

## A common framework for using and reporting consumer purchase data (CPD) in foodborne outbreak investigations in Europe

Solveig Jore <sup>a</sup>, Uffe Christian Braae<sup>b</sup>, Frederik Trier Møller<sup>b</sup>, Ingrid Friesema<sup>c</sup>, Karthik Paranthaman<sup>d</sup>, Katri Jalava <sup>e</sup>, Nathalie Jourdan-DaSilva<sup>f</sup>, Emma Löf<sup>g</sup>, Moa Rehn<sup>g</sup> and Steen Ethelberg<sup>b,h</sup>

<sup>a</sup>Department of Infection Control & Preparedness, Norwegian Institute of Public Health, Oslo, Norway; <sup>b</sup>Department of Infectious Disease Epidemiology and Prevention, Statens Serum Institut, Copenhagen, Denmark; <sup>c</sup>Centre for Infectious Disease Control, National Institute for Public Health and the Environment, Bilthoven The Netherlands; <sup>d</sup>Field Service, National Infection Service, Public Health England, Ashford UK; <sup>e</sup>Department of Mathematics and Statistics, University of Helsinki, Helsinki, Finland; <sup>f</sup>Département des maladies infectieuses Santé Publique France (French National Public Health Agency), Saint-Maurice, France; <sup>g</sup>Communicable Disease Control and Health Protection, Public Health Agency of Sweden, Solna, Sweden; <sup>h</sup>Department of Public Health, Global Health Section, University of Copenhagen, Copenhagen, Denmark

### ABSTRACT

Consumer purchase data (CPD) can be a powerful tool in the investigation of foodborne outbreaks through analyses of electronic records of food that individuals buy. The objective of this study was to develop a common framework for use of CPD in foodborne outbreak investigations using the expertise of European public health professionals from 11 European countries. We also aimed to describe barriers and limitations preventing CPD utilization. CPD are mainly gathered from supermarket loyalty programmes, smaller consortia, and independent supermarkets. Privacy legislation governing CPD was perceived as the most crucial barrier for CPD usage, but still resolvable. The main practical challenges were obtaining consumer consent for CPD usage, the associated workload, data access, format, and analysis. Harmonising methods and reporting across countries, standardised consent forms and electronic consent methods were identified as solutions.

This guideline was developed to support outbreak investigators in overcoming barriers in using CPD, thereby increasing public health professionals' application and value of this powerful investigation tool. In addition, we hope this framework will lead to more public health institutions, in collaboration with food safety authorities, making use of CPD in outbreak investigations in the future.

### ARTICLE HISTORY

Received 15 July 2021  
Accepted 15 November 2021

### KEYWORDS

Consumer purchase data;  
foodborne outbreaks;  
epidemiology

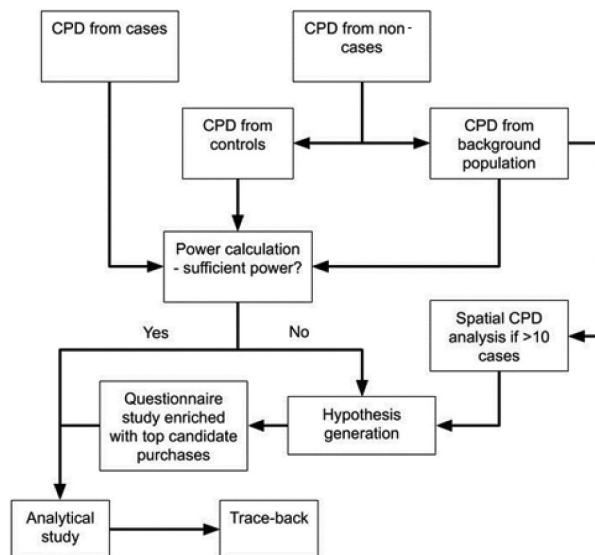
## Introduction

Consumer purchase data (CPD) are electronic records of consumer purchases, collected by the retailers (e.g. supermarkets) and stored in secure databases with key customer and product details [1]. CPD can be divided into two types; individual and population-level data. Individual-level data may be collected anonymously with unique purchases or as individual customer identifiable data. This data contains information on store location, purchase time, and product names. The population-level data are aggregated data, including store location, product type, and purchase time. For foodborne outbreak investigations, the individual-linked CPD is the primary focus. CPD can be subsequently linked to individuals by loyalty or membership cards, credit or debit cards, or by printed receipts or bank statements crosschecked with retailer databases.

Foodborne diseases are a global, growing challenge with economic and public health impacts [2]. Early detection, implementation of control measures, notification, and investigation of foodborne outbreaks are

key actions in identifying correct public health measures to reduce the burden of foodborne illness in the population. Outbreak investigations follow standardised steps. One of the critical challenges for analytical questionnaire data is the recall bias for the case patients and healthy referral participants to remember what they have purchased and eaten during the incubation period. Therefore, CPD are increasingly used as a tool in aiding foodborne outbreak investigations. However, there has been heterogeneity in the CPD usage methods among public health and food safety professionals. CPD is infrequently used in foodborne outbreak investigations [1], and if utilized, CPD methodology is often poorly described in the literature. **Figure 1** shows pathways of foodborne outbreak investigations using consumer purchase data.

CPD is commonly used for hypothesis generation and gathers further evidence for any existing hypothesis [3–8]. CPD has also been used to conduct trace-back or perform analytical studies [9–14]. CPD has been shown to be a powerful tool in severe outbreaks with a single or



**Figure 1.** Pathways of foodborne outbreak investigations using consumer purchase data (CPD).

few cases (e.g. botulism, listeriosis) when standard, statistical methods have failed to identify the causative vehicle [14,15]. Overall, heterogeneous ways to report on outcomes, description of CPD methodology and how CPD was collected, utilised or analysed have been major challenges [1]. Thus, the low use of CPD in foodborne investigations could be due to lack of awareness, knowledge and standardised protocols.

## Materials and methods

We conducted a questionnaire survey among public health and food safety professionals across Europe to elucidate barriers on the use of CPD in foodborne outbreak investigations under the EU-wide Horizon 2020 programme of the NOVA consortium (Novel approaches for design and evaluation of cost-effective surveillance across the food chain, <https://onehealth.eu/jrp-nova/>). Investigators from 11 European countries completed the survey, seven countries reported applying CPD in outbreak investigations, typically within national or regional public health institutes, or national food safety authorities. We also conducted focus group discussions among European disease outbreak public health specialists to further illuminate barriers and map existing usage of CPD in outbreak scenarios. These specialists were identified following a broad invitation to participate distributed via the Epiet Alumni Network (EAN) (<https://www.ecdc.europa.eu/en/epiet-euphem/about/intro>). The discussions were done as a series of three online meetings in 2019.

## Results

### General barriers in the use of CPD

Several barriers preventing the use of CPD exist on an either regional, national or global scale. Barriers can be

categorised as data availability and format constraints, feasibility constraints, or legislative constraints. Steps to be considered and reported if CPD were used in the outbreak investigation are presented in Figure 2.

### Data availability and format constraints

CPD are primarily gathered from supermarket loyalty programmes, mainly from larger supermarket chains, but smaller consortia and individual supermarkets may also hold valuable CPD. The availability of CPD differed by the size of the retailer, with smaller regional retailers less likely to store CPD, and more prone to unavailability issues. In contrast, large nationwide retailers have centralized servers, database storage facilities, and more employees with technical knowhow to assist the request. The format of the stored data varied from one retailer to another, complicating data handling and flexibility of obtaining data, especially if the data acquirement process is to be automated. Digitalized CPD from large, national supermarket chains will most likely become available for use in the near future. However, smaller retail chains or individual retailers might only have local or regional coverage and are therefore less likely to have CPD readily available.

### Feasibility constraints

Obtaining and analysing the data can be time consuming depending on several factors such as ease of data access, data format, and patient compliance, especially when CPD was in analogue format or used for hypothesis generation. Seven out of 11 European countries considered practical constraints or technical obstacles for obtaining and using CPD major issues, with workload being the primary obstacle. The number of steps

## Recommendations for reporting outbreaks using CPD

**Title/Abstract/Keywords**

- 1) Consumer Purchase Data should be mentioned in full in the Title or Abstract or as a minimum included in abbreviated form 'CPD' in the keywords if possible

**Introduction/Background**

- 2) Explain the scientific basis for using CPD in the current outbreak and generally
- 3) State the specific objectives for using CPD

**Materials and Methods**

- 4) Present the step-by-step procedure of how CPD was obtained and used
- 5) State the inclusion and exclusion criteria for study participants in regards to CPD
- 6) Describe the consumer level of the CPD e.g. personal/household level
- 7) Describe the setting in which the CPD was used
- 8) Present the time period covered by CPD
- 9) Describe the product level e.g. food category (beef), brand (company x), product type made by company x (product a, b, etc.)
- 10) Describe the retailer coverage (market and geographical) within the CPD, percentage of cash purchases or other information that might affect the power of the study to detect significant differences in the purchase patterns between for instance cases and controls.
- 11) Clearly define all variables achieved in the CPD and provide the data source for each of the variables
- 12) Provide details of the methods used e.g. hypothesis generation, analytical study, or trace-back

**Results**

- 13) Describe the CPD coverage among cases
- 14) Provide descriptive results of data
- 15) State whether batch information was obtainable if relevant

**Discussion**

- 16) Summarise key results of the CPD analyses
- 17) Describe the potential bias of the study specifically relating to using CPD

**Figure 2.** Standard operating procedures for reporting consumer purchase data (CPD) in regards to foodborne outbreak investigations.

and persons/institutions required before data access is granted is challenging when working with retailers. Subsequently, data handling and the use of different formats are significant obstacles in using CPD from different supermarkets. Optimising and automating the processes are considered priorities, as the workload on handling and processing data was perceived as important barriers to implementing CPD in outbreak investigations.

**Legislative constraints**

The consensus was that insecurity regarding privacy legislation governing CPD is perceived as the most important barrier together with the ability to share consumer-level data with public authorities. Individual public health agencies need to work within the existing national legal position and public acceptability around data sharing. Obtaining consent from consumers can be complicated and time-consuming.

### **Country specific differences**

Some of the barriers encountered were partly country specific and determined the ease of using CPD. If a country had low coverage of supermarket chains and many small retailers, the challenge to gather CPD is much greater than in a country with fewer supermarket chains. The presence of a third-party collecting data from several supermarket chains in one database facilitates the use of CPD. Such a third party currently exists in Denmark, Sweden, Norway, and Finland. However, collection of grocery-related CPD was only widely used in Denmark. Within the EU, there were large differences in the proportion of cash-based transactions versus digital transactions, though loyalty card schemes often allow customers to register cash purchases as well.

We provide a practical guideline for using CPD in foodborne outbreak investigations; in summary, the steps include establishing data availability, ensuring informed consent from cases, acquiring data from retailers or third party providers, data formatting and background information issues and conducting the analysis.

### **Guideline for using consumer purchase data (CPD) in outbreak investigations**

This guide provides an overview of the required processes when using consumer purchase data (CPD) in foodborne outbreak investigations.

#### **(1) Establish data availability**

Food purchases need to be paid by cards or payments identifiable by loyalty cards or equivalent to enable use of CPD in a foodborne outbreak investigation. The data needs to be stored on servers at the retailers or third-party providers. The data and data delivery mechanisms should be curated before an outbreak occurs to avoid discussions on what data to include, such as purchase date and time, quantity, and items of purchases.

#### **(2) Secure informed consent**

The requirements for informed consent vary between countries and retailers and should be legally sound and not delay the rapid use of CPD. Where required by national legal standards, informed consent on the usage of CPD must be obtained from cases and controls included in the outbreak investigation. In addition, the consent needs to cover the pathogens' incubation time, disease reporting delay, shelf life of suspected products, and extended storage (e.g. freezing) before consumption.

#### **(3) Acquire data from retailers or third-party provider**

Data may be obtained directly from retailers or via a third-party provider given that the customers have obtained informed consent. Depending on data storage system, loyalty cards, and bank statements containing information on the purchases obtained from

the consumers, it might be necessary to link consumers and their CPD. Third-party providers may collect CPD from selected supermarket chains based on bank/credit cards enabling linkage of CPD directly from a range of retailers simultaneously.

#### **(4) Data formats and background information**

The data may have various formats, from only product type to detailed information about each specific product purchased. Nevertheless, information on the proportion of retailers included in the investigation should be recorded. Depending on the data source, the level of product information will be recorded at three levels: level 1 – product type, level 2 – product name (commercial and generic), and level 3 – a unique barcode/item number/GTIN/EAN/UPC that identifies product ingredients, and storage information. The data available at card user/household level might require household-level consent if the household consists of more than one person performing the grocery shopping.

#### **(5) Conduct analysis**

Depending on data formats, percentages of cases consuming each product should be recorded. To find shared food items between cases hypothesis generation or analytical studies can be done. Based on CPD findings, the investigators should consider further investigations if needed. This might include food sampling investigations, analytical studies or trace-back and trace-forward studies. Conducting especially microbiological analysis on suspected food items is of vital importance.

## **Discussion**

### **Standardised reporting for outbreaks using CPD**

Past studies have reported CPD usage in very different ways [1]. This clearly illustrates the need for a standardized framework for investigating and reporting. Our study presents a direct and practical approach in applying and reporting CPD under various outbreak settings. Figure 2 illustrates the steps that should be considered and reported if CPD is used in the outbreak investigation. First, it is important to report the market coverage (how large a proportion and geographical coverage of the monetary market the retailers providing CPD cover), unless other data indicates a specific supermarket chain as the source of the outbreak. In addition, retailer coverage might be important in assessing the geographical spread of cases and the strength of an association. Also, the variables typically available in the CPD such as product type, product name, batch number are crucial to understand the strength and limitations of the investigation. In a previous review [1], we could identify only one study [9], which thoroughly described the CPD variables.

We would encourage national authorities to arrange dedicated workshops with retailers and public health/food safety agencies to agree and establish general procedures needed for outbreak investigations. Ideally, such agreed procedures should be published in relevant outbreak response documents/plans in the public domain to ensure transparency and public access. Our framework can assist wide implementation and ease access to CPD usage while reporting guidelines provide consistency in publications. However, despite a structured framework (Figure 1), practical obstacles like country-specific barriers need to be addressed. Nonetheless, we perceive this framework as a necessary and significant step forward. Purchases are increasingly cash free and the possibilities of applying CPD in foodborne outbreak investigations and wider consumer health studies, are thus increasing. CPD data has proven powerful when used, even with limited access to CPD [1]. As CPD will cover a larger percentage of the population and from multiple retailers, CPD will likely gain importance as a starting tool after hypothesis generation in foodborne outbreak investigations, rather than as a tool used when all other sources seem exhausted. As the CPD becomes readily available, foodborne outbreak investigation teams should explore the possibilities of this type of data more frequently.

### **Strength and limitations of CPD used in outbreak investigations**

Usage of CPD is also beneficial for retailers as it allows for rapid identification of implicated products and robust, targeted control measures. CPD can also assist in excluding incorrectly hypothesized products. CPD are potentially very powerful in overcoming recall bias regarding foods or product type, especially when multiple types of the same food are suspected.

Naturally, there are, also inherent limitations when using CPD. Regardless of the CPD quality, the purchase of a product does not necessarily equal consumption. More importantly, lack of purchase does not necessarily mean lack of consumption; CPD tend to primarily capture foods consumed in the household. In addition, defining the CPD exposure period can also be a challenge, as purchase date does not necessarily equal consumption date. Public health and food control professionals need to consider the exposure period of interest, based on disease epidemiology, hypothesized food product, and shopping patterns. Another challenge is the sometimes long shelf life of products and the potential to freeze certain products. Finally, it cannot be ruled out that the consumers have bought products at retailers not providing purchases or not using their loyalty card every time, variably paid with cash only, or conducted otherwise non-registered purchases.

CPD is a relatively new methodology, but considerable differences between countries exist in the application of CPD. Current practice has evolved in different countries based on legal and regulatory policies, relationships between public health/food safety agencies and retailers, and capacity to undertake such investigations. This framework for use of CPD is based on a broad consensus among experienced outbreak investigators and were critically reviewed by experienced public health and food safety professionals. Outbreaks necessitate rapid access and analysis to inform control measures. Public health authorities and food safety authorities should work with their retailers and other data holders to consider how CPD could be used rapidly in outbreak situations within the legal framework. Systematically developed guidelines can improve accessibility and enhance use of CPD. Further, it can also aid in improving the transparency and comparability of studies and thus impact the quality of research done in this domain.

### **Acknowledgments**

To all the outbreak investigators who provided valuable inputs: Eleni Galanis, Marsha Taylor, Tanis Kershaw, Nicole Haywood, Luise Müller, Celine Gossner, Joanna Nederby Öhd, Hendrik Wilking, Karin Nygard, Maja Ostojic, Eva Grilc, Luigi Ianetti, Vera Cantaert, Marie Jansson Mörk, Katja Alt, and Natalie Adams.

### **Disclosure statement**

No potential conflict of interest was reported by the author(s).

### **Funding**

This work was done as part of work package 2 on use of consumer purchase data for reduction of foodborne diseases in the NOVA project under the One Health European Joint Programme, funded by the European Union Horizon 2020 research and innovation programme, Grant Agreement No 773830. FTM received grants from Helsefonden, Nordforsk, and EHEN project HEAP, European Union's Horizon 2020 research and innovation programme under Grant Agreement No 874662.

### **Notes on contributors**

**Solveig Jore** (DVM, BV.Sc, PhD): Is an epidemiologist and part of the covid-19 response team in Norway and has broad experience in both public health and veterinary public health from surveillance, outbreak and research into zoonotic diseases. Main interests are the impact of climate and environmental change upon zoonotic diseases (PhD).

**Uffe Christian Braae** (PhD): Is a One Health Epidemiologist and Public Health Specialist with broad experience in clinical trials and research project



management with focus on intervention and control studies of foodborne pathogens and zoonotic diseases.

**Frederik Trier Møller** (MD, Ph.D): Is a public health specialist and part of the covid-19 response team in Denmark. Current work includes research into how consumer purchase data can be used to improve public health through individual and systemic prevention approaches. Previous work includes register-based research within familial aggregation of inflammatory bowel disease and the establishment of a biobank containing a broad range of data.

**Ingrid Friesma** (PhD): works at the Centre for Infectious Disease Control (CIb), National Institute for Public Health and the Environment (RIVM) as epidemiologist. She does research in and surveillance of several infectious diseases related to gastroenteritis, foodborne diseases and zoonoses, and has coordinated several national outbreak investigations.

**Karthik Paranthaman** (FFPH): Consultant Epidemiologist in Public Health England.

**Katri Jalava** (PhD): Is an epidemiologist and statistician with an academic interest on public health crisis and emergency response and environmental epidemiology. Also keen on statistical and mathematical models applicable to address key public health issues. Adjunct professor (title of docent), associated with Department of Food Hygiene and Environmental Health and Department of Mathematics and Statistics, University of Helsinki, Finland.

**Nathalie Jourdan-Da Silva** (MD, PhD): Is a medical doctor and epidemiologist working with surveillance, food safety and AMR at French Institute for Public Health.

**Emma Löf** (PhD) Epidemiologist, Analyst at Public Health Agency of Sweden.

**Moa Rehn** (PhD) Epidemiologist at the Public Health Agency of Sweden.

**Steen Ethelberg** (PhD, Professor) Head of the section for water and foodborne and zoonotic infections in the Department of Infectious Epidemiology at Statens Serum Institut and Professor at the University of Copenhagen at the Section for Global Health. Epidemiological research on water- and foodborne infections, including methods of disease outbreak investigation, with the overall aim of limiting these often preventable diseases in a national and international context.

## Disclaimer

The comments, discussions and recommendations contained in this article have been summarised from the direct participant feedback. They do not necessarily reflect the views and opinions of the affiliated organisations of authors.

## Authors' contributions

SJ and UCB drafted and wrote the manuscript with notable input from all co-authors. The questionnaire and focus group discussions were planned and done by SE, UCB and IF. All authors edited, revised and approved the final manuscript.

## ORCID

Solveig Jore  <http://orcid.org/0000-0001-8527-4892>

Katri Jalava  <http://orcid.org/0000-0001-9106-3186>

## References

- [1] Møller FT, Mølbak K, Ethelberg S. Analysis of consumer food purchase data used for outbreak investigations, a review. *Euro Surveill.* 2018;23(24). DOI:10.2807/1560-7917.ES.2018.23.24.1700503
- [2] Havelaar AH, Kirk MD, Torgerson PR, et al., World health organization foodborne disease burden epidemiology reference group; 2015.
- [3] Ethelberg S, Smith B, and Torpdahl M, et al. Outbreak of non-O157 shiga toxin-producing *Escherichia coli* infection from consumption of beef sausage. *Clin Infect Dis.* 2009;48(8):e78–81.
- [4] Müller L, Kjelsø C, Frank C, et al. Outbreak of *Salmonella* Strathcona caused by datterino tomatoes, Denmark, 2011. *Epidemiol Infect.* 2016;144:2802–2811.
- [5] Taylor J, Galanis E, Wilcott L, et al., *Salmonella* Chester Outbreak Investigation Team. An outbreak of salmonella chester infection in Canada: rare serotype, uncommon exposure, and unusual population demographic facilitate rapid identification of food vehicle. *J Food Prot.* 2012;75:738–742.
- [6] Ung A, Baidjoe AY, Van Cauteren D, et al. Disentangling a complex nationwide *Salmonella* Dublin outbreak associated with raw-milk cheese consumption, France, 2015 to 2016. *Euro Surveill.* 2019;24(3). <https://doi.org/10.2807/1560-7917.ES.2019.24.3.1700703>.
- [7] Gossner CM, van Cauteren D, Le Hello S, et al. Nationwide outbreak of *Salmonella enterica* serotype 4,[5],12: i:-infection associated with consumption of dried pork sausage, France, November to December 2011. *Euro Surveill.* 2012 Feb 2;17(5):20071.
- [8] Byrne L, Kaindama L, Bentley M, et al. Investigation into a national outbreak of STEC O157: h7 associated with frozen beef burgers, UK, 2017 [published online ahead of print, 2020 Jul 16]. *Epidemiol Infect.* 2020;1–23. DOI:10.1017/S0950268820001582.
- [9] Barret A-S, Charron M, Mariani-Kurkdjian P, et al., group of investigators. Shopper cards data and storage practices for the investigation of an outbreak of Shiga-toxin producing *Escherichia coli* O157 infections. *Med Mal Infect.* 2013;43(9):368–373.
- [10] Gieraltowski L, Julian E, Pringle J, et al. Nationwide outbreak of *Salmonella* Montevideo infections associated with contaminated imported black and red pepper: warehouse membership cards provide critical clues to identify the source. *Epidemiol Infect.* 2013; 141(6):1244–1252.
- [11] Jones G, Lefèvre S, Donguy M-P, et al. Outbreak of Shiga toxin-producing *Escherichia coli* (STEC) O26 paediatric haemolytic uraemic syndrome (HUS) cases associated with the consumption of soft raw cow's milk cheeses, France, March to May 2019. *Euro Surveill.* 2019;24(22). DOI:10.2807/1560-7917.ES.2019.24.22.1900305
- [12] Miller BD, Rigdon CE, and Robinson TJ, et al. Use of global trade item numbers in the investigation of a *Salmonella* Newport outbreak associated with blueberries in Minnesota, 2010. *J Food Prot.* 2013;76(5):762–769.

- [13] Wilking H, Götsch U, Meier H, et al. Identifying risk factors for shiga toxin-producing *Escherichia coli* by payment information. *Emerging Infect Dis.* **2012**;18:169–170.
- [14] Keto-Timonen R, Markkula A, Halkilahti J, et al. Shopping detail information and home freezer sampling confirmed the role of commercial, modified-atmosphere packaged meatballs as a vehicle for Listeriosis in Finland. *Front Public Health.* **2019**;7:216.
- [15] Jalava K, Selby K, Pihlajasaari A, et al. Two cases of food-borne botulism in Finland caused by conserved olives, October 2011. *Euro Surveill.* **2011**;16(49):20034.