Pre-analysis Plan: Vietnam Pork Vendor Grading

Vendor food safety, practice, and knowledge outcomes:

Primary:

1) Log of total bacterial count (TBC), pork sample
2) Detection of Salmonella, pork sample
3) Total points for actions in the table below (range from 1 to 6)

Actions *The enumerator marks the actions of the vendor based on the clandestine observation with the following criteria:*		
1.	EITHER washes hands with soap or disinfects hands OR hand swab is clean NOTE: EITHER WASHING OR DISINFECTION IS SUFFICIENT.	1. 1 (Only mark this if s2q5 = 1 or s2q6 =1 or s4q40 = 1 or s4q38a=1) 2. 0
2.	EITHER washes OR disinfects the knife OR knife swab is clean NOTE: EITHER WASHING OR DISINFECTION IS SUFFICIENT.	1. 1 (Only mark this if s2q8=1 or s2q9=1 or s4q8b=1) 2. 0
3.	Washes or disinfects the cutting board	1. 1 (Only mark this if s2q11 = 1 or s2q12=1) 2. 0
5.	Pork is displayed on an easy to clean surface	1. 1 (Only mark this if s2q3=1) 2. 0
6.	Different products are kept on separate trays	1) 1 (Only mark this if s2q3a=1 or s2q3a = 97) 2) 0

Secondary:

4) Log of total bacteria count (TBC), cutting board
5) Score on the 5-point rating scale used for signage (based on points for actions + hardware)
6) Total points for hardware in the table below (range from 0 to 4)

Hardware *After all discrete observation are complete, the enumerator comes back with the manager and reveal identity to the vendor. The enumerator comments on any deviations from recommended practice observed and complete observation checklist for hardware if not already complete.*		
1.	Has a way to wash hands without recontamination (container with spigot or tap)	0/1
2.	Has soap / detergent for cleaning surfaces and washing hands	0/1
3.	Has disinfectant for surfaces / utensils	0/1
4.	Are there separate knives for separate products?	1/0 (1 if only one type of product is sold)

7) Actions during 30-minute observation (sum of binary indicators for ANY of action, AND individual variables)
    1. Number of times washed hands in any way
    2. Times washed hands with soap
    3. Times cleaned cutting board in any way (incl scraped, washed, disinfected)
    4. Times washed cutting board with soap or sprayed disinfectant
    5. Times washed or disinfected knife
    6. Maintained separation of tools (did NOT use same knife, cutting board, weighing plate/scale, or cloth for different products) (0/1)

8) Answer to "what do you think consumers look for when deciding from whom to buy pork" is one of the following:
    a. stall is clean
    b. vendor is clean
    c. food is safe
    d. vendor has a high rating

9) Binary (correct/incorrect) or proportion of correct answers for each question below and the sum of these values for the section.

1. What are some things you can you do while selling to make sure your pork stays clean? SELECT ALL RESPONSES MENTIONED BY THE RESPONDENT.	1. Wash food preparation surfaces regularly 2. Disinfect food preparation surfaces regularly 3. Use easy to clean food preparation surfaces 4. Store raw viscera and raw meat separately 5. Use different knives for raw meat, raw viscera, and cooked (meat and/or viscera) 6. Wash hands regularly 99. Other, specify *(responses 1-6 are correct; score is the proportion of these given)*
2. What is the best surface on which to display or store meat?	1. Metal 2. Plastic 3. Wood 4. Cardboard 5. Woven mats 6. Cloth 7. Granite 99 Other, specify *(responses 1 and 7 are correct; either of these result in score of 1, 0 otherwise)*
3. How should you display viscera? SELECT MULTIPLE	1. Separately from meat 2. In a tray 3. Together with meat 99. Other, specify 98. Don't know *(response 2 is correct; 0 otherwise)*
4. What are some ways that pork can become unsafe to eat / contaminated while selling? SELECT ALL RESPONSES MENTIONED BY THE RESPONDENT.	1. Flies 2. Dirty hands 3. Dust from ground 4. Dirty knives 5. Dirty surface 6. Contact with offal 7. Dirty water 8. Hot weather 98. Don't know 99. Other, specify: *(responses 1-8 are correct; score is the proportion of these given)*
5. When should you wash your hands while you are at work?	1. After each customer 2. After using the toilet

SELECT ALL RESPONSES MENTIONED BY THE RESPONDENT.	3. After handling money 4. At the beginning of the day 5. At the end of the day 6. When hands feel dirty 7. After eating 8. Before eating 99. Other, specify: *(responses 1-8 are correct; score is the proportion of these given)*

Estimating equation for food safety, practice and knowledge outcomes:

$$y_{ij1} = \alpha + \beta_1 X_{ij0} + \beta_2 MKT\_Treat_j + \beta_3 IND\_Treat_i + \beta_4 y_{ij0} + S_j + \varepsilon_{ij} \quad (1)$$

Where X_{ij0} are baseline characteristics and personnel and environmental factors at endline selected via pdslasso, S_j are stratification (province) dummies, MKT_Treat_j is the market-level treatment indicator, and IND_Treat_i is the individual-level indicator that the vendor was selected for provision of equipment. y_{ij0} is the baseline value of the outcome variable. Standard errors are clustered at the market level. The primary hypotheses to be tested are $\beta_2 = 0$ and $\beta_2 + \beta_3 = 0$.

Vendor business outcomes:

10) Log revenue from pork products sold at the market today (based on purchased amount, prices, and pork remaining at the end of the day), measured based on the value calculated from reported purchases and measured remaining pork. Missing values of this variable will be imputed with the vendor's self-reported revenue.
11) Log cost of pork purchased for sale today
12) Total time spent selling at the market today (if not yet finished selling, use expected leaving time)
13) Price of sample modal pork product sold
14) Log non-pork costs last month
15) Water and cleaning supply costs last month
16) Daily revenue minus pork costs
17) Monthly profits (based on today's revenue minus pork costs, times number of days per week selling pork based on past week, minus monthly costs)

Estimating equation for business outcomes (ITT)

$$y_{ij1} = \alpha + \beta_1 X_{ij0} + \beta_2 MKT_{Treat_j} + \beta_3 IND_{Treat_i} + \beta_4 y_{ij0} + S_j + \varepsilon_{ij} \quad (2)$$

Estimating equation for business outcomes (2SLS)[1]

First stage, where r_{ij1} is the 5-point rating score measured at endline.

$$r_{ij1} = \alpha + \beta_1 X_{ij0} + \beta_2 MKT_{Treat_j} + \beta_3 IND_{Treat_i} + \beta_4 r_{ij0} + S_j + \varepsilon_{ij} \quad (3)$$

Second stage (excluding IND_{Treat_i})

$$y_{ij1} = \alpha + \beta_1 X_{ij0} + \beta_2 MKT_{Treat_j} + \beta_3 \widehat{r_{ij1}} + \beta_4 y_{ij0} + S_j + \varepsilon_{ij} \quad (4)$$

Estimating equation for business outcomes (exploratory: correlation with current practices)

$$y_{ij1} = \alpha + \beta_1 X_{ij0} + \beta_2 r_{ij1} + \beta_3 MKT_{Treat_j} + \beta_4 r_{ij1} * MKT_{Treat_j} + \beta_4 y_{ij0} + S_j + \varepsilon_{ij} \quad (5)$$

Where r_{ij1} is the 5-point score as measured at endline.

Consumer outcomes:

Primary

18) Response to "why did you choose this vendor today" is one of the following:
    a. stall is clean
    b. vendor is clean
    c. food is safe
    d. vendor has a high rating

Secondary

19) Number of times bought pork for household consumption in past 7 days
20) Proportion of times bought pork from this market
21) Amount spent on pork in past 7 days
22) Amount spent on other meat and seafood past seven days
23) Has seen information regarding safe pork at this market

Estimating equation for consumer outcomes:

[1] The 2SLS specification will not be applied to outcomes 14 (non-pork costs), 15 (cleaning costs), or 17 (monthly profits) as non-pork costs could be directly affected by the provision of equipment.

$$y_{ij1} = \alpha + \beta_1 \bar{X}_{ij0} + \beta_2 X_{ij1} + \beta_3 MKT_{Treat_j} + \beta_4 \bar{y}_{j0} + S_j + \varepsilon_{ij} \quad (6)$$

Where $\bar{y}_{j0}$ is the average outcome at baseline among consumers interviewed at the same market, $\bar{X}_{ij0}$ is a set of market-level variables, including the market-level means of consumer characteristics measured at baseline and selected via pdslasso, and X_{ij1} is a set of individual-level time-invariant consumer characteristics measured contemporaneously with y_{ij1}.

Controls:

We will select baseline controls via PDS-LASSO from among those listed below. In addition, we will run all models with only the baseline version of the outcome (market-level mean for consumers) as a control and report any substantial differences across models. Missing values of controls will be imputed with the sample mean.

For each continuous variable included below, the following transformations will also be included as separate candidate variables:

1. inverse hyperbolic sine transformation
2. square
3. tercile dummies

For each variable indicating a scale of satisfaction, cooperation, agreement, etc. the following will be included as candidate controls

1. indicators for each response (including missing / N.A.)
2. indicators for "at least" each response from the second one (e.g., at least somewhat disagree, at least neutral, at least somewhat agree, at least strongly agree)

Market-level controls will be included for both vendor and consumer outcomes. Vendor-level controls will be included for vendor outcomes. Consumer-level averages per market will be included for consumer outcomes.

Vendor-level control variables

- All vendor outcome variables (at baseline) described in this pre-analysis plan
- Educational attainment of respondent (binary variables per category)
- Age of baseline respondent
- Gender of baseline respondent
- Number of people involved in the business
- Proportion of women involved in the business
- Average age of people involved in the business
- Hours per day worked by the baseline respondent
- Average hours per day worked by others involved in the business

- Total wages paid per day
- Years the respondent has sold pork
- Respondent control over income (input into at least some decisions; input into most or all)
- Job satisfaction (at least unsatisfied, at least neutral, at least satisfied, very satisfied; also indicators for each level of satisfaction)
- Purchased stomach or intestines for sale
- Prepared any pork products
- Prepared any pork products containing stomach or intestines
- Prepared any pork products
- Average price of pork cuts sold (per KG)
- Number of days sold pork at this market during past 7 days
- Average hours open per day in this market (on days sold at this market)
- Average number of customers served
- Sold pork elsewhere (binary)
- Number of days sold pork elsewhere
- Average number of hours sold elsewhere (on days sold elsewhere)
- Binary indicator of any expenditure per cost category last month
- Amount spent per cost category last month
- Total costs past 30 days
- Perceptions of food safety risk
    1. Cumulative indicators (at least somewhat unlikely, at least neutral, up to highly likely, treating "don't know" as neutral)
    2. Individual categorical indicators (don't know, highly unlikely, somewhat unlikely, etc.)
- Ever attended food safety training
- Time since attended food safety training
- Indicator variables for each response to question on selection of safe pork
- Number of responses to question on selection of safe pork
- Indicator variables for each response to question on keeping pork clean when selling
- Number of responses to question on keeping pork clean when selling
- Indicator variables for each response to question on when to wash hands
- Number of responses to question on when to wash hands
- Availability of safe, clean toilet facility at market
- Total weight remaining at weighing
- Products remaining at weighing (indicators)

- Market level indicator: level 1 (wholesale), level 2, level 3, other
- Days per week the market is open
- For microbiological outcomes only (log TBC on pork sample and cutting board; detection of Salmonella on pork sample), the following candidate controls will be included:
    - time at which sample was taken
    - time elapsed between taking the sample and receipt of sample by laboratory
    - mean temperature between 6 am and time of sampling at the sampling location, if available, otherwise mean temperature on the day of sampling
    - interaction of temperature with time of sampling
    - interaction of temperature and time between sampling and laboratory receipt

Market-level control variables (market manager survey unless specified as census form)

- Number of pork stalls in the market
- Number of pork stalls in the market
- Water is available at the market
- Water source (tap, well, nearby homes or businesses, other)
- Vendors can store equipment securely overnight
- Market manager is employed by government
- Indicator for level of government by which manager is employed (commune, district, province, other)
- Indicator variables and "at least" indicators (at least didn't interfere, at least neutral, etc. for observation about market manager's level of cooperation (market census form)

Consumer-level control variables

** indicates variables included as potential contemporaneous controls*

- Age *
- Gender *
- Indicator for living in a couple (= 1 if married or living together) *
- Household size *
- Number of children under 5 years living in household *
- Indicator for any children under 5 years living in household *
- Employment type *
- Educational attainment (indicator for highest completed) *
- Has a working refrigerator *
- Is able to show enumerator the vendor from whom purchased pork

- Number of different cuts purchased
- Purchased any intestines or stomach
- Purchased any processed products (sausage, Vietnamese sausage, fried mince, mince)
- Purchased any cooked pork products
- Total amount spent on pork today
- Share of total value purchased cooked
- "Some pork will still be stored at": A set of indicators which = 1 if the latest time at which any pork product purchased today will be eaten or cooked is *greater than* each category. For example, if the maximum response for the consumer to Q14c. "When do you plan to [cook/eat] this pork type" is "4 = Tomorrow", the indicator for "within the next 4 hours" = 1, for "within the next 8 hours" =1, for "by the end of the day" =1, for "tomorrow" = 0, "within the next 2 days" = 0, and so on.
- Indicator for whether any of the pork purchased will be stored in a freezer
- Indicator for whether any of the pork purchased will be stored in a fridge
- Indicator for whether any of the pork purchased will be stored at room temperature
- Interaction of the storage duration variables with the indicator for storage at room temperature
- Interaction of the storage duration variables with the indicator for ownership of a working refrigerator
- Over past 30 days, customer bought from same vendor either exclusively or if they were at the market
- Over past 30 days, customer bought from vendor "most of the time"
- Does not normally buy from this market
- Reasons for choosing vendor (indicators for each)
- Most important reason for choosing vendor (indicators for each)
- Days since started buying from this vendor (0 if first time)
- Amount spent on food in typical week
- Number of times bought pork past 7 days
- Share of times bought pork from this market
- Amount spent on pork past 7 days
- Amount spent on other meat past 7 days
- Amount spent on seafood past 7 days
- Checks any hygiene practice when purchasing pork
- Number of hygiene practices checked when purchasing pork
- Likelihood of getting sick from pork purchased from typical vendor in this market: indicators for each response option and cumulative indicators: at least somewhat likely, at least neither likely nor unlikely, etc.

- Likelihood of getting sick from pork purchased from: indicators for each response option and cumulative indicators: at least somewhat likely, at least neither likely nor unlikely, etc.
- Indicator for having seen any information (of any form) regarding safe pork in this market
- Prefers to buy pork from female vendors
- Prefers to buy pork from male vendors
- Reason prefers to buy pork from vendors of a particular gender

Multiple hypothesis corrections

The false discovery rate (the proportion of hypotheses of those we reject, which should not in fact have been rejected) will be controlled using sharpened q values using the two-step procedure described by Benjamini, Krieger, and Yekutieli (2006), as applied by Anderson (2008), for the three primary outcomes using specification (1) and the primary consumer outcome using specification (6).

References:

Anderson, M.L., 2008. Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. *Journal of the American statistical Association*, *103*(484), pp.1481-1495.

Benjamini, Y. and Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, *57*(1), pp.289-300.

Benjamini, Y., Krieger, A.M. and Yekutieli, D., 2006. Adaptive linear step-up procedures that control the false discovery rate. *Biometrika*, *93*(3), pp.491-507.