KANTAR

Food Standards Agency

Food allergen communication in businesses feasibility trial

Developed for the **Food Standards Agency** by **Kantar's Behavioural Practice**

Food Allergen Communication In-business Feasibility Trial:

Authors: McPhedran, R; Patel, K; Toombs, B; Menon, P; Patel, M; Disson, J; Porter, K; John, A; Rayner A

Abstract

Background:

Clear allergen communication in food business operators (FBOs) has been shown to have a positive impact on customers' perceptions of businesses (Barnett et al., 2013). However, the precise size and nature of this effect is not known: there is a paucity of quantitative evidence in this area, particularly in the form of randomised controlled trials (RCTs).

The Food Standards Agency (FSA), in collaboration with Kantar's Behavioural Practice, conducted a feasibility trial to investigate whether a randomised cluster trial – involving the proactive communication of allergen information at the point of sale in FBOs – is feasible in the United Kingdom (UK).

Objectives:

The trial sought to establish: ease of recruitments of businesses into trials; customer response rates for in-store outcome surveys; fidelity of intervention delivery by FBO staff; sensitivity of outcome survey measures to change; and appropriateness of the chosen analytical approach.

Method:

Following a recruitment phase – in which one of fourteen multinational FBOs was successfully recruited – the execution of the feasibility trial involved a quasi-randomised matched-pairs clustered experiment. Each of the FBO's ten participating branches underwent pair-wise matching, with similarity of branches judged according to four criteria: Food Hygiene Rating Scheme (FHRS) score, average weekly footfall, number of staff and customer satisfaction rating. The allocation ratio for this trial was 1:1: one branch in each pair was assigned to the treatment group by a representative from the FBO, while the other continued to operate in accordance with their standard operating procedure.

As a business-based feasibility trial, customers at participating branches throughout the fieldwork period were automatically enrolled in the trial. The trial was single-blind: customers at treatment branches were not aware that they were receiving an intervention. All customers who visited participating branches throughout the fieldwork period were asked to complete a short in-store survey on a tablet affixed in branches. This survey contained four outcome measures which operationalised customers': perceptions of food safety in the FBO; trust in the FBO; self-reported confidence to ask for allergen information in future visits; and overall satisfaction with their visit.

Results:

Fieldwork was conducted from the 3 – 20 March 2020, with cessation occurring prematurely due to the closure of outlets following the proliferation of COVID-19. n=177 participants took part in the trial across the ten branches; however, response rates (which ranged between 0.1 - 0.8%) were likely also adversely affected by COVID-19.

Intervention fidelity was an issue in this study: while compliance with delivery of the intervention was relatively high in treatment branches (78.9%), erroneous delivery in control branches was also common (46.2%).

Survey data were analysed using random-intercept multilevel linear regression models (due to the nesting of customers within branches). Despite the trial's modest sample size, there was some evidence to suggest that the intervention had a positive effect for those suffering from allergies/intolerances for the 'trust' ($\beta = 1.288$, $p<0.01$) and 'satisfaction' ($\beta = 0.945$, $p<0.01$) outcome variables. Due to singularity within the fitted linear models, hierarchical Bayes models were used to corroborate the size of these interactions.

Conclusions:

The results of this trial suggest that a fully powered clustered RCT would likely be feasible in the UK. In this case, the primary challenge in the execution of the trial was the recruitment of FBOs: despite high levels of initial interest from four chains, only one took part. However, it is likely that the proliferation of COVID-19 adversely impacted chain participation – two other FBOs withdrew during branch eligibility assessment and selection, citing COVID-19 as a barrier. COVID-19 also likely lowered the on-site survey response rate: a significant negative Pearson correlation was observed between daily survey completions and COVID-19 cases in the UK, highlighting a likely relationship between the two.

Limitations:

The trial was quasi-random: selection of branches, pair matching and allocation to treatment/control groups were not systematically conducted. These processes were undertaken by a representative from the FBO's Safety and Quality Assurance team (with oversight from Kantar representatives on pair matching), as a result of the chain's internal operational restrictions.

Introduction

Background

Food intolerances and allergies are negative reactions from one's immune system to proteins found in different kinds of food (NHS, 2019). Approximately two million people in the UK have clinically diagnosed food allergies (Wearne, 2017), while the proportion that experience intolerances remains unclear. The severity of food allergies varies considerably; as a result, there are relatively few fatalities directly attributable to food-related anaphylaxis per annum (Wearne, 2017; BBC, 2019). However, anaphylactic hospitalisation is increasing: according to NHS data, the number of finished admission episodes with a primary diagnosis of food allergies have increased by several hundred cases each year since 2013.

In the UK, FBOs – encompassing restaurants, cafes, takeaways and businesses that produce, manufacture or pre-pack food – must inform customers if their products contain any of 14 key allergens (FSA, 2020). The mode of this communication differs significantly across FBOs: while some proactively ask about customers' allergies at the point of sale, others communicate allergen information in writing (Soon, 2018). Among FBOs who do not prominently provide allergen information, high direct costs and ramifications in terms of service efficiency and customer confidence are viewed as key barriers (Smeaton, 2013).

While customers themselves are generally aware of their own food allergies, some are reticent to proactively seek information on allergens. Research suggests that this hesitance may stem from two factors. Firstly, precautionary messaging and broad statements about allergens at FBOs – for example, 'may contain', 'ask staff' or 'cannot guarantee allergen free' – inherently suggests that the business may lack reliable information; and, to some, can also imply that asking is not a social norm (Barnett et al., 2017). Further, customers may mentally discount the risks associated with consumption of allergens to avoid feelings of social discomfort, an effect known as 'courtesy bias' (Barnett et al., 2019).

With these barriers in mind, behavioural interventions where allergen information is provided to consumers as a default are likely to be effective in facilitating information transfer. For FBOs, such interventions may have additional benefits, including increases in consumer confidence and likelihood to patronise (Barnett et al., 2017).

FSA commissioned Kantar's Behavioural Practice to carry out a feasibility trial to examine whether shifting the responsibility of allergen communication from customers to FBO staff – by proactively asking all customers if they have any food allergies/intolerances, or want any information about allergens before a purchase is made – augments consumers' confidence in food safety, trust in the business, confidence to ask about allergens in food products and customer satisfaction.

The results of this feasibility trial will be used to inform a future, fully powered RCT. The primary research questions in this RCT will include:

1) To investigate whether customers who have any food allergies/intolerances harbour different perceptions of food safety regarding food and drink sold in FBOs, or have differential levels of trust in FBOs.
2) To investigate whether the intervention[1] positively impacts the perceptions of consumers who experience allergies/intolerances, including their: perceptions of food safety regarding food and drink sold in FBOs; trust in FBOs; self-reported confidence in asking about allergens in future; and satisfaction with FBOs.
3) To investigate whether there is a halo effect associated with the intervention. That is, whether the intervention positively impacts the views of all customers, including their: perceptions of food safety regarding food and drink sold in FBOs; trust in FBOs; self-reported confidence in asking about allergens in future; and satisfaction with FBOs.

Objectives

While there is qualitative research relating to FBOs' allergen communication (see Barnett et al., 2017; Barnett et al., 2019), there is a relative scarcity of quantitative evidence. To date, there has not been an RCT relating to provision of allergen information conducted in the UK.

Therefore, the aim of this feasibility trial was to investigate the practicability of a fully powered clustered RCT in the UK. Specifically, the trial aimed to examine the feasibility of the methodological approach, focusing particularly upon five aspects.

[1] proactively asking customers if they have any food allergies/intolerances - or want any information about allergens - before a purchase is made

1) **Objective 1**, ease of recruitment of businesses: are large FBOs amenable to participation in an RCT?
2) **Objective 2**, FBO customer engagement: are FBO customers willing to complete an in-store survey following their order?
3) **Objective 3**, intervention fidelity: is the experiment's intervention correctly delivered by staff in treatment branches, and not delivered by staff in control branches?
4) **Objective 4**, appropriateness of the survey outcome measures: are the outcome survey measures sensitive enough to detect treatment effects in an RCT?
5) **Objective 5**, appropriateness of the analytical approach: is the use of hierarchical models in analysis – to detect treatment effects – appropriate?

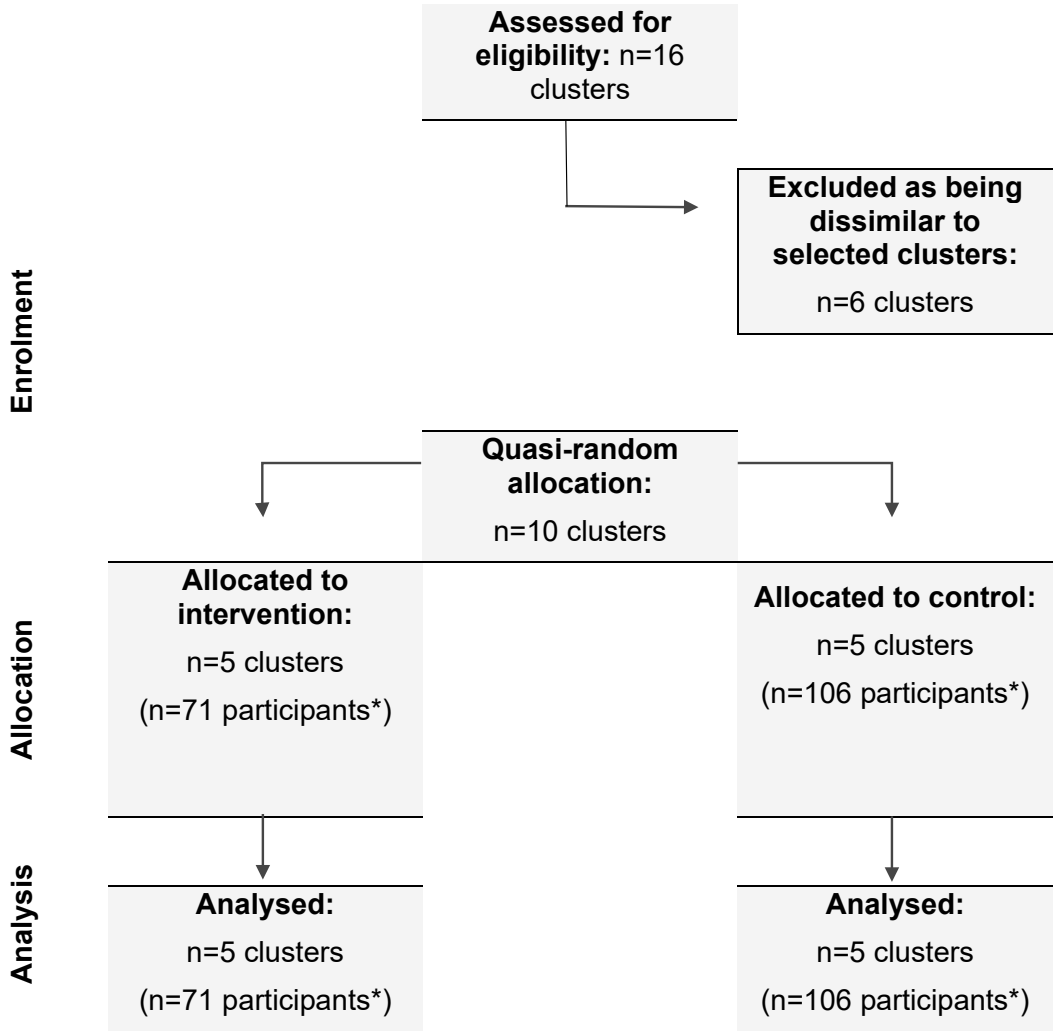
Methods

Trial design

The trial was conducted on-site at ten branches of a multinational FBO. It involved a matched pairs cluster quasi-randomised experiment, with five pairs of branches selected for participation.

The trial design can be seen below. As contained in Figure 1, six additional branches were considered for inclusion but were ultimately excluded from the trial due to relative dissimilarity with other participating branches (including differential FHRS scores and weekly customer footfall).

Figure 1: Overview of trial design



*** Participants refer to customers who successfully completed the survey; partial responses were not included in analysis**

Branch selection was conducted by a representative from the FBO's Safety and Quality Assurance team in consultation with Kantar, due to the chain's operational and data privacy restrictions.

Despite the shortcomings associated with clustered RCTs – including lower statistical power (Cornfield, 1978) – this design was chosen for two reasons.

1) A cluster design was thought to attenuate risk of contamination: that is, 'control' customers would be less likely to be exposed to the intervention in a cluster trial than in an RCT in which randomisation would occur at customer level (that is, delivery of the intervention to every N^{th} customer).
2) Randomisation at customer level was thought to be operationally challenging, increasing the burden upon participating FBOs.

As for branch selection, pair matching was conducted by the FBO representative and overseen by the Kantar team. Four key criteria used to adjudge branch similarity: FSA FHRS score, weekly customer footfall, number of staff and customer satisfaction score

(CSS). The matched pairs can be seen below in Table 1 (branch locations have been redacted to preserve the FBO's anonymity).

Table 1a: The feasibility trial's matched pair 1

Criteria	Treatment	Control
Branch	B1	B6
Mean weekly footfall	2,025	2,023
Mean CSS score	54.4	56.5
FHRS rating	5	5
Staff (n)	27	37

Table 2b: The feasibility trial's matched pair 2

Criteria	Treatment	Control
Branch	B2	B7
Mean weekly footfall	2,718	2,674
Mean CSS score	62.3	62.3
FHRS rating	5	5
Staff (n)	50	42

Table 3c: The feasibility trial's matched pair 3

Criteria	Treatment	Control
Branch	B3	B8
Mean weekly footfall	2,104	2,107
Mean CSS score	57.3	56.9
FHRS rating	5	5
Staff (n)	333	36

Table 4d: The feasibility trial's matched pair 4

Criteria	Treatment	Control
Branch	B4	B9
Mean weekly footfall	2,314	2,305
Mean CSS score	63.0	62.0
FHRS rating	5	5
Staff (n)	31	38

Table 5e: The feasibility trial's matched pair 5

Criteria	Treatment	Control
Branch	B5	B10
Mean weekly footfall	2,482	2,510
Mean CSS score	62.1	59.8
FHRS rating	5	5
Staff (n)	38	43

Pair-matching has been shown to increase efficiency of sample size by reducing variance between clusters (Rutterford et al., 2015).

The allocation ratio was 1:1; one branch per pair was assigned to the treatment group, while the other continued operating in accordance with their standard operating practice. As for selection and pair-matching, the FBO representative performed this allocation.

Participants

Businesses

Prior to recruitment, FBOs (that is, potential participating organisations) were screened for eligibility. Eligibility was based on three criteria:

1) No proactive provision of information on allergens by staff at the point-of-sale.
2) Sufficient branches (more than N=50) to enable the selection of n=10 branches (five pairs) and pair matching.
3) Sufficient footfall (approximately N=200 per day or more, on average) to ensure a sample size with adequate statistical power to detect treatment effects.

Invitations to participate in the feasibility trial were emailed to fourteen FBOs who met these criteria, and spanned different business categories (dine-in, take-away, coffee shops). A trial information sheet was disseminated following the initial email communication; subsequent teleconferences were conducted with 11 of the 14 FBOs to provide them with more detailed information about the trial.

For chains who maintained interest following these teleconferences, short bespoke PowerPoint decks (highlighting the benefits of participation in the trial) were prepared and distributed via email.

Customers

The trial ran continuously throughout business hours over the fieldwork period (3rd – 20th March 2020). Fieldwork was scheduled for a period of two months; however, it was truncated due to the proliferation of COVID-19 and the consequent UK Government directives regarding non-essential business operation (which resulted in the temporary closure of the FBO).

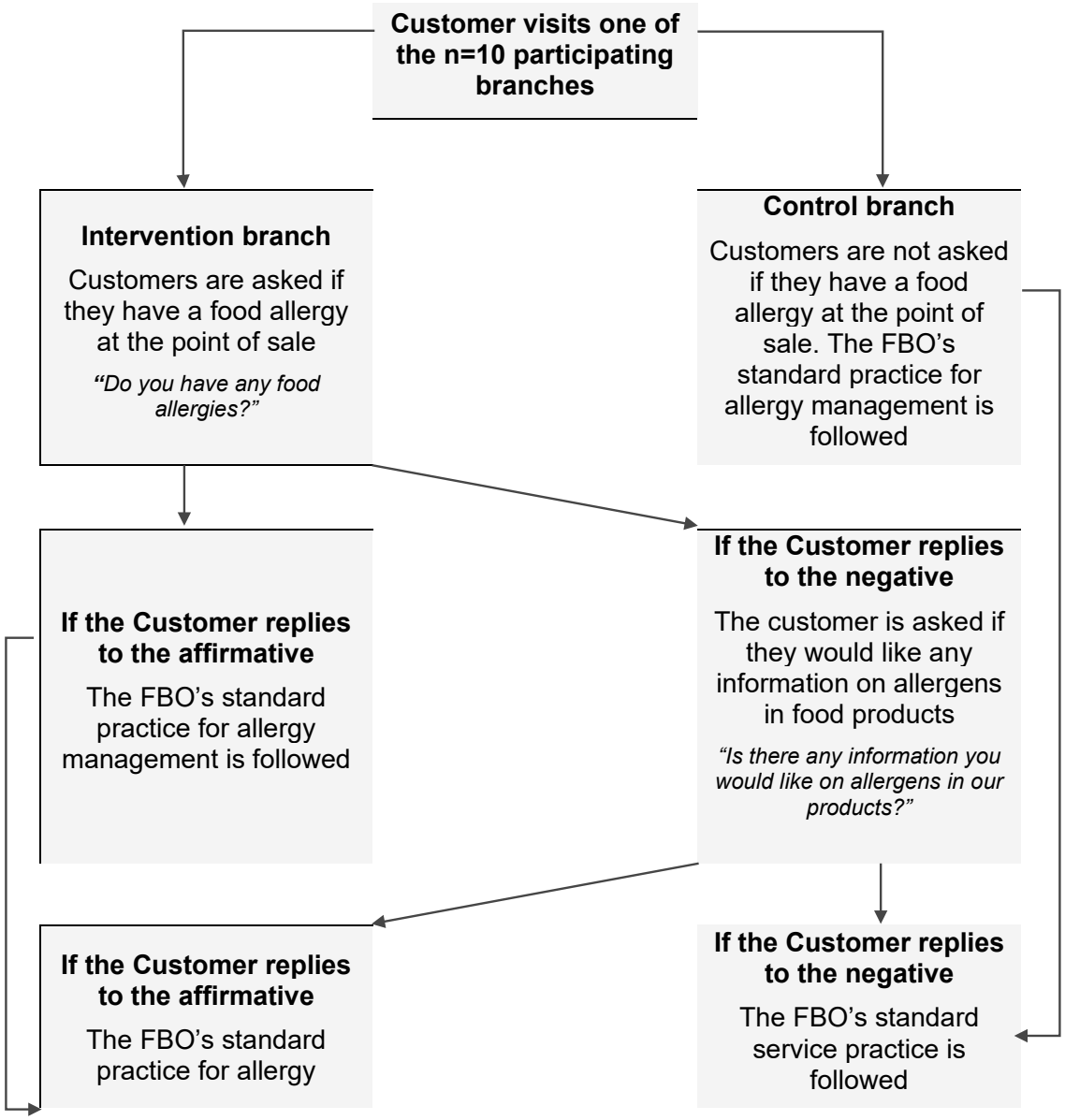
Throughout fieldwork, all customers visiting FBO branches were eligible for the trial and were therefore automatically enrolled as participants. Nested within the total population of customers was the primary population of interest: customers who indicated that they have food allergies or food intolerances.

No information that could be used to personally identify participants was collected.

Procedure

In treatment branches, the intervention was delivered by staff. The process for delivery is outlined in the flowchart below.

Figure 2: Flow diagram illustrating intervention delivery in the feasibility trial



In order to reduce the potential for incomplete surveys - which could have prevented the participation of subsequent customers - a time-out function was embedded in the survey script. If a question was left incomplete for a period of 30 seconds, the survey was re-routed to the home page.

All customers were able to withdraw from the trial at any point (either by not starting or completing the survey).

Outcomes

To establish the feasibility of a fully powered clustered RCT, the trial objectives were measured using one or more discrete outcomes, as follows.

1) **Objective 1**, ease of recruitment of businesses: are large FBOs amenable to participation in an RCT?

   Outcome measure

   Result of recruitment, in terms of number of chains who participated in the trial, number of chains who initially indicated interest and number of chains who refused to participate.

2) **Objective 2**, FBO customer engagement: are FBO customers willing to complete an in-store survey following their order?

   Outcome measure

   Completed survey response rate = number of survey completions divided by the number of eligible participants

3) **Objective 3**, intervention fidelity: is the experiment's intervention correctly delivered by staff in treatment branches, and not delivered by staff in control branches?

   Outcome measure

   Consistency of delivery of the intervention = number of survey participants in treatment branches who indicated that they were asked if they have an allergy or intolerance (Q4[2]) divided by total number of survey participants in treatment branches

4) **Objective 4**, appropriateness of the survey outcome measures: are the outcome survey measures sensitive enough to detect a treatment effect in an RCT?

   Outcome measure

   Detection of significant differences between participant groups – particularly treatment effects – using multilevel linear regression models (with branch specified as a random variable)

5) **Objective 5**, appropriateness of the analytical approach: is the use of hierarchical models in analysis – to detect treatment effects – appropriate?

   Outcome measure

   Defining unknown parameters that could impact the trial's statistical power, such as the intra-cluster correlation coefficients for each outcome measure (ρ)

Objectives 2-5 all relate to the in-store customer survey, which was the instrument used to measure treatment effects. Survey measures that map onto the trial research questions are outlined in Table 2.

[2] Thinking back to when you were served, did the [INSERT CHAIN NAME] employee ask you whether you have a food allergy or intolerance before you made your purchase?

Table 6: Research questions and their operationalisation in the outcome survey

Research question	Construct	Survey measure	Survey scale
Does asking customers if they suffer from any food allergies/intolerances – or want any information about allergens – before a purchase is made positively impact consumers' perceptions of food safety?	Perception of safety of food sold in the FBO	How concerned are you about the safety of the food that is sold in **[INSERT CHAIN NAME]** for consumption by those with food allergies and intolerances?	• Very unconcerned • Fairly unconcerned • Neither concerned nor unconcerned • Fairly concerned • Very concerned
Does asking customers if they suffer from any food allergies/intolerances – or want any information about allergens – before a purchase is made positively impact levels of consumer trust?	Trust in the FBO	And how much do you trust or distrust **[INSERT CHAIN NAME]** as a business responsible for the sale of food and drinks?	• I distrust it a lot • I distrust it • I neither trust nor distrust it • I trust it • I trust it a lot
Does asking customers if they suffer from any food allergies/intolerances – or want any information about allergens – before a purchase is made positively impact confidence to ask about allergen information in future visits?	Confidence to ask about allergens in future visits	How confident would you feel in asking a member of staff for information about the ingredients in the foods they are selling, because of a concern about possible allergens/food intolerances?	• Not at all confident • Not very confident • Neither confident nor unconfident • Somewhat confident • Very confident
Does asking customers if they suffer from any food allergies/intolerances – or want any information about allergens – before a purchase is made positively impact levels of consumer satisfaction?	Satisfaction with visit to the FBO	Thinking of your experience at **[INSERT CHAIN NAME]** today, how satisfied or dissatisfied were you?	• Not at all satisfied • Not very satisfied • Neither satisfied nor dissatisfied • Somewhat satisfied • Very satisfied
All	Allergy status	Have you experienced either of the following adverse reactions after consuming certain foods or drinks?	• A food allergy • A food intolerance • I haven't experienced either of these

The survey outcome measures (aside from 'satisfaction') were based on questions in the Public Attitudes Tracker, FSA's flagship survey. These measures were previously cognitively tested to ensure their construct validity (Bryson and Purdon, 2010). The 'satisfaction' measure was designed using a Likert scale in line with the 'confidence' measure, and was phrased in line with other validated measures of satisfaction (see Hero et al., 2016; Green et al., 2008).

As outlined in the trial protocol, it was initially intended that survey measure sensitivity would be assessed using standardised response mean. However, in practice, the implementation of the intervention was launched prior to any pre-trial data collection, rendering this measure infeasible. For this reason, this objective was conducted using hierarchical linear mixed regression models (the selected form of analysis).

Sample size

The target sample size for fieldwork was approximately n=600 responses per cluster (branch). This target was predicated on the assumption of a two-month trial, with a 2-5% response rate across thirty branches (ten branches each from three chains), and an intra-cluster correlation coefficient of $\rho = 0.03$ (in line with those observed in other human studies, see Killip et al., 2004).

Under these assumptions, the feasibility trial would have had sufficient statistical power ($1-\beta = 0.84$ and 0.96) to detect treatment effects of 0.2 and 0.25 respectively (consistent with other studies evaluating the effects of the provision of information regarding food and beverages, see: Raats et al., 2015; Vasiljevic et al., 2018).

Randomisation

It was intended that randomisation – allocation of one branch within each pair to the treatment group – would be executed by Kantar via stratified random assignment, conducted using R's *randomizr* package. However, due to unavoidable constraints – operational restrictions within the participating FBO – this was not possible.

As such, allocation was conducted by a representative from the FBO's Safety and Quality Assurance team. Allocation was quasi-random, as it involved a degree of self-selection: participation was dependent upon store managers' willingness to participate in the trial and alter their internal operations to deliver the intervention.

On this basis, the trial was single-blind: the researchers, store managers and store employees were aware of the trial design and intervention delivery; however, the participants (customers) were not.

Statistical methods

Moerbeek (2006) recommends the use of hierarchical linear models for analysis of cluster randomised trials.

As such, a series of two-level hierarchical linear models were used in analysis. The specification of the linear mixed models used can be seen below:

$$y_{ij} = \beta_0 + \beta_{10}(\text{allergy/intolerance status})_i + \beta_{20}(\text{intervention delivery})_i + \beta_{30}(\text{age})_i + \beta_{40}(\text{sex})_i + \beta_{01}(\text{allocation to treatment branch})_j + \beta_{50}(\text{allergy/intolerance status})_i * (\text{intervention delivery})_i + \beta_{11}(\text{allergy/intolerance status})_i * (\text{allocation to treatment branch})_j + \beta_{21}(\text{intervention}$$

$$\text{delivery})_i * (\text{allocation to treatment branch})_j + \beta_{51}(\text{allergy/intolerance status})_i * (\text{intervention delivery})_i * (\text{allocation to treatment branch})_j + u_{0j} + \varepsilon_{ij}$$

$$\varepsilon_{ij} \sim N(0, \sigma^2),\ u_{0j} \sim N(0, \tau^2)$$

Where:

- y_{ij} represents the score on the outcome variables (perceptions of food safety, trust, confidence to ask about food allergens in products, and satisfaction) for the i^{th} customer in branch j;
- β_i and β_j represent the unique influence of level 1 (e.g. allergy/intolerance status) and level 2 (e.g. allocation to treatment branch) variables on the outcome variable, respectively;
- u_{0j} is the error term of branch j from the mean outcome in its treatment condition; and
- ε_{ij} is the individual error term, normally and independent identically distributed.

In these models, age and sex were individual-level categorical covariates:

- Age (1 = 16-25; 2 = 26-35; 3 = 36-49; 4 = 50-65; 5 = 66+; 6 = I'd prefer not to say), and
- Sex (1 = male, 0 = female).

Additionally:

- Allergy and intolerance (individual-level characteristics) were conflated in a dummy variable (1 = food allergy or intolerance; 0 = non-allergy population).
- Allocation to treatment branch (a branch-level variable) was a dummy variable that reflected whether a branch was allocated to treatment/control groups (1 = treatment, 0 = control).
- Intervention delivery (an individual-level variable) was a dummy variable that reflected whether the intervention was delivered to a customer (1 = asked; 0 = not asked).

These variables were considered fixed effects in the linear models. Branch was the cluster in the experimental design; it was therefore considered a random effect.

Prior to models being run, outcome variables were recoded using a consistent mapping system, such that '1' represented the most negative option and '5' represented the most positive option. This coding was employed to ensure the results of models could be more easily interpreted.

Linear models were run first using a base model (comprising only the demographic covariates and branch's random intercept), which were then compared to full factorial models. Analysis was run using R statistical software (R Core Team, 2018), with the *lme4* package (Bates, Maechler, Bolker & Walker, 2015) using the *lmer* function. Contrasts were computed using the *linearHypothesis* function. Homogeneity of variances was tested using the *levenesTest* function in R, and results indicated equal variances across all outcome variables for those in treatment and control branches ($p > 0.05$), as well as both intervention categories ($p > 0.05$).

Outputs from analysis indicated singularity in the variance-covariance matrices of the linear models (suggesting overfitting). Therefore, follow-up hierarchical Bayes models were run to corroborate the parameter estimates generated by the linear models, as

recommended by Gelman and Hill (2006). Hierarchical Bayes models were run with the *rstanarm* package (Goodrich, 2020), using the *stan_lmer* function. Bayesian estimates were conducted using Markov Chain Monte Carlo estimates, involving ten chains of 2,000 iterations. Default – weakly informative – prior distributions were specified for the model hyperparameters (Lee et al., 2018).

All reporting is for the population-level coefficients.

Results

Recruitment

With the proliferation of COVID-19 in the UK throughout March, fieldwork was restricted to the 3rd – 20th March 2020. On March 21 2020, fieldwork was suspended due to the regulations from the Secretary of State regarding business closure (Department of Health and Social Care, 2020).

Baseline data

Throughout fieldwork, n=177 participants (Male = 77, Female = 81, undisclosed = 19) were recruited across the ten branches. Number of survey completions varied across branches, ranging from n=4 to n=39 (see Table 3).

The demographic composition of each branch differed significantly according to gender ($\chi(18) = 29.314$, $p = 0.045$); no significant differences were observed for age.

Table 7a: Number of survey completions in treatment branch (gender)

Branch	B1	B2	B3	B4	B5
Number of survey completions	4	9	23	9	26
Male	2	2	17	6	8
Female	2	5	5	2	12
Prefer not to say	0	2	1	1	6

Table 8b: Number of survey completions in control branch (gender)

Branch	B6	B7	B8	B9	B10
Number of survey completions	39	10	5	20	32
Male	14	1	3	7	17
Female	22	7	1	11	14
Prefer not to say	3	2	1	2	1

Table 9c: Number of survey completions in treatment branch (age)

Branch	B1	B2	B3	B4	B5
Number of survey completions	4	9	23	9	26
16-25	1	2	11	3	9
26-35	1	6	5	5	4
36-49	2	1	6	1	7
50-65	0	0	1	0	2
66+	0	0	0	0	1
Prefer not to say	0	0	0	0	3

Table 10d: Number of survey completions in control branch (age)

Branch	B6	B7	B8	B9	B10
Number of survey completions	39	10	5	20	32
16-25	14	3	2	11	9
26-35	13	3	2	2	11
36-49	6	3	1	2	6
50-65	5	1	0	3	3
66+	1	0	0	2	1
Prefer not to say	0	0	0	0	2

Outcomes

Objective 1, ease of recruitment of businesses: are large FBOs amenable to participation in an RCT?

Outcome measure: result of recruitment, in terms of number of chains who participated in the trial, number of chains who initially indicated interest and number of chains who refused to participate.

As mentioned in the Methods Section, 14 national or multinational FBOs were contacted for participation in the trial by telephone, with an accompanying email also sent to them (see Table 4 below; names have been redacted to preserve FBOs' anonymity). Additional details of recruitment follow:

- 11 of the 14 FBOs requested the provision of more detailed information about the trial, with the other three refusing to participate following the first communication.
- A teleconference – in which further information on the trial was provided, and clarifying questions fielded – was held with nine of the 14 FBOs.
- Additional information-sharing teleconferences were held with four chains.

- Three chains gave verbal confirmation of their willingness to participate and proceeded to branch selection.

In early March 2020 two of the FBOs withdrew, citing COVID-19 as an unanticipated and immovable barrier to their participation.

Table 11: Recruitment of FBOs, by stage

Chain	Stage 1 – Follow-up email requested and sent	Stage 2 – Teleconference	Stage 3 – Further engagement	Stage 4 – verbal agreement and branch selection
1	Yes	Yes	Yes	Yes
2	Yes	Yes	Yes	Yes
3	Yes	Yes	Yes	Yes
4	Yes	Yes	Yes	No
5	Yes	Yes	No	No
6	Yes	Yes	No	No
7	Yes	Yes	No	No
8	Yes	Yes	No	No
9	Yes	Yes	No	No
10	Yes	No	No	No
11	Yes	No	No	No
12	No	No	No	No
13	No	No	No	No
14	No	No	No	No

Objective 2, FBO customer engagement: are FBO customers willing to complete an in-store survey following their order?

Outcome measure

Completed survey response rate = number of survey completions divided by the number of eligible participants

n=4 to n=39 surveys were completed across branches; with response rates ranging between 0.1-0.8% due to variable footfall (see Table 3). The overall survey response rate was 0.3%.

Table 12: Survey response rate, per branch

Treatment branches

Branch	B1	B2	B3	B4	B5
Number of transactions	4,535	8,044	5,292	5,763	5,809
Number of survey completions	4	9	23	9	26
Survey response rate	0.1%	0.1%	0.4%	0.2%	0.4%

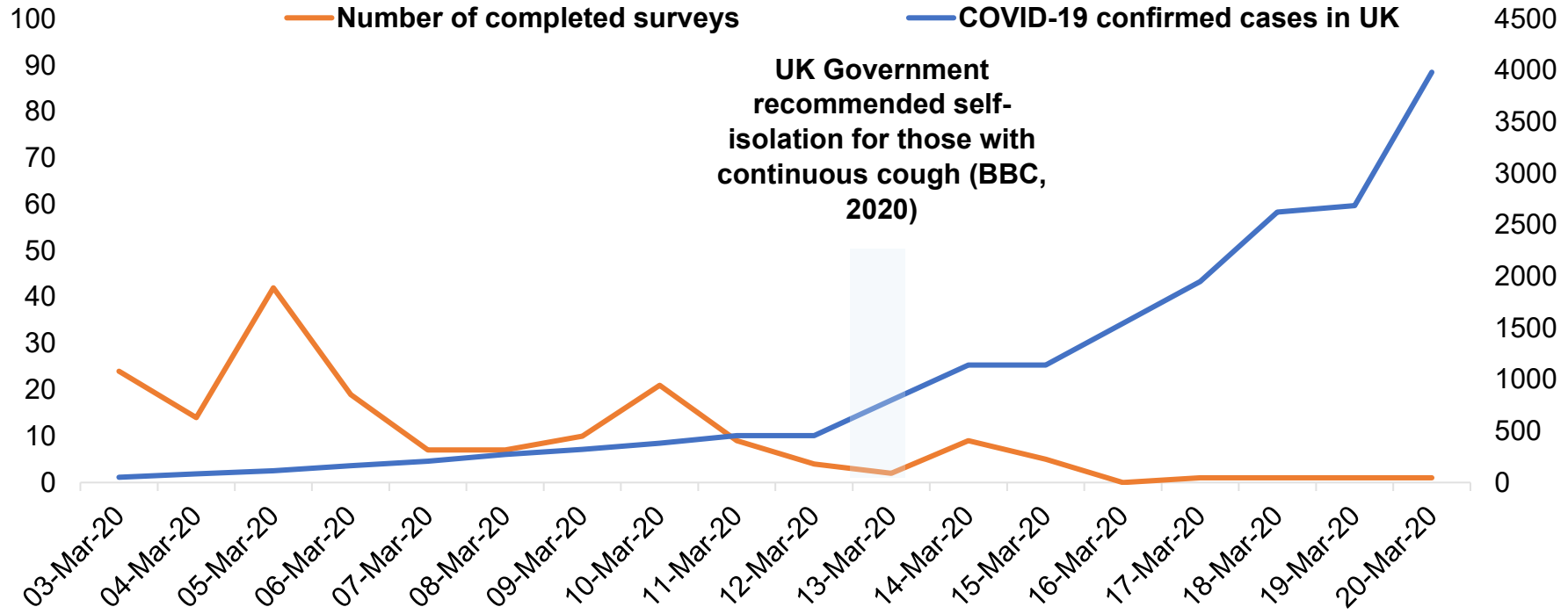
Control branches

Branch	B6	B7	B8	B9	B10
Number of transactions	4,701	5,665	5,025	5,830	5,369
Number of survey completions	39	10	5	20	32
Survey response rate	0.8%	0.2%	0.1%	0.3%	0.6%

While the total and branch-level survey response rates were much lower than projected – anticipated to be a minimum of 2% – they were almost certainly adversely affected by COVID-19, likely due to hesitance to complete the outcome survey via tablet.

Figure 3 below illustrates the number of daily survey completions mapped against the number of diagnosed COVID-19 cases in the UK (Johns Hopkins University, 2020). The negative Pearson correlation between daily survey completions and COVID-19 cases in the UK is highly significant ($r = –0.58$, $p = 0.015$), highlighting a likely relationship between the two.

Figure 3: Number of survey completes throughout fieldwork, mapped against the number of COVID-19 death in UK hospitals (NHS, 2020)



Objective 3, intervention fidelity: is the experiment's intervention correctly delivered by staff in treatment branches, and not delivered by staff in control branches?

Outcome measure

Consistency of delivery of the intervention = number of survey participants in treatment branches who indicated that they were asked if they have an allergy or intolerance (Q4[3]) divided by total number of survey participants in treatment branches.

Intervention fidelity was a clear issue in this feasibility trial. While survey responses indicated that intervention was delivered relatively consistently in the branches allocated to the treatment group (78.9%), it was also often erroneously delivered in control branches (46.2%). This result suggests that, while store managers and their staff were fully briefed (and provided guiding materials) on intervention delivery and survey completion prior to the trial, this approach was not adequate to ensure consistent compliance.

In future RCTs, monitoring of intervention delivery should be more frequent, and additional briefings should be conducted if erroneous delivery remains prevalent.

Table 13: Proportion of consumers who indicated that they were delivered the intervention, treatment vs control branches

Treatment branches

Branch	B1	B2	B3	B4	B5
Intervention on delivery (%)	100	77.8	95.7	66.7	65.4

Control branches

Branch	B6	B7	B8	B9	B10
Intervention on delivery (%)	30.8	40	40	50	65.6

Objective 4, appropriateness of the survey outcome measures: are the outcome survey measures sensitive enough detect treatment effects in an RCT?

Outcome measure: Detection of significant differences between participant groups – particularly treatment effects – using multilevel linear regression models (with branch specified as a random variable)

Two linear models – base and full factorial – were run for each of the four survey outcome variables (customers' perceptions of food safety, 'trust', confidence to ask about food allergens in products, and 'satisfaction'). In addition, as mentioned in the Methods section, hierarchical Bayes models were run to confirm model results.

[3] Thinking back to when you were served, did the **[INSERT CHAIN NAME]** employee ask you whether you have a food allergy or intolerance before you made your purchase?

Results from these models should be interpreted with caution given the trial's relatively small sample size. Future – higher powered – trials will be able to provide more definitive evidence about the impact of the intervention of the outcome measures.

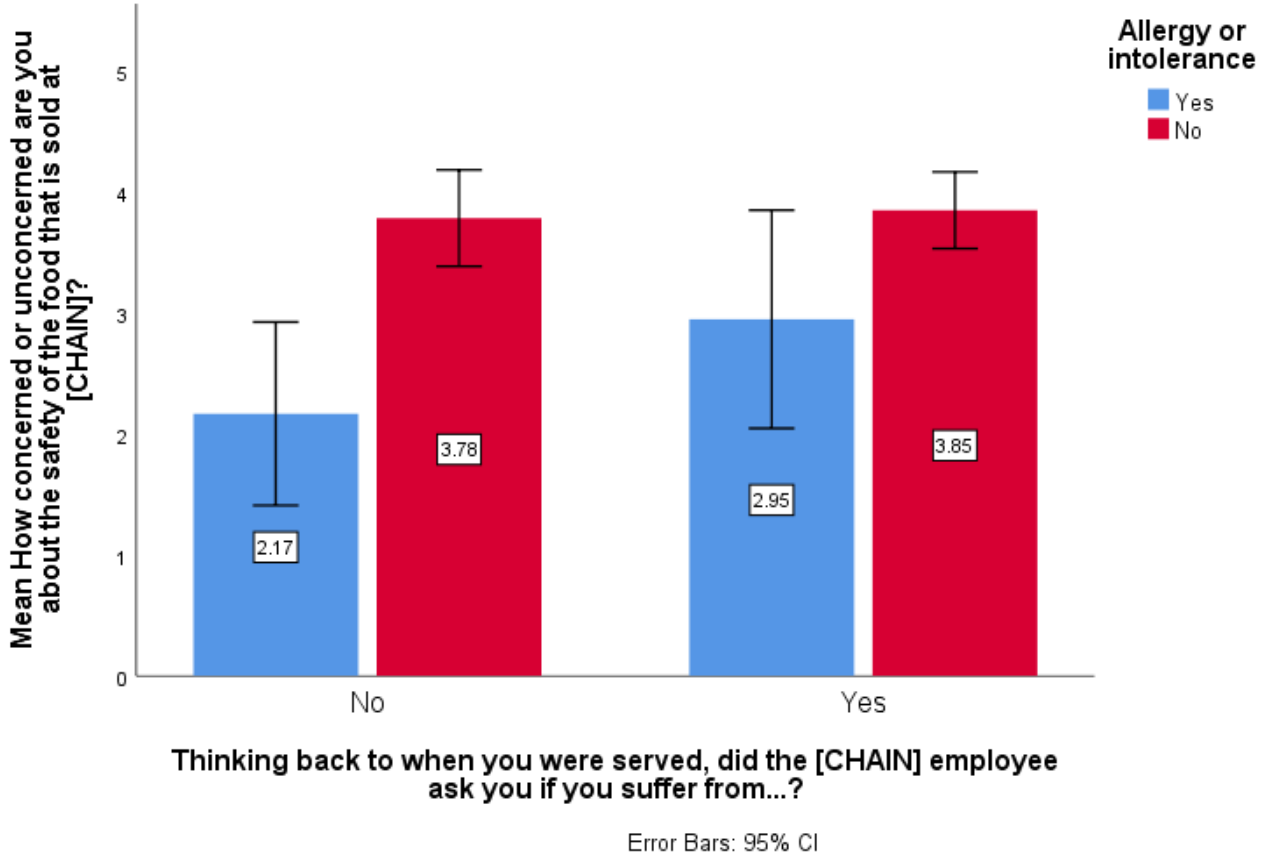
Outputs from these models follow.

1) Customers' perceptions of food safety

There was no evidence of a treatment effect associated with delivery of the intervention for the first outcome variable – customers' perceptions of food safety. This was the case both for those with allergy/intolerances and the total customer population.

The mean 'concern' scores for those with and without allergies/intolerances can be seen below in Figure 4.

Figure 4: Mean concern about food safety in FBO (where 5 is 'very concerned' and 1 is 'very unconcerned'), by intervention delivery and allergy/intolerance status



However, in the full factorial linear model[4] (see Table 7), there was a significant negative coefficient observed for the food allergies and/or intolerances variable (β = -1.493, p = 0.004). This suggests a lower perception of food safety among this subpopulation.

[4] Comparisons between the base model and the full factorial model indicated a significant improvement in fit for the full model ($\chi(7)$ = 23.933, p = 0.001).

° = 95% Bayesian Credible Interval does not contain zero

A similar sized mean coefficient was also observed in the hierarchical Bayes model (μ = -1.515°).

Table 14: Model #1 – Customers' perception of food safety outputs

Base model

Variable	***Coefficient β***	***Std error***
Intercept	4.010***	0.264
Age	-0.126	0.096
Sex	-0.288	0.185
Allocation of treatment	-	-
Intervention delivery	-	-
Allergy/intolerance status	-	-
Allocation to treatment: intervention delivery	-	-
Allocation to treatment: Allergy/intolerance status	-	-
Intervention delivery: Allergy/intolerance status	-	-
Allocation to treatment: Intervention delivery: Allergy/intolerance status	-	-

Full factorial linear model

Variable	***Coefficient β***	***Std error***
Intercept	4.172***	0.321
Age	-0.153	0.097
Sex	-0.352	0.180
Allocation of treatment	0.290	0.470
Intervention delivery	0.392	0.334
Allergy/intolerance status	-1.493**	0.511
Allocation to treatment: intervention delivery	-0.729	0.583
Allocation to treatment: Allergy/intolerance status	-1.404	1.692
Intervention delivery: Allergy/intolerance status	0.358	0.729
Allocation to treatment: Intervention delivery: Allergy/intolerance status	1.584	1.858

Hierarchical Bayes model

Variable	***Coefficient β***	***Std error***
Intercept	4.828°	0.351
Age	-0.157	0.099
Sex	-0.324	0.182
Allocation of treatment	0.281	0.510
Intervention delivery	0.383	0.336
Allergy/intolerance status	-1.515°	0.506
Allocation to treatment: intervention delivery	-0.667	0.578
Allocation to treatment: Allergy/intolerance status	-1.010	1.443
Intervention delivery: Allergy/intolerance status	0.402	0.702
Allocation to treatment: Intervention delivery: Allergy/intolerance status	1.147	1.583

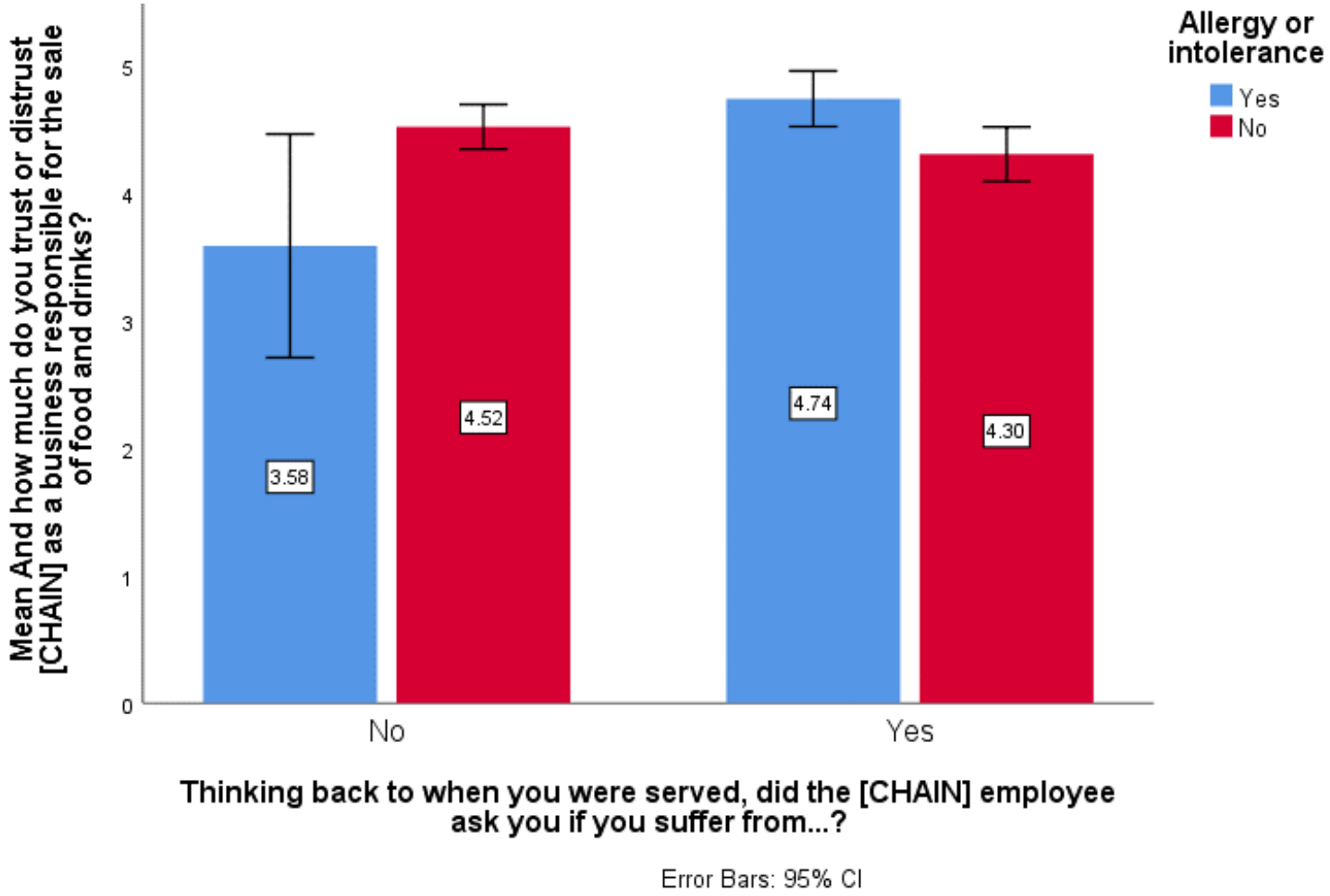
*** = p < 0.001; ** = p < 0.01; * = p < 0.05; ° = 95% Bayesian Credible Interval does not contain zero

2) Customers' trust in business

Positively, there was evidence to suggest that the delivery of the intervention was effective in increasing trust among those with allergies/intolerances.

The mean 'trust' scores for each of these groups can be seen below in Figure 5. As illustrated, 'trust' scores were highest among customers with allergies/intolerances who were delivered the intervention ($\bar{x}$ = 4.74, $\sigma_{\bar{x}}$ = 0.104).

Figure 5: Mean trust in FBO (where 5 is 'I trust it a lot' and 1 is 'I distrust it a lot'), by intervention delivery and allergy/intolerance status



Further, in the full linear model[5], a significant interaction between intervention delivery and allergy/intolerance status was observed (β = 1.288, p = 0.002), again suggesting that the intervention was effective in increasing trust among those with allergies/intolerances.

A significant negative coefficient observed for the food allergies and/or intolerances variable, (β = -0.714, p = 0.011), suggesting lower levels trust among this subpopulation.

These results were corroborated by output of the hierarchical Bayes model, in which similar sized mean coefficients were observed (see Table 8).

Table 15: Model #2 – Customers' trust outputs

Base model

Variable	**Coefficient β**	**Std error**
Intercept	4.797***	0.147
Age	0.183***	0.055
Sex	-0.065	0.104
Allocation of treatment	-	-
Intervention delivery	-	-
Allergy/intolerance status	-	-
Allocation to treatment: intervention delivery	-	-
Allocation to treatment: Allergy/intolerance status	-	-
Intervention delivery: Allergy/intolerance status	-	-
Allocation to treatment: Intervention delivery: Allergy/intolerance status	-	-

[5] Comparisons between the base model and the full factorial model indicated a significant improvement in fit for the full factorial model ($\chi(7)$ = 18.696, p = 0.009).

Full factorial linear model

Variable	Coefficient β	Std error
Intercept	4.838***	0.18
Age	-0.154**	0.055
Sex	-0.075	0.101
Allocation of treatment	0.249	0.263
Intervention delivery	-0.281	0.187
Allergy/intolerance status	-0.714*	0.286
Allocation to treatment: intervention delivery	-0.058	0.327
Allocation to treatment: Allergy/intolerance status	-2.527**	0.948
Intervention delivery: Allergy/intolerance status	1.288**	0.408
Allocation to treatment: Intervention delivery: Allergy/intolerance status	2.145*	1.041

Hierarchical Bayes model

Variable	Coefficient β	Std error
Intercept	4.852°	0.195
Age	-0.164°	0.054
Sex	-0.068	0.103
Allocation of treatment	0.214	0.284
Intervention delivery	-0.301	0.187
Allergy/intolerance status	-0.733°	0.287
Allocation to treatment: intervention delivery	0.016	0.324
Allocation to treatment: Allergy/intolerance status	-1.975°	0.820
Intervention delivery: Allergy/intolerance status	1.328°	0.393
Allocation to treatment: Intervention delivery: Allergy/intolerance status	1.551	0.897

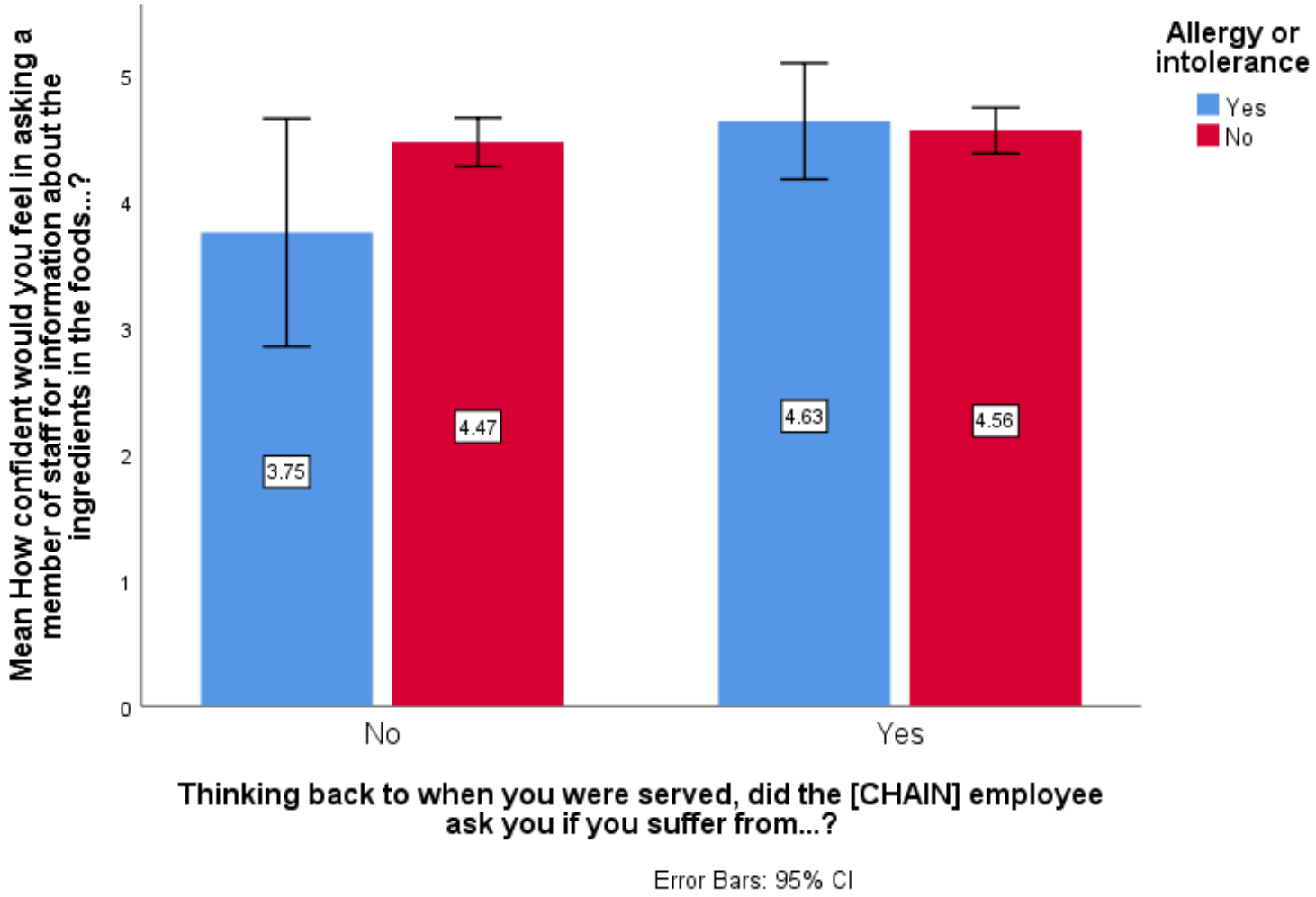
*** = $p < 0.001$; ** = $p < 0.01$; * = $p < 0.05$; ° = 95% Bayesian Credible Interval does not contain zero

3) Customers' confidence to ask about food allergens in products

There was some evidence to suggest that the intervention could help to increase the confidence of consumers with allergies/intolerances to ask about product allergens.

The mean ‘confidence’ scores for each of these groups can be seen overleaf in Figure 6. As illustrated, the mean scores were highest among customers with allergies/intolerances who were delivered the intervention ($\bar{x}$ = 4.63, $\sigma_{\bar{x}}$ = 0.219).

Figure 6: Mean customer confidence to ask about food allergens in products, (where 5 is 'very confident' and 1 is 'not at all confident'), by intervention delivery and allergy/intolerance status



In the full factorial linear model[6], the three-way interaction between allocation to treatment branch, intervention delivery and allergy/intolerance status was significant (β = 3.184, p = 0.002). A similar result was also observed in the hierarchical Bayes model (see Table 9 below).

However, there was some uncertainty associated with this result: specifically, the two-way interaction between intervention delivery and allergy/intolerance status was not significant (β = 0.243, p = 0.380).

Table 16: Model #3 – Customers' confidence to ask about food allergens outputs

Base model

Variable	Coefficient β	Std error
Intercept	4.864***	0.144
Age	-0.171**	0.054
Sex	-0.025	0.101
Allocation of treatment	-	-
Intervention delivery	-	-
Allergy/intolerance status	-	-
Allocation to treatment: intervention delivery	-	-
Allocation to treatment: Allergy/intolerance status	-	-
Intervention delivery: Allergy/intolerance status	-	-
Allocation to treatment: Intervention delivery: Allergy/intolerance status	-	-

[6] Comparison of linear models highlighted a significant improvement in fit for the full factorial model ($\chi(7)$ = 16.751, p = 0.019).

Full factorial linear model

Variable	Coefficient β	Std error
Intercept	4.797***	0.176
Age	-0.158**	0.054
Sex	-0.051	0.099
Allocation of treatment	0.231	0.258
Intervention delivery	0.317	0.184
Allergy/intolerance status	-0.490	0.281
Allocation to treatment: intervention delivery	-0.577	0.321
Allocation to treatment: Allergy/intolerance status	-2.639**	0.931
Intervention delivery: Allergy/intolerance status	0.243	0.401
Allocation to treatment: Intervention delivery: Allergy/intolerance status	3.184**	1.023

Hierarchical Bayes model

Variable	Coefficient β	Std error
Intercept	4.812°	0.185
Age	-0.165°	0.054
Sex	-0.038	0.101
Allocation of treatment	0.175	0.271
Intervention delivery	0.297	0.183
Allergy/intolerance status	-0.572°	0.277
Allocation to treatment: intervention delivery	-0.488	0.319
Allocation to treatment: Allergy/intolerance status	-1.873°	0.806
Intervention delivery: Allergy/intolerance status	0.355	0.380
Allocation to treatment: Intervention delivery: Allergy/intolerance status	2.346°	0.884

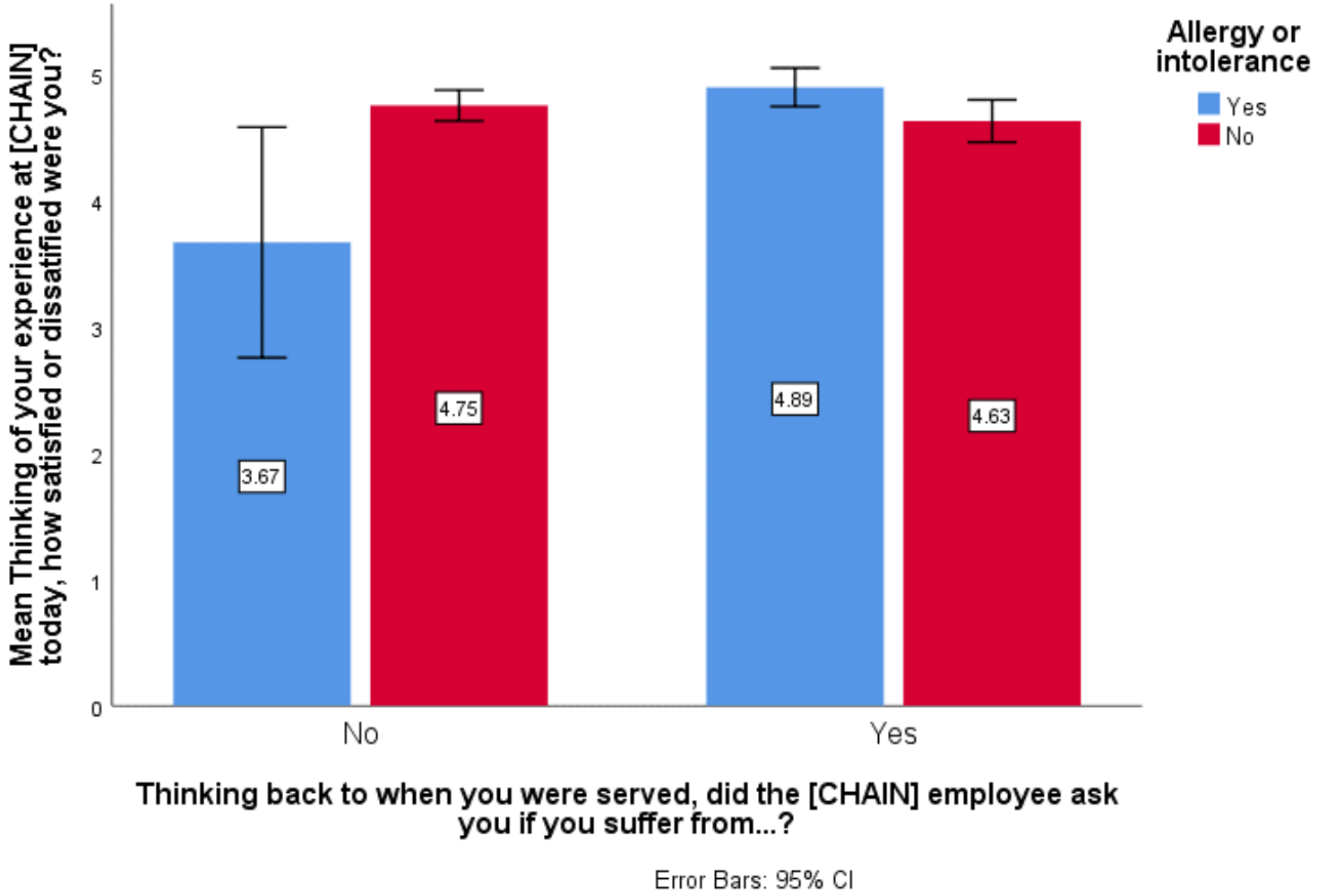
*** = $p < 0.001$; ** = $p < 0.01$; * = $p < 0.05$; ° = 95% Bayesian Credible Interval does not contain zero

4) Customer Satisfaction

As for trust, there was evidence to suggest that delivery of the intervention was effective in increasing satisfaction with visits among those with allergies/intolerances.

The mean ‘satisfaction’ scores for each of these groups can be seen below in Figure 5. As illustrated, scores were highest among customers with allergies/intolerances who were delivered the intervention ($\bar{x}$ = 4.89, $\sigma_{\bar{x}}$ = 0.072).

Figure 7: Mean customer satisfaction with their visit to the FBO, (where 5 is 'very satisfied' and 1 is 'not at all satisfied'), by intervention delivery and allergy/intolerance status



In the full linear model[7], both the two-way interaction between intervention delivery and allergy/intolerance status (β = 0.945, p = 0.004); and the three-way interaction between intervention delivery, allergy/intolerance status and allocation to treatment branch (β = 2.986, p = 0.000) were highly significant.

These results – which suggest that the intervention increased satisfaction among the allergy/intolerance subset – were also corroborated by the outputs of the respective hierarchical Bayes model (see Table 10).

In addition, there was a significant negative coefficient observed for food allergy/intolerance status (β = –0.815, p = 0.001), suggesting that those from this subpopulation had lower levels of satisfaction with their visit to the FBO.

Table 17: Model #4 – Customers' satisfaction outputs

Base model

Variable	**Coefficient β**	**Std error**
Intercept	4.931***	0.130
Age	-0.130**	0.046
Sex	-0.068	0.088
Allocation of treatment	-	-
Intervention delivery	-	-
Allergy/intolerance status	-	-
Allocation to treatment: intervention delivery	-	-
Allocation to treatment: Allergy/intolerance status	-	-
Intervention delivery: Allergy/intolerance status	-	-
Allocation to treatment: Intervention delivery: Allergy/intolerance status	-	-

[7] Model comparisons indicated a significant improvement in fit for the full factorial model, compared the base model ($\chi(7) = 33.538$, $p < 0.000$).

Full factorial linear model

Variable	Coefficient β	Std error
Intercept	4.949***	0.149
Age	-0.113*	0.044
Sex	-0.005	0.081
Allocation of treatment	0.287	0.219
Intervention delivery	0.072	0.151
Allergy/intolerance status	-0.815***	0.232
Allocation to treatment: intervention delivery	-0.531*	0.263
Allocation to treatment: Allergy/intolerance status	-2.778***	0.757
Intervention delivery: Allergy/intolerance status	0.945**	0.326
Allocation to treatment: Intervention delivery: Allergy/intolerance status	2.986***	0.831

Hierarchical Bayes model

Variable	Coefficient β	Std error
Intercept	4.967°	0.165
Age	-0.119°	0.045
Sex	0.006	0.085
Allocation of treatment	0.240	0.233
Intervention delivery	0.062	0.149
Allergy/intolerance status	-0.855°	0.229
Allocation to treatment: intervention delivery	-0.466	0.260
Allocation to treatment: Allergy/intolerance status	-2.112°	0.669
Intervention delivery: Allergy/intolerance status	1.008°	0.315
Allocation to treatment: Intervention delivery: Allergy/intolerance status	2.255°	0.733

*** = $p < 0.001$; ** = $p < 0.01$; * = $p < 0.05$; ° = 95% Bayesian Credible Interval does not contain zero

Objective 5, appropriateness of the analytical approach: is the use of hierarchical models in analysis – to detect treatment effects – appropriate?

Outcome measure: Defining unknown parameters that could impact the trial's statistical power, such as the intra-cluster correlation coefficients for each outcome measure (ρ)

There are two factors that could render a clustered design inappropriate for a future RCT and/or incompatible with the chosen analytical approach:

- large intra-cluster correlation coefficients, and
- vastly discrepant response rates across branches (and chains).

The intra-cluster correlation coefficients (ICCs; measured using one-way random effects models as outlined in Lohr, 1999) for the four outcome variables by branch are contained overleaf in Table 11.

As can be seen, the observed ρ values were small. This result indicates that clustering had a minimal impact on outcome variables; therefore, statistical power to detect treatment effects was not drastically affected by the design (Moerbeek, 2006).

Additionally, the two negative ICC values in Table 11 indicate that variance within clusters (branches) exceeds that between clusters, likely an artefact of the relatively small number of clusters and modest cluster sizes (Kahlia, 2015). Importantly, negative ICC values do not preclude the use of multi-level linear mixed effects models (Chen, 2017).

Table 18: Intra-cluster correlation coefficients for the linear model dependent variables

Survey measure	ICC (ρ)
How concerned are you about the safety of the food that is sold in **[INSERT CHAIN NAME]** for consumption by those with food allergies and intolerances?	0.023
And how much do you trust or distrust **[INSERT CHAIN NAME]** as a business responsible for the sale of food and drinks?	-0.031
How confident would you feel in asking a member of staff for information about the ingredients in the foods they are selling, because of a concern about possible allergens/food intolerances?	-0.028
Thinking of your experience at **[INSERT CHAIN NAME]** today, how satisfied or dissatisfied were you?	0.001

As highlighted in the results for Objective 2, variance in sample sizes across branches is likely to be more of a challenge for the future RCT and its statistical power (Kerry & Bland, 2001; Rutterford et al, 2015). Based on the results of this feasibility trial, there is evidence to suggest that response rates would naturally differ significantly across branches, perhaps as a result of disparate customer profiles or staff engagement with the trial.

However, the timing of the fieldwork period in this trial – in which an anomalous population-level health crisis impacting response rates occurred – means that the true extent of this potential problem is uncertain. Nevertheless, the use of stopping guidelines in future trials to ensure similar sample sizes across clusters should be strongly considered.

Discussion

This study aimed to assess the feasibility of a fully powered randomised cluster trial – involving the proactive communication of allergen information at the point of sale – in UK FBOs.

The results of this study suggest that a future fully powered clustered RCT in the UK would likely be feasible. However, for the execution of such an RCT to occur, the three primary issues encountered in this feasibility trial would require amelioration:

1) Difficulty recruiting businesses into the trial,

2) A low in-store survey response rate, and
3) Lower than expected intervention fidelity.

The former two issues were likely negatively impacted by the trial's timing: COVID-19 was rapidly proliferating while recruitment was being finalised for two FBOs, and fieldwork was beginning for the participating business. To this point, the two FBOs who withdrew their participation during the selection of branches cited COVID-19 as a barrier. With regard to the profound influence of this event upon survey response rate, the Pearson correlation indicated a significant negative association between the number of daily survey responses and the number of COVID-19 cases in the UK.

The last of these three issues – intervention fidelity – is a challenge unrelated to COVID-19 that will require redress in the future fully powered RCT. We recommend a fortification of briefing procedures, including an increased frequency of communication with store managers, to enhance compliance with trial procedures (such as the delivery of interventions), with further checks implemented to measure and monitor intervention fidelity.

This study's approach to data analysis – hierarchical linear mixed models – appears to be suitable for use in a fully powered RCT. The intra-cluster correlation coefficient (ICC) values observed in this study were low, broadly in line with those observed in other human cluster trials. With such low ICC values in mind, future trials should seek to further investigate whether linear mixed models (in which branch is specified as a random effect) or more parsimonious fixed-effects models are most appropriate. As an alternative to the analytical approach chosen for this trial, the use of hierarchical Bayes models also appears feasible.

Based on the analysis conducted, the survey outcome measures appear to be sensitive enough to detect treatment effects, as significant differences were observed across population groups and treatment allocation categories.

Positively, the significant differences observed in the linear models directly related to the fully powered trial's research questions, as follows.

1) Customers with food allergies/intolerances had significantly higher levels of concern around food safety; and significantly lower ratings of trust in the FBO.
2) Significant interactions were observed between allergy/intolerance status and delivery of the intervention for the 'trust' and 'satisfaction' outcome variables, which suggests that the intervention may have a positive impact upon these outcomes.
3) Halo effects – that is, whether the intervention positively impacted the views of all customers – were not observed for any of the outcome measures.

However, as previously noted, the limited sample size of this study precludes definitive conclusions about the impact of the intervention on the trial's outcome measures.

Limitations

This trial had several key limitations. As we have consistently noted in this paper, the rapid proliferation of COVID-19 throughout the UK in the trial execution period had severe ramifications, both in terms of FBOs' participation and survey fieldwork (including the length of fieldwork and the survey response rate).

Another of the limitations of this feasibility trial was the participation of a single FBO, rather than the three intended. With only one FBO having participated, the extent to which the results of this study can be generalised to other chains – including those of other types and sizes – is unclear. There may also be additional unforeseen barriers to trial implementation which were not identified in this study.

Quasi-random branch selection and non-systematic random allocation to treatment and control groups were other limitations in this trial. It may be that these aspects of execution are inherently compromised in business trials in the UK: based on conversations held with FBOs throughout recruitment, participation appeared to be partly contingent upon ability to select participating branches. However, with the successful execution of in-business trials such as this one – which will inevitably help to build working relationships with, and trust among FBOs – FSA may be able to be more involved in these processes moving forward.

With these limitations in mind, the execution of fully randomised business trials in the UK will continue to be challenging. However, the extent of this challenge will become clearer as additional field trials are conducted over the coming years.

Bibliography

Barnett, J., et al. (2017). *The preferences of those with food allergies and/or intolerances when eating out. Final Report*. Accessed 22/01/2020.

Barnett, J., Vasileioua., J. & Lucas, J. (2019). Conversations about food allergy risk with restaurant staff when eating out: A customer perspective. *Food Control*; 108, pp. 1-9.

Bates, D., Maechler, M., Bolker, B. & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*; 67(1), pp. 1-48.

BBC News (2020). *Coronavirus: People with fever or 'continuous' cough told to self-isolate*. Accessed 14/4/2020.

BBC News (2019). Owen Carey: Allergy deaths register 'could save lives'. Accessed 14/4/2020.

Kerry, S. and Bland, J. (2001). Unequal cluster sizes for trials in English and Welsh general practice: implications for sample size calculations. *Statistics in Medicine*; 20, pp. 377–90.

Killip, S., Mahfoud, Z. & Pearce, K. (2004). What is an intracluster correlation coefficient? Crucial concepts for primary care researchers. *Annals of family medicine*; 2(3), pp. 204–208.

Chen, G., Taylor, P., Haller, S., Kircanski, K., Stoddard, J., Pine, D., Leibenluft, E., Brotman, M., & Cox, R. (2018). Intraclass correlation: Improved modeling approaches and applications for neuroimaging. *Human brain mapping*; 39(3), pp. 1187–1206.

Cornfield, J. (1978). Randomization by group: a formal analysis. *American Journal of Epidemiology*; 108, pp. 100–102.

Food Standards Agency (2020). *Allergy and intolerance*; accessed 14/4/2020.

Gelman, A. and Hill, J. (2006). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press.

Goodrich, B (2020). *Package 'rstanarm'*. Accessed 10/4/2020.

Green, J., McDowall, Z. & Potts, H. (2008). Does Choose & Book fail to deliver the expected choice to patients? A survey of patients' experience of outpatient appointment booking. *BMC Medical Informatics and Decision Making*; 8, pp. 36 - 44.

Hero, J., Blendon, R., Zaslavsky, A., & Campbell, A. (2016). Understanding What Makes Americans Dissatisfied With Their Health Care System: An International Comparison. *Health Affairs*; 35(3), pp. 502-509.

John Hopkins University (2020). *COVID-19 Dashboard*. Accessed 13/4/2020.

Killip, S., Mahfoud, Z., & Pearce, K. (2004). What is an intracluster correlation coefficient? Crucial concepts for primary care researchers. *Annals of family medicine*; 2(3), pp. 204–208.

Lee, J., Sim, N., Feng J. & Rabe-Hesketh, S. (2020). *Introduction to multilevel modeling using rstanarm: A tutorial for education researchers*. Accessed 13/4/2020.

Lohr, S. (2019). *Sampling: design and analysis*. Duxbury Press, New York, NY.

Moerbeek, M. (2006). 'Cluster Randomised Trials: Design and Analysis'. In Pham, H. Springer Handbook of Engineering Statistics: Springer-Verland London Limited.

NHS (2020). *COVID-19 daily deaths*. Accessed 13/4/2020.

Raats, M. (2015). Reference amounts utilised in front of package nutrition labelling; impact on product healthfulness evaluations. *European Journal of Clinical Nutrition*; 69(5), pp. 619-25.

Rutterford, C., Copas, A., Eldridge, S. (2015). Methods for sample size determination in cluster randomized trials. *International Journal of Epidemiology*; 44(3), pp. 1051–1067

Soon, J. (2018). 'No nuts please': Food allergen management in takeaways. *Food Control*; 91, pp. 349 - 356

Smeaton, D. (2013). Baseline Study on the Provision of Allergy Information to Consumers for Foods which are not prepacked. Accessed 14/4/2020.

Vasiljevic, M. et al. (2018). Impact of lower strength alcohol labeling on consumption: A randomized controlled trial. *Journal of Health Psychology*; 37(7), pp. 658-667.

Wearne, S. (2017). Food allergy and intolerance programme. Accessed 14/4/2020.

CONSORT 2010 checklist of information to include when reporting a pilot or feasibility trial*

Title and abstract

1a) Identification as a pilot or feasibility randomised trial in the title, page 2

1b) Structured summary of pilot trial design, methods, results, and conclusions (for specific guidance see CONSORT abstract extension for pilot trials), page 2

Backgrounds and objectives

2a) Scientific background and explanation of rationale for future definitive trial, and reasons for randomised pilot trial, page 3

2b) Specific objectives or research questions for pilot trial, page 3

Methods

Trial design

3a) Description of pilot trial design (such as parallel, factorial) including allocation ratio, page 4

3b) Important changes to methods after pilot trial commencement (such as eligibility criteria), with reasons

Participants

4a) Eligibility criteria for participants, page 4/5

4b) Settings and locations where the data were collected, page 5

4c) How participants were identified and consented, page 6

Interventions

5) The interventions for each group with sufficient details to allow replication, including how and when they were actually administered, page 6

Outcomes

6a) Completely defined prespecified assessments or measurements to address each pilot trial objective specified in 2b, including how and when they were assessed, page 7

6b) Any changes to pilot trial assessments or measurements after the pilot trial commenced, with reasons, page 8

6c) If applicable, prespecified criteria used to judge whether, or how, to proceed with future definitive trial

Sample size

7a) Rationale for numbers in pilot trail, page 8

7b) When applicable, explanation of any interim analyses and stopping guidelines

Randomisation and sequence generation

8a) Method used to generate the random allocation sequence, page 9

8b) Type of randomisation(s); details of any restriction (such as blocking and block size), page 9

Allocation concealment mechanism

9) Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned

Implementation

10) Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions, page 9

Blinding

11a) If done, who was blinded after assignment to interventions (for example, participants, care providers, those assessing outcomes) and how, page 9

11b) If relevant, description of the similarity of interventions

Statistical methods

12) Methods used to address each pilot trial objectives whether qualitative or quantitative, page 9

Results

Participant flow (a diagram is strongly recommended)

13a) For each group, the numbers of participants who were approached and/or assessed for eligibility, randomly assigned, received intended treatment, and were assessed for each objective, page 6

13b) For each group, losses and exclusions after randomisation, together with reasons, page 6

Recruitment

14a) Dates defining the periods of recruitment and follow-up, page 10

14b) Why the pilot trial ended or was stopped, page 10

Baseline data

15) A table showing baseline demographic and clinical characteristics for each group, page 10

Numbers analysed

16) For each objective, number of participants (denominator) included in each analysis. If relevant, these numbers should be by randomised group, page 10 – 19

Outcomes and estimation

17) For each objective, results including expressions of uncertainty (such as 95% confidence interval) for any estimates. If relevant, these results should be by randomised group, page 10 – 19

Ancillary analyses

18) Results of any other analyses performed that could be used to inform the future definitive trial, page 10 – 19

Harms

19a) All important harms or unintended effects in each group (for specific guidance see CONSORT for harms)

19b) If relevant, other important unintended consequences

Discussion

Limitations

20) Pilot trial limitations, addressing sources of potential bias and remaining uncertainty about feasibility, page 20

Generalisability

21) Generalisability (applicability) of pilot trial methods and findings to future definitive trial and other studies, page 20

Interpretation

22a) Interpretation consistent with pilot trial objectives and findings, balancing potential benefits and harms, and considering other relevant evidence, page 20

Other information

Registration

23) Registration number for pilot trial and name of trial registry

Protocol

24) Where the pilot trial protocol can be accessed, if available

Funding

25) Sources of funding and other support (such as supply of drugs), role of funders

Other

26) Ethical approval or approval by research review committee, confirmed with reference number

Citation: Eldridge SM, Chan CL, Campbell MJ, Bond CM, Hopewell S, Thabane L, et al. CONSORT 2010 statement: extension to randomised pilot and feasibility trials. BMJ. 2016;355.

*We strongly recommend reading this statement in conjunction with the CONSORT 2010, extension to randomised pilot and feasibility trials, Explanation and Elaboration for important clarifications on all the items. If relevant, we also recommend reading CONSORT extensions for cluster randomised trials, non-inferiority and equivalence trials, non-pharmacological treatments, herbal interventions, and pragmatic trials. Additional extensions are forthcoming: for those and for up to date references relevant to this checklist, see www.consort-statement.org.