

An Evaluation Design for Food Hygiene Rating Schemes

Policy Studies Institute

Social Science Research Unit

Food Standards Agency

March 2011

Unit Report 14





Policy Studies Institute

An Evaluation Design for Food Hygiene Rating Schemes

Fatima Husain and Stephen Morris

This report was produced by Policy Studies Institute on behalf of the Food Standards Agency.

© Crown Copyright 2011

This report has been produced by Policy Studies Institute under a contract placed by the Food Standards Agency (the Agency). The views expressed herein are not necessarily those of the Agency. Policy Studies Institute warrants that all reasonable skill and care has been used in preparing this report. Notwithstanding this warranty, Policy Studies Institute shall not be under any liability for loss of profit, business, revenues or any special indirect or consequential damage of any nature whatsoever or loss of anticipated saving or for any increased costs sustained by the client or his or her servants or agents arising in any way whether directly or indirectly as a result of reliance on this report or of any error or defect in this report.

Contents

Executive Summary	3
Introduction	7
National Food Hygiene Rating Scheme: Theories of Change	8
Developing theories of change	8
The overarching model of behaviour change	9
Target groups and theories of change	11
A Preliminary Evaluation Design	17
Purpose and nature of programme evaluation	17
An impact evaluation design	18
General assumptions and constraints	18
Research hypotheses and the counterfactual	19
Unit of analysis	21
A framework for estimating impacts	22
An approach to estimating the counterfactual	23
A Process Study Design	33
Developing research questions for a process study	33
Contextual considerations	35
The proposed approach	37
Research methods	38
Key challenges	39
Taking the process study forward	41
Concluding comments: Is it possible to meaningfully evaluate the national Food Hygiene Rating Scheme?	43
References	45

Executive Summary

This report examines potential approaches to evaluating the Food Standards Agency's (FSA) national Food Hygiene Rating Scheme that is due to be implemented across England, Wales and Northern Ireland and the Food Hygiene Information Scheme in Scotland.¹ It also considers how existing food hygiene rating schemes (often referred to generically as 'scores on the doors' schemes), the forerunners of the national scheme, can be evaluated. It commences by setting out the 'theories of change' or the underlying programme logic which underpin the national scheme. These theories were developed through exploration of project documentation and through holding workshops with policy makers and analysts from the Agency. The theories of change form a foundation upon which discussion of approaches to both impact and process evaluation are then developed.

An impact evaluation design

It would appear that developing a longitudinal panel data set of all local authorities with data stretching back in time to before the introduction of existing food hygiene rating schemes offers the prospect of a rigorous and convincing assessment of the schemes' impacts. If the required data are available, it may be possible to use a difference-in-differences (DiD) estimator² to identify the effects of existing food hygiene rating schemes on foodborne illness and hygiene standards in food businesses, extending the analysis to incorporate the effects of the national programme in due course. It is our view that DiD offers the best prospects of being able to estimate the effects of schemes in a rigorous and convincing manner.

We need to remain cautious, however, because it has not yet been possible to determine precisely whether the data necessary to implement a DiD approach are available. Our initial explorations do suggest that data of the type and quality required may well be available but we have not yet confirmed this. Crucially it is not clear at this stage whether the Agency can determine precisely when schemes were introduced across all local authorities. This is important because for the DiD estimator to be applied effectively schemes will need to have been introduced over a reasonably short time period. There is also work ongoing within the Agency to derive measures which can be used to capture important outcomes at the local authority level and the results of this

¹ Although the report focuses on the national Food Hygiene Rating Scheme and existing food hygiene rating schemes, the FSA will also be taking forward these recommendations for the evaluation of the Food Hygiene Information Scheme in Scotland.

² The DID estimator measures average impacts at 2 time periods, before and after the introduction of a policy, by comparing the impact on both the treatment and control groups (that is, those local authorities with and without the national scheme respectively).

work are critical in determining whether a convincing impact evaluation can be implemented. It is also worth pointing out that even if all the data required were available and were of sufficient quality, the schemes themselves may not produce impacts which can be measured – a rigorous evaluation design does not guarantee that schemes will be found to be effective. Moreover, if outcome measures are particularly susceptible to measurement error and other sources of variance this may frustrate our attempts to identify programme impacts. In other words, it may prove impossible to separate out the policy ‘signal’ from the background ‘noise’. This issue is of particular concern if schemes’ impacts are expected to be modest. Finally, there is also the possibility that existing schemes will have been implemented in a variety of different ways and further thought will be required as to how to account for this diversity within a quantitative impact evaluation design. It is suggested that the Agency divide the impact evaluation into three separate stages. The first would comprise a detailed examination of the types of data held by the Agency and available from public data sources in order to determine whether they are sufficient to deliver a convincing impact evaluation. This stage of the work would also seek to identify whether a longitudinal data set of the type required could be constructed and whether policy variables which identify when schemes were introduced are available for all authorities. This stage in the work would also determine whether data could be used in order to sample local authorities for a process evaluation.

The second stage of the work would only take place if it were felt that the data were of sufficient breadth and quality to permit a rigorous impact evaluation and/or enable a sampling frame to be constructed for the process study. The work would involve extracting and cleaning data, and creating a longitudinal panel data set of local authorities containing relevant outcome measures, policy and control variables.

The final stage of the work – the estimation of programme impacts - would only take place if it was felt at Stage 1 that a credible assessment of the schemes’ impacts was possible.

A process study design

The theories of change for the national scheme identifies three key target groups – local authorities, food businesses and consumers - that the initiative intends to influence through a variety of mechanisms. The process study design proposed in this report incorporates an investigation of programme processes as they affect the key target groups. The design also involves a consideration of the range of factors and characteristics that add to the complexity of the initiative’s implementation and delivery landscape. The anticipated staged roll-out of the scheme which comprises ‘early adopters’ followed by a gradual national roll-out phase involving other local authorities, has been incorporated into the process study design.

As the voluntary take-up of the scheme is at the local area level, we suggest a process study that utilises a case study approach with Local Authorities’ area boundaries as the study sites. The purposive selection of case study local

authorities is an important first step for the process study. Case study selection needs to be justified based on available data related to food hygiene standards and information on local socio-economic and other contextual characteristics.

A number of considerations can influence the timing of a process study. In the first instance commissioners would need to consider if the focus of the study is to be an exploration of early implementation of an initiative or an assessment of long-term delivery. The timing of a process study is determined also by the nature of an initiative and the length of time that is needed for the initiative to become embedded within its contextual surroundings. We suggest that for the national scheme the study should be timed so that fieldwork takes place six months after implementation. Usually, this is sufficient time for any early implementation issues to have been addressed and for the scheme to be functioning in a consistent way.

We recommend that a process study of the national scheme is taken forward in the following two stages:

Stage 1:- A pilot study of 4-5 'early adopter'³ case study sites. The fieldwork would take place six months after implementation. The emerging findings would elaborate on the strengths and weaknesses of the initiative and highlight best practice. The findings and lessons learned during this pilot phase could be used to influence the national roll-out of the scheme.

Stage 2:- This study of national implementation would take place six months after the national roll-out of the scheme.⁴ Replicating the methodology used at Stage 1, the study would be focused on 7 case study local authorities, 3 of which would be early adopters. Incorporating a longitudinal element by studying a small number of early adopters at this stage would provide depth of detail on the longer term delivery of the scheme.

The evaluation timetable

The timetable that we have set out for an evaluation of the national scheme, presumes that the distinct stages of both the impact and process studies can be conducted in a sequential manner. Both studies can take place

³ Early entrants who will adopt the national scheme before the end of June 2011. This could include either/both local authorities migrating to the national scheme from their own existing scheme, and local authorities who have not previously operated a scheme.

⁴ The FSA officially launched the scheme on 30 November 2010 and the roll-out of the scheme is a gradual, fluid process. It is now envisaged that 'Stage 2' of the process evaluation will take place about 18 months – 2 years after the official launch.

concurrently and an ideal scenario would be to complete Stages 1 and 2 of the impact study prior to commencing a process study. This is because the construction of a panel dataset during Stage 2 of the impact study would greatly facilitate the case study selection process.

Whilst the estimation of programme impacts at Stage 3 of the proposed impact study can only take place after the creation of a dataset at Stage 2, there are fewer barriers to conducting a process study. In the absence of the panel dataset, a process study can still be carried out using available contextual data to select case study sites.

In our opinion, the Agency should consider commissioning a process study regardless of whether an impact study appears achievable.

Introduction

The Policy Studies Institute was commissioned by the Food Standards Agency (FSA) to advise the Agency on how it might evaluate a key policy intervention, food hygiene rating schemes. Food hygiene rating schemes are designed to provide consumers with information about the hygiene standards found in food business premises so that consumers are able to make informed choices. The national Food Hygiene Rating Scheme is due to be implemented across England, Wales and Northern Ireland, with the Food Hygiene Information Scheme rolling out in Scotland.

The key aims of this study were to:

- understand the background and nature of the intervention
- determine the main hypotheses to be tested through the evaluation – the key evaluation questions
- identify the most appropriate evaluation approaches to measure the impact and effectiveness of the initiative
- identify constraints, including timescales and resources.

Our Approach

To meet the Agency's aims and objectives a 'programme theory' approach was used to develop a detailed understanding of the initiative. The programme theory approach entailed an initial desk-based review of programme documentation. This was followed by a consultation workshop with key policy and research staff at the FSA who are involved in the design, piloting and roll-out of the national Food Hygiene Rating Scheme.⁵ The document review and analysis of the workshop discussion were instrumental in providing:

- a framework and structure for the study of the programme
- an explicit outline of the intervention and the programme design which will generate positive outcomes
- important insights as to which evaluation strategies would meet the Agency's needs.

This programme theory approach yielded a series of testable research hypotheses or questions. A range of evaluation methods that could potentially answer these research questions were considered and the approaches that are proposed in this report were driven by the nature of these research questions.

⁵ However, it should be noted that the FSA also considers the resulting theory of change to be applicable to the Scottish Food Hygiene Information Scheme.

Structure of the report

This report is divided into three sections. The first section sets out the programme theories of change for the national scheme and discusses the take-up and delivery context of the initiative. The second section proposes an evaluation design that incorporates both impact and process studies. The report concludes with a short discussion on the extent to which it is possible to meaningfully evaluate this initiative.

National Food Hygiene Rating Scheme: Theories of Change

Developing theories of change

A theory of change approach holds that programme interventions, in almost all cases, comprise an implicit theory or set of assumptions about how the interventions will bring about change (Pawson, et al., 2004). At its core, this approach is an explanation of how a group of stakeholders expects to reach a commonly understood goal (Anderson, 2007). Therefore facilitating a discussion with key FSA stakeholders to articulate a programme's theories of change was helpful in describing the underlying causal mechanisms and understanding how these are activated to produce desired short-term and longer-term outcomes.

The aim of the stakeholder discussion, which took place in November 2009, was to understand policymakers' intentions and reasoning, and to draw-out the 'theory' underpinning the intervention. FSA policy and research analysts and staff from the communications team participated in a structured discussion which focused on developing a detailed understanding of the following:

- The nature of the issue and the behaviour change the initiative seeks to bring about
- The target groups that are the focus of the intervention
- The outcomes (short- and long-term) that the initiative aims to influence
- The steps and processes within the initiative that will bring about the desired change
- The identification of data sources and measures.

This collaborative process allowed for the articulation of an intervention logic that separated the initiative into its key components and then systematically brought them together to produce a causal chain of links and relationships between activities, outcomes and the overarching policy goals of the intervention.

Analysis of the information gathered during this structured discussion led to the identification of an overall behavioural change model for the initiative as

well as target-group-focused theories of change embedded within the overarching model. Furthermore, the discussion resulted in the identification of a number of assumptions that underpin the intervention leading to a deeper understanding of the social and structural systems (the context) within which the national Food Hygiene Rating Scheme is to be implemented.

Embedding our development of an evaluation framework for the national Food Hygiene Rating Scheme within a programme theory approach is helpful in formulating the hypotheses or research questions for a national evaluation and in framing a discussion on indicators, data sources and how programme impact can be measured.

Diagrams that depict the different steps, causal links and relationships between programme activities leading to the ultimate programme goal, are the most practical way to present the theories of change for the initiative. In the next section we present first the overarching theory of behaviour change for the initiative. This is followed by theories of change that were articulated for the key target groups the initiative aims to influence. It is to be noted that these diagrams were developed based on the discussions that took place during the stakeholder workshop in November 2009.

The overarching model of behaviour change

The overarching policy goal of the national Food Hygiene Rating Scheme is to reduce the rate of foodborne illnesses by providing consumers with information on food hygiene inspections in accessible formats.

The theories of change that emerged from the analysis of the workshop discussion relate to a desired behaviour change principally for the main subject of the intervention – the consumer. Apart from consumers, the scheme is designed to influence behaviour change outcomes in two other groups: food businesses and local authorities.

At the highest level, the national scheme aims to publicise the results of statutory hygiene inspections of food businesses in a particular way. The key outputs of the scheme are a website and a sticker / certificate both showing ratings intended to make consumers aware of the hygiene rating of a particular food business. The stickers / certificates are given to food businesses who will be encouraged to display their food hygiene ratings but are not legally obliged to do so.

The overarching behaviour model is reliant on the consumer interacting with the public outputs of the inspection regime. It is expected that a consumer will look at the food hygiene ratings of food businesses, understand them and then alter their behaviour to make a decision that prioritises food hygiene. The pressure on food businesses to improve food hygiene will come primarily through customer behaviour change and also through competitive pressure from other food businesses.

From an evaluation viewpoint, a number of anticipated outcomes were specified during the consultation. Some of these relate to greater consistency in the inspections process, the scoring process and, in the longer term, aspirations in relation to local resource allocation, all requiring significant buy-in from local authorities. Another aspect of this model is focused on a behaviour change outcome for food businesses. It is expected that once hygiene inspection results are publicised, food businesses will focus on increasing their hygiene rating, resulting in an overall improvement in food hygiene standards. Other outcomes are more specific to consumers as the ultimate target group and involve a range of activities and interim outputs and outcomes.

The initiative has been designed in such a way that all desired behaviour change outcomes lead to the ultimate goal of the scheme which is to reduce the incidence of foodborne illnesses. The high level model for change for the national initiative is represented in the following flow chart:

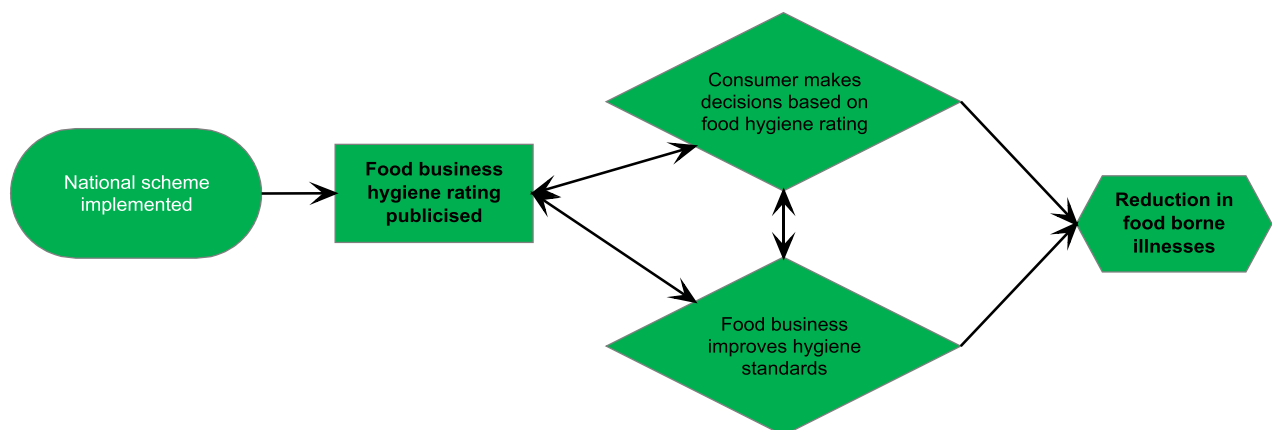


Diagram 1. The overarching Theory of Change

A number of key assumptions are made at this level of behaviour change. The most fundamental assumption relates to the adoption of the national Food Hygiene Rating Scheme which is dependent on voluntary take-up by local authorities. Local authority 'buy-in' is a critical first step requiring the commitment of sustained resources by the local authority to deliver and administer the scheme. Other assumptions at this level of desired change are that:

- consumers will be aware of the national Food Hygiene Rating Scheme
- consumers will be able to access information about hygiene standards in food businesses
- consumers' understand the rating system and this understanding will trigger a response
- consumers will respond by altering their behaviour and will make choices that prioritise food hygiene

- changes in consumer choices will impact on business sales and profits, increasing competitive pressures and triggering possible changes in attitudes to food hygiene.

There is evidence to suggest that informing consumers of the hygiene standards of food businesses does lead to behaviour change among both consumers and businesses which in turn results in the reduction of foodborne illnesses (Thompson et al., 2005; Worsfold, 2005; Zhe Jin and Leslie, 2003).

Target groups and theories of change

To fully understand the process of change that is expected from this initiative we have 'unpacked' this top level theory of change into its component parts as they relate to the key target groups or 'actors' that the FSA aims to influence. These are:

- Local authorities
- Food businesses
- Consumers.

In the following sections, we identify the theory of change for each of these groups by exploring in detail the pathways of change that would be necessary in order to achieve the desired long-term outcome of reducing the incidence of foodborne illness.

In presenting the theories of change as they relate to each 'actor' we take an activities approach that connects the actors through a range of activities and outputs to the desired result. An activities approach makes it easier to identify the stages or key intervention points in the process of change as well as indicate where monitoring systems may need to be set up to assess progress towards change.

Local Authorities

The implementation of the national Food Hygiene Rating Scheme is dependent on the voluntary take-up by local authorities. National legislation obliges local authorities to conduct hygiene inspections of food outlets, but public access to the results of these inspections is at the discretion of local authorities, although they are required to release information if a request is made under the Freedom of Information Act (2000).

The key inputs that the FSA intends to make available to local authorities that adopt the national Food Hygiene Rating Scheme are guidance documents and training programmes for both Food Hygiene Managers and Inspectors⁶. It

⁶ Subsequent to the stakeholder discussion the Agency identified additional inputs including grant funding, free IT platform, national promotional activities and support for local promotional activities, template materials for

is expected that successful participation in the training programmes⁷ will lead to common understanding of the national scheme and related rating system, enabling food inspectors to conduct consistent inspections across all food businesses in their local authority. The long-term intent is that all local authorities will adopt the national scheme resulting in a consistent understanding of the rating system. The local authority theory of change can be expressed diagrammatically as follows:

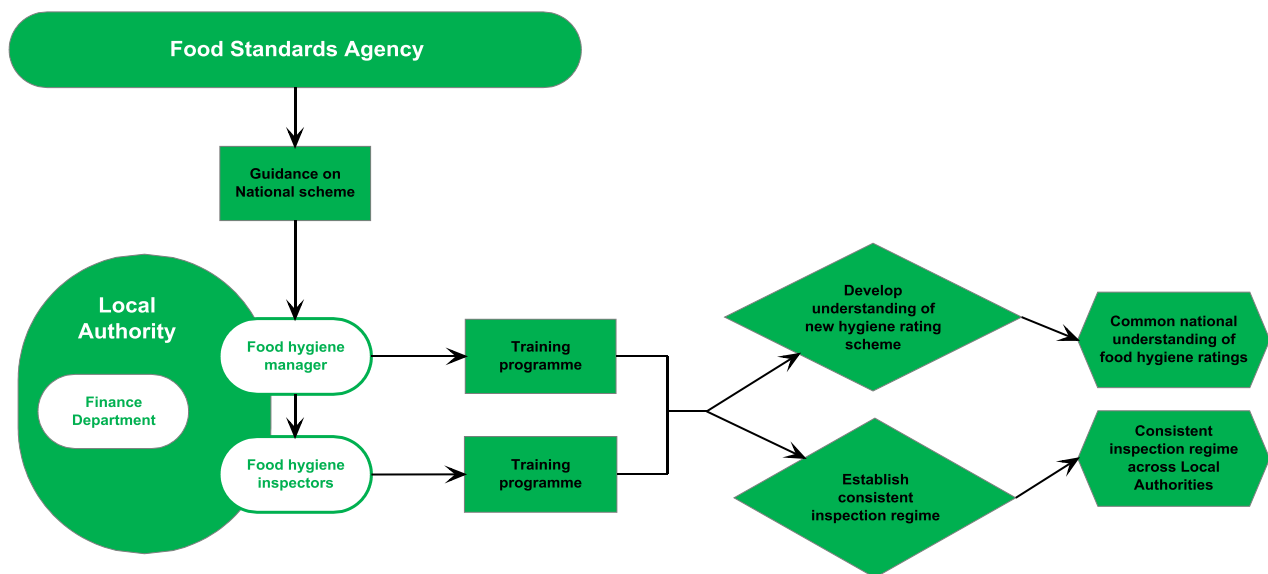


Diagram 2. The Theory of Change - Local Authorities

The theory of change we have identified for local authorities assumes that:

- local authorities which adopt the scheme will be able to allocate sufficient resources to properly publicise and implement the programme
- resources will be allocated for marketing and communications activities to raise awareness of the rating scheme as well as of food hygiene⁸

communicating to businesses and council members, Early Adopters forum and IT advisory forum

⁷ Subsequent to the stakeholder workshop separate training programmes for managers and inspectors will not be provided. Instead all local authorities in England, Wales and Northern Ireland will receive consistency training on rating using Annex 5 of the Food Law Code of Practice (regardless of whether they run a food hygiene rating scheme or not).

⁸ Subsequent to the stakeholder workshop, the FSA indicated its intention to provide support for promotion of the national Food Hygiene Rating Scheme including promotional materials, campaign guidance and grant funding.

- the guidance and training programme will lead to the requisite level of understanding needed to conduct consistent inspections
- that over time all local authorities in England, Wales and Northern Ireland will adopt the scheme.

The national 'rating scheme context' is varied as a range of other rating schemes are already being used in many local authorities. In Scotland a rating scheme based on a two-tier approach (Pass/Improvement required) is being implemented on a voluntary basis. As the schemes are adopted on a voluntary basis, the rating scheme/no rating scheme landscape may influence the desired outcomes.

There is some concern that local authority priorities may differ from those of the FSA and the new scheme may be viewed as another enforcement tool that requires additional resources. Irrespective of whether local authorities already have a rating scheme, there may be reluctance at the local level to allocate resources towards a non-mandatory scheme. This would conflict with FSA ambitions for consistent rating and a common national understanding of the rating system.

Food Businesses

The theory of change for food businesses centres on the effective implementation of the national scheme as it relates to inspection and the rating process at the local authority level. Based on the national Food Hygiene Rating Scheme and subsequent to inspection, each food business will be allocated a rating based on a six-tier system. Apart from the ratings being reported back to the local authority (as per existing reporting requirements), the rating would be displayed on a national website and a sticker / certificate displaying the rating would be issued to the food business. In Scotland, the Food Hygiene Information Scheme will involve stickers / certificates issued to the food business and ratings on the national website based on a two tier rating system.

It is at this level of programme activity that consumer engagement with the food inspection process takes place. The consumer, having accessed or seen the rating of a food business makes decisions based on food hygiene alongside other factors (that is, whether to purchase food from a food business with a low rating or choose a higher rated alternative food business).

The consumer's 'food hygiene-influenced' decision making leads to the improvement of hygiene standards as food businesses aim to maintain or increase their consumer base. Over the longer term, it is hoped that driving up hygiene standards in this way will contribute to the overarching goal of reducing the number of cases of foodborne illnesses.

Another process that could drive up hygiene standards is dependent on the extent to which food businesses think they need to compete (in non-price, food safety/hygiene terms) with other food businesses in their area. The extent to which a food business may compare their ratings with those of other

businesses is dependent on the existence of similar types of food businesses and their proximity to each other. The degree to which the national scheme will improve hygiene ratings will depend on the extent to which food businesses face competition, whether their competitors rate highly, and whether they decide to compete on the basis of food safety and hygiene as well as price. The chart below summarises the inspection process and its links to the overarching policy goal of the initiative.

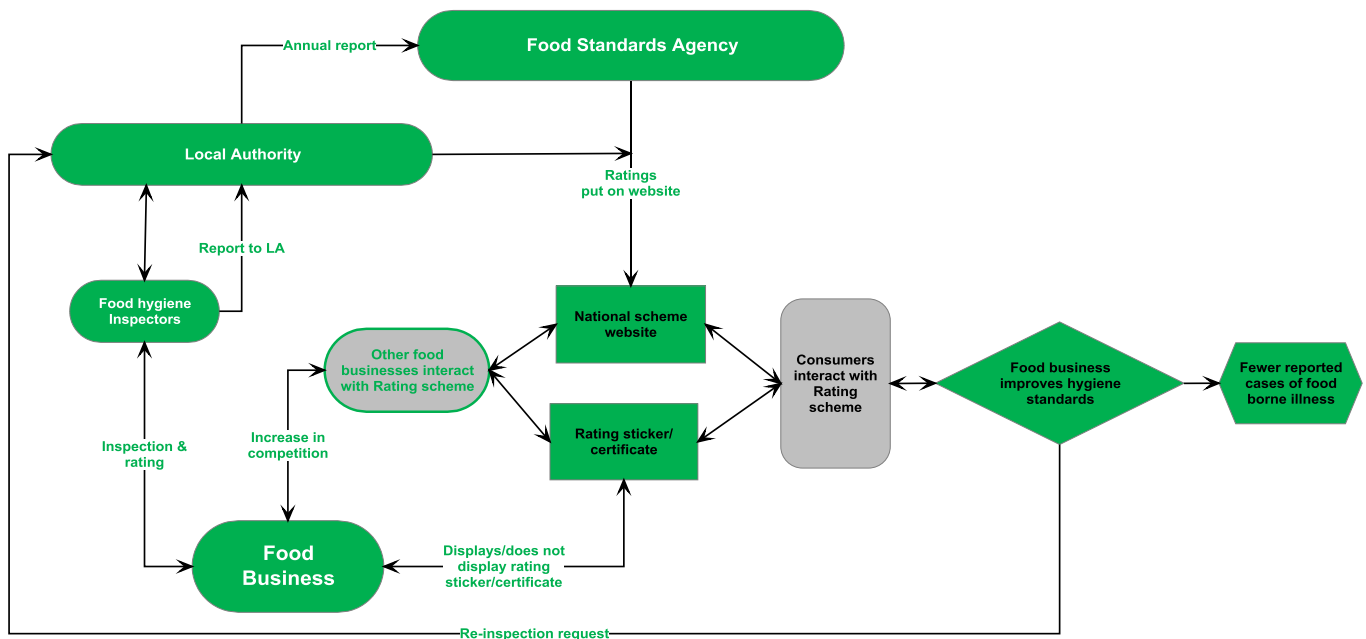


Diagram 3. The Theory of Change - Food Businesses

A key assumption is that food businesses will engage actively with the national rating system, that is, food businesses will voluntarily display their rating where it is easily visible to consumers and choose to attend to the food safety and hygiene of their products. Another underlying supposition is that where the rating is not displayed, consumers will assume that the business may be unwilling to display a low rating and will make an appropriate food hygiene related decision. A number of other potential consequences of the food hygiene rating scheme were highlighted during the workshop. These included:

- Food businesses with low ratings going out of business
- Food businesses with low ratings appealing
- Food businesses requesting re-inspections after improving standards
- Food businesses competing on food safety/hygiene and price (or on price alone).

The human interaction or the relationship between inspectors and food business owners/managers can also influence the outcome of inspections. It may be that inspectors may alter how they rate businesses because they know that food business hygiene ratings will be available to the public.

The key target group within the pathway to change outlined above is the consumer whose activities in the grey box in the diagram above are critical to the success of the scheme. The activities and anticipated outcomes for consumers are addressed in the next section.

Consumers

Consumers are the key target group for the national Food Hygiene Rating Scheme. The theory of change for consumers centres on food hygiene rating information being made public in a variety of ways. The FSA intends to make food business hygiene ratings available via a national website and through the voluntary display of rating stickers / certificates by businesses. The publicly available rating system is to be backed by communications and marketing activities conducted by the FSA that raises awareness and understanding of food hygiene. The theory of change for consumers is depicted in the diagram below:

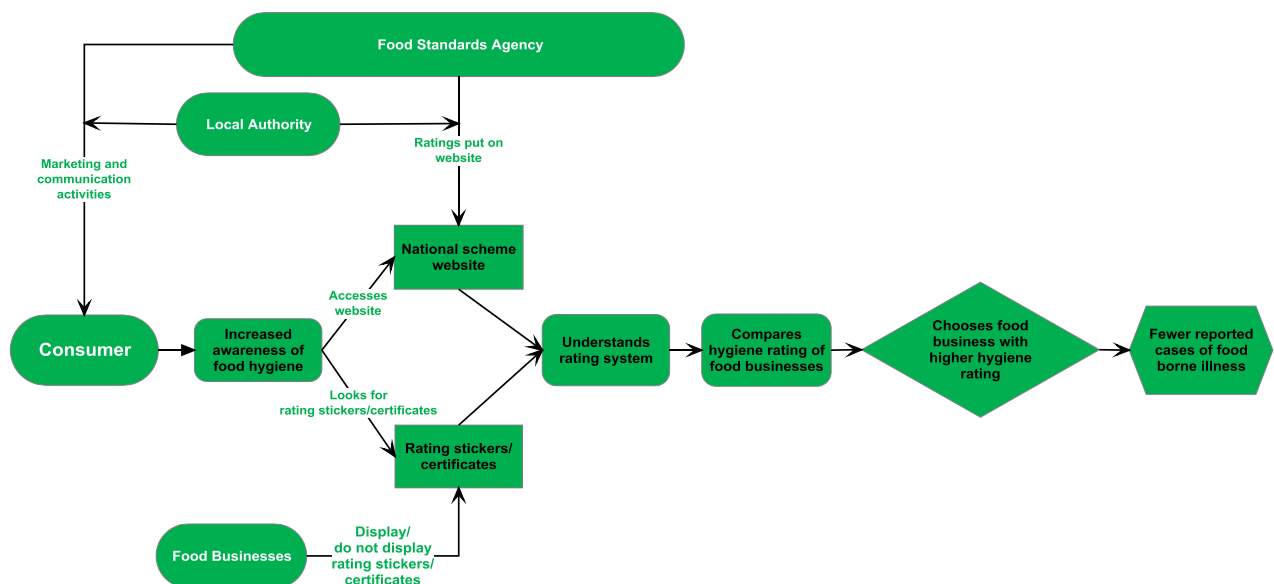


Diagram 4. The Theory of Change - Consumers

As change in consumer behaviour is crucial to the overall success of the scheme, a number of assumptions that underpin the consumer theory of change have already been mentioned in the context of the other theories of change (above). Factors such as consumer diversity and neighbourhood characteristics will further influence the effectiveness of the scheme. This is discussed in more detail in the section on contextual variations.

The design of the scheme assumes that when consumers are made aware of food hygiene issues and the national Food Hygiene Rating Scheme rating system they will seek out information on food hygiene. It is expected that this will result in customers prioritising food hygiene alongside other considerations. Another underlying supposition is that there will be similar food business options for consumers to choose from. The notion of similarity

is highly subjective as decisions about which food businesses are perceived as being similar can depend on consumer perceptions of food types, the variations in prices as well as how far consumers are willing to travel.

Another underlying assumption is that all consumers will engage with the scheme in a similar way irrespective of differences in incomes levels or of differences in how food businesses price their products (for example, if food businesses with higher ratings raise prices, some consumers may no longer be able to afford using these businesses and may choose a business with a lower rating and lower prices).

Lastly, access to the internet is assumed for consumers who want to check hygiene ratings. When consumers do not have access to the national Food Hygiene Rating Scheme website, that is they do not have internet access or do not know how to use a computer or the internet, will they look for/ask for certificate displays in food outlets and then 'walk the extra mile' to find a food outlet that has a higher rating on display?

The models of behaviour change presented in the sections above show how the national initiative was designed to work. Each target group focused theory of change presents the causal link between activities and desired change in a linear fashion. While the separate theories of change are useful in understanding the different aspects of the national scheme, a three-dimensional model that embeds each target group theory of change within the overarching theory of change would effectively depict the complex and intricate map of causal pathways and mechanisms of the scheme.

Our analysis of the theories of change was led by the core aim of opening a discussion on possible evaluation approaches and informing our proposals for a preliminary evaluation design. The theories of change that we developed for the national initiative provide a detailed understanding of the scheme's mechanisms, of the causal relationships that are aimed at influencing behaviour change and how these lead to desired policy outcomes.

These theories of change set out how the scheme will operate if the target groups react to the scheme in particular ways. The scheme is reliant on voluntary take-up and once it is implemented, its success is dependent on specific attitudes and behaviour change among food businesses and, more importantly, consumers.

An evaluation can shed light on whether the theories of change embedded within the policy initiative are 'true' or if the underlying assumptions guiding the scheme need to be reassessed and the policy redesigned.

A Preliminary Evaluation Design

Purpose and nature of programme evaluation

Programme evaluation can be defined as “the systematic assessment of the operation and/or the outcomes of a program or policy” (Weiss, 1998: 4). In the previous section of this report, we outlined programme theories which describe the essence of the national scheme; what it seeks to achieve and how it is designed to bring about change. The purpose of doing so was to identify some of the most important issues and challenges that face us in attempting to evaluate the national scheme. The end objective for this present study, therefore, is to facilitate the development of approaches that will ‘systematically assess’ the *operation* of the national Food Hygiene Rating Scheme (and its forerunners, existing food hygiene rating schemes), and permit us to identify its impacts.

Broadly speaking evaluation research can be divided into two separate sets of activities that seek to address fundamentally different questions about policy. The first of these approaches can be grouped together under the heading *summative* methods. Summative methods essentially involve judgements regarding a programme’s effectiveness and value for money (Rossi, et al., 2004). They identify the extent to which we can attribute any change in an outcome of interest to the programme or scheme itself, rather than to competing explanations. We refer in this present discussion to these approaches as *impact evaluation*. Summative methods, however, also embrace cost-benefit analysis that assesses the net costs and benefits of a programme. We do not discuss cost-benefit analysis in this paper.

Formative methods essentially describe the way a scheme or programme operates. They permit us to understand or explain how the programme being evaluated brings about or generates change. They test the programme theory in its constituent parts examining the extent to which the programme model actually ‘holds-up’ in reality. They describe and analyse the contexts in which the programme is implemented, and explore how target groups respond and perceive the intervention. Shadish, et al., (2002) refer to formative approaches as providing *causal explanation*. In this present study we conceive of formative methods as involving the study of programme processes. *Process evaluation* is an essential component of any evaluation, in that it helps explain the effects of the programme identified through *impact evaluation*. In many instances, however, process evaluations are conducted where impact evaluations are not required or not feasible.

In what follows, we set out our initial thoughts as to the most promising approaches to evaluating both the impact of food hygiene rating schemes and scheme processes. We start by considering a preliminary design for assessing scheme impact. Following this, we turn our attention to process evaluation.

An impact evaluation design

As we have noted, impact evaluation informs judgements regarding the effectiveness of a programme. It seeks to enable evaluators and policymakers to draw conclusions of a causal nature; that is the extent to which the policy or programme under investigation has brought about the desired change in behaviour.

This section commences with a discussion of some of the general assumptions we have made in developing a credible approach to identifying the impacts of existing food hygiene rating schemes should local schemes prove effective, prior to extending the evaluation to consider the national scheme. In discussing impact evaluation, when we refer to the national scheme we refer to all authorities entering the scheme regardless of when they do so. By contrast the process evaluation distinguishes a group of early entrants who will adopt the national scheme before the end of June 2011. This section then introduces the concept of the counterfactual and discusses the main research hypotheses the impact study will test. The discussion then moves on to present some initial thoughts on the most promising approach to evaluating the impact of the scheme given the constraints faced. It is important to note at the outset that there are significant uncertainties, particularly regarding the types of data which may be available and the manner in which schemes were implemented, which may render a convincing assessment of the effects of existing food hygiene rating schemes impossible. At this stage there is also doubt as to the quality of data which may be available and whether it will be possible to detect the impacts of existing food hygiene rating schemes if its effects are modest.

General assumptions and constraints

In order to develop a convincing impact evaluation design some broad assumptions need to be made regarding priorities and resource constraints. These are not analytical constraints but rather those that impinge upon our design considerations which are more practical in nature.

The following assumptions guide our design work:

- Although the national Food Hygiene Rating Scheme is a policy developed by the Food Standards Agency it is implemented by local authorities. The national scheme will be non-mandatory and in the final analysis it is Local Authorities which determine whether a scheme is introduced.⁹

⁹ Although launched as a voluntary scheme, following Lord Young's recommendations in the 'Common Sense, Common Safety' report (http://www.number10.gov.uk/wp-content/uploads/402906_CommonSense_acc.pdf) the FSA plans to implement mandatory participation for all English local authorities in the future.

- The scope and nature of any evaluation is almost always constrained (or enabled) by the types of data that are available and their quality, or the data that could feasibly be collected. It is our understanding that resources that would permit large-scale quantitative survey data collection among individual consumers and businesses are unlikely to be available, and that there is generally a wish to avoid expensive primary quantitative data collection.
- We assume that data recording the incidence of foodborne illnesses among *individual* consumers will not be available to the evaluation. Our proposed approach does require that such data are available aggregated to the level of the Local Authority.
- We assume that data recording the outcome of hygiene inspections of individual food businesses will not be available to the evaluation. We understand that aggregated data on the outcome of inspections across local authorities will be available in the future, and that there are some potential aggregate measures available now and in the recent past.
- Data from the FSA's Food and You Survey and the Tracker Survey will be available to provide contextual information and an indication of the level of awareness of food hygiene rating schemes among the general population of consumers.

Research hypotheses and the counterfactual

In order to set out the main hypotheses to be tested through the impact evaluation, we first need to introduce the idea of the counterfactual. At first sight, the concept of a counterfactual appears artificial and rather abstract. It is, however, an essential notion that guides our considerations.

Impacts are always defined relative to some other condition (Holland, 1986). To help see this, imagine a 'unit', this could be an individual or a local authority, or some other organisation, the behaviour of which we wish to alter. We introduce a policy in order to influence or change the behaviour of this unit. Furthermore, the behaviour we wish to alter or change is captured in some outcome measure which we denote as 'Y'.

To determine whether our policy actually alters the behaviour of this 'unit' and thus changes 'Y', theoretically, we need to observe 'Y' in two conditions or states. First, we need to observe 'Y' in the state where the unit *is exposed* to the policy. Second we need to observe 'Y' in the state where the unit *is not exposed* to the policy. The complication comes from the fact that we need to make these observations simultaneously, at the same point in time. The effect of the policy is then the difference between 'Y' in the exposed and unexposed condition or state. So we measure the effectiveness of exposure to the policy relative to non-exposure¹⁰.

¹⁰ Note that non-exposure need not necessarily mean exposure to no policy. It could for example mean exposure to an existing scheme or policy where interest is in understanding the impact of some new programme or scheme.

The obvious problem we have to address is that we do not observe 'Y' for the unit in both the exposed and unexposed states simultaneously. We can observe 'Y' in either state (exposed to the policy or remaining unexposed) but not in both at the same point in time. Thus if a unit *is exposed* to a scheme, we have *missing information* – we do not know what the outcome 'Y' is in the absence of the policy or programme. The state or condition we do not directly observe is referred to as the *counterfactual* – that is 'counter to the fact'. The challenge of impact evaluation is devising a method of *estimating* this missing information – or estimating the counterfactual. The degree to which any approach to impact estimation is deemed convincing depends on the plausibility of the approach to estimating the counterfactual adopted.

Research hypotheses

Bearing in mind the concept of the counterfactual, we can now specify the main research hypotheses which the impact evaluation will address. These hypotheses are directly related to the theories of change discussed in earlier sections of this report.

The main hypotheses that the impact evaluation will test are:

- The incidence of foodborne illness is reduced in areas where any existing food hygiene rating scheme operates compared to those areas where no scheme is in operation
- The incidence of foodborne illness is reduced in areas where the national Food Hygiene Rating Scheme operates compared to areas where there is no scheme and areas with existing schemes
- Standards of food hygiene among businesses are higher in areas where a food hygiene rating scheme is in operation than in areas where there are no schemes
- Standards of food hygiene among businesses are higher in areas where the national scheme operates compared to areas where there is no scheme and areas with existing schemes.

A number of important points need to be made about these hypotheses. First, they recognise that the national Food Hygiene Rating Scheme is being introduced into a context in which around half of local authorities already run an existing food hygiene rating programme. These schemes are referred to as 'existing schemes' throughout this report. The research hypotheses articulated above recognise the importance of understanding the impact of these schemes first, prior to attempting to evaluate the impact of the national policy. This is because the impacts of the national programme within a local authority will depend, in our view, on whether that authority already runs a scheme. For example, the national scheme is likely to have a much greater impact in areas without an existing scheme than it will in those with a scheme. This may be particularly so from the perspective of food businesses. For businesses the change from no scheme to the national scheme will be a more significant change than moving from an existing scheme to the national scheme.

The theories of change set out earlier in this report revealed the importance of two main outcomes. These are that the national scheme will reduce the incidence of foodborne illness among consumers and lead to improvements in food hygiene standards across food businesses. Note that the hypotheses set out above imply that the impact study would first consider the effects of existing schemes on the incidence of foodborne illnesses and hygiene standards, and then the effects of the national scheme given the presence (or otherwise) of existing schemes.

For existing schemes the comparison will be with the condition where no-scheme is in operation – so put crudely, the effect of having some food hygiene rating programme rather than nothing. In evaluating the national scheme, the impact study will compare outcomes across four separate states or conditions. These are:

- Operating the national scheme having previously operated a similar food hygiene rating programme
- Operating the national scheme having *not* previously operated a similar food hygiene rating programme
- Not operating the national scheme but continuing to operate an existing similar food hygiene rating programme
- Not operating any scheme.

The degree to which these various effects can be identified in the data will depend on the number of observations in the dataset, the differential effects of the various programme combinations, the number of units (in our case local authorities – see below) exposed to the various treatment combinations (described at the bullet points above) and the degree of statistical precision required, among other considerations.

The approach to measuring programme impacts we put forward here can in theory be extended from first measuring the effects of existing schemes to incorporating the introduction of the national programme. However, as our discussion here illustrates there will remain some uncertainty regarding how effectively this can be done (we discuss this further below).

Unit of analysis

Thus far our discussion has left unaddressed the important issue of the unit of analysis in which the impact evaluation should be conducted. This is clearly of major importance. The unit of analysis determines not only how results from an impact study are to be interpreted and understood but also practical issues such as the availability of data.

Given the hypotheses outlined above, two obvious units of analysis would be individuals, in terms of exposure to foodborne illness, and business, in terms of food hygiene standards. However, as previously discussed, it is unlikely that the evaluation will have access to data at the level of the individual or food business.

As a result, we propose conducting impact analysis at the level of the local authority. This is for the following two reasons:

- Local authorities are responsible for delivering food hygiene rating schemes, and there will be variation in implementation across authorities which can be exploited in estimating counterfactuals
- It is our understanding that the FSA has access to data which could be used to measure outcomes of interest at the local authority level or which could be developed in order to do so.

The main implications of defining the local authority as the unit of analysis are that all data used in our analysis need to be aggregated to, or available at, this level of geography and that impacts and outcomes are defined at this level.

A framework for estimating impacts

In order to be able to estimate a counterfactual we need to identify sources of variation in the way that existing food hygiene rating schemes and the national scheme are implemented. Thinking firstly about the effects of existing schemes, we can identify two important sources of variation in the way they have been implemented. Note that in this discussion we will focus on existing schemes but emphasise that our approach can, in theory, be extended to incorporate an evaluation of the national scheme.

First, at the level of the local authority there is variation over time in the implementation of schemes. That is, the rather obvious point, that there were periods of time prior to the introduction of schemes where we could feasibly observe outcome measures of interest; that is the incidence of foodborne illnesses and food business hygiene standards. This is important because the incidence of foodborne illnesses and food hygiene standards prevailing in local authority areas prior to the introduction of schemes may tell us something about what would have happened subsequently in local authorities had they not introduced schemes. In other words, historic trends in key outcomes prior to the introduction of existing schemes may yield important information regarding counterfactual outcomes.

The second source of variation is that across local authorities. Not all local authorities have implemented schemes. In fact around 50 per cent have no food hygiene rating scheme. This means that we can examine outcomes among local authorities with no scheme and use these to estimate the outcomes which would have prevailed among local authorities with schemes had they instead not introduced a food hygiene rating programme.

This discussion leads us to conclude that any approach to identifying the impact of food hygiene rating schemes should harness *both* these sources of variation in order to be credible.

An approach to estimating the counterfactual

In order to provide estimates of the impact of food hygiene rating schemes which exploits both the sources of variation discussed above we recommend using a Difference-in-Differences (DiD) estimator (Blundell and Costa Diaz, 2009). The DiD estimator is an econometric approach to estimation which possesses a number of desirable qualities. In sociology and psychology this approach is often referred to as the ‘gain-score’ estimator or the ‘fixed effect’ estimator (Oakes and Feldman, 2001; and Allison, 1994).

In this section we discuss the types of data that will be required in order to permit the use of the DiD estimator as well as its desirable features. The DiD estimator has been used in a number of applications, using *aggregated* UK data, that are not too dissimilar to that envisaged here. For example, Machin and Marie (2005) used DiD to examine the effect of the Street Crime Initiative on robbery, importantly using data aggregated at the level of the Crime and Disorder Reduction Partnership (analogous to local authorities). Sabates and Feinstein (2007) used DiD to explore the effects of Educational Maintenance Allowance and the Reducing Burglary Initiative on burglary convictions with data defined at the level of Local Education Authorities. In the US a number of studies have used a similar estimator to identify the impacts of hygiene grade score cards for restaurants in the Los Angeles area though in this case analysts had access to individual and business-level data (for example see Jin and Leslie, 2003).

The Difference-in-Differences estimator

Table 1 helps illustrate, in a simple way, how the DiD estimator can provide a measure of impact.

Table 1: Illustration of the DiD method of estimating scheme impacts with hypothetical data

	Average food hygiene rating prior to the scheme's launch	Average food hygiene rating after the scheme's launch	Difference
Local authorities with a scheme	105	85	-20
Local authorities without a scheme	110	105	-5
DiD estimate			-15
Note: a lower score is a better hygiene rating			

The basic approach is to measure average outcomes before and after schemes are introduced, in areas which do go on to introduce schemes as well as those areas which do not. This leaves us with four measures – average outcomes before and after in areas which *do* introduce schemes and average outcomes before and after in areas which *do not* introduce schemes. Next we take the after measure among those areas with schemes and subtract from this the average before measure. For the example in Table 1 we can see that the result of this calculation is ‘-20’ ($105-85=-20$) for areas with schemes. This means that hygiene ratings have fallen by 20 points. We then repeat this calculation for authorities which do not introduce schemes. In Table 1, we can see that this calculation yields a result of ‘-5’ ($110-105=-5$). Finally, we subtract the result obtained from areas without schemes from the result for areas with schemes to obtain the DiD estimate of impact. In our example, this final calculation yields a result of ‘-15’ points ($-20 - -5=-15$).

The important point to note is that the difference in outcomes obtained from authorities without schemes *is the counterfactual estimate*. In other words, in the absence of the scheme, we assume that hygiene ratings would have decreased in this example by 5 points. However, because hygiene ratings actually fell by 20 points in areas with schemes, we can attribute the difference-in-differences (‘15’ points) to the scheme itself – hygiene ratings would have fallen by 5 points if we had done nothing, so put crudely, policy cannot take the credit for this!

The attraction of the DiD approach is that it controls for unobserved permanent differences between local authorities. Thus, for example, it *does not* matter whether authorities that introduce schemes tend to have permanently higher or lower hygiene ratings. Moreover, the approach also controls for trends in outcomes common to both authorities with and without schemes.

In practice DiD impact estimates are obtained through implementing a linear regression model¹¹. This enables us to include control variables. These are variables whose values vary over time and across local authorities, and which also help explain variation in the outcomes we are interested in, thus potentially reducing any bias in our estimates and improving precision. Furthermore, this approach can be extended to incorporate matching. Matching can be used in combination with DiD where we are concerned that the data do not support all the covariate patterns which arise as a result of the inclusion of control variables. Matching in this respect enables us to relax some of the assumptions underpinning the linear regression model approach to implementing DiD and in some senses can be considered a semi-parametric approach to estimation (Heckman, Ichimura and Todd, 1997). Through matching we can directly enforce common support.

What data do we need?

In order to implement the DiD estimator we recommend constructing a local authority panel data set. Ideally the panel should commence in, for example, about 2001, several years before schemes were introduced. Pre-scheme data is essential for the proper implementation of DiD. What this effectively means is that for each local authority in the country we collect a range of data items, defined consistently over time (since 2001), updated on a regular basis. These data can be weekly, monthly, quarterly but at a minimum would have to be annual. The point to emphasise is that data items would need to be defined consistently over the time periods covered by the data set, or converted in some way to make them consistent¹².

It is critical to bear in mind that our ability to conduct a rigorous assessment of the effects of food hygiene rating schemes is determined by the availability

¹¹ In this case an equation of the following form would be estimated $Y_{it} = \alpha_i + \phi S_i * T_t + \phi X_{it} + \varepsilon_{it}$, where Y represents the outcome of interest, S a dummy variable coded '1' for authorities with schemes and zero otherwise, T a dummy variable coded '1' for periods of time following the introduction of schemes and zero otherwise, and X a series of time varying controls. The equation would be estimated using linear OLS. In theory this equation can then be expanded to incorporate the effects of the national scheme.

¹² For example, it will be necessary to consider carefully the effects of food hygiene legislation introduced in 2006 on potential data series and the research design more widely.

and quality of data. There is still much uncertainty regarding the types and quality of data that are and will be available, and therefore whether it will prove possible to undertake an impact evaluation of the type discussed here.

It is very important to note, that a data set such as that described here, will be useful for designing qualitative and process study samples, irrespective of whether it is used for impact estimation.

The panel data set would need to contain the following items:

Outcome variables: variables which, for example, capture the incidence of foodborne illnesses at the level of the local authority, and variables which measure, in aggregate, food hygiene standards. Our preliminary discussions with FSA officials suggest that variables do exist which could be used to measure such outcomes or are in the process of being developed, and that such variables are available historically¹³ (though with some concern over consistency).

Policy variables: A variable which captures the point in time at which schemes were introduced across authorities. This will typically take the form of simple dummy variables coded '1' at points in time where a scheme was in place and '0' where no scheme was in place (capturing before and after variation). Secondly, a dummy variable indicating which local authorities have schemes and which do not. A further point to consider is the identification and collection of variables which capture something of the possible variation in the ways in which local schemes have been implemented. At the time of writing it is difficult to anticipate precisely what variation in implementation is likely to be important, beyond the presentation or depiction of inspection scores to consumers. Qualitative research could help shed light on these issues.

Contextual variables: these are variables which enable us to capture other aspects of local authorities which are likely to influence and account for variation in the outcomes of interest. For example, local authorities in which there are a large numbers of food businesses (areas which attract a large number of tourists) may display different patterns of foodborne illness to those areas where there are fewer food businesses. For this reason it may be advisable to collect contextual variables which measure, for example, the proportion of the local workforce employed in leisure and tourism industries, or develop a measure of the number of registered food businesses within a local authority. Variables that describe the nature of food businesses within a local authority will also be of importance as might variables that measure the demographic structure of the local population – the proportion of elderly people, the proportion of families with children, its ethnic composition, and so on.

¹³ For example, data which will become available through the LAEMS system should be explored to determine how far it could be used in the evaluation of the national scheme.

Practical challenges

Complex schemes, such as food hygiene rating schemes, which seek to change the behaviour of different groups of agents, are never easy to evaluate. There are a whole range of potential threats to conducting a successful evaluation quite apart from the issue of whether schemes are likely to generate impacts that can be detected. In this section, we seek to outline some of the main foreseeable challenges to successfully evaluating the impact of food hygiene rating schemes using DiD. In the following section we focus more on analytical challenges. These practical challenges include:

Data sets need to be complete: we need consistent data series for inclusion in the local authority panel data set. Candidate data items for inclusion should be examined to ensure that any definitional changes can be adjusted to achieve consistency over time. Many contextual variables can be obtained from consistent, publicly available data series, such as the Annual Population Survey, Labour Force Survey and other annually published data disaggregate to the local authority level. Of greater concern are the data series which are administered by the FSA or other organisations such as the Health Protection Agency. For example, those series which are being developed that seek to capture the incidence of foodborne illnesses at the local authority level, as well as those series which record the nature of and interventions among food businesses.

Consistency over time: for the DiD estimator to be applied to the data it is absolutely essential for consistent data to be available prior to the introduction of schemes. We recommended in the first instance to construct a panel going back to at least 2001 if possible.

Policy innovation: as has already been noted, another essential component of any data series is variables which track the introduction of schemes across authorities. Furthermore, in order for the impact of schemes to be identified they will need to have been introduced at roughly the same points in time. In others words, the estimator requires that the gap in time between authorities which first implemented existing schemes and authorities which were last to do so, is not too great, ideally not greater than 18-24 months. This requirement is necessary in order for analysts to identify clear before and after periods in the data. At the time of writing, it is not apparent whether the FSA knows when existing schemes were introduced by authorities. In theory, however, this data must be available or could be obtained. In assessing the evaluability of the programme this area might be one of the first that will require detailed investigation. The FSA has recently inserted questions on the uptake of schemes in the local authority omnibus survey. This will be useful but will not give complete coverage of all local authorities due to survey non-response.

Local government reorganisation: it is important to note that local government reorganisation may affect the extent to which consistent series can be derived. The effects of any re-organisation will need to be examined carefully.

Take-up and timing of the national scheme: in order to evaluate the effects of the national scheme the rate and timing of its take-up among those who do and do not operate existing schemes needs to be carefully considered. It is difficult to be precise, but ideally, from the perspective of the evaluation only (clearly FSA will want to maximise take-up), we would want about half of authorities which operate existing schemes to introduce the national scheme and half of authorities with no existing scheme to do so. As the proportion taking-up the new scheme diverges from roughly 50/50 the standard error associated with the impact estimate will increase all else being equal. This will make it harder to detect the effects of schemes, at conventional levels of statistical significance, should they generate impacts greater than zero. In the extreme case of all local authorities taking-up the national scheme at about the same point in time, a DiD approach will become infeasible. In such a case some form of interrupted time series design might be considered but this is a substantially weaker approach than DiD.

We would also prefer it, from an analytical perspective, if all those authorities who sign-up to the national scheme introduce it over a relatively short time period, say within a single year. This would enable us to clearly establish before and after measures. Having said this, it is obviously not tenable for the evaluation to constrain attempts to introduce the national scheme more widely. Therefore the precise timing and uptake of schemes is clearly a risk to the evaluation.

Measuring outcomes: any impact evaluation relies on the valid and reliable measurement of outcomes. Our initial discussions with Agency officials suggest that existing data held by the Agency do offer the potential for successful measurement of outcomes but that further work is required in order to confirm the position.

Work is underway within the Agency, for example, to develop measures of foodborne illness reported to general practitioners at the level of the local authority. However, even if such measures were successfully developed, they would suffer from some detracting factors for our purposes. First, the measures would capture foodborne illnesses contracted in the home as well as from food businesses. Second, data reported at the local authority level will include illnesses acquired from food businesses located outside the local authority and vice versa. Third, such illnesses can be contracted from a wide range of other sources including farm animals and pets. Fourthly, the weather is also a factor and FSA analysts suggest controlling for it in any analysis. Fifth, not all disease is reported and thus recorded. Despite these problems, similar data

has been used in the evaluation of the Los Angeles Hygiene Grade Cards scheme (Jin and Leslie, 2003)¹⁴.

In terms of outcomes which measure aggregate levels of food hygiene amongst businesses at the level of the local authority, the FSA appear to collect annual data from which useful outcome measures *might* be obtained. Some of the potential outcome measures are aggregated and cover warnings, notices, suspensions, seizures and prohibitions for both food standards and hygiene violations. Practices in relation to formal enforcement action appear to vary from local authority to local authority. FSA do have data on prosecutions for hygiene reasons by establishment type, which would seem a potentially promising measure to explore. It is also worth investigating whether trend data on expenditure on food hygiene inspection by authorities can be obtained. We understand that the Department of Communities and Local Government may collect such data already.

It is important to note that if outcomes measures are particularly susceptible to measurement error and other sources of variance then this may frustrate our attempts to identify statistically significant programme impacts at conventional levels. In other words, it may prove impossible to separate out the policy 'signal' from the background 'noise'. Control variables of sufficient explanatory power are important in this respect. This issue is of appreciable concern if scheme impacts are expected to be modest.

Analytical challenges

In order for the DiD estimator to yield unbiased results certain conditions need to be met. The first and most important of these is the so called assumption of 'common trends'.

Common trends basically holds that the difference or trend in outcomes observed among authorities *without* schemes is that which would have prevailed among those authorities *with* schemes had they not introduced

¹⁴ Jin and Leslie derived an outcome measure from hospital admissions data within which they isolate diagnoses principally relating to foodborne illness. They had monthly data and it appears information about the zip code of the individual admitted (though it is not clear if they mean zip code of the hospital, the individual's address, or where the disease was contracted – though the latter seems unlikely). They were able to separate out admissions relating to foodborne and non-foodborne illness. Many of the other detractors of the data discussed above in the case of food hygiene rating schemes apply, however, to Jin and Leslie's data. Interestingly, despite these limitations, the researchers did find that the Los Angeles' scheme had a negative impact on hospital admissions relating to foodborne illness. The authors do use the change in hospital admissions for non-food related digestive disorders as an additional control. It might be possible to replicate this approach with data available from HPA via FSA.

them. Clearly it is impossible to know whether this condition actually holds or not. We can, however, obtain evidence on whether the common trends condition is *likely* to be met by comparing trends in outcomes between authorities with and without schemes in the period of time prior to their introduction. If the relative trends in outcomes across the two groups of authorities (those with and without schemes) are stable over this period, this provides evidence that common trends are likely to hold. It is partly for this reason that a series of pre-scheme historic outcome data are required.

One concern related to that of common trends is the motivation among local authorities for introducing schemes. For example, if authorities introduce a scheme in response to a sudden spike in the incidence of foodborne illness this can frustrate our attempt to recover unbiased impact estimates. This concern stems from the potential for mean reversion as the level of foodborne illness reverts to that more typically seen post the introduction of the scheme and this 'reversion' is conflated with the effect of the programme. The extent of this problem can again be identified by exploring trends in outcomes prior to the introduction of schemes. If necessary, and if the data allow, the problem can be counteracted by redefining the baseline – the period of time over which we observe our pre-scheme outcome measure.

Another assumption which must hold is that of 'common composition'. This assumption means that we assume the populations of consumers or food businesses do not alter significantly in ways which are related to the existence of schemes over the period of the evaluation. For example, we would be concerned if there was evidence that food businesses were relocating to areas where schemes were not operating from areas where they were in order to avoid extra burdens being placed upon them. To some extent such shifts, if they occur, can be controlled for through including time varying covariates in the regression model used to estimate impacts. Nonetheless, it is worth considering the degree to which compositional effects might be important, particularly any changes resulting from the recent economic downturn.

Another matter which will require careful investigation if results from DiD are to be considered unbiased is the degree to which food businesses may have anticipated the introduction of schemes and adjusted their behaviour pre-emptively. In such cases it is possible that anticipatory behaviour change may be captured in the baseline or before measures, and thus bias estimates. Anecdotal evidence suggests that businesses may start to improve hygiene practices when they are aware of changes that are to be introduced. Evidence from the process evaluation can help determine whether anticipatory effects are likely to be significant. If they are, baseline measures may need to be adjusted.

Finally, the existence of panel data implies that the DiD estimator in this case will be applied to data with multiple pre- and post-scheme measures. This raises the rather technical problem of period-specific serial correlation in the transitory error components of the regression model, first raised by Bertrand, et al., (2004). Bertrand and her colleagues show that such problems can lead

to incorrect statistical inferences, most notably a tendency to over-reject the typical null hypotheses of no programme impact. Fortunately, Bertrand and colleagues also put forward a number of potential solutions to this problem which will need to be considered carefully.

Taking the impact evaluation forward

In order to take forward the impact evaluation we identify three separate sets of tasks which need to be undertaken:

Step 1 – data exploration

- This step could commence immediately and will involve detailed discussion with Agency staff, both policy makers and operational researchers, in order to examine the types of data they hold and the types of data they are currently seeking to construct. This would cover particularly the outcome measures available as well as the policy variables that might be constructed
- This stage will also involve exploring publicly available sources of data which might be available
- It will also involve provisional analysis of the data which are available to determine their quality
- It will provide an assessment of the suitability of the available data for undertaking DiD analysis and their suitability for use as a sampling framework for the process study. As the end of the work it should be possible to say whether a DiD approach is possible
- The work will require knowledge of publicly available data sources, experience in data capture, manipulation, analysis of secondary data, knowledge of DiD estimators and data set construction.

Step 2 – constructing panel data set

- The precise nature of the activities envisaged under this Step will depend on the outcome of Step 1.
- If Step 1 finds that a DiD approach is viable, then Step 2 will involve constructing a longitudinal panel data set of local authorities. This data set will include outcome measures, policy variables and control variables as described above. The work will involve extracting data from FSA sources and public data repositories. Derived variables will be created, the data cleaned and checked. The data will then be matched to individual local authorities and time series created. At this stage it will be possible to determine the size of impacts a DiD approach is likely to be able to measure
- If Step 2 finds that a DiD approach is not possible then the work will focus on creating a sampling frame for the process study and facilitation of contextual analysis. For example, looking at what types of authorities have schemes and how they differ to those who do not
- Careful consideration will need to be given as to the consistency of data, time intervals over which data are measured and other data quality issues.

Step 3 – Estimating the effects of existing schemes

- If the conclusion of Step 1 above is that a DiD approach to estimating programme effects is viable and the panel data set has been constructed successfully under Step 2, then Step 3 will involve the estimation of programme effects for existing schemes
- The work would involve data manipulation, identification of pre- and post-outcome measures, data cleaning and checking, estimation itself, specification and robustness checks as well as report writing and presentations.

A Process Study Design

A process study focuses on how a programme was implemented and how it operates. It also helps to explain which elements of a programme work well and the conditions under which this is achieved. A process study is also useful in understanding the links between programme processes and delivery contexts, allowing for a comparison of programme delivery across multiple sites. In order to design a process study an evaluator must first identify all the components of a programme including activities, materials and outputs as well as the key groups that the programme is attempting to influence. This is useful in setting out the main research questions for such a study.

In the next section, key research questions that a process study of the national scheme might answer are set out. This is followed by a discussion of practical considerations in designing a process study. The section concludes with a proposal for a preliminary design.

Developing research questions for a process study

In setting out the research questions for a process study, it is particularly useful to ask how, why, for whom and under what conditions a policy intervention works, or fails to work? The overarching questions that a process study for the national scheme would answer are:

- How is the national scheme being implemented?
- Is the scheme operating as designed?
- What are the experiences of the key groups that the scheme is attempting to influence?
- What are the perceived changes that are emerging as a result of the key target groups' interactions with the scheme?

To answer these questions, it is important to explore how the different components of the scheme are linked together and how each component is experienced by the relevant target groups. The theories of change articulated for each target group are a useful guide in identifying the thematic areas or categories under which detailed research questions can be developed. In the next sections we consider each target group in turn, identifying relevant areas of inquiry in a process study.

Local Authorities

Referring back to Diagram 2 which identifies the theory of change for local authorities, the thematic areas that a process study might explore are:

- The adoption and implementation of the national Food Hygiene Rating Scheme (rationale and process)
- Marketing and communications strategies to publicise the initiative
- Inspection regime (changes, consistency etc))
- Inspectors' skills and training
- Inspectors' perceptions, attitudes and behaviour

- Inspection reporting – both inspectors' reports and the local authority annual reporting process for the FSA.

Food Businesses

The theory of change for food businesses described in Diagram 3 helps to establish the following thematic areas for possible exploration in a process study:

- Food businesses' knowledge and awareness of new rating system
- Their perceptions of the inspection process
- Food businesses' reactions to the national Food Hygiene Rating Scheme (including attitudes to the website and rating certificate)
- Effects of the new rating system on food hygiene-related behaviour
- Their perceptions of the national Food Hygiene Rating Scheme
- Food businesses reactions to the ratings of competitors.

Consumers

Referring to Diagram 4, the key thematic areas in relation to consumers that can be explored in a process study are:

- Consumer characteristics in relation to the use of food businesses
- Consumers' awareness and knowledge of the national Food Hygiene Rating Scheme
- Use of the national Food Hygiene Rating Scheme website
- Reactions of consumers to rating certificates
- Consumers' understanding and use of the rating system
- Consumer decision making – exploration of factors influencing choice and the importance of food hygiene in that choice
- Effects of the national scheme on food hygiene related behaviour.

For each target group and under each thematic area a series of questions can be developed. For example, in order to explore consumers' understanding and use of the rating system in more detail, the following questions would be a useful starting point:

- a. How well do they understand the ratings?
- b. What does a low rating (as opposed to the top rating) mean to them?
- c. What do they think of the rating system? (Is it useful? Why/Why not?)
- d. Do they compare ratings of food businesses? (Why/why not?)
- e. If yes, then what do they do (specific steps)? How often do they do this?
- f. Do they make trade-offs between price and food safety/hygiene? Is this related to their income?

Covering each thematic area for each target group by developing a set of detailed questions and related 'probes' will build a comprehensive multi-

perspective understanding of how, why, for whom and under what circumstances or conditions the national scheme works or fails to work.

Contextual considerations

Before proposing an approach for a process study, it is important to consider the range of possible variations or contextual factors that can influence the implementation and the overall desired outcome of a programme or initiative. The complex weave of contextual social factors that can influence the successful delivery of the national scheme and affect the desired ultimate outcomes, is an important consideration when determining the most appropriate methodological approach for a process study.

Taking an approach similar to the one used to develop thematic areas above, the contextual implementation and delivery variables are considered in turn for each target group.

Local Authorities

A local authority's history of similar food rating schemes is useful in understanding the background context within which the national scheme is implemented. The rationale for adopting the national scheme may differ between those local authorities that have had schemes in place and those that have not previously had a food hygiene rating scheme. The existence of previous schemes may also influence the level and extent of behaviour change among food businesses and consumers, determining also the content of any local public information campaign.

In areas where there is no rating scheme in place, food businesses may have been awarded food hygiene certificates by national bodies. In such cases, if a national scheme is adopted, the display of multiple or different food hygiene certificates could result in consumer confusion, which in turn could have implications for the scheme's intended effects on attitudes and behaviour.

The socio-economic characteristics of a local area including population density and variations in the level of affluence and deprivation between wards will most likely affect consumer behaviour. These variations may also influence the proliferation of particular types of food businesses as well as the size of the leisure and tourism industry.

In areas with tourist attractions, the size of the leisure and tourism industry may be skewed and the clustering of food businesses in the proximity of tourist attractions will be quite marked. This supposition would be more likely for urban areas and less so for tourist attractions in more rural locations.

Business Types

The wide range of food businesses that are subject to food hygiene inspections adds an additional layer of complexity that may affect the success of the scheme in relation to consumer decision making and behaviour

(consumer consideration of food hygiene may vary depending on the type of food business they use).

The location of food businesses and their clustering by type could influence competition between food businesses based on price or food hygiene standards. For example, if a similar type of food business is clustered (such as, take-aways on one street in a deprived urban neighbourhood) businesses with a lower hygiene rating may decide to cut prices in order to compete with a business with a higher food hygiene rating or they may decide to compete by raising hygiene standards and/or request re-inspection. An unintended consequence of the success of the scheme could be a greater burden on local authorities in terms of re-inspection requests or appeals.

Another unexpected result could be a reduction in competition if food businesses with lower hygiene ratings close down. Alternatively prices could either rise in response to increased costs associated with higher hygiene standards, or could fall if businesses respond to increased non-price competition (for example, a low hygiene rating) by lowering their prices. Similarly, consumers' reactions to food prices may also correspond to hygiene ratings. Consumers may think that a better hygiene rating justifies higher prices and may therefore, be willing to pay more for the same type of food at food business with a higher rating than at one with a lower rating.

Consumers

As consumers are the primary target group for this initiative, a consideration of key differences among consumers is important in establishing how well the initiative works for different consumer groups. Two broad categories which can be used to consider consumer variations are 'characteristics' relating specifically to socio-economic dimensions and 'types' relating to a consumer typology based on interactions with food businesses. This is explained in more detail below.

It is highly likely that differences in socio-economic characteristics will influence consumer behaviour in terms of the types of food businesses that are regularly used, the emphasis placed on food hygiene in general (particularly in relation to price) and the distance that consumers might be willing to travel to use a particular food business. Relevant consumer socio-economic characteristics include:

- Household income/level of disposable income
- Household composition (couple parents, single adults, lone parents)
- Age of consumers (elderly, young adults)
- Age of family members (households with young children)
- Level of education (particularly level of literacy but also proficiency in English)
- Gender
- Ethnicity.

Consumer ‘types’ are defined based on the nature of interactions that consumers have with food businesses, influencing not just which food businesses consumers use but determining also the nature of consumers’ interactions with the scheme’s activities and outputs. Types of consumers include transient populations, such as students and tourists, whose presence or temporary residence in a local area influences not only the types of food businesses that flourish but also consumer behaviour in relation to food hygiene. Their level of understanding of food rating initiatives may be poor and their behaviour may be determined more by contextual opportunity (for example, a food business close to a museum in the case of tourists).

Other identified consumer types include the ‘planner’ – a consumer who makes food choice decisions on a number of factors such as type of cuisine, location, price and food hygiene and the ‘spontaneous’ consumer whose behaviour in relation to food businesses is not based on a clear decision making process.

These consumer types are not however mutually exclusive and a consumer can shift between the types based on their immediate context. This fluidity in consumer types is based on factors such as time, place, occasion and opportunity and may affect consumers’ decisions in relation to food hygiene.

A process study of the national scheme should consider identified consumer types and incorporate new types that emerge into the existing typology. Analysis based on a such a typology can further understanding of how different types of consumers interact with the schemes’ outputs leading to possible refinement of how food hygiene messages or scheme outputs are targeted to different consumer types.

The proposed approach

The range of factors and characteristics discussed above identify a complex multi-layered landscape within which the national scheme is to be implemented. In light of this complexity and the range of stakeholders whose interactions are key to the success of the scheme, a focus or point of access through which to conduct the process study needs to be considered. This can be done by thinking about the ‘layers’ that the scheme affects. The top tier consists of local authorities which adopt and deliver the scheme, the middle layer comprises food businesses and the bottom layer consists of consumers with the greatest number of variations and characteristics.

It makes sense therefore to develop a process study design by initially focusing on the top layer, in this case, local authorities. This is particularly appropriate for the national scheme as its adoption and delivery is the responsibility of local authorities as is the legislative requirement to conduct and report on hygiene standards in food businesses.

We suggest that administrative areas that adopt the scheme would be the most effective unit of analysis for a process study of the national scheme. Using area boundaries to define an area of study for the national scheme

means that local contextual factors relevant to the scheme can be considered within clearly defined parameters. Furthermore this complements the suggestion of using local authorities as the unit of analysis for the impact assessment (refer to page 14). This focus further suggests that a process study should use a case study approach to develop a holistic understanding of national scheme processes, activities and interactions within specified and relatively fixed implementation areas.

In general, the aim of a case study approach is not to abstract an initiative from the socio-economic reality within which it is being implemented and delivered but to consider it in “terms of its embeddedness within the ambient context” (Snow and Anderson 1991: 153). In particular, a case study approach can be useful in understanding the relationship between programme context and programme processes which can then be compared across multiple delivery sites.

A case study approach entails a multi-perspective understanding of the causal links between activities along with the interactions between the relevant target groups. In practical terms this means that researchers undertaking a case study approach should consider that range of perspectives and voices to build a comprehensive case study. Given the complex and layered design of the national scheme, a process study using a case study approach would be the most appropriate method for investigating the effectiveness of the scheme.

Research methods

A characteristic of the case study approach is the gathering of information from multiple sources. The data from these sources is then brought together in an analytical process referred to as triangulation, which leads to the development of a detailed understanding of the ‘particulars of the case in its complexity’ (Stake 1994).

Case study approaches can use both qualitative and quantitative methods to gather data from relevant perspectives. The range of perspectives that a case study approach to evaluating the national scheme can incorporate, includes:

- A document review that covers guidance documents, training guides and marketing and communication materials
- A review and analysis of socio-economic factors in case study areas to establish a picture of the local context
- An assessment of activities and outputs utilising both qualitative and survey methods to capture use of the website, consumer (and food business) feedback on website, participant feedback on training programmes, feedback from public information campaigns
- Capturing the perspectives and attitudes of target groups including food hygiene managers, food inspectors, food business owners and managers and consumers using qualitative interviewing techniques such as in-depth interviews and focus groups.

Key challenges

Selecting case study areas

The selection of local authority areas for a case study based process study of the national scheme requires an “information oriented selection” (Flyberg 2006). This means that selecting case study areas is not based on achieving national representation but rather on selecting a small number of case study areas based on differing socio-economic characteristics (such as the existence of previous schemes) or observed particularities of outcomes (such as a high/low in incidence of foodborne illnesses). This type of purposive selection can either use a strategy that aims to maximise variation across a small number of selected contextual factors or one that intends to study unusual cases that vary considerably in either contextual factors or outcomes.

The construction of a panel dataset as proposed for the impact evaluation would greatly facilitate this selection process.

How many perspectives?

Two key characteristics of a case study approach that affect reliability are maintaining consistency in method across all case study areas and ensuring that an initiative is studied from a sufficient number of perspectives. However, case study researchers have noted that it is not possible to understand everything in a comprehensive manner (Becker, 1970; Snow and Anderson, 1991), and the key is to ensure that case studies are selective, focusing on specific issues and processes that are fundamental to understanding the initiative.

Although a number of methods of gathering case study data from multi-perspectives were outlined above, it may be possible to focus a process study solely on the perspectives of consumers as the key target group to understand and assess decision making and behaviour change as a result of the scheme. Or, a more complex case study might explore not just the perspectives of specific target groups but also how they interact with each other.

Often decisions regarding the number of case study sites (the scope) and the number of perspectives (the depth) from which an initiative is studied depend on the overall aims and objectives of the process study as defined by commissioners as well as the cost and time frame within which the study is to be conducted. The key issue for evaluators is to propose a scope and depth that maximise what can be learned in the time available for the study.

Selecting research participants from the key target groups

In our opinion, multi-perspective case studies that are developed for the national scheme process study should consult individuals from all three target groups. The selection process would have to be targeted taking into account factors that are most relevant for the scope and aims of the study. This type

of targeted selection is also referred to as purposive sampling. Purposive sampling is often used for process studies to select participants that evaluators consider to be the most appropriate for the study.

To capture the local authority perspective, the selection of stakeholders would be the most straightforward, involving consultation with staff that have responsibility for hygiene standards, such as food hygiene managers and inspectors. We suggest that 3 in-depth interviews per case study area would be sufficient.

In comparison to the selection of local authority staff the process for selecting food businesses is slightly more complex. The full range of food businesses in a case study area would have to be considered and the availability of data on the number and types of food businesses in an area would be useful in guiding the selection. A simpler targeted approach would be select food businesses based only on one or two relevant factors. For example, food businesses could be invited to participate in the research based solely on their hygiene ratings. Selection criteria might also consider selecting food businesses that have improved their food safety and hygiene ratings over a specified period of time. The aim would be to select a small number of food businesses that would allow evaluators to capture a range of opinions. The best method in our opinion would be to conduct in-depth interviews with around 7 food businesses in each area.

Identifying consumers is more complicated as there is a wide range of contextual factors to consider. Information about the socio-economic characteristics of local area residents and about consumer 'types' would be useful in determining a sampling strategy. Bearing in mind that a process study sample is not intended to be representative but aims to gather a small number of detailed opinions, a sampling strategy for consumers should utilise a targeted approach. For example, one approach would be to invite consumers who use the food businesses that have agreed to participate in the study.

Focus groups discussions would be the best method to gain insight into the attitudes and food hygiene related behaviour of consumers, to explore their understanding of food hygiene, the national scheme and their food business choices. Focus groups provide a dynamic space where differing opinion can be expressed and where discussion can reveal consumers' opinions on different aspects of the scheme. We would suggest conducting two focus groups in each area with each group comprising 8-10 participants.

These suggestions are one approach to consulting multiple target groups and potential evaluators would need to carefully consider the best balance between contextual factors (again the construction of a panel data set would be useful in informing this process), the scope and budget of the study and practical fieldwork considerations to develop a sampling strategy which provides the best fit for the specified aims of the study.

The timing of the study

Other factors that can influence evaluation findings of a programme are the timing and duration of the evaluation. Determining the ideal point at which evaluators can conduct fieldwork depends on the nature and length of the intervention. In the case of the national scheme, we would suggest that a process study should begin at least six months after the initiative has been implemented in a local authority. As it takes time for a programme to be implemented and start functioning as expected, a study at an earlier point in time might reveal more about the early process of implementation rather than the consistent delivery of the initiative.

The duration of a process study is often determined by commissioners reporting deadlines. It can be that the length of time for reporting on evaluation findings may be so restricted as to impact on the methodological approach, which for case study approaches means restricting the number of perspectives from which a case is studied. Planning of the research commissioning process by working backwards from ultimate policy deadlines and ensuring a realistic timescale for the research is an important first step in ensuring the effectiveness of a process study.

Taking the process study forward

Using the case study approach outlined above, and with the knowledge that local authorities that implement the scheme will be either 'early adopters' or part of the national roll-out, our proposal would be to conduct a two stage process study. Unlike the impact evaluation discussed previously, the process study design does make a distinction between early adopters and others as part of the design for evaluating the national scheme.

The first stage would be a study of early adopters of the national scheme and would commence once the scheme has been in place for at least six months. The early adopters selected for the pilot study could include local authorities that have had another scheme in place or selection could be restricted to those local authorities that have not had any other rating scheme in place prior to the national scheme. This study of early adopters would provide detailed information on the implementation and delivery of the national scheme across a number of case study sites (4-5 sites would be sufficient). The findings of such a pilot study would elaborate on the strengths and weaknesses of the initiative and would be a useful way to highlight best practice. This information could then be used by the FSA to inform and support the national roll-out of the initiative.

The second stage of the process study would take place six months after the national roll-out of the scheme.¹⁵ Stage 2 would replicate the case study approach and the range of methods used in Stage 1. The case study sites should comprise both a small number of early adopters and national roll-out local authorities. We would suggest a sample of 7 local authorities, 3 of which would be early adopters. The Stage 2 study of the national roll-out of local authorities would provide detail on the delivery of the scheme and could incorporate an investigation of promotion of best practice and lessons learned from the Stage 1 study. Incorporating a longitudinal element by selecting some early adopters from Stage 1 local authorities would provide depth of detail of the continuous delivery over time of the scheme and provide details of how well the initiative is becoming embedded within local authority contexts.

At its core the national scheme is about how the different groups the scheme is aiming to influence interact with each other to bring about behaviour change. For this reason we would suggest that case studies for the national scheme process study should be multi-perspective incorporating as many 'voices' as possible. Including the perspectives of local authority food hygiene managers and inspectors, food businesses as well as consumers would strengthen the robustness of each area study and consistency in the methods across all case study areas would lend added weight to the findings and recommendations of the evaluation. The selection of food businesses and consumers for each case study area will require careful consideration of the contextual factors and the use of a range of targeted sampling techniques may be needed to capture relevant 'voices'.

¹⁵ The FSA officially launched the scheme on 30 November 2010 and the roll-out of the scheme is a gradual, fluid process. It is now envisaged that 'Stage 2' of the process evaluation will take place about 18 months – 2 years after the official launch.

Concluding comments: Is it possible to meaningfully evaluate the national Food Hygiene Rating Scheme?

As is probably evident from the preceding discussion, achieving a rigorous and convincing evaluation of the national scheme (and its predecessors - existing food hygiene rating schemes) is a complex task accompanied by much uncertainty. Broadly speaking, however, we would conclude that the schemes are amenable to rigorous and scientifically credible evaluation. Having said, this, such a statement is unfortunately subject to qualification. A rigorous evaluation design does not guarantee that schemes will be found to be effective – a modest intervention may not produce effects that can be identified statistically. One of the key sets of qualifying criteria that will determine whether existing food hygiene rating schemes (and by extension the national scheme) can be meaningfully evaluated is the availability and quality of data, and we have stressed this throughout. It is critical to bear in mind that the Agency's ability to conduct a rigorous assessment of the effects of schemes is data dependent. There is still much uncertainty regarding the types and quality of data that are and will be available, and therefore whether it will prove possible to undertake an impact evaluation of the type discussed here. Other than in the way hygiene scores are presented to consumers, there is also some uncertainty about how existing schemes have actually been implemented in practice and how implementation varies from authority to authority.

From the perspective of the impact evaluation three questions will need to be addressed reasonably promptly. These are: can we convincingly measure the incidence of foodborne illnesses at the level of the local authority or are there reasonable prospects of being able to do so?¹⁶ Are there reliable measures of standards of food hygiene amongst businesses at the level of the local authority? And, can we obtain accurate information about which local authorities have existing schemes, and importantly, when precisely schemes were introduced? As has been emphasised, data needs to be available historically extending backward in time, in a consistent manner, to periods prior to the introduction of schemes. These data will also need to, ideally, be available consistently into the future. If the historic pre-scheme data are unavailable then a DiD approach will not be possible. Some primary data collection among local authorities may be required in order to supplement data already held by the Authority. It is important to note that if outcomes measures are particularly susceptible to measurement error and other sources of variance then this may frustrate our attempts to identify programme impacts which are statistically significant at conventional levels. In other words, it may prove impossible to separate out the policy 'signal' from the

¹⁶ There may be work-arounds which can be explored – for example using aggregation at geographies other than local authority but this would require complex mapping, conversion and manipulation of data over spatial units.

background 'noise'. This issue is of appreciable concern if scheme impacts are expected to be modest.

One further point to make regarding the impact evaluation is that the introduction of existing schemes will need to have taken place over a reasonably short period of time in order for baselines (or before measures) to be clearly identified as well as follow-up or after measures to be defined. If the period over which schemes are introduced is too elongated, there will also be asymmetries in the time period for which post-scheme outcomes might materialise among authorities. Even where such a problem of elongation is found, it may still be possible to estimate impacts on a subset of local authorities (such as early-adopters of existing schemes) but such an approach will require careful thought.

As far as the process study is concerned, the availability of data is also of importance. The data set which is required for the impact evaluation could be used to map local authorities for the purpose of selecting case study sites. Moreover, a data set which does not entirely meet the demands of the impact study will still be of use to the process study.

Generally speaking there are fewer barriers to successfully conducting a process study. The Agency should consider commissioning a process study regardless of whether an impact study appears achievable. There will be much to learn, particularly lessons around successful implementation as well as what it takes to deliver a scheme which is accepted by businesses, local authorities and consumers alike.

References

- Allison, P. D. (1994) Using panel data to estimate the effects of events. *Sociological Methods and Research*, 23, 174-99.
- Anderson, A. (2007) *The Community Builder's Approach to Theory of Change: A practical guide to theory development*. N.Y., N.Y.: The Aspen Institute.
- Becker, H. (1970) *Sociological Work: Method and Substance*. Chicago: Aldine.
- Bertrand, M., Duflo, E. and Mullainathan, S. (2004) How much should we trust differences-in-differences. *The Quarterly Journal of Economics*, February, 249-275
- Blamey, A. and Mackenzie, M. (2007) Theories of Change and Realistic Evaluation: Peas in a Pod or Apples and Oranges. *Evaluation* 13 (4), 439-455.
- Blundell, R. and Costa Dias, M. (2009) Alternative Approaches to Evaluation in Empirical Microeconomics. *The Journal of Human Resources*, 44(3), 565-640.
- Flyberg, B. (2006) Five misunderstandings about case study research. *Qualitative Inquiry* 12(2), 219-245.
- Heckman, J. H., Ichimura, H. and Todd, P. (1997) Matching as an econometric evaluation estimator. *Review of Economic Studies*, 65, 261-294.
- Holland, P. (1986) Statistics and causal inferences. *Journal of the American Statistical Association*, 81 (396), 945-960.
- Jin, G. Z. and Leslie, P. (2003) The effect of information on product quality: Evidence from restaurant hygiene grade cards. *The Quarterly Journal of Economics*, 118(2), 409-451.
- The Kellogg Foundation (2004) *Logic Model Development Guide: Using logic models to bring together planning, evaluation and action*. Battle Creek, MI: The Kellogg Foundation.
- Machin, S. J. and Marie, O. (2005) *Crime and police resources: The Street Crime initiative*. CERP Discussion Paper 5390, London School of Economics – Centre for Economic Performance
- Oakes, J. M. and Feldman, H. A. (2001) Statistical Power for Nonequivalent Pretest-Posttest Designs: The Impact of Change-Score versus ANCOVA models. *Evaluation Review*, 21(3), 3-28

Pawson, R. Greenhalgh, T, Harvey, G. and Walshe, K. (2004) *Realist Synthesis: an introduction*. Paper 2, ESRC Research Methods Programme, University of Manchester.

Pawson, R and Tilley, N. (2004) *Realistic Evaluation*. Retrieved from http://www.communitymatters.com.au/RE_chapter.pdf

Rossi, P. H., Lipsey, M. W. and Freeman, H. E. (2004) *Evaluation: A systematic approach*, (Seventh edition). Thousand Oakes: Sage Publications

Sabates, R. and Feinstein, L. (2007) Effects of government initiatives on youth crime. *Oxford Economic Papers*, 60, 462–483.

Shadish, W. R., Cook T. D. and Campbell D. T. (2002) *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston: Houghton-Mifflin.

Sjoberg, G., Williams, N., Vaughan, T. R. and Sjoberg, A. F. (1991) The Case Study Approach in Social Research: Basic Methodological Issues. In J.R. Feagin, A.M. Orum and G. Sjoberg (Eds.), *A case for the case study* (pp. 27-79). Chapel Hill, N.C.: The UNC Press.

Snow and Anderson (1991) Researching the homeless: the characteristic features and virtues of the case study. In J.R. Feagin, A.M. Orum and G. Sjoberg (Eds.), *A case for the case study* (pp. 148-173). Chapel Hill, NC: The UNC Press.

Stake, R. E. (1994) Case Studies. In N. K. Denzin and Y.S. (Eds.) *Handbook of Qualitative Research* (pp. 236-248). Thousand Oaks, CA.: Sage.

Sullivan, H. (2004) *Developing a 'Theory of Change' to evaluate local public service agreements*. London: DCLG.

Tellis, W. (1997) Introduction to Case Study. *The Qualitative report*, 3(2), retrieved from www.nova.edu/ssss/QR/QR3-2/tellis1.html

Tellis, W. (1997) Application of a case study methodology. *The Qualitative report* 3(3) retrieved from www.nova.edu/ssss/QR/QR3-3/tellis2.html

Thompson, S., de Burger, R., and Kadri, O. (2005) The Toronto food inspection and disclosure system: a case study. *British Food Journal*, 107(3), 136-45.

Weiss, C. H. (1998) *Evaluation*, 2nd Edition), New Jersey: Prentice Hall

Worsfold, D. (2005) Promoting excellence in food hygiene. *Nutrition & Food Science*, 35(1), 6-14.

Yin R. K. (2009) *Case Study Research: Design and Methods*, 4th Edition. Newbury Park, CA: Sage