## Evidence of relationship between food business hygiene compliance and measures of food safety

September 2018

Shamim Rahman, David Millson, and Laura Thomson

Operational Research, Analytics Unit, Science Evidence and Research Division



## Acknowledgments

The FSA is grateful to Public Health England who provided outbreak data used in this analysis

## **Contents**

1		Gloss	ary	5					
2		Executive Summary7							
3		Background8							
4		The r	elationship between food business compliance and microbiological contamination sample						
οι	tc	omes		)					
	4.:	1 I	ntroduction	)					
		4.1.1	Sampling Data	)					
	4.2	2 1	Methodology11	L					
		4.2.1	LAEMS data12	<u>)</u>					
		4.2.2	UKFSS data12	<u>)</u>					
		4.2.3	Statistical tests	3					
	4.3	3 I	Results14	ţ					
		4.3.1 prem	The relationship between FHRS rating and proportion of unsatisfactory samples at a ises (LAEMS)14	1					
		4.3.2 prem	The relationship between FHRS rating and proportion of unsatisfactory samples at a ises by product group (LAEMS)	5					
		4.3.3 prem	The relationship between FHRS rating and proportion of unsatisfactory samples at a ises by business type (LAEMS)17	7					
		4.3.4 unsat	The relationship between component scores that derive FHRS and proportion of isfactory samples at a premises (LAEMS)17	7					
		4.3.5 (UKFS	Food business compliance and microbiological contamination samples outcomes (S) 19						
		4.3.6 (UKFS	Food business compliance and microbiological contamination samples outcomes S) with TVC30 removed21	L					
		4.3.7 unsat	The relationship between component scores that derive FHRS and proportion of isfactory samples at a premises (UKFSS)22	L					
	4.4	4 (	Conclusions and Caveats23	3					
5 01	ıtb	The rore	elationship between food business compliance and identified foodborne disease	5					
	5.3	1 I	ntroduction25	5					
	5.2	2 1	Nethodology25	5					
	5.3	3 I	26 Results	5					
	5.4	4 (	Conclusions and Caveats27	7					
6		Estim	ate of impact of the FHRS scheme on foodborne disease	3					
	6.:	1 I	ntroduction	3					

	6.2	Methodology	. 28			
	6.3	Results	. 28			
	6.4	Conclusions and Caveats	. 29			
7	Con	clusions	. 30			
Aı	nnexe 1 – Address Matching					

## **1** Glossary

**Broadly Compliant (BC)** Premises whose compliance levels have been assessed as equivalent to an FHRS rating of 3, 4 or 5 at their most recent food hygiene inspection.

**Error bars** A graphical representation of the variability of data used to show the error, or uncertainty in a reported measurement. If repeated samples were taken and the 95% confidence interval was computed for each sample, 95% of the intervals would contain the actual value of the parameter being estimated. Errors bars illustrated in this report show the 95% confidence interval.

**FHRS equivalent rating** As FHRS ratings (ranging from 0 at the bottom to 5 at the top) are determined on the basis of compliance levels at the time of the last inspection, equivalent ratings can be calculated for all inspected premises, regardless of whether the FHRS was rolled out in the relevant Local Authority at the time or whether the premises type is covered by the FHRS scheme.

**Food business compliance** The requirements of compliance with food hygiene law for food businesses are outlined in the Food Law Code of Practice<sup>1</sup>. Environmental health officers inspect food premises, assessing compliance level against three criteria: compliance with food safety and hygiene procedures; compliance with structural requirements; and confidence in management and control procedures. Numerical scores are given for each of these criteria (the lower the score, the more compliant that premises is). These three scores determine the compliance level.

**Foodborne disease (FBD)** An illness resulting from the ingestion of foodstuffs contaminated with microorganisms or chemicals.

**Food Hygiene Rating Scheme (FHRS)** Any food business premises that serves or supplies food direct to consumers is covered by the Food Hygiene Rating Scheme. When a premises is inspected, it is given a hygiene rating from '0' (Urgent Improvement Necessary) to '5' (Very Good), based on the hygiene standards found at the time. These hygiene standards are determined by converting the numerical scores of the three component compliance criteria from the inspection into a single rating: the FHRS rating. The lower the criteria scores, the higher the FHRS rating will be. <u>Further details of the scheme</u>.

**Fully Compliant (FC)** Premises whose compliance levels have been assessed as equivalent to an FHRS rating of 5 at their most recent food hygiene inspection. Fully compliant premises are a subset of the broadly compliant category i.e. all fully compliant premises are also broadly compliant.

**Indicator species** Due to challenges in accurately identifying low population levels, laboratory tests generally do not endeavour to count the numbers of specific bacterial pathogens in a sample. Instead, indicator bacteria, which do exist in test samples in sufficient numbers to be counted, are used to test for microbiological quality. For example, the numbers of *Enterobacteriaceae* present in a test sample can be thought of as a general indicator of the degree of contamination acquired from faecal material.

<sup>&</sup>lt;sup>1</sup> <u>https://www.food.gov.uk/about-us/food-and-feed-codes-of-practice</u>

**Local Authority Enforcement Monitoring System (LAEMS)** A system used to report local authority food law enforcement activities to the Food Standards Agency (FSA). Local authorities upload premises level data generated from their local food law enforcement activities data systems to LAEMS. Once uploaded to LAEMS, the local authority data are aggregated to the pre-defined categories including 'interventions', 'sampling' and 'enforcement'. Local authorities can submit amendments to the aggregated data, if required, to correct it. Once local authorities are content that the aggregate data are correct, they are required to confirm the accuracy of the data before it is submitted for evaluation and publication by the FSA at an LA level.

Outbreak of foodborne disease illness Two or more linked cases of the same disease.

**Pathogen** An organism that causes disease. They include microorganisms such as bacteria, viruses, and fungi. Examples include *Salmonella*, *Clostridium Perfringens* and *Staphylococcus Aureus*.

**Percentage point (ppt)** The unit for the difference between two percentages. For example, if the proportion of broadly compliant premises in England is 90% in 2013 and 92% in 2015, the growth in broadly compliant premises is 2 percentage points (ppts) over two years.

**Poorly Compliant (PC)** Premises whose compliance levels have been assessed as equivalent to an FHRS rating of either 0 or 1 at their most recent food hygiene inspection. Poorly compliant premises are a subset of the not broadly compliant category i.e. all poorly compliant premises are also not broadly compliant.

**Premises** A food business can have one or more establishments or outlets at different locations. A premises refers to one specific location for that business. For example, for a food business that consists of a chain of takeaways, each takeaway at each location is considered a separate premises.

**P-value** The probability of finding the observed, or more extreme, results when the null hypothesis (usually the hypothesis that there is no observed difference or relationship) is true. A small p-value (typically  $\leq 0.05$ ) indicates strong evidence that the results did not occur purely by chance. In this case, the result would be significant at the 95% confidence level.

**Statistical significance** The probability that a pattern or relationship found in the data occurred purely by chance. The lower the significance level (or equivalently, the higher the confidence level) the higher the likelihood that the result observed is true. For example, if a result is said to be statistically significant at the 0.05 (5%) significance level (or equivalently, the 95% confidence level), the probability is high that the result is true. All references to statistical significance in this report are at the 0.05 (5%) significance level).

**UK Food Surveillance System (UKFSS)** A national database for central storage of analytical results from feed and food samples taken by enforcement authorities (local authorities and port health authorities) as part of their official controls. Not all local authorities use this system for microbiological samples.

## 2 Executive Summary

Estimating the direct impact of food business hygiene compliance on foodborne disease is complex and challenging. This is due to a variety of reasons including underreporting, the difficulty of linking foodborne disease cases to particular food establishments if reported cases are isolated or symptoms delayed, the differences in reporting practices and access to health care between local authorities, and the difficulty determining whether the illness was contracted through food or human contact.

In order to build up evidence for the relationship between compliance with food hygiene law and food safety, a series of analyses was conducted using proxy measures of food safety:

- The relationship between food business compliance and microbiological contamination sample outcomes (recorded in LAEMS and in UKFSS).
- The relationship between food business compliance and identified foodborne disease outbreaks

Although compliance with food hygiene law does not eliminate the risk of outbreaks or unsatisfactory samples results, samples results from LAEMS and UKFSS, and analysis of outbreaks data indicate that premises with higher FHRS ratings are less likely to have unsatisfactory results or encounter outbreaks.

For microbiological contamination samples recorded in LAEMS, broadly compliant (FHRS rating 3, 4, or 5) premises are likely to have a smaller proportion of unsatisfactory samples than not broadly compliant (FHRS rating 0, 1, and 2) and poorly compliant (FHRS rating 0 and 1) premises. Fully compliant (FHRS rating 5) premises are likely to have a smaller proportion of unsatisfactory samples than all other premises. Despite the differences in data recording, the direction of results from samples recorded in UKFSS is the same as LAEMS i.e. for high FHRS ratings, the proportion of unsatisfactory samples at a Premises is smaller than for low FHRS ratings. However, for UKFSS samples, only one result was statistically significant: fully compliant premises (FHRS rating 5) are likely to have a smaller proportion of unsatisfactory samples than all other premises (FHRS rating 5) are likely to have a smaller proportion of unsatisfactory samples than all other premises (FHRS rating 5) are likely to have a smaller proportion of unsatisfactory samples than all other premises (FHRS rating 5) are likely to have a smaller proportion of unsatisfactory samples than all other premises (FHRS rating 5) are likely to have a smaller proportion of unsatisfactory samples than all other premises (FHRS rating 5) are likely to have a smaller proportion of unsatisfactory samples than all other premises (FHRS rating 5) are likely to have a smaller proportion of unsatisfactory samples than all other premises

The food safety proxy measure outcomes for different compliance categories were used to estimate the effect on foodborne disease when compliance levels change. Comparing to 2009-10 (before FHRS) to 2014-15, the model estimates that there are now approximately:

 2.6% to 6.4% fewer cases of FBD from food businesses (based on the increase in proportion of broadly compliant premises)

If broad compliance increased by 1% from 2014-15 levels, it is estimated FBD cases from food businesses would fall by 0.4% to 1%.

The evidence presented may show that the food safety proxy measures change as FHRS ratings change, but cannot be used to conclude that the changes in ratings actually drive the changes in the food safety proxy measures. Although each analysis in isolation does not always yield statistically significant results, all results are in the same direction, and combine to provide an evidence base that indicates that higher food business compliance and food safety have a positive relationship.

## **3 Background**

Estimating the direct impact of food business hygiene compliance on foodborne disease is complex and challenging. There are several reasons for this. Often, we can't link foodborne disease cases to particular food establishments because reported cases are isolated, or symptoms of the illness are delayed. Additionally, illnesses may be reported in a different local authority to where they were contracted, for example if someone is visiting friends in another city and eats there but doesn't develop or report symptoms until they return home. There are also differences in reporting practices and access to health care between local authorities, which may artificially reduce or inflate the reported incidence of foodborne disease cases. In general, foodborne disease cases are actually under-reported. Finally, it is difficult to determine what proportion of reported illness is due to food compared to other causes, as many of the illnesses can also be contracted through human contact.

As an alternative, we can use look at other measures to estimate the relationship between food business compliance and food safety. This report builds up evidence for the correlation between higher FHRS ratings and food safety through a series of analyses:

- The relationship between food business compliance and microbiological contamination sample outcomes
- The relationship between food business compliance and identified foodborne disease outbreaks

This is a collection of observational evidence to investigate any association between FHRS rating and proxy measures of foodborne disease. Throughout the report, the term relationship refers to association rather than causation.

Only results that are statistically significant or required to clarify what has been investigated and discounted are presented here.

# 4 The relationship between food business compliance and microbiological contamination sample outcomes

## 4.1 Introduction

Sampling of food products is an integral part of local authorities work in carrying out official controls and each local authority must publish a food sampling policy and make it available to businesses and consumers<sup>2</sup>.

Environmental health officers will make a risk based professional judgement as to the samples they choose to take and send for testing. The number of samples per premise will depend on many factors, such as scale of the operation, types of food and handling and any potential risk. In addition some samples will be taken as part of wider surveys for a particular premise and food type. Samples are then sent to an official laboratory for testing. All laboratories will be accredited by the United Kingdom Accreditation Service.

Each sample is tested for one or more microbiological organism. These can be for either a pathogen or an indicator bacteria (used to test for microbiological quality). The outcomes of these sample tests can be used as a proxy for food safety. Whether a microbiological contamination sample test outcome is considered satisfactory or not is determined by the limits and procedures set by EU regulation on microbiological criteria of foodstuffs (Commission Regulation (EC) No 2073/2005)<sup>3</sup>.

### 4.1.1 Sampling Data

While the detailed test results will be provided to local authorities from the laboratories, the data held centrally by the FSA is less complete. The FSA has two different systems which hold sample results.

The first is Local Authority Enforcement Monitoring System (LAEMS). Each year all Local authorities upload data that has been generated from their local system(s) on which they record information on food law enforcement activities, onto the LAEMS. Once uploaded onto LAEMS, the local authority data are aggregated to the pre-defined categories required by the FSA. For each sample the data returned gives an anonymised premise identifier, the food product sampled and the sample result, either satisfactory or unsatisfactory. Although a sample may be tested for a number of different microbiological organisms only one result is given, which is based on whether it is satisfactory for all microbes tested for or not. No detail is given for the organisms tested for nor for which specific test(s), if any, were unsatisfactory. Also, although details of product type are available e.g dairy, fish & shellfish, this are quite broad and do no state at what stage in the preparation process they were taken. Despite obvious limitations, the number of samples taken are far more than would be available from any survey and there are not believed to be any systematic biases. Therefore analysis undertaken using these results, with sufficient caveats can be insightful.

<sup>&</sup>lt;sup>2</sup> Further details of sampling protocols and requirements can be found in Chapter 6 of Food Law Code of practice (England): <u>https://www.food.gov.uk/about-us/food-and-feed-codes-of-practice</u>. Similar Code of Practices are also published by the other 3 UK countries

<sup>&</sup>lt;sup>3</sup> <u>http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2005:338:0001:0026:EN:PDF.</u>

The second source of sample data available centrally to the FSA is the UK Food Surveillance System (UKFSS). This is a national database for central storage of analytical results from feed and food samples taken by enforcement authorities (local authorities and port health authorities) as part of their official controls. These database does hold details on test results for specific organisms and as such is more detailed that LAEMs. However, entering results in this system is not compulsory and to date is used by mainly Northern Irish and Scottish local authorities (the later are out of the scope of this work) in terms of microbiological samples (other Local authorities use UKFSS for other types of tests).

As mentioned above sample results are recorded in the Local Authority Enforcement Monitoring System (LAEMS), UK Food Surveillance System (UKFSS), or sometimes both. Table 1 shows the key differences between data used in the analysis from the two systems.

	UKFSS	LAEMS
Time period	Jan 2013 to June 2015	2009-10 to 2013-14
Number of years	2.5 years	5 years
Local Authorities	Northern Ireland	England, Northern Ireland and Wales

Table 1: Differences in data used in analysis from UKFSS versus LAEMS

For the periods covered fewer local authorities used UKFSS to record sample outcomes for microbiological tests compared to those that used LAEMS.

Separate analyses were carried out for samples in UKFSS and in LAEMS. The number of years used was determined by what data was available at time of analysis. Although some samples results from England, Northern Ireland and Wales are recorded in UKFSS, once these were narrowed to microbiological samples, only local authorities in Northern Ireland had sufficient numbers to include in the analysis.

The analysis using LAEMS data was based on 37,304 samples data from 11,891 premises from 190 Local Authorities in England and Wales over 5 years, from 2009-10 to 2013-14. The analysis using matched FHRS and UKFSS data used 7,115 samples from 2,774 premises and 26 Local Authorities from Northern Ireland were gathered from January 2013 to June 2015.

In LAEMS, a sample test outcome is recorded as either satisfactory or unsatisfactory, and it does not show the individual tests that were conducted on the sample to come to this conclusion. UKFSS records the specific tests conducted on each sample. These individual test outcomes on a sample can be satisfactory, borderline, or unsatisfactory. Table 2 shows an anonymised example of what information is recorded in UKFSS for sample results. The sample taken from 'Cooked Diced Chicken' is considered 'Unsatisfactory' overall because one of the organisms tested for, *Enterobacteriaceae*, had an unsatisfactory outcome. The sample taken from 'Frozen Chicken Pieces' is considered 'Satisfactory' overall because all pathogen tests have satisfactory outcomes.

Table 2: Anonymised example of the information recorded on UKFSS for microbiological contamination samples taken at a Premises

Premises	remises Sample No. Food Determination		Determination	Test Outcome	Sample	
		Description	name		Outcome	
EAT-A-LOT CAFE, 502	801HQ01780001829	Cooked Diced	TVC30	RTE Guidelines - Acceptable/Borderline	Unsatisfactory	
HIGH		Chicken	B. cereus	Satisfactory		
ANTRIM.			Bacillus Species	Satisfactory		
BT41 4XY			C. perfringens	Satisfactory		
			E. coli	Satisfactory		
			Enterobacteriaceae	RTE Guidelines – Unsatisfactory		
			Listeria species	Satisfactory		
			Salmonella	Satisfactory		
			Staphylococcus aureus and other species	Satisfactory		
	801HQ01780001842	Frozen	TVC30	Satisfactory	Satisfactory	
		Chicken	B. cereus	Satisfactory		
		Pieces	Bacillus Species	Satisfactory		
			C. perfringens	Satisfactory		
			E. Coli	Satisfactory		
			Enterobacteriaceae	Satisfactory		
			Listeria species	Satisfactory		
			Salmonella	Satisfactory		
			Staphylococcus	Satisfactory		
			aureus and other			
			species			

Local authorities also appear to differ in how they judge the borderline category for an organism. In some local authorities, a borderline test outcome would automatically mean that sample is considered unsatisfactory overall. In other local authorities, a borderline test outcome would not make a sample unsatisfactory, only an unsatisfactory test outcome would do so. Table 3 shows three example local authorities that record their samples in both LAEMS and UKFSS.

Table 3: A comparison of the proportion of samples considered unsatisfactory in LAEMS and UKFSS for three example local authorities

		UKF	SS
Local Authority	LAEMS	Borderline and Unsatisfactory	Unsatisfactory
Local Authority 1	26%	46%	12%
Local Authority 2	9%	42%	8%
Local Authority 3	11%	46%	12%

Local Authorities 2 and 3 appear to record borderline samples as satisfactory in LAEMS, while Local Authority 1 may be doing a mixture - sometimes borderline is considered unsatisfactory and sometimes satisfactory.

## 4.2 Methodology

The aim of the analysis was to examine the relationship between compliance of a food business Premises with food hygiene law and microbiological contamination sample test outcomes. The mean percentage of unsatisfactory samples at premises with different compliance levels (as measured by FHRS equivalent rating) was compared.

### 4.2.1 LAEMS data

The analysis was carried out with underlying premises level data from LAEMs returns for five years: 2009-10, 2010-11, 2011-12, 2012-13, 2013-14. Each year was considered separately as there were enough premises and samples in each year to allow the drawing of statistical significant findings.

An FHRS equivalent rating can be calculated for all inspected premises from compliance levels recorded in LAEMS, regardless of whether FHRS was rolled out at the time in the LA in which the premises is located, or whether the premises type is one covered by FHRS. This was used to determine the FHRS rating at the premises where the sample was taken, and no linking was required between the LAEMS and FHRS database.

To ensure that the compliance scores given to a premises were not influenced by the microbiological sample results, data from consecutive years was combined:

- The compliance scores were taken from the first year, and the microbiological sample results from the second year.
- For two consecutive years, a premises was considered to be the same between years where it had the same:
  - Unique premises ID
  - Local authority code
  - Risk score given for Type of food and method of handling, Method of processing, Consumers at risk and Vulnerable groups.
  - Risk band at the end of the first year and risk band at the start of the second year.
- Premises that had ceased trading in the first year were excluded.

Where there were small sample numbers (e.g. when looking at samples for particular product types), data from all years were combined to create one dataset, with the time variable eliminated. As the same premises in different years was then analysed as two separate premises, this may affect the independence required for the validity of some of the statistical tests used in the analysis. These results should be used with caution.

All results presented in this report used the data from consecutive years method, apart from the analysis by product type (Figure 2) and individual compliance component scores (Figures 3-5) which use data from all years combined.

### 4.2.2 UKFSS data

Unlike LAEMS, UKFSS does not contain information on compliance levels, so an equivalent FHRS rating cannot be determined from UKFSS alone. The FHRS database contains FHRS ratings for all premises covered by the scheme. Therefore, premises in UKFSS and the FHRS database were linked using fuzzy address matching<sup>4</sup>. For each sample at each premises in UKFSS, the most recent FHRS rating for that premises before the sample date was matched from the FHRS database. This ensured the FHRS ratings given to a premises were not influenced by the microbiological sample results.

All years of data in UKFSS were considered together as separating into individual years would mean the dataset sizes were too small to draw statistically significant conclusions.

<sup>&</sup>lt;sup>4</sup> See Annex 1

Only those samples that had been tested for the main 9 specific pathogens or indicators were considered. These are detailed in Table 4.

Test	Satisfactory (count)	Borderline (count)	Unsatisfactory (count)	Unsatisfactory (proportion)	Unsatisfactory and Borderline (proportion)
B. cereus	7060	262	32	0.4%	4.0%
Bacillus Species	7051	256	32	0.4%	3.9%
Clostridium perfringens	7316	23	1	0.0%	0.3%
E. coli	7205	79	55	0.7%	1.8%
Enterobacteriaceae	6019	957	364	5.0%	18.0%
Listeria species	7738	3	-	0.0%	0.0%
Salmonella	7337	0	2	0.0%	0.0%
Staphylococcus aureus and other species	7185	146	8	0.1%	2.1%
Total Variable Count 30 (TVC30)	5133	1829	377	5.1%	30.1%

Table 4: Tests conducted on the samples included in the analysis, and their outcome

It is also important to note that some organisms such as *Enterobacteriaceae* and total *E. coli* are indicators of faecal contamination, while Total Variable Count 30 (TVC30) is an indicator of quality, not safety. Also some of the pathogens most prevalent as the cause of foodborne disease (e.g. *Campylobacter* and Norovirus) are not investigated routinely so could not be included in this analysis.

2,755 out of 7,115 samples (39%) are unsatisfactory or borderline when tested for the 9 pathogens and/or indicators in Table 4. Each sample just need to be unsatisfactory or borderline for one test to be regarded as unsatisfactory or borderline overall. In practice many samples are unsatisfactory or borderline for more than one test. This means the overall unsatisfactory and borderline rate for microbiological samples is higher than the rates for individual pathogens (e.g. higher than the 30% for TVC30).

Most food samples are classified as unsatisfactory due to indirect evidence such as excess total bacterial counts or measures of faecal contamination, rather than due to direct evidence from the detection of pathogens. Although the microorganisms tested are not necessarily the ones that are likely to cause diseases, some of these microorganisms are associated with other microorganisms that cause disease.

The Health Protection Agency Guidelines for Assessing the Microbiological Safety of Ready-to-eat Foods (2009)<sup>5</sup> does note that TVC30 cannot contribute towards a safety assessment of ready-to-eat foods. Therefore the analysis also considered these samples with the TVC30 results removed when determining overall satisfactory outcome of the sample.

### 4.2.3 Statistical tests

T-tests were undertaken to determine if there was a statistically significant relationship between equivalent FHRS scores (0, 1, 2, 3, 4, 5) and sample test results (satisfactory/unsatisfactory). FHRS equivalent scores were also grouped by the authors into the following categories:

<sup>&</sup>lt;sup>5</sup>https://www.gov.uk/government/uploads/system/uploads/attachment\_data/file/363146/Guidelines\_for\_ass essing\_the\_microbiological\_safety\_of\_ready-to-eat\_foods\_on\_the\_market.pdf

- Fully Compliant (FC) Premises whose compliance levels have been assessed as equivalent to an FHRS rating of 5 at their most recent food hygiene inspection. Fully compliant premises are a subset of the broadly compliant category i.e. all fully compliant premises are also broadly compliant.
- Broadly Compliant (BC) Premises whose compliance levels have been assessed as equivalent to an FHRS rating of 3, 4 or 5 at their most recent food hygiene inspection.
- Poorly Compliant (PC) Premises whose compliance levels have been assessed as equivalent to an FHRS rating of either 0 or 1 at their most recent food hygiene inspection. Poorly compliant premises are a subset of the not broadly compliant category i.e. all poorly compliant premises are also not broadly compliant.

For LAEMs data where there were more samples, statistical tests where undertaken on samples of different product types. These were

- Materials in contact (e.g. cutting board)
- Protein foods (e.g. meat, game, poultry, dairy, fish eggs)
- Complex foods (e.g. ices & deserts, prepared dishes, broths & soups)

## 4.3 Results

## 4.3.1 The relationship between FHRS rating and proportion of unsatisfactory samples at a premises (LAEMS)

Figure 1 shows the relationship between the proportions of unsatisfactory samples at a premises and food business compliance, as measured by FHRS rating.



Figure 1: The relationship between FHRS rating and proportion of unsatisfactory samples at a premises (LAEMS)

Broadly compliant premises are likely to have a smaller proportion of unsatisfactory samples than not broadly compliant and poorly compliant premises. These results were statistically significant for all years other than for 2009 where there was no statistically significant difference. Fully compliant premises are likely to have a smaller proportion of unsatisfactory samples than all other premises. These results were statistically significant for all years. There was a rise in the average percentage of unsatisfactory samples in broadly compliant and fully compliant premises from 2011-12 to 2013-14. However, 2009-10 had similar levels to 2013-14 so it is not necessarily a trend. There was also a rise in the average percentage of unsatisfactory samples in poorly compliant premises from 2009-10 to 2013-14, with a dip in 2011-12.

## 4.3.2 The relationship between FHRS rating and proportion of unsatisfactory samples at a premises by product group (LAEMS)

Figure 2 shows the relationship between the proportions of unsatisfactory samples at a premises by product group from which the sample is taken, and food business compliance, as measured by FHRS rating.

Figure 2: The relationship between FHRS rating and proportion of unsatisfactory samples at a premises by product group (LAEMS)



For material in contact (e.g. cutting boards), the proportion of unsatisfactory samples was nearly 12 ppts greater in not broadly compliant premises compared to fully compliant premises. Across all compliance levels, the proportion of unsatisfactory samples for material in contact was at least 17 ppts greater than for protein and complex foods samples.

For protein foods (e.g. meat, poultry, dairy), not broadly compliant premises had a larger proportion of unsatisfactory samples than all other premises. The proportion was 7.5 ppts greater in not broadly compliant premises compared to fully compliant premises.

For complex foods (e.g. soups, desserts), fully compliant premises had a smaller proportion of unsatisfactory samples than all other premises.

Although the different product groups have different proportions of unsatisfactory samples due to different levels of risk associated with the product groups, the relationship with FHRS rating is in the same direction i.e. the proportion of unsatisfactory samples is lower for more compliant premises for all three product groups.

## 4.3.3 The relationship between FHRS rating and proportion of unsatisfactory samples at a premises by business type (LAEMS)

Figure 3 shows the relationship between the proportions of unsatisfactory samples at a premises by business type and FHRS rating.

The proportion of unsatisfactory samples in a premises is generally smaller for retailers than for restaurants and caterers, but the difference is only statistically significant for fully complaint and broadly compliant premises.

Figure 3: The relationship between FHRS rating and proportion of unsatisfactory samples at a premises by business type (LAEMS) in 2009-10 to 2013-14



The difference in unsatisfactory samples between not broadly and broadly compliant premises is statistically significant for restaurants and caterers but not for retailers. However, the difference in unsatisfactory samples between not broadly compliant and fully compliant premises is statistically significant for both business types.

## 4.3.4 The relationship between component scores that derive FHRS and proportion of unsatisfactory samples at a premises (LAEMS)

Figures 4-5 show the relationship between the individual component scores that determine the overall FHRS rating, and the proportion of unsatisfactory microbiological contamination sample outcomes. It is important to remember here that the higher the individual component score, the less compliant the premises is.

Figure 4: The relationship between level of compliance: hygiene score and proportion of unsatisfactory samples at a premises (LAEMS)



Figure 4 shows there is a statistically significant increase in the proportion of unsatisfactory samples for each increase in a 'Level of compliance: hygiene' premises' score, apart from between scores 15 and 20.

Figure 5: The relationship between 'Level of compliance: structural' score and proportion of unsatisfactory microbiological samples at a premises (LAEMS)



Figure 5 shows the proportion of unsatisfactory microbiological samples is smaller for scores 0 and 5 than all other scores (except 25). This result is statistically significant. Figure 6 shows there is a statistically significant increase in the proportion of unsatisfactory microbiological samples for each increase in a premises' 'Confidence in management' score.



Figure 6: The relationship between 'Confidence in management' score and proportion of unsatisfactory microbiological samples at a premises (LAEMS)

#### 4.3.5 Food business compliance and microbiological contamination samples outcomes (UKFSS)

Figure 7 shows the relationship between the proportions of unsatisfactory samples at a premises and food business compliance, as measured by FHRS rating. Borderline samples are classed as satisfactory here.

Figure 7: The relationship between FHRS rating and proportion of unsatisfactory samples at a premises (UKFSS) - when borderline samples are classed as satisfactory



Mean proportion of unsatisfactory samples in a premises by FHRS rating

FHRS rating 0 has been excluded due to very small sample size. Apart from FHRS rating 1, the proportion of unsatisfactory samples falls as FHRS rating increases. However, the differences in unsatisfactory rates between individual FHRS ratings are not statistically significant. The distribution of samples taken at premises of different FHRS ratings over time was examined (see Section 4.4) to explore if there was any targeted sampling of premises rated 1, but this was not evident in the data.

If borderline samples are included in the unsatisfactory category, then unsatisfactory rates rise to 30-40% for all FHRS ratings, but the pattern remains the same as in Figure 6. In the remainder of the results presented (Figures 7-11), borderline samples were classed as satisfactory.

The differences in unsatisfactory rates between individual FHRS ratings are not statistically significant. However, despite the differences in data recording, the direction of results in UKFSS is the same as LAEMS i.e. for high FHRS ratings, the proportion of unsatisfactory samples at a premises is smaller than for low FHRS ratings.

Figure 8: The relationship between full compliance and proportion of unsatisfactory samples at a premises (UKFSS) - when borderline samples are classed as satisfactory



As shown in Figure 8, the average proportion of unsatisfactory samples is lower in premises rated 5 (6.4%) than in premises rated 0-4 (8.2%). This difference is statistically significant.

However, other results (e.g. broadly versus not broadly compliant as shown in Figure 9) are in the same direction but not statistically significant.

Figure 9: The relationship between broad compliance and proportion of unsatisfactory samples at a premises (UKFSS) - when borderline samples are classed as satisfactory



## 4.3.6 Food business compliance and microbiological contamination samples outcomes (UKFSS) with TVC30 removed

As stated in Section 4.2.2, some of the tests are better measures of preparation and food spoilage than of contamination with pathogens. To test the impact TVC30 may have had on the analysis, the results were also analysed with TVC30 results removed when determining overall satisfactory outcome of the sample. In this case, neither fully compliant nor broadly compliant showed a statistically significant difference from not fully compliant and not broadly compliant premises, respectively.

## 4.3.7 The relationship between component scores that derive FHRS and proportion of unsatisfactory samples at a premises (UKFSS)

Figures 10-12 show the relationship between the individual component scores that determine the overall FHRS rating, and the proportion of unsatisfactory microbiological contamination sample outcomes. Unlike FHRS rating, here, the lower the individual component score, the more compliant the premises is.

Figure 10: The relationship between 'Level of compliance: hygiene' score and proportion of unsatisfactory samples at a premises (UKFSS) - when borderline samples are classed as satisfactory



Level of compliance: hygiene

As shown in Figure 10, the difference in proportion of unsatisfactory samples between 'Level of compliance: hygiene' scores is not statistically significant, but is in the same direction as previous results i.e. the proportion of unsatisfactory samples decreases with an increase in compliance.

Figure 11: The relationship between 'Level of compliance: structural' score and proportion of unsatisfactory samples at a premises (UKFSS) - when borderline samples are classed as satisfactory



Level of compliance: structural

For both 'Level of compliance: structural' and 'Confidence in management, as shown in Figures 11 and 12 respectively, the difference in proportion of unsatisfactory samples between scores is not statistically significant, and there is no clear pattern.

Figure 12: The relationship between 'Confidence in Management' score and proportion of unsatisfactory samples at a premises (UKFSS) - when borderline samples are classed as satisfactory



**Confidence in management** 

### 4.4 Conclusions and Caveats

For microbiological contamination samples recorded in LAEMS, broadly compliant premises are likely to have a smaller proportion of unsatisfactory samples than not broadly compliant and poorly compliant premises. Fully compliant premises are likely to have a smaller proportion of unsatisfactory samples than all other premises. Across all compliance levels, the proportion of unsatisfactory samples for material in contact was over 17ppts greater than for protein and complex foods samples. For each increase in a premises's level of compliance: hygiene score, the proportion of unsatisfactory samples rises, apart from between scores 15 and 20. There is no consistent change in the proportion of samples taken at premises of different FHRS ratings over time, as shown in Figure 13, so the other results found are not affected by any changes in sampling strategy.



Figure 13: Distribution of samples taken at premises of different FHRS ratings over time (LAEMS)

For microbiological contamination samples recorded in UKFSS, the average proportion of unsatisfactory samples is lower in premises rated 5 (6.4%) than in premises rated 0-4 (8.2%). This difference is statistically significant. Other results (e.g. broadly and poorly compliant) are in a similar direction but not statistically significant. Unsatisfactory samples rates at premises are 30-40% when borderline samples are classed as unsatisfactory. If borderline is classed as satisfactory, unsatisfactory rates fall to under 10%.

Despite the differences in data recording, the direction of results in UKFSS is the same as LAEMS i.e. for high FHRS ratings, the proportion of unsatisfactory samples at a premises is smaller than for low FHRS ratings.

Local authorities differ in how they classify borderline samples (as satisfactory or unsatisfactory) in LAEMS, whereas UKFSS distinguishes between borderline and unsatisfactory. This explains why unsatisfactory rates may differ between LAEMS and UKFSS. Additionally, the trend over 5 separate years could be examined in LAEMS whereas all data for 2.5 years was combined in UKFSS.

# 5 The relationship between food business compliance and identified foodborne disease outbreaks

## 5.1 Introduction

An outbreak of foodborne illness is defined as two or more linked cases of the same disease. Outbreaks are a measure of the actual relationship with human health. However, the data on outbreaks is limited (around 60 per year<sup>6</sup>) and only a proportion of these can be matched to restaurants. The aim was to see if there is a relationship between a food business's FHRS rating and the chance of an outbreak.

## 5.2 Methodology

The outbreaks data was obtained from Public Health England (PHE), and covered the time period 2010 to 2014. There were 298 outbreaks over these 5 years, and 261 or these occurred at "Restaurants" (i.e. "Caring Premises", "Hotel/Guest House", "Pub/Club", "Restaurant/Cafe/Canteen", "Restaurants and Caterers - Other", "School/College", or "Take-Away").

The outbreaks data was matched to FHRS historical data using fuzzy premises name and address matching.<sup>7</sup> The most recent FHRS rating prior to the outbreak was taken as a measure of compliance at time of outbreak. This was to ensure the FHRS rating was not influenced by outbreak occurring. Some restaurants could not be matched by name and address, while others could be matched but did not have an FHRS rating prior to the outbreak. Table 5 shows the numbers and proportions of outbreaks at restaurants that were successfully matched, and those that weren't. The increase in the proportion of matched premises over time reflects the increased take up of FHRS by Local Authorities.

Year	Unmatched	Matched (numbers shown by FHRS Rating)					Matched	
		0	1	2	3	4	5	Proportion
2010	35	0	1	1	1	4	7	0.29
2011	38	0	3	3	2	5	20	0.46
2012	10	0	1	2	9	10	16	0.79
2013	6	0	2	0	1	8	10	0.78
2014	3	1	3	4	5	9	22	0.94

Table 5: No of matched and unmatched outbreaks at restaurants

Within the period of study (2010 to 2014), four premises saw repeat outbreaks.

The rate of outbreaks per 10,000 premises was used to compare the likelihood of an outbreak between broadly compliant and non-broadly compliant premises. A confidence interval for the difference in rates was estimated using bootstrapping<sup>8</sup>. This approach can estimate the properties of a statistic through repeated sampling from the observed data. It does not rely on an assumption of normal distribution.

<sup>&</sup>lt;sup>6</sup> Most cases of foodborne disease cannot be linked to an outbreak and are classed as sporadic

<sup>&</sup>lt;sup>7</sup> See Annex 1 for methodology.

<sup>&</sup>lt;sup>8</sup> Bootstrapping is a method for estimating the sampling distribution of a statistic by resampling with replacement from the original sample. Effectively, the approach uses the observed sample to estimate of the distribution of the true population. Repeated samples are drawn from this distribution and the statistic of interest is estimated from each one. This simulates the sampling distribution for the statistic, and values from its low and high tails provide the limits for its confidence interval.

## 5.3 Results

Figure 14 shows that although it seems as if businesses rated 3, 4 or 5 have fewer outbreaks per restaurant than those businesses rated lower, this result is not statistically significant at the 95% confidence level, and therefore we cannot draw any conclusions about the relationship between the chance of an outbreak and individual FHRS score.

Figure 14: The relationship between a food business's FHRS rating and the chance of an outbreak



When comparing businesses that are broadly compliant (FHRS rating 3, 4, 5) against those that are not broadly complaint (FHRS rating 0, 1, 2), there is a statistically significant relationship at the 95% confidence level. Broadly compliant premises have fewer outbreaks per 10,000 restaurants than lower rated premises, as shown in Figure 15.



Figure 15: The relationship between a food business's FHRS compliance category and the chance of an outbreak

Although Figure 15 suggests broadly compliant have almost half the chance of a foodborne disease outbreak than those that are not broadly compliant, it should be stressed that the actual level of risk to an individual consumer is still very low indeed.

### 5.4 Conclusions and Caveats

The analysis shows that broadly compliant premises have a smaller chance of a foodborne disease outbreak than those that are not broadly compliant. It is important to note some of the limitations of the analysis, however. There is not enough data to perform the analysis on a single year, so the five years of data is treated as one. Nor is there enough data to split out the different pathogens causing the outbreaks. Also, no account was taken of the time between inspection and outbreak, whereas our confidence in the accuracy of the rating reduces with time since inspection.

As the outbreaks data available at the time of analysis was limited in size, this analysis should be treated with caution and could be repeated in future, when more years' of data is available, and therefore the number of outbreaks that can be matched to FHRS ratings increases.

## 6 Estimate of impact of the FHRS scheme on foodborne disease

### 6.1 Introduction

The food safety proxy measure outcomes for different compliance categories can be used to estimate the impact of foodborne disease when compliance levels change. This assumes changes in foodborne disease are similar to those seen in the proxy measures.

## 6.2 Methodology

The compliance levels used in the model are based on:

- The change from pre-FHRS levels (2009-10) to the most current data available at time of analysis (2014-15). This is a proxy of the impact of FHRS on compliance levels
- A small incremental change (+1% for broad compliance or +1% for full compliance).

Only those measures that showed a statistically significant impact in the previous analysis have been included in the estimations.

The proportion of FBD attributed to foodservice settings is notoriously challenging to estimate due to difficulty of tracing illness to source. In the analysis we used results from a recent report published by the FSA titled "Systematic review of the relative proportion of foodborne disease associated with food preparation or handling practices within the home"<sup>9</sup>. This found that estimates of the proportion of foodborne illness outbreaks attributed to foodservice settings in the UK is between 44% to 85%. The report heavily caveated the results and stressed that sporadic cases were more likely to occur in the home. With this in mind we used the more conservative scenario of 44%.

## 6.3 Results

The FSA estimates there are 1 million foodborne disease cases per year. Table 6 shows how foodborne disease cases change as proportion of broadly compliant businesses changes.

Proxy measures of foodborne disease	Proportion of broadly compliant businesses	Change in proportion of broadly compliant businesses compared to 2014- 15	Change in number of cases assuming 44% of foodborne disease cases are from food businesses	% Change in number of foodborne disease cases compared to 2014-15
2014-15	94%			
2009-10	88%	-6%		
Unsatisfactory sample outcomes (LAEMS)			10,864	2.5%
Unsatisfactory sample outcomes (UKFSS)	Unable to use in	estimation as food p	roxy impact not statis	tically significant
Outbreaks			28,150	6.4%
+1% broad compliance from 2014-15	95%	1%		
Unsatisfactory sample outcomes (LAEMS)			- 1,752	-0.4%
Unsatisfactory sample outcomes (UKFSS)	Unable to use in	estimation as food p	roxy impact not statis	tically significant
Outbreaks			- 4,540	-1.0%

Table 6: Estimation of the change in foodborne disease cases change as proportion of broadly compliant businesses changes

<sup>&</sup>lt;sup>9</sup> <u>https://www.food.gov.uk/research/foodborne-diseases/systematic-review-of-the-relative-proportion-of-foodborne-disease-caused-by-food-preparation-or-handling-within-the-home</u>

Table 7 shows how foodborne disease cases change as proportion of fully compliant businesses changes.

Proxy measures of foodborne disease	Proportion of fully compliant businesses	Change in proportion of fully compliant businesses compared to 2014- 15	Change in number of cases assuming 44% of foodborne disease cases are from food businesses	% Change in number of foodborne disease cases compared to 2014-15
2014-15	60%			
2009-10	43%	-18%		
Unsatisfactory sample outcomes (LAEMS)			23,153	5.3%
Unsatisfactory sample outcomes (UKFSS)			19,391	4.4%
Outbreaks	Unable to use in	estimation as food p	roxy impact not statis	tically significant
+1% full compliance from 2014-15	61%	1%		
Unsatisfactory sample outcomes (LAEMS)			- 1,323	-0.3%
Unsatisfactory sample outcomes (UKFSS)			- 1,108	-0.3%
Outbreaks	Unable to use in	estimation as food p	roxy impact not statis	tically significant

Table 7: Estimation of the change in foodborne disease cases change as proportion of fully compliant businesses changes Conclusions and Caveats

Comparing to 2009-10 (before FHRS) to 2014-15, the model estimates that there are now approximately:

- 2.5% to 6.4% fewer cases of FBD from food businesses (based on the increase in proportion of broadly compliant premises)
- 4.4% to 5.3% fewer cases of FBD from food businesses (based on the increase in proportion of fully compliant premises)

If broad compliance increased/decreased by 1% from 2014-15 levels, it is estimated FBD cases from food businesses would fall/rise by 0.4% to 1%. If full compliance increased/decreased by 1% from 2014-15 levels, it is estimated FBD cases from food businesses would fall/rise by 0.3%.

It is also important to note that the analysis investigated only association. Consequently, action specifically designed to improve food business hygiene compliance may not be the most effective way of impacting on levels of foodborne illness.

## 7 Conclusions

High FHRS ratings do not eliminate risk of outbreak or unsatisfactory samples results. However, samples results from LAEMS and UKFSS indicate premises with higher FHRS ratings are less likely to have unsatisfactory results. Similarly outbreaks are less likely to occur at broadly compliant premises.

The evidence presented consists of observational studies. This approach is suitable for identifying an association between variables but cannot be used to determine causation. Therefore, the evidence presented may show that the food safety proxy measures change as FHRS ratings change, but cannot be used to conclude that the changes in ratings actually drive the changes in the food safety proxy measures. Causation can only be determined by intervention studies that deliberately increase or decrease the scope of the FHRS programme, keeping all other relevant variables constant and recording the impact on food safety proxy measures.

Although each analysis in isolation does not always yield statistically significant results, all results are in the same direction, and combine to provide an evidence base that indicates that higher food business compliance and food safety have a positive relationship.

## Annexe 1 – Address Matching

The address matcher links two datasets based on their address fields. It looks for *potential matches* in one dataset that match a *target address* in the other. For instance, when matching UKFSS to FHRS, for each UKFSS entry (*target address*) it will identify a list of potential matches in FHRS (*potential addresse*).

To do this the target address is split into *tokens*: the individual words that make up the address. It then performs a *full text search*<sup>10</sup> using the tokens to identify a list of potential matches. It begins by searching for all terms, then deletes the most specific and searches on the remainder, and so on. This is illustrated in Figure 1.



Figure 1: Illustrated example of full text searching using tokens.

For each potential match a score is computed that in some limited sense represents the inverse likelihood that this address and the target address match. Note that this is not a true probability, but the smaller the number the better the match.

Part of the scoring involves computing the Levenshtein distance between potential and target tokens. This is a string metric used for measuring the difference between two sequences. There are three edit operations, each increasing the score by one: deletion, substitution and insertion. This is illustrated in Figure 2.

<sup>&</sup>lt;sup>10</sup> Full text searching is a feature implemented in PostgreSQL. See the documentation for a full explanation: <u>http://www.postgresql.org/docs/9.1/static/textsearch-intro.html</u>

Figure 2: Illustrated example of calculating the Levenshtein distance (LD) between two tokens.



The process computes an overall score that represents the "distinctness" or "discriminativeness" of each token; that is, how much it helps in narrowing down the full list of all addresses. For example, a score of one means that the token doesn't narrow down the full list at all, while a score of 0.01 means that it cuts it down by 99%. Note that this is not a true probability because if the token cannot be found in the target address, the code attempts to find a fuzzy match. The probability here is ill-defined.

For each potential address, the scores are combined by multiplying the individual token scores. The potential matches are then sorted from the "best" to the "worst" matches: the lowest to the highest "probability".

As well as this "probability", a score is assigned to each potential match, referred to as *match stats* in the code. This is a score between 0 and 1, where 1 is excellent and 0 is very bad. Further to this a description is assigned to the match score based on the following scale:

- Match score ≥ 0.9: "Good match"
- 0.75 ≤ match score < 0.9: "Average match"
- 0.6 ≤ match score < 0.75: "Poor match"
- 0.3 ≤ match score < 0.6: "Very poor match"
- 0 ≤ match score < 0.3: "No match"

If no matches are found, it is possible to try to match the address using random combinations of the tokens.