaller sample sizes, leading to high standard deviation (SD) and some bias.⁴ In the case of FBDO counts from Saudi Arabia, the main goal was to assess the average mean of the population rather than each region, for which GEE is appropriate. However, GLMM was also analyzed to obtain the final results as the findings are similar and both methods are appropriate for longitudinal data.⁶

Statistical analysis. The SAS software was carried out and descriptive statistics (mean±SD) were calculated for each outcome from 2017-2023.

Results. Foodborne disease outbreak data from 13 regions in Saudi Arabia were analyzed and discussed. Each region was considered a cluster or subject observed

from 2017-2023. A total of 5 outcomes were analyzed in 5 separate models to determine the change in FBDO counts over the years 2017-2023 based on 5 factors: counts, cases, gender, nationality, and age groups. For the first outcome (FBDO counts) from home and public sources, the public source mean was 14.69 in 2017, which decreased in 2020 and then increased to 15.31 in 2023, with a high SD of 23.03. The results showed that the mean values of FBDO counts were higher from public sources than home sources. The same trend was observed in the second outcome (counts of FBDO cases), where the mean values were higher from public sources than home sources. The statistics decreased from 2017-2020 and increased until 2023. For the gender variable in outcome 3, the FBDO case counts did not show a clear change between men and women over the years. In contrast, the nationality variable in outcome 4 exhibited higher mean values for Saudi citizens than non-Saudi citizens. The counts of FBDO cases based on age groups indicated that higher mean values for FBDO counts were for the middle groups 5-9 years and 20-49 years. The main point highlighted in [Table 1](#) is that the calculated mean values decreased from 2017-2020 and then increased until 2023. In addition, the SDs were high, which led to the selection of an appropriate distribution to model the outcomes, which was an NB distribution instead of a Poisson distribution.¹⁹

In [Figure 2](#), every outcome profile from 2017-2023 is plotted with the interval mean $\pm$ SD. The mean values

decreased until 2020 and then increased until 2023. The FBDO counts and cases were higher from the public source and Saudi cases for the 5-19 and 20-49 age groups. The observed trends in outcome profiles suggested potential impacts from the COVID-19 pandemic, with a notable dip in 2020, followed by a recovery period. This pattern may reflect changes in reporting practices, healthcare access, and actual disease incidence during the pandemic.

[Table 2](#) presents the results of the GEE, and [Table 3](#) provides the findings of the GLMM for the 5 outcomes using an NB distribution for both methods. As mentioned in [Table 1](#), most SDs were high, the variance that doubled the SD was high, and the NB was the appropriate choice for running the models. In [Tables 2 & 3](#), the intercept, covariate parameter estimations $b \pm$ standard error (SE), p -values, and relative risk ([RR]=exp[b]) are presented for the GEE and GLMM methods. Model 1 included 2 categorical covariates: source type (public or home) and year. The year was considered a categorical variable, and 2017 was taken as a reference to compare the FBDO count for each year relative to 2017. As inferred from [Table 2](#), the estimation parameter for the home source using the GEE method was -0.29, and the reference was the public source. In contrast, in [Table 3](#), the estimation from the GLMM method was -0.198. In general, the results of the 2 methods were similar, with some differences owing to the sample size.^{4,20,21} Both methods were employed because the goal of the GEE method is

Table 1 - The descriptives statistics for the outcomes of foodborne disease outbreaks through 2017-2023.

Variables	2017	2018	2019	2020	2021	2022	2023
<i>Counts of FBDOs</i>							
Public source	14.69 $\pm$ 15.48	14 $\pm$ 15.81	10.69 $\pm$ 12.74	6 $\pm$ 4.9	8.08 $\pm$ 6.76	9.31 $\pm$ 5.57	15.31 $\pm$ 23.02
Home source	15.08 $\pm$ 12.66	14.67 $\pm$ 11.46	10.77 $\pm$ 9.68	4.31 $\pm$ 3.33	5.92 $\pm$ 3.84	4.23 $\pm$ 3.39	7.46 $\pm$ 5.98
<i>Counts of FBDOs cases</i>							
Public source	163.77 $\pm$ 252.14	97.15 $\pm$ 111.89	93.23 $\pm$ 114.49	78.38 $\pm$ 88.97	60.85 $\pm$ 67.1	72.85 $\pm$ 61.95	110.62 $\pm$ 134.68
Home source	77.77 $\pm$ 68.33	77.33 $\pm$ 58.56	53.31 $\pm$ 45.96	19.31 $\pm$ 15.56	26.54 $\pm$ 15.7	18.54 $\pm$ 13.54	34.85 $\pm$ 27.18
<i>Counts of FBDOs cases based on gender</i>							
Male	114.15 $\pm$ 133.07	88.38 $\pm$ 88.64	68.46 $\pm$ 69.45	56.92 $\pm$ 58.43	43.31 $\pm$ 46.32	48.38 $\pm$ 35.93	79.92 $\pm$ 83.6
Female	127.38 $\pm$ 150.21	80.15 $\pm$ 70.72	78.08 $\pm$ 68.85	40.77 $\pm$ 41.44	44.08 $\pm$ 38.62	43 $\pm$ 29.6	65.54 $\pm$ 76.15
<i>Counts of FBDOs cases based on nationality</i>							
SA	168 $\pm$ 145.12	125.38 $\pm$ 90.77	112.15 $\pm$ 113.34	83.38 $\pm$ 80.77	73.69 $\pm$ 69.25	67.69 $\pm$ 47.62	105.08 $\pm$ 135.61
Non-SA	73.54 $\pm$ 148.26	43.15 $\pm$ 81.57	34.38 $\pm$ 53.14	14.31 $\pm$ 16.24	13.69 $\pm$ 17	23.69 $\pm$ 32.67	40.38 $\pm$ 62.52
<i>Counts of FBDOs cases based on age</i>							
<1 years	0.92 $\pm$ 1.26	1.23 $\pm$ 1.36	0.38 $\pm$ 0.51	0.15 $\pm$ 0.38	0.69 $\pm$ 0.95	0.15 $\pm$ 0.55	0.62 $\pm$ 1.12
1-4 years	18.77 $\pm$ 12.83	19.31 $\pm$ 17.43	13 $\pm$ 9.64	7.62 $\pm$ 6.34	9.77 $\pm$ 6.55	6.85 $\pm$ 5.61	7.85 $\pm$ 7.71
5-19 years	99.54 $\pm$ 84.32	67.85 $\pm$ 49.61	62 $\pm$ 55.99	36.31 $\pm$ 29.13	41.38 $\pm$ 32.87	38.54 $\pm$ 28.64	53.15 $\pm$ 61.39
20-49 years	104.46 $\pm$ 151.93	67.15 $\pm$ 74.03	66.77 $\pm$ 74.37	50.62 $\pm$ 65.56	32 $\pm$ 39.51	39.46 $\pm$ 28.7	72.92 $\pm$ 85.5
50+ years	17.85 $\pm$ 44.77	13 $\pm$ 31.57	4.38 $\pm$ 5.24	3 $\pm$ 4.16	3.54 $\pm$ 4.14	6.38 $\pm$ 14.21	10.92 $\pm$ 16.73
Values are presented as mean $\pm$ standard deviation (SD). FBDOs: foodborne disease outbreaks, SA: Saudi Arabian							

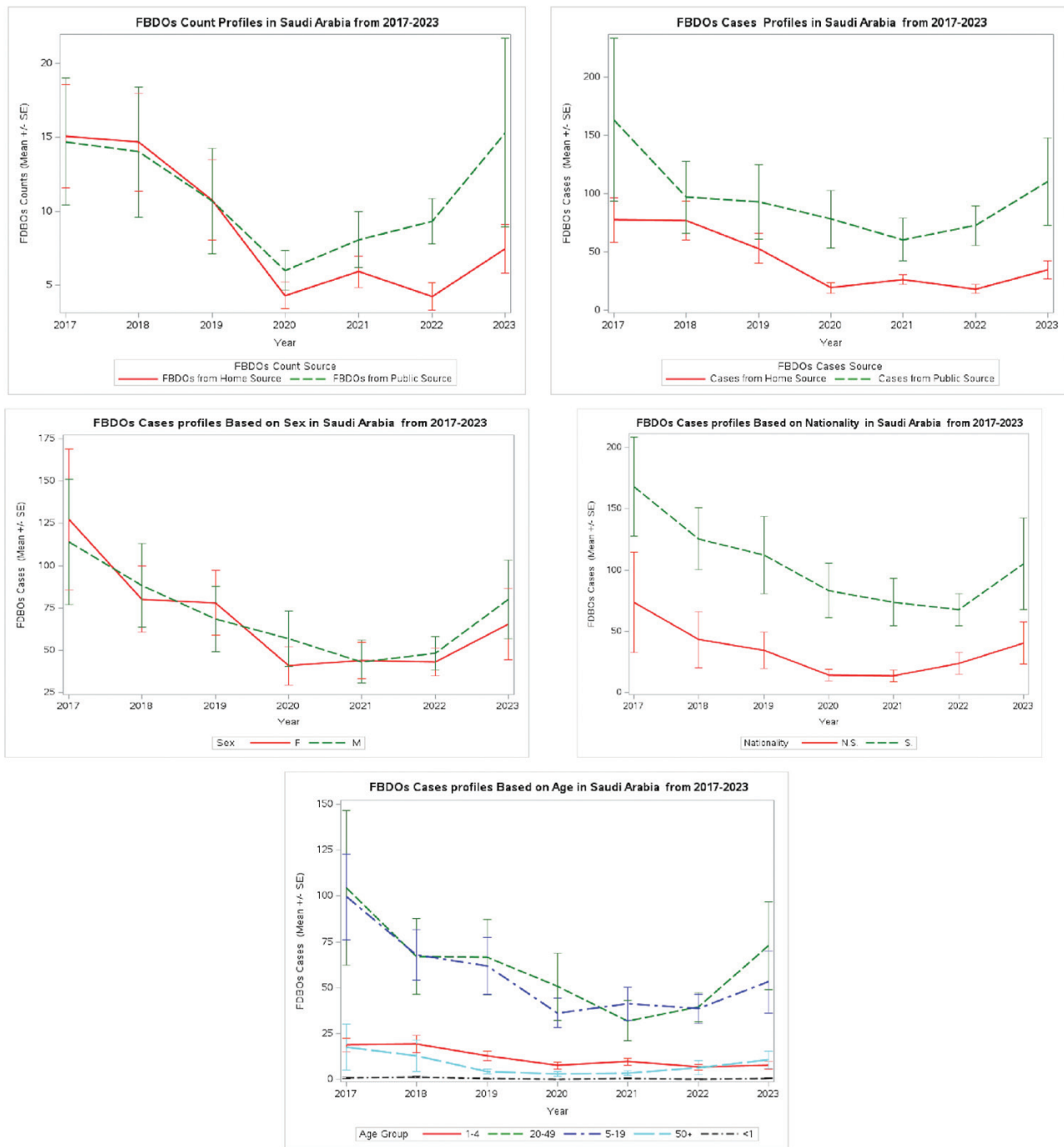


Figure 2 - Plots of foodborne disease outbreaks outcomes profiles through 2017-2023 in Saudi Arabia. FBDOS: foodborne disease outbreaks, SE: standard error

to describe covariate effects on the population means rather than the subject (here, the region), which is more applicable to our FBDOS case counts. The RR from the GEE method for the home source was 0.749,

which signifies that the estimated RR was 0.749 times higher than that for the public source. The parameter estimations for 2018-2020 decreased significantly and then increased until 2023, compared with the reference

Table 2 - The generalized estimating equations models, the parameters estimation, *p*-values and relative risks used in the study.

Variables	b±SE	P-values	Relative risks
<i>Y=counts of FBDOs, model 1</i>			
Intercept	2.858±0.258	<0.0001 [†]	17.427
Ref=public	-	-	-
Source=home	-0.29±0.23	0.207	0.749
Ref=2017	-	-	-
Year=2018	-0.061±0.085	0.473	0.941
Year=2019	-0.33±0.1	0.0009 [†]	0.719
Year=2020	-1.083±0.137	<0.0001 [†]	0.339
Year=2021	-0.776±0.202	0.0001 [†]	0.461
Year=2022	-0.836±0.112	<0.0001 [†]	0.434
Year=2023	-0.315±0.249	0.206	0.73
<i>Y=counts of FBDOs cases, model 2</i>			
Intercept	5.181±0.294	<0.0001 [†]	177.772
Ref=public	-	-	-
Source=home	-0.901±0.172	<0.0001 [†]	0.407
Ref=2017	-	-	-
Year=2018	-0.236±0.142	0.0966*	0.791
Year=2019	-0.46±0.18	0.0104**	0.632
Year=2020	-1.034±0.278	0.0002 [†]	0.356
Year=2021	-1.036±0.276	0.0002 [†]	0.356
Year=2022	-1.095±0.22	<0.0001 [†]	0.335
Year=2023	-0.592±0.228	0.0093 [†]	0.554
<i>Y=Counts of FBDOs cases based on gender, model 3</i>			
Intercept	4.832±0.303	<0.0001 [†]	125.349
Ref=male	-	-	-
Gender=female	-0.07±0.05	0.1615	0.933
Ref=2017	-	-	-
Year=2018	-0.364±0.175	0.0374**	0.696
Year=2019	-0.5±0.253	0.0482**	0.607
Year=2020	-0.913±0.367	0.0128**	0.402
Year=2021	-1.019±0.367	0.0055**	0.362
Year=2022	-0.976±0.273	0.0003 [†]	0.377
Year=2023	-0.513±0.261	0.0491**	0.6
<i>Y=counts of FBDOs cases based on nationality, model 4</i>			
Intercept	4.124±0.539	<0.0001 [†]	61.788
Ref=non-SA	-	-	-
Nationality=SA	1.214±0.315	0.0001 [†]	3.366
Ref=2017	-	-	-
Year=2018	-0.429±0.182	0.0181**	0.652
Year=2019	-0.6±0.252	0.0172**	0.549
Year=2020	-1.146±0.425	0.0069**	0.318
Year=2021	-1.24±0.448	0.0056**	0.29
Year=2022	-1.037±0.272	0.0001 [†]	0.355
Year=2023	-0.545±0.186	0.0033 [†]	0.58
<i>Y=counts of FBDOs cases based on age (years), model 5</i>			
Intercept	0.065±0.238	0.7871	1DE067
Ref= age <1	-	-	-
Age=1-4	3.014±0.113	<0.0001 [†]	20.369
Age=5-19	4.591±0.115	<0.0001 [†]	98.515
Age=20-49	4.68±0.157	<0.0001 [†]	107.749
Age=50+	2.598±0.437	<0.0001 [†]	13.433
Ref=2017	-	-	-
Year=2018	-0.202±0.155	0.1929	0.818
Year=2019	-0.649±0.28	0.0204**	0.523
Year=2020	-1.089±0.339	0.0013 [†]	0.337
Year=2021	-0.98±0.36	0.0065 [†]	0.376
Year=2022	-1.022±0.214	<0.0001 [†]	0.36
Year=2023	-0.544±0.228	0.0169**	0.581

[†]Significant at *p*=0.005, *significant at *p*=0.05, **significant at *p*=0.10.
FBDOs: foodborne disease outbreaks

Table 3 - The generalized linear mixed models, the parameters estimation, *p*-values and relative risks used in the study.

Variables	b±SE	P-values	Relative risks
<i>Y=counts of FBDOs, model 1</i>			
Intercept	2.572±0.214	<0.0001 [†]	13.089
Ref=public	-	-	-
Source=home	-0.198±0.092	0.0311**	0.821
Ref=2017	-	-	-
Year=2018	-0.058±0.156	0.7133	0.945
Year=2019	-0.355±0.158	0.0259**	0.702
Year=2020	-1.015±0.172	<0.0001 [†]	0.363
Year=2021	-0.678±0.166	<0.0001 [†]	0.508
Year=2022	-0.727±0.166	<0.0001 [†]	0.484
Year=2023	-0.321±0.159	0.045**	0.726
<i>Y=counts of FBDOs cases, model 2</i>			
Intercept	4.9±0.235	<0.0001 [†]	134.25
Ref=public	-	-	-
Source=home	-0.777±0.121	<0.0001 [†]	0.461
Ref=2017	-	-	-
Year=2018	-0.262±0.216	0.2253	0.77
Year=2019	-0.474±0.214	0.0283**	0.624
Year=2020	-0.939±0.222	<0.0001 [†]	0.392
Year=2021	-0.949±0.218	<0.0001 [†]	0.388
Year=2022	-0.916±0.219	<0.0001 [†]	0.401
Year=2023	-0.615±0.215	0.0048 [†]	0.541
<i>Y=counts of FBDOs cases based on gender, model 3</i>			
Intercept	4.597±0.232	<0.0001 [†]	99.107
Ref=male	-	-	-
Gender=female	-0.062±0.097	0.5297	0.941
Ref=2017	-	-	-
Year=2018	-0.424±0.182	0.0205**	0.655
Year=2019	-0.547±0.183	0.0032 [†]	0.58
Year=2020	-0.843±0.191	<0.0001 [†]	0.431
Year=2021	-0.944±0.187	<0.0001 [†]	0.39
Year=2022	-0.813±0.187	<0.0001 [†]	0.444
Year=2023	-0.598±0.184	0.0013 [†]	0.551
<i>Y=counts of FBDOs cases based on nationality, model 4</i>			
Intercept	3.469±0.297	<0.0001 [†]	32.079
Ref=non-SA	-	-	-
Nationality=SA	1.657±0.144	<0.0001 [†]	5.24
Ref=2017	-	-	-
Year=2018	-0.483±0.248	0.0525*	0.618
Year=2019	-0.539±0.249	0.0319**	0.584
Year=2020	-0.888±0.262	0.0009 [†]	0.412
Year=2021	-1.02±0.257	0.0001 [†]	0.361
Year=2022	-0.798±0.255	0.0021 [†]	0.451
Year=2023	-0.44±0.249	0.0795*	0.645
<i>Y=counts of FBDOs cases based on age (years), model 5</i>			
Intercept	-0.248±0.272	0.3809	0.781
Ref= age <1	-	-	-
Age=1-4	3.094±0.19	<0.0001 [†]	22.061
Age=5-19	4.664±0.187	<0.0001 [†]	106.06
Age=20-49	4.596±0.186	<0.0001 [†]	98.999
Age=50+	2.329±0.192	<0.0001 [†]	10.266
Ref=2017	-	-	-
Year=2018	-0.279±0.151	0.0640*	0.757
Year=2019	-0.569±0.153	0.0002 [†]	0.567
Year=2020	-0.876±0.158	<0.0001 [†]	0.417
Year=2021	-0.842±0.157	<0.0001 [†]	0.432
Year=2022	-0.844±0.155	<0.0001 [†]	0.431
Year=2023	-0.498±0.153	0.0012 [†]	0.608

[†]Significant at *p*=0.005, **significant at *p*=0.05, *significant at *p*=0.10.
FBDOs: foodborne disease outbreaks

year 2017, for both methods, except in 2018 and 2023 for the GEE method (p -values were not significant).

For model 2, the outcome was the number of FBDO cases applied to the 2 methods. The results were the same as those for model 1; the RR for the home source was 0.407 times higher than that for the public source, and the parameter estimations for the FBDO case counts decreased significantly toward 2020 and then increased until 2023 relative to the year 2017. The only exception was 2018, which was not significant. For model 3, the gender variable did not significantly affect the number of FBDO cases; all parameter estimations for the years were significant and decreased until 2021 and then increased until 2023. The results of model 4 were similar to those of model 3, but the nationality variable had a significant effect on FBDO cases for Saudi citizens compared with the reference non-Saudi citizens. Finally, model 5 included 2 covariates: age and year. All parameter estimations and RR were significantly higher than those for the reference age group (<1), and the parameter estimation decreased significantly toward 2020 and then increased toward 2023.

Discussion. This comprehensive study examining FBDOs in Saudi Arabia from 2017-2023 has shown significant trends and patterns. Using GEE and GLMM, 5 primary outcomes, namely, the number of FBDOs, FBDO cases, and FBDO cases categorized by gender, nationality, and age, were analyzed. The investigation revealed that public sources consistently reported higher FBDO counts and cases than home sources. In addition, a prominent decline in FBDO counts and cases was observed from 2017-2020, followed by an increase until 2023. The study found no significant disparity in FBDO case counts between men and women; however, Saudi citizens exhibited higher FBDO case counts than non-Saudi residents. The age groups 5-19 and 20-49 demonstrated the highest FBDO case counts. This comprehensive analysis provides valuable insights for public health officials and policymakers in Saudi Arabia to develop targeted interventions and preventive strategies. The observed trends in FBDO counts and cases across demographics highlight the need for tailored approaches to food safety education and regulations.

Study limitations. Despite providing valuable insights into the trends and patterns of FBDOs, this study's reliance on retrospective surveillance information might have led to potential underreporting or misclassification of incidents, which could have affected the precision of the estimates. Furthermore, the research did not

explicitly account for possible confounding factors, such as differences in food safety regulations, reporting methods, and access to healthcare, which might have influenced the observed patterns.

The findings of this study provide valuable insights for improving FBDO surveillance and prevention strategies in Saudi Arabia. Advanced statistical models, such as GEE and GLMM, should be integrated into routine epidemiological analyses to better account for temporal and demographic variations in foodborne illnesses. Moreover, targeted public health interventions, including risk-based food safety regulations, public awareness campaigns, and enhanced inspection protocols, are essential for reducing the outbreak frequency and severity. Future research should explore spatial modeling and climate-related factors to refine risk prediction models and inform proactive food safety policies. Collaboration among health authorities, food industry stakeholders, and policymakers is crucial for implementing evidence-based strategies that mitigate FBDO risks and enhance public health protection. However, further studies should incorporate real-time surveillance data, spatial modeling, and additional covariates to enhance the robustness of the findings. This study recommends further investigation into the factors underlying these trends and the development of tailored interventions for high-risk groups and settings.

In conclusion, these findings emphasize the need for ongoing surveillance and targeted interventions in public food establishments. The observed patterns, especially the decrease toward 2020 and the subsequent increase, underscore the need for adaptable public health strategies to alleviate FBDO counts and cases. This research contributes to an in-depth understanding of foodborne illness epidemiology in Saudi Arabia and can inform future prevention and control measures. The implications of these findings extend beyond Saudi Arabia, potentially offering insights for other countries grappling with similar food safety challenges.

Acknowledgment. The author gratefully acknowledge ContentConcepts (www.contentconcepts.com) for the English language editing.

References

1. Al-Asmari F, Ismail AIH. Evaluating food safety knowledge and practices among Saudi women in Al-Ahsa Region, Saudi Arabia. *Ital J Food Saf* 2023; 12: 10716.
2. Gosadi IM. Epidemiology of communicable diseases in Jazan region: situational assessment, risk characterization, and evaluation of prevention and control programs outcomes. *Saudi Med J* 2023; 44: 1073-1084.

3. Springer. Marginal versus random-effects models. [Updated 2005; accessed 2024 Nov 25]. Available from: https://link.springer.com/chapter/10.1007/0-387-28980-1_16#chapter-info
4. Paul S, Zhang X. Small sample GEE estimation of regression parameters for longitudinal data. *Stat Med* 2014; 33: 3869-3881.
5. SDAIA. Open source data in Saudi Arabia. [Updated 2025; accessed 2024 Sep 1]. Available from: <https://open.data.gov.sa/ar/datasets>
6. Trindade D, Ospina R, Amorim LD. Choosing the right strategy to model longitudinal count data in epidemiology: an application with CD4 cell counts. [Updated 2022; accessed 2024 Nov 25]. Available from: <https://riviste.unimi.it/index.php/ebph/article/view/17752>
7. Yirga AA, Melesse SF, Mwambi HG, Ayele DG. Negative binomial mixed models for analyzing longitudinal CD4 count data. *Sci Rep* 2020; 10: 16742.
8. Pittman B, Buta E, Krishnan-Sarin S, O'Malley SS, Liss T, Gueorguieva R. Models for analyzing zero-inflated and overdispersed count data: an application to cigarette and marijuana use. *Nicotine Tob Res* 2018; 22: 1390-1398.
9. Agresti A, Kateri M. Categorical data analysis. [Updated 2012; accessed 2024 Nov 25]. Available from: https://link.springer.com/referenceworkentry/10.1007/978-3-642-04898-2_161
10. Diggle PJ, Heagerty PJ, Liang K-Y, Zeger SL. Analysis of longitudinal data. [Updated 2002; accessed 2024 Nov 25]. Available from: <https://academic.oup.com/book/54026?login=false>
11. Fitzmaurice G, Davidian M, Verbeke G, Molenberghs G. Longitudinal data analysis. [Updated 2008; accessed 2024 Nov 25]. Available from: <https://www.taylorfrancis.com/books/edit/10.1201/9781420011579/longitudinal-data-analysis-geert-molenberghs-marie-davidian-geert-verbeke-garrett-fitzmaurice>
12. Stroup WW. Generalized linear mixed models: modern concepts, methods and applications. [Updated 2012; accessed 2024 Nov 25]. Available from: <https://www.taylorfrancis.com/books/mono/10.1201/b13151/generalized-linear-mixed-models-walter-stroup>
13. Dobson AJ, Barnett AG. An introduction to generalized linear models, 3rd edition. [Updated 2018; accessed 2024 Nov 25]. Available from: <https://academic.oup.com/biometrics/article-abstract/65/2/672/7331783?redirectedFrom=fulltext&login=false>
14. Bozdogan H. Model selection and Akaike's information criterion (AIC): the general theory and its analytical extensions. *Psychometrika* 1987; 52: 345-370.
15. West BT, Welch KB, Galecki AT. Linear mixed models: a practical guide using statistical software. [Updated 2022; accessed 2024 Nov 25]. Available from: <https://www.taylorfrancis.com/books/mono/10.1201/9781003181064/linear-mixed-models-brady-west-kathleen-welch-andrzej-galecki>
16. Cary N. The SAS system for windows: SAS version 93.0 procedure guide. [Updated 2011; accessed 2024 Nov 25]. Available from: https://www.google.com/url?sa=t&source=web&rct=j&opi=89978449&url=https://support.sas.com/documentation/onlinedoc/base/procstat93m1.pdf&ved=2ahUKEwi1pbT5j9yMAxUHRvEDHVkqB-4QFnoECBYQAQ&usg=AOvVaw17na5YLlymXuYP_xO5jhSe
17. Der G, Everitt BS. Applied medical statistics using SAS. [Updated 2012; accessed 2024 Nov 25]. Available from: <https://www.taylorfrancis.com/books/mono/10.1201/b12738/applied-medical-statistics-using-sas-geoff-der-brian-everitt>
18. Shoukri MM. Analysis of correlated data with SAS and R. [Updated 2018; accessed 2024 Nov 25]. Available from: <https://www.taylorfrancis.com/books/mono/10.1201/9781315277738/analysis-correlated-data-sas-mohamed-shoukri>
19. Zhang X, Mallick H, Tang Z, Zhang L, Cui X, Benson AK, et al. Negative binomial mixed models for analyzing microbiome count data. *BMC Bioinformatics* 2017; 18: 4.
20. Han P, Song PX, Wang L. Longitudinal data analysis using the conditional empirical likelihood method. *Can J Stat* 2014; 42: 404-422.
21. Dahmen G, Rochon J, König IR, Ziegler A. Sample size calculations for controlled clinical trials using generalized estimating equations (GEE). *Methods Inf Med* 2004; 43: 451-456.