Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Developing Estimating Incidences of Foodborne Diseases: Probabilistic Profiling Modelling

Manal A. Almusa

mamusa@sfda.gov.sa

Saudi Food and Drug Authority (SFDA)

Omar A. Alhumaidan

Saudi Food and Drug Authority (SFDA)

Abdullah A. Alajlan

Saudi Food and Drug Authority (SFDA)

Ayidh Almansour

Saudi Food and Drug Authority (SFDA)

Najla A. Albaridi

Princess Nourah bint Abdulrahman University

Research Article

Keywords: Foodborne Diseases, Computational Model, Salmonella, Shigella, Actual Reported Food illnesses

Posted Date: August 13th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10556900/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Developing Estimating Incidences of Foodborne Diseases: Probabilistic Profiling Modelling

Manal A. Almusa[1 *], Omar A. Alhumaidan[1], Abdullah A. Alajlan[1] Ayidh Almansour[1], Najla A. Albaridi[2]

[1] Saudi Food and Drug Authority (SFDA), Riyadh 11671, Saudi Arabia

[2] Department of Health Science, College of Health and Rehabilitation, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia (ORCID :0000-0001-8888-4098).

[*] Correspondence should be addressed to Manal A. Almusa (mamusa@sfda.gov.sa)

Abstract

Background: Foodborne diseases (FBD) pose a significant public health and economic threat globally. In Saudi Arabia, underreporting and a lack of predictive surveillance techniques hinder efficient management of these outbreaks. **Objective:** This study developed a new computational probabilistic model for FBD case estimation, which provide better insights into the incidence of these infections. **Methods:** The proposed model adapted the conceptual principles of the WHO envelope approach into a probabilistic modelling framework. Historical surveillance data from Riyadh (2015–2018) and WHO Foodborne Disease Burden Epidemiology Reference Group estimates were combined to estimate the burden of *Salmonella* spp., *Shigella* spp., *Staphylococcus aureus*, and *Bacillus cereus*. Model performance was evaluated using epidemiological surveillance data together with mean absolute error (MAE), root mean square error (RMSE), bias factor (BF), and sensitivity analysis. **Results:** The model demonstrated robust predictive performance, with prediction errors ≤11% for three of the four pathogens. Overall agreement between predicted and observed values was high, yielding a mean absolute error of 2.26 cases and a bias factor of 1.015. While predictions for *Bacillus cereus* were associated with greater uncertainty because of limited surveillance data, sensitivity analysis confirmed that the model remained stable and robust across plausible input variations. **Conclusion**: The proposed probabilistic model provides a practical framework for estimating foodborne disease burden under limited surveillance conditions and may support evidence-based food safety decision-making. Further validation across additional regions and pathogens is needed.

Key words:

Foodborne Diseases, Computational Model, *Salmonella*, *Shigella,* Actual Reported Food illnesses,

1. Introduction:

Foodborne diseases (FBD), commonly referred to as food poisoning, constitute a major global public health concern affecting both industrialized and developing countries (Faour-Klingbeil & Todd, 2019; Grace, 2023). The World Health Organization (WHO) estimates that approximately 600 million people contract foodborne diseases each year. These illnesses result in around 420,000 deaths annually. Children under the age of five account for nearly 30% of these fatalities (WHO, 2021).

The burden of foodborne diseases extends beyond morbidity and mortality. It also imposes substantial economic costs through productivity losses and healthcare expenditures. A comprehensive global assessment of 42 foodborne infectious and chemical hazards between 2000 and 2021 estimated productivity losses of US$310 billion in nominal terms. When adjusted for purchasing power parity (PPP), the estimated losses increased to US$647 billion (Lake et al. ,2026).

To mitigate the impact of foodborne diseases, it is essential to engage in various activities, such as ongoing risk assessment, surveillance, and implementation of evidence-based regulatory interventions (Codex Alimentarius Commission, 2013). Mathematical and statistical models are crucial for facilitating these activities by providing insights into surveillance strategies, risk assessment, and regulatory decision-making (Alvarenga, 2023; Buchanan et al., 2018; Lubbock & Lopez, 2021; Tarlak, 2023).

In food surveillance systems, risk-based sampling plans are integral these systems, facilitating monitoring and regulatory control (Niesters et al., 2013). Their design is contingent upon specific objectives, which may encompass verifying compliance with food safety criteria or identifying contamination events, as outlined in the Codex Alimentarius guidelines (Codex Alimentarius Commission, 2023). Statistical methods, such as one-sample significance tests and Student’s t-tests, are employed in

designing and evaluating food monitoring programs (Alvarenga, 2023; Codex Alimentarius Commission, 2023).

In the context of microbial risk assessment, modelling is integral to food safety risk evaluation and the formulation of regulatory decisions (Alvarenga, 2023; Stavropoulou & Bezirtzoglou, 2019). Techniques such as scenario analysis, predictive microbiology, and quantitative microbial risk assessment (QMRA) are extensively employed to assess food safety interventions and to estimate their potential impact on public health (Teunis & Schijven, 2019).

One of the established risk analysis techniques is the Food Safety Objective (FSO) approach (Buchanan et al., 2018). This framework simulates food production scenarios and assesses control measures throughout the food chain (Gorris, 2005). Similarly, QMRA models simulate a microbial hazard pathway from production to consumption, integrating exposure assessment with dose–response modelling to estimate health outcomes under scenarios (Teunis & Schijven, 2019).

In Saudi Arabia, national surveillance systems have generated a growing body of epidemiological data on foodborne diseases and food poisoning incidents. Several national and regional studies have examined temporal trends, demographic patterns, and geographic variations in reported cases (Alamri et al., 2026; Alfaleh et al., 2025; Alsayeqh, 2015; Alzahrani, 2025). Other studies have focused on suspected food poisoning incidents and laboratory findings at the regional level (Alajlan et al., 2025). Together, these data provide an important evidence base for public health surveillance and food safety policy development.

In parallel, mathematical modeling has become an important tool for supporting food safety decision-making in Saudi Arabia. National initiatives have developed quantitative models to support risk assessment and exposure estimation. These include a QMRA model for *Escherichia coli* O157 in sheep meat and the NNC-V2 dietary exposure model for estimating dietary intake

among the Saudi population (Alhumaidan et al., 2026; Almasaud et al., 2026). These models support exposure assessment, risk characterization, and the evaluation of potential regulatory interventions.

While these methods are crucial for estimating exposure and characterizing risks, the burden of foodborne diseases is also affected by factors such as healthcare-seeking behavior, access to healthcare, laboratory testing, case confirmation, and reporting practices (Belbasis & Bellou, 2018; Goel et al., 2023; Kaydos-Daniels et al., 2013). Consequently, individuals affected by foodborne diseases may not seek medical care, undergo laboratory testing, or be reported, leading to reported cases representing only a small fraction of the actual disease burden.

Surveillance pyramid approaches address this limitation (WHO, 2021b). They describe the sequence of events required for a foodborne disease case to be recorded by a surveillance system (WHO, 2021b; Peleg et al., 2011; Peleg & Normand, 2014). The probability associated with each stage is then estimated. These stages include symptom development, healthcare seeking, sample collection, laboratory testing, case confirmation, and formal case reporting.

An extension of the surveillance pyramid is the envelope approach, which estimates the total burden of foodborne diseases by combining the probabilities associated with each event in the reporting pathway (WHO, 2021b). These probabilities are used to estimate the proportion of cases captured by the surveillance system (WHO, 2021b). The inverse of this proportion is then applied to estimate the number of unreported cases.

The WHO report on the global burden of foodborne diseases (WHO, 2015) and the surveillance report by the Saudi Ministry of Health (Ministry of Health Saudi Arabia, 2018) serve as foundational resources for estimating the incidence of foodborne diseases in Saudi Arabia. The data and reports available facilitate the construction of a probabilistic profiling model aimed at

estimating the incidence of foodborne diseases in the region. This model employs both the surveillance pyramid and envelope approach to ascertain the true incidence of both reported and unreported cases. The proposed model seeks to explore the feasibility of applying computational approaches to support disease burden estimation, address surveillance underreporting, and enhance future foodborne disease monitoring and public health decision-making.

2. Materials and Methods

2.1 Model Overview

The model was developed based on the conceptual principles of the envelope approach used in foodborne disease burden estimation, as described in the WHO Foodborne Diseases Burden Epidemiology Reference Group guidelines (WHO, 2021b; Peleg et al., 2011; Peleg & Normand, 2014). The underlying principles were adapted into a probabilistic computational framework to estimate the incidence of foodborne diseases.

As shown in figure 1 , the model operated by: 1) estimating the total number of potential foodborne illness cases from all microorganisms, including both reported and unreported cases; 2) deriving the estimated number of reported and unreported cases attributed to the microorganisms included in this study (*Salmonella* spp., *Staphylococcus aureus*, *Shigella*, and *Bacillus cereus*); 3) isolating the number of reported cases specifically associated with the modeled microorganisms; and 4) estimating the expected number of cases within a defined geographical region.

Model inputs were represented using probability distributions and evaluated through Monte Carlo simulation (10,000 iterations). The model combined surveillance data with microorganism-specific parameters to estimate reported and unreported foodborne disease cases

associated with *Salmonella* spp., *Staphylococcus aureus*, *Shigella*, and *Bacillus cereus*. The estimated incidence was then calculated for the target geographical region.

The overall data flow and computational sequence applied within the probabilistic framework are summarized in Table 1, which present the foundational data structure and model input specifications.

2.2 Estimating Total Foodborne Illness Cases

In the first step of the estimation process, the probability of an individual contracting a foodborne illness, donated as FBA_i, was represented by a Beta distribution, as shown in Equation 1. This probability was then multiplied by the total number of populations in Saudi Arabia (N) to estimate the total number of foodborne illness cases (A_i), as shown in Equation 2.

$$FBA_i \sim Beta((TC \times n) + 1, n - (TC \times n) + 1) \quad (1)$$

$$A_i = N \times FBA_i \quad (2)$$

As shown in Equation 1, the key parameter in this distribution is TC, which represents the number of foodborne illness cases per cohort (n = 100,000). The value of TC was derived from the global median per 100,000 population reported in Table A 8.2 of the Foodborne Diseases Burden Epidemiology Reference Group (WHO,2015*)*. Since this model is a probabilistic model, the calculation was repeated over multiple iterations i following the Monte Carlo approach.

2.3 Estimating Cases Attributable to Target Microorganisms

In this step, the model derives the number of underlying reported and unreported cases attributed to the microorganisms (D_{ji}) included in this study *Salmonella* spp., *Staphylococcus aureus*, *Shigella* and *Bacillus cereus*. This is done by taking the previously estimated total number of cases

A_i and multiplying it by the probability that an individual will contract a foodborne illness from a given microorganism, denoted as FB_{ji}.

As shown in Equation 3, this probability follows a Beta distribution. In each iteration, a random value is drawn from this distribution and multiplied by the corresponding A_i value to obtain the estimated number of cases for that iteration, as shown in Equation 4.

$$FB_{ji} \sim Beta((CC_j \times (n \times TC)) + 1, (n \times TC) - (CC_j \times (n \times TC)) + 1) \quad (3)$$

$$D_{ji} = A_i \times FB_{ji} \quad (4)$$

Where CC_j is the proportion of foodborne illness cases attributed to each microorganism. The value of CC_j was extracted from the *Foodborne Diseases Burden Epidemiology Reference Group 2007–2015 report*, which includes the distribution of cases out of a total of 7,968 foodborne illness cases. This yielding $CC_{Salmonella\ spp}$ (1279/7968), $CC_{Staphylococcus\ aureus}$ (50/7968), $CC_{Shigella}$ (741/7968), and $CC_{Bacillus\ cereus}$ (6/7968).

2.4 Isolating Reported Cases of Modelled Microorganisms

As not all individuals affected by foodborne formally report their cases, the number of reported case (RD_{ji}) was estimated by applying probability of reporting Foodborne illness (R_i). This probability was obtained by Equation 5, and 6.

$$RC_i = ARC_{2015-2018} \div A_i \quad (5)$$

$$R_i \sim Beta((RC_i \times A_i) + 1, A_i - (RC_i \times A_i) + 1) \quad (6)$$

Where RC_i is the ratio of reported to total infection, and $ARC_{2015-2018}$ is the average number of reported cases from 2015 to 2018. Finally, the probability derived from Equation 6 was applied to D_{ji} to obtain the estimated outcome for each iteration, as presented in Equation 7.

$$RD_{ji} = D_{ji} \times R_i \quad (7)$$

2.5 Identifying Expected Cases by Geographical Area

After isolating the reported cases, the model estimated total reported foodborne illness in a specific region (RDC_{ji}). using the regional weight. This was mathematically represented in Equations 8–9.

$$W_i \sim U(LB, UB) \quad (8)$$

$$RDC_{ji} = RD_{ji} \times W_i \quad (9)$$

Where W_i is probability of reporting foodborne illness in a specific region. The W_i calculated using the region's share from the total number of cases, as derived from the Ministry of Health data (*MOH Publications - Annual Report on Foodborne Disease Outbreaks-2018,* n.d.). For Riyadh, the share was 16% of national cases ($RW = 0.16$). This was mathematically represented in Equations 10–11.

$$LB = RW - (LU \times RW) \quad (10)$$

$$UB = RW + (LU \times RW) \quad (11)$$

To capture yearly variation, an uncertainty level (LU) of ±25% was applied, defining a uniform distribution for the probability of reporting foodborne illness in a specific region as shown in Equation 8.

2.6. Model Validation

Model performance was evaluated by comparing the estimated number of cases to the epidemiologically reported data as reported in Alajlan et al. (2025). Three quantitative metrics

were computed as follows: Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and Bias Factor (BF) are given as:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n} | P_i - O_i | \quad (12)$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - O_i)^2} \quad (13)$$

$$\bar{BF} = \frac{1}{n}\sum_{i=1}^{n}\frac{P_i}{O_i} \quad (14)$$

Where (P_i) is the predicted number of cases, (O_i) is the observed number of cases, and (n) is the number of pathogens included in the evaluation. A bias factor (BF) of 1 indicates perfect agreement between predicted and observed values, whereas values greater than 1 indicate overestimation and values less than 1 indicate underestimation.

3. Results

3.1 Estimated foodborne illness cases

As shown in Table 2, the estimated total number of foodborne illness cases from all etiologies was 2,630,542 (95% CI: 2,574,769–2,687,001). In modeling the specific pathogens *Salmonella*, *Staphylococcus aureus*, *Shigella*, and *Bacillus cereus*, the predominant estimated cases were linked to *Salmonella* (422,579, 95% CI: 400,145–445,941), followed by *Shigella* (244,949, 95% CI: 227,622–262,622), *Staphylococcus aureus* (16,717, 95% CI: 12,588–21,628), and *Bacillus cereus* (2,197, 95% CI: 932–4,380).

3.2 Comparison Between Model Predictions and Epidemiological Data

As shown in Figure 2, the model predicted the distribution of reported foodborne illness cases in Riyadh as follows: *Salmonella spp.* 62 cases (95% CI: 47–78), *Staphylococcus aureus* 3 cases (95% CI: 2–4), *Shigella spp.* 36 cases (95% CI: 27–45), and *Bacillus cereus* 0.3 cases (95%

CI: 0.1–1). The corresponding epidemiologically reported values extracted from (Alajlan et al., 2025b) were 70, 3, 37, and 0.25 (1 case every 4 years), respectively. When comparing the model estimates to the reported data, the difference ranged from 0% for *Staphylococcus aureus* to 11% for *Salmonella* spp., 3% for *Shigella* spp., and 20% for *Bacillus cereus*, indicating that the predicted values were consistent with the observed epidemiological records within the confidence intervals.

The MAE was 2.26 cases, indicating minor error in predicting cases across all four pathogens (Table 3). Conversely, the RMSE was 4.03 cases, higher than the MAE because of the larger absolute error for *Salmonella* spp. (8 cases) and the greater weight assigned to squared errors. The mean BF for all pathogens was 1.015, suggesting no significant prediction bias. For *Bacillus cereus*, the percentage error was 20%, corresponding to an absolute difference of approximately 0.05 cases (predicted: 0.30; observed: 0.25). The individual BF was 1.200, and the predicted value was within the 95% confidence interval (CI) of the observed estimate (0.1–1). The corresponding BF, MAE, and RMSE values are presented in Table 3.

4. Discussion

4.1 Main Findings

In the current study, a computational model has been designed to estimate the incidence of foodborne diseases by leveraging multiple datasets, as detailed in the methodology section. Using historical outbreak data, and that the infection originates from a microbe-specific probability from the burden report (*WHO*, 2015). The annual incidence rate of FBDs has been effectively estimated by our model.

The findings suggest that the model effectively provides accurate estimates of foodborne disease incidence by efficiently using existing surveillance data. The strong correlation between

the model's predictions and the actual reported cases underscores its reliability. Specifically, the model forecasted 62 cases of Salmonella compared to 70 reported cases, 36 cases of Shigella versus 37 reported cases, and accurately predicted 3 cases of Staphylococcus aureus.

Notably, the prediction errors were ≤11% for *Salmonella* spp., *Shigella* spp., and *Staphylococcus aureus*, with a mean absolute error (MAE) of cases and a bias factor of. This indicates that the model tends to predict slightly lower incidences, likely due to underreporting and underestimation of the disease burden, consistent with the application of global burden parameters (WHO, 2015).

One key finding is the discrepancy between the predicted and observed total foodborne disease (FBD) burden. While approximately 2.63 million cases were anticipated, Ministry of Health (MOH) surveillance data report an average of only 2,410 cases per year, resulting in a ratio exceeding 1,000:1. It is crucial to note that this latter figure does not accurately represent the true burden due to underreporting in surveillance, as documented in numerous countries (Alsayeqh, 2015; Havelaar et al., 2015). Consequently, the explicit and quantitative illustration of the gap between observed and actual burdens should be regarded as a contribution of the proposed framework, as it enhances public health agencies' comprehension of the magnitude of the undetected FBD burden.

The model indicates the predominance of *Salmonella* and *Shigella* among the selected bacteria, attributed to Riyadh's high ambient temperature and consumption of fresh, minimally processed foods, which create conditions conducive to bacterial proliferation in food matrices (Preetha & Narayanan, 2020). Conversely, in northern Europe, such as Denmark, a different pathogen prevails, and noroviruses constitute most of the burden due to person-to-person

transmission and adaptation to cold climates. Consequently, a burden modelling approach must consider local conditions and pathogens.

4.2 Relationship to WHO Global Burden Estimates

The estimates generated by the FERG of the WHO represent the most detailed and accurate assessment of global FBD burden, although they cannot be directly used at the national/sub-national levels. The WHO burden estimates are derived from pooled multi-country data (Havelaar et al., 2015), and there are no country-specific estimates available for Saudi Arabia. Hence, applying WHO FERG median parameters to Riyadh will lead to the loss of their inherent uncertainties and will produce only one point estimate without accounting for regional specifics.

The framework addresses issues in two ways. First, model parameters were treated as informed priors, with calibrated reporting probability and regional weight from MOH surveillance data, integrating WHO burden estimates into an approach combining vetted priors with regional information. Second, the Monte Carlo approach accounts for uncertainty by generating distributions for each parameter combination and 95% confidence intervals across 10,000 simulations, quantifying uncertainty better than extrapolating WHO median estimates.

4.3 Practical Application of the model

The proposed methodology enables limited surveillance data to be translated into quantitative burden estimates that support evidence-based decision-making by public health authorities, including the SFDA and Ministry of Health. This framework can be applied to estimate surveillance gaps for specific foodborne pathogens. It also enables a comparison between the estimated total burden (reported and unreported cases) and the number of cases recorded by the surveillance system. This comparison helps identify pathogens that are likely to be substantially under-detected. For example, the model estimated that only a small proportion of *Salmonella* and

Shigella infections were captured by the official surveillance system. These findings may support the prioritization of laboratory-based surveillance and targeted public health interventions for pathogens with the greatest level of under-detection.

Second, the pathogen-specific burden estimates generated by the model can be used to support the prioritization of food safety inspection and testing activities by the SFDA. Rather than allocating inspection resources equally across all pathogens, the model provides a quantitative basis for prioritizing pathogens with the highest estimated burden in the target region. In the present study, *Salmonella* and *Shigella* were identified as the pathogens with the highest estimated burdens. This approach may improve resource allocation, particularly in areas where inspection capacity is limited.

Third, the proposed approach can provide a quantitative baseline for evaluating changes in the foodborne disease burden over time. When applied to updated surveillance data, the estimated burden can be compared with baseline estimates generated in the present study. This comparison can be used to assess the potential impact of surveillance improvements and public health interventions. This approach may also support the evaluation of evidence-based food safety policies developed by the SFDA.

Fourth, the proposed model requires relatively limited input data, making it suitable for application in other regions of Saudi Arabia and in countries across the Middle East and North Africa (MENA) region with a similar surveillance capacity. The model can be recalibrated using locally available surveillance data to estimate the burden of foodborne diseases in different settings. This allows for comparisons across regions with different food safety profiles and surveillance systems. The relatively low data requirements of the model may facilitate its application in data-limited, public health settings.

There are several opportunities for further development of the proposed framework in the future. Incorporating environmental variables, such as ambient temperature, humidity, and seasonal variation, may improve the model by accounting for the influence of climatic conditions on pathogen growth and transmission, particularly in the high-temperature environment of Riyadh (Preetha & Narayanan, 2020; Rolfe & Daryaei, 2020). The framework can also be extended by incorporating food supply chain data. For example, gravity-based food flow models (Schlaich et al., 2020) may support source attribution and the identification of potential contamination hot spots. In addition, integrating the framework with machine learning methods and real-time data sources, such as emergency department syndromic surveillance or social media data, may enable a timelier estimation of the foodborne disease burden and support outbreak detection and public health response.

4.3 Limitations

Several limitations should be considered when interpreting the findings of this study. First, the limited number of *Bacillus cereus* cases posed a challenge during model evaluation. Only a few cases were available over the study period. Consequently, the estimates for this pathogen were more sensitive to small changes in the input data and showed greater uncertainty than those for the other pathogens. This limitation reflects the scarcity of epidemiological data for *Bacillus cereus* and highlights the need for more comprehensive surveillance of low-frequency foodborne pathogens.

Second, the model was developed and calibrated using surveillance data from Riyadh. Its application to other regions of Saudi Arabia may therefore require recalibration to account for differences in population characteristics, food handling practices, environmental conditions, and pathogen distribution.

Third, the current framework does not include calibration coefficients or correction factors that could improve model performance in other settings. Finally, the model is based primarily on epidemiological surveillance data. It does not account for factors such as socioeconomic characteristics, healthcare-seeking behavior, or food supply chain variables, all of which may influence the burden of foodborne diseases. Future studies should incorporate these variables and externally validate the model using surveillance data from other regions of Saudi Arabia.

5. Conclusion

This study developed a probabilistic computational model for estimating foodborne disease burden using surveillance and WHO burden data. The proposed framework provides a practical approach for estimating reported and unreported foodborne disease cases under limited surveillance conditions and may support evidence-based food safety decision-making. Future studies should focus on external validation, recalibration, and application across additional regions and pathogens.

Author contributions: CRediT

Manal A. Almusa: Conceptualization, methodology, writing—original draft preparation, writing—review and editing**,** investigation, visualization, resources. **Omar A. Alhumaidan**: Conceptualization, methodology, software, validation, investigation, visualization, writing—original draft, writing—review and editing**,** resources, formal analysis. **Abdullah A. Alajlan:** writing—review and editing, supervision, Conceptualization, resources. project administration. **Ayidh M. ALMansour**: writing—review and editing. **Najla A. Albaridi**: writing—review and editing.

Funding

This study did not receive any financial support. The authors confirm that the study design, data analysis, decision to publish, and manuscript preparation were conducted independently.

Acknowledgement

The authors are grateful to the Research and Laboratories Sector at the SFDA for their keen interest in this study and their instrumental support. The authors also would like to thank Princess Nourah bint Abdulrahman University Researchers supporting project number (PNURSP2024R130), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia

Conflicts of Interest:

The authors declare no conflicts of interest.

Disclaimer

The views expressed in this paper are those of the author(s) and not do not necessarily reflect those of the SFDA or its stakeholders. Guaranteeing the accuracy and the validity of the data is a sole responsibility of the research team

Availability of Data and Materials:

Data supporting the findings of this study are available upon reasonable request. Responsibility for the accuracy and validity of the data rests solely with the research team. Please contac Manal A. Almusa via email at mamusa@sfda.gov.sa . The model code and related computations can be accessed via the Kaggle platform at:

https://www.kaggle.com/code/omaraalhumaidan/fbd-model

Consent to Publish declaration: Not applicable.

Ethics and Consent to Participate: Not applicable.

References

Alajlan, A. A., Almansour, A., Alhumaidan, O. A., Alwetaid, M. Y., Al-Shehri, A. A., Alajel, S. M., Almusa, M., & Albaridi, N. A. (2025). *Analysis of suspected food poisoning incidents in Riyadh, Saudi Arabia: An epidemiological perspective*. *International Journal of Microbiology, 2025*(1). https://doi.org/10.1155/IJM/4803121

Alamri, A., Alwetaid, M., Alshawaish, N. et al. (2026). *Temporal and regional trends of foodborne diseases in Saudi Arabia (2015–2023): A nationwide retrospective analysis*. *Journal of Epidemiology and Global Health*. *Journal of Epidemiology and Global Health.* https://doi.org/10.1007/s44197-025-00514-7

Alfaleh, A., Alkattan, A., Alzaher, A., Sagor, K., Almarshoud, S., & Ibrahim, M. H. (2025). Infectious foodborne disease outbreaks and sporadic cases in Saudi Arabia: A 2017–2023 trend analysis. Infectious Diseases Now, 55(5), 105079. https://doi.org/10.1016/j.idnow.2025.105079

Alhumaidan, O., Alakeel, S., Alkhunein, S., & Fallata, G. (2026). Estimating dietary intake and nutrient adequacy among adults in Saudi Arabia: A population-based assessment. BMC Nutrition, 12, 118. https://doi.org/10.1186/s40795-026-01341-0

Almasaud, L., Alfaifi, S., Almadani, O., & Alhumaidan, O. (2026). A quantitative microbial risk assessment of Escherichia coli O157 in sheep meat in Saudi Arabia. Research Square. https://doi.org/10.21203/rs.3.rs-9449338/v1

Alzahrani, H. A. (2025). *Longitudinal analysis of foodborne disease outbreaks in Saudi Arabia: Trends and demographic patterns (2017–2023). Saudi Medical Journal, 46*(5), 507–514. https://doi.org/10.15537/smj.2025.46.5.20241010

Alsayeqh, A. F. (2015). *Foodborne disease risk factors among women in Riyadh, Saudi Arabia*. *Food Control, 50*, 85–91. https://doi.org/10.1016/j.foodcont.2014.08.036

Alvarenga, V. O. (2023). *Basic protocols in predictive food microbiology*. Springer. https://doi.org/10.1007/978-3-031-20788-7

406 Belbasis, L., & Bellou, V. (2018). Introduction to epidemiological studies. *Methods in Molecular*
407 *Biology, 1793*, 1–6. https://doi.org/10.1007/978-1-4939-7868-7_1

408 Buchanan, R. L., Anderson, W., Anelich, L., Cordier, J.-L., Dewanti-Hariyadi, R., Ross, T., &
409 Zwietering, M. H. (2018). *Microorganisms in foods 7: Microbiological testing in food*
410 *safety management*. Springer.

411 Centers for Disease Control and Prevention. (2019). *National Outbreak Reporting System (NORS).*
412 https://www.cdc.gov/nors

413 Codex Alimentarius Commission. (2013). *Principles and guidelines for the establishment and*
414 *application of microbiological criteria related to foods (CXG 21-1997).* FAO & WHO.
415 https://www.fao.org/fao-who-codexalimentarius/

416 Codex Alimentarius Commission. (2023). *General guidelines on sampling (CXG 50-2004).* FAO
417 & WHO.

418 Faour-Klingbeil, D., & Todd, E. C. D. (2019). Prevention and control of foodborne diseases in
419 Middle-East North African countries: Review of national control systems. *International*
420 *Journal of Environmental Research and Public Health, 17*(1), 70.
421 https://doi.org/10.3390/ijerph17010070

422 General Authority for Statistics. (n.d.). *General Authority for Statistics*.
423 https://www.stats.gov.sa/en/home

424 Goel, V., Mathew, S., Gudi, N., Jacob, A., & John, O. (2023). A scoping review on laboratory
425 surveillance in the WHO Southeast Asia Region: Past, present and the future. *Journal of*
426 *Global Health, 13*, 04028. https://doi.org/10.7189/jogh.13.04028

427 Gorris, L. G. M. (2005). Food safety objective: An integral part of food chain management. *Food*
428 *Control, 16*(9), 801–809. https://doi.org/10.1016/j.foodcont.2004.06.028

429 Grace, D. (2023). Burden of foodborne disease in low-income and middle-income countries and
430 opportunities for scaling food safety interventions. *Food Security, 15*(6), 1475–1488.
431 https://doi.org/10.1007/s12571-023-01391-3

Havelaar, A. H., Kirk, M. D., Torgerson, P. R., Gibb, H. J., Hald, T., Lake, R. J., Praet, N., Bellinger, D. C., de Silva, N. R., Gargouri, N., Speybroeck, N., Cawthorne, A., Mathers, C., Stein, C., Angulo, F. J., Devleesschauwer, B., & Zeilmaker, M. (2015). World Health Organization global estimates and regional comparisons of the burden of foodborne disease in 2010. *PLOS Medicine, 12*(12), e1001923. https://doi.org/10.1371/journal.pmed.1001923

Kaydos-Daniels, S. C., Smith, L. R., & Farris, T. R. (2013). Biosurveillance in outbreak investigations. *Biosecurity and Bioterrorism: Biodefense Strategy, Practice, and Science, 11*(1), 20–28. https://doi.org/10.1089/bsp.2011.0109

Lake, R. J., Devleesschauwer, B., Majowicz, S. E., Robertson, L. J., Jakobsen, L. S., Agudo, A., ... & Minato, Y. (2026). WHO estimates of the global, regional, and national burden of 42 foodborne infectious and chemical hazards, 2000–21: An updated data synthesis. *The Lancet Global Health*.

Lubbock, A. L. R., & Lopez, C. F. (2021). Programmatic modeling for biological systems. *Current Opinion in Systems Biology, 27*, 100343. https://doi.org/10.1016/j.coisb.2021.05.004

Ministry of Health Saudi Arabia. (2018). *Annual report on foodborne disease outbreaks 2018*. https://www.moh.gov.sa/en/Ministry/MediaCenter/Publications/Pages/Publications-2019-11-12-001.aspx

Niesters, H. G., Rossen, J. W., van der Avoort, H., Baas, D., Benschop, K., Claas, E. C., Kroneman, A., van Maarseveen, N., Pas, S., van Pelt, W., Rahamat-Langendoen, J. C., Schuurman, R., Vennema, H., Verhoef, L., Wolthers, K., & Koopmans, M. (2013). Laboratory-based surveillance in the molecular era: The TYPENED model, a joint data-sharing platform for clinical and public health laboratories. *Euro Surveillance, 18*(4). https://doi.org/10.2807/ese.18.04.20387-en

O'Brien, S. J., Larose, T. L., Adak, G. K., Evans, M. R., & Tam, C. C. (2016). Modelling study to estimate the health burden of foodborne diseases: Cases, general practice consultations and hospitalisations in the UK, 2009. *BMJ Open, 6*(9), e011119. https://doi.org/10.1136/bmjopen-2016-011119

Oliveira-Filho, E. R., Campos-Silva, R., & Hanson, A. D. (2025). Running Fermi calculations as a superpower to gauge reality. *Plant Physiology, 198*(3), 347. https://doi.org/10.1093/plphys/kiae347

Peleg, M., & Normand, M. D. (2014). Expanded Fermi solution to estimate health risks or benefits from interacting factors. *Food Research International, 64*, 371–379. https://doi.org/10.1016/j.foodres.2014.07.001

Peleg, M., Normand, M. D., Horowitz, J., & Corradini, M. G. (2011). Expanded Fermi solution for estimating the survival of ingested pathogenic and probiotic microbial cells and spores. *Applied and Environmental Microbiology, 77*(1), 312–319. https://doi.org/10.1128/AEM.01448-10

Preetha, S. S., & Narayanan, R. (2020). Factors influencing the development of microbes in food. *Shanlax International Journal of Arts, Science and Humanities, 7*(3), 57–77. https://doi.org/10.34293/sijash.v7i3.473

Rolfe, C., & Daryaei, H. (2020). Intrinsic and extrinsic factors affecting microbial growth in food systems. *Food Engineering Series*, 3–24. https://doi.org/10.1007/978-3-030-42660-6_1

Schlaich, T., Horn, A. L., Fuhrmann, M., & Friedrich, H. (2020). A gravity-based food flow model to identify the source of foodborne disease outbreaks. *International Journal of Environmental Research and Public Health, 17*(2), 444. https://doi.org/10.3390/ijerph17020444

Stavropoulou, E., & Bezirtzoglou, E. (2019). Predictive modeling of microbial behavior in food. *Foods, 8*(12), 654. https://doi.org/10.3390/foods8120654

Tarlak, F. (2023). The use of predictive microbiology for the prediction of the shelf life of food products. *Foods, 12*(24), 4461. https://doi.org/10.3390/foods12244461

Teunis, P., & Schijven, J. F. (2019). *Generic guidance to quantitative microbial risk assessment for food and water*. National Institute for Public Health and the Environment (RIVM).

World Health Organization. (2015). *WHO estimates of the global burden of foodborne diseases: Foodborne Disease Burden Epidemiology Reference Group 2007–2015*. World Health Organization. https://www.who.int/publications/i/item/9789241565165

World Health Organization. (2021a). *Estimating the burden of foodborne diseases*. https://www.who.int/activities/estimating-the-burden-of-foodborne-diseases

World Health Organization. (2021b). *Estimating the burden of foodborne diseases: A practical handbook for countries: A guide for planning, implementing and reporting country-level burden of foodborne disease: Executive summary*. World Health Organization.

Table 1. Data and parameters for the Model

Variable	Description	Calculation / Input	Data source	Units
Total Number of Foodborne Illness Cases				
A_i	Total number of foodborne illness cases for giving iteration i.	$A_i = N \times FBA_i$		Cases
FBA_i	Probability of an individual contracting a foodborne disease for giving iteration i.	$FBA_i \sim Beta((TC \times n) + 1, n - (TC \times n) + 1)$		Fraction
TC	Total number of cases from foodborne illness in cohort.	$TC = 7968$ /n		Fraction
n	Size for cohort of 100,000 individual.	$n = 100{,}000$		Cases
N	Total number of populations in Saudi Arabia.	$N = 33013253$	Data extract form General Authority for Statistics (GASTAT) database (*General Authority for Statistics* , nd)	individuals
i	Number of iterations during modeling process.	$i = \{1,2,3, \ldots, 10000\}$		
Total Number of Foodborne Illness Cases from a given Microbe				
D_{ji}	Total number of foodborne illness cases from a given microbe for giving iteration i.	$D_{ji} = A_i \times FB_{ji}$		Cases
FB_{ji}	Probability that an individual will contract a foodborne disease from a given microbe j	$FB_{ji} \sim Beta((CC_j \times (n \times TC)) + 1, (n \times TC) - (CC_j \times (n \times TC)) + 1)$		Fraction
CC_j	Number of cases from foodborne illness attributed from a given microbe j.	$CC_{Salmonella\ spp}$ = 1279/ 7968 $CC_{staphylococus\ aureus}$ = 50/ 7968 $CC_{Shigella}$ = 741/ 7968 $CC_{Bacillus\ cereus}$ = 6 / 7968.	Data extract form Foodborne Diseases Burden Epidemiology Reference Group 2007-2015 report (*WHO*, 2015), 2020)	Fraction
j	Index for the pathogen during modeling process.	$j \in \{Salmonella\ spp,\ staphylococus\ aureus\ ,\ Shigella\ ,\ Bacillus\ cereus\}$		-
Total Reported Cases for A Given Foodborne Microbe				
RD_{ji}	Total reported cases for a given foodborne microbe for giving iteration i.	$RD_{ji} = D_{ji} \times R_i$		Cases
R_i	Probability of reporting Foodborne illness from a given microbe j	$R_i \sim Beta((RC_i \times A_i) + 1, A_i - (RC_i \times A_i) + 1)$		Fraction
RC_i	Ratio of reported cases to total infected papulation for giving iteration i.	$RC_i = ARC_{2015-2018} \div A_i$		Fraction
$ARC_{2015-2018}$	Average Foodborne illness reported case from a given microbe j	$ARC_{2015-2018} = \frac{(2196 + 2113 + 3140 + 2191)}{4} = 2410$	MOH Annual Reports on Foodborne Disease Outbreaks, 2015–2018 (*MOH Publications - Annual Report on Foodborne Disease Outbreaks-2018*, n.d.)	Cases
Total Reported Foodborne Illness in A Specific Region				

RDC_{ji}	Total reported foodborne illness in a specific region for giving iteration i.	$RDC_{ji} = RD_{ji} \times W_i$		Cases
W_i	Probability of reporting foodborne illness in a specific region for giving iteration i.	$W_i \sim U(LB, UB)$ $LB = RW - (LU \times RW)$ $UB = RW + (LU \times RW)$		Fraction
LU	Level of uncertainty for 25%.	$LU = 0.25$	MOH Annual Reports on Foodborne Disease Outbreaks, 2015–2018 (*MOH Publications - Annual Report on Foodborne Disease Outbreaks-2018*, n.d.)	Fraction
RW	The proportion of individuals infected in Riyadh in Saudi Arabia.	$RW = 0.16$	MOH Annual Reports on Foodborne Disease Outbreaks, 2015–2018 (*MOH Publications - Annual Report on Foodborne Disease Outbreaks-2018*, n.d.)	Fraction

Table 2. Estimated foodborne illness cases (total, reported, and reported in Riyadh region) by pathogen.

		Salmonella	***Staphylococcus aureus***	***Shigella***	***Bacillus cereus***
Total number of cases	**2.5th**	400145	12588	227622	932
	50th	422579	16717	244949	2197
	97.5th	445941	21628	262622	4380
Number of reported cases	**2.5th**	363	11	206	1
	50th	387	15	224	2
	97.5th	412	19	242	4
Number of reported cases in Riyadh	**2.5th**	47	2	27	0.1
	50th	62	3	36	0.3
	97.5th	78	4	45	1

Table 3. Validation of model-predicted versus epidemiologically observed foodborne illness cases in Riyadh for four bacterial pathogens.

Pathogen	P_i(Predicted)	O_i(Observed)	$\| \boldsymbol{P_i} - \boldsymbol{O_i} \|$	$(\boldsymbol{P_i} - \boldsymbol{O_i})^2$	$\mathbf{BF}_i$
Salmonella spp.	62	70	8	64	0.886
S. aureus	3	3	0	0	1.000
Shigella spp.	36	37	1	1	0.973
B. cereus	0.3	0.25	0.05	0.0025	1.200
Mean	—	—	—	—	**0.965**

525

526

527

528

529

530

531

532

533

534

535

536

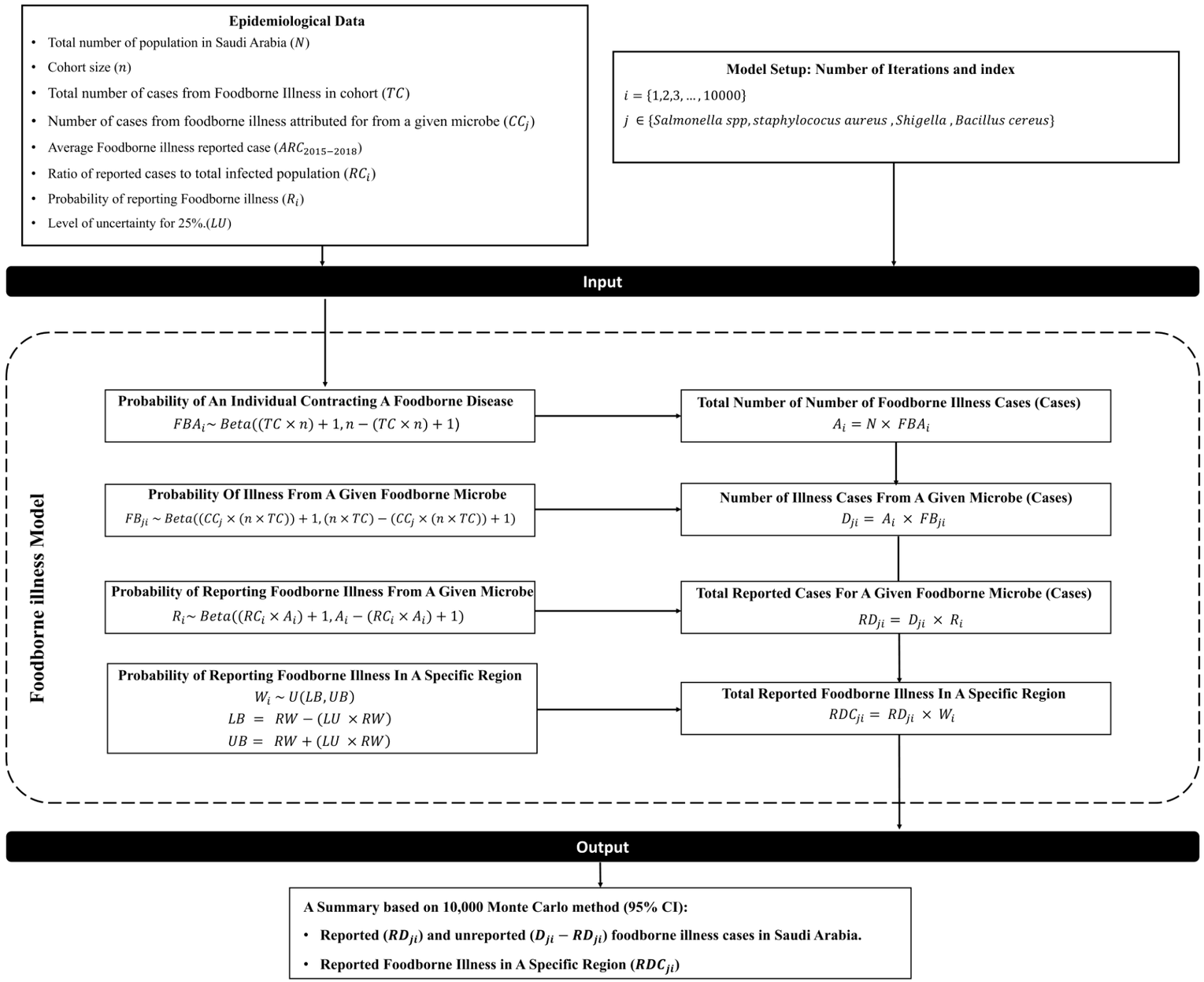
537

538

539

540

541



542

543 **Figure. 1**. The conceptual model for estimating the number of foodborne illnesses

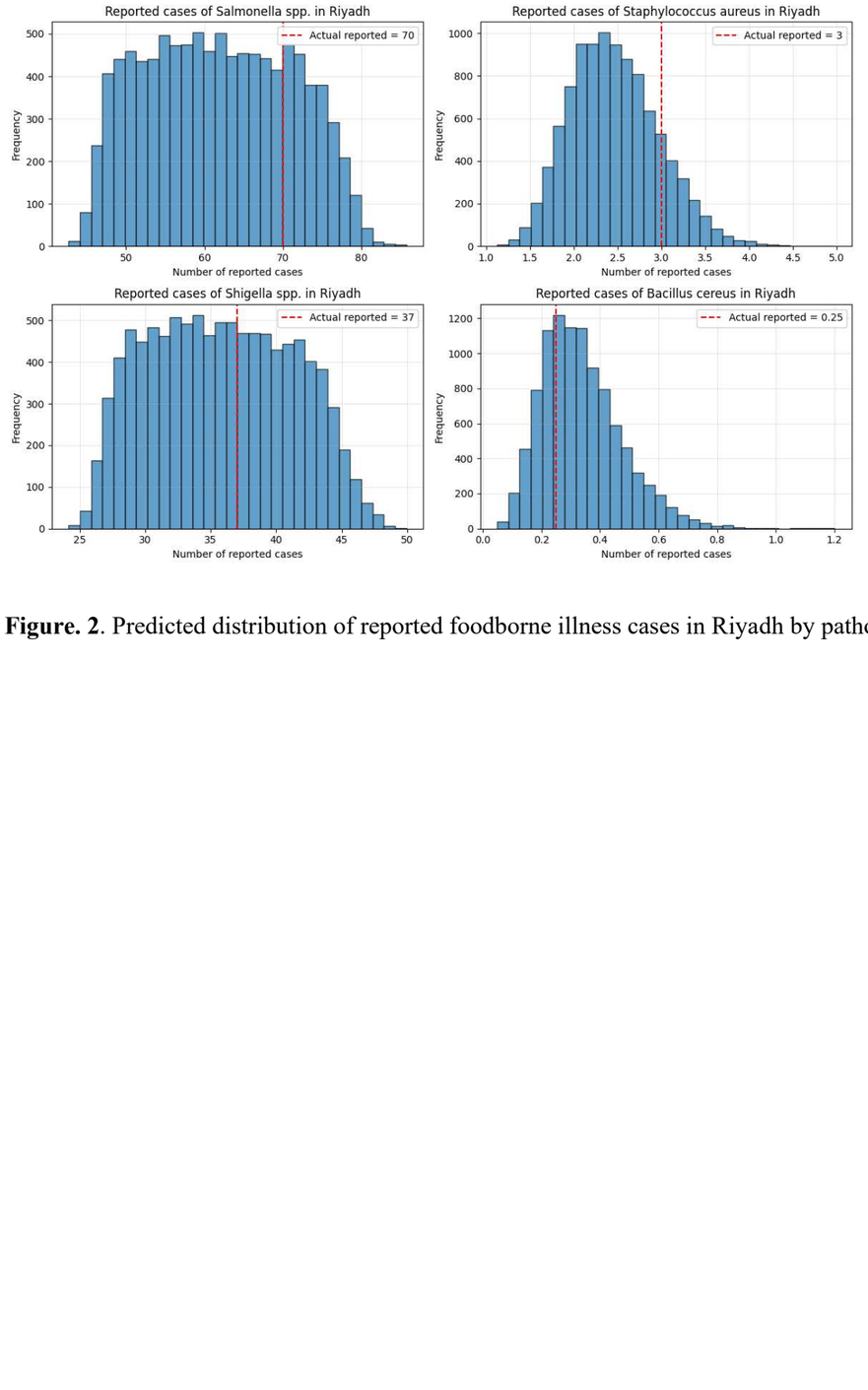


Figure. 2. Predicted distribution of reported foodborne illness cases in Riyadh by pathogen, compared with epidemiologically reported values.