

**Written evidence submitted by Dr Anya Skatova, Dr Michelle Morris, Dr James Goulding, Prof Mark Birkin, Dr Nik Lomax, Dr Georgiana Nica-Avram and Prof Andrew Smith (DDA0045)**

**About the authors**

**Dr Anya Skatova**, UKRI Future Leaders Fellow and Director at Digital Footprints Lab; MRC Integrative Epidemiology Unit at the University of Bristol; Population Health Sciences, Bristol Medical School, University of Bristol; Turing Fellow, Alan Turing Institute.

**Dr Michelle Morris**, Associate Professor Nutrition and Lifestyle Analytics, Consumer Data Research Centre & School of Medicine, University of Leeds; Turing Fellow, Alan Turing Institute.

**Dr James Goulding**, Associate Professor of Analytics, Co-Director, Neodemographics Lab for Analytics in Business (N/LAB), Nottingham University Business School, University of Nottingham.

**Prof Mark Birkin**, Professor of Spatial Analysis and Policy, University of Leeds; Co-Director of the Consumer Data Research Centre and Leeds Institute for Data Analytics; Programme Director for Urban Analytics and Turing Fellow, Alan Turing Institute.

**Dr Nik Lomax**, Associate Professor Data Analytics for Population Research, Co-Director of the Consumer Data Research Centre, School of Geography, University of Leeds; Turing Fellow, Alan Turing Institute.

**Dr Georgiana Nica-Avram**, Assistant Professor in Business Analytics, Neodemographics Lab for Analytics in Business (N/LAB), Nottingham University Business School, University of Nottingham.

**Prof Andrew Smith**, Professor of Consumer Behaviour, Director, Neodemographics Lab for Analytics in Business (N/LAB), Nottingham University Business School, University of Nottingham.

## **Our reasons for submitting evidence**

Research using shopping data to better understand and improve population health is a rapidly emerging area, spanning multiple academic disciplines. It has significant potential benefits to society if undertaken responsibly, transparently, and in partnership with British citizens.

The authoring team are deeply embedded in this growing academic domain, reflecting world-leading expertise from the behavioural sciences, nutritional epidemiology, health geography, artificial intelligence research and data science.

In this document we share a broad range of evidence demonstrating the potential national benefit of responsible use of shopping history data as a novel source for health research; to provide social good; and to underpin healthcare innovations aimed at saving lives. This report focuses on the UK context of transparent sharing of such data between research institutions and the commercial organisations responsible for its collection.

## **Executive Summary**

This submission directly addresses the Committee's request for points on "*the potential benefits, including to research, to effectively use and share data between and across Government, other public bodies, research institutions and commercial organisations, and the existing barriers to such data sharing*".

It focuses on the use of shopping data in health research for public good. We identify the following key stakeholders: Academia, Industry, and the Public; and provide evidence to support our recommendations to establish a framework for effective sharing of digital footprint data, in the form of transactional shopping logs, between research institutions and retailers, mediated by consent of the public.

We empathise the importance of establishing guidelines and best community research practices for both academic and industry data controllers and the ethical responsibility to use these data for public good. All references not available through open access can be provided on request.

### **Key points**

- Shopping data hold a wealth of information relating to both individual and population health and wellbeing.
- The public has the rights to their personal data, including shopping history data recorded by the supermarkets and other retailers.
- The public are positive about donating such shopping data to research that helps improve the nation's health and wellbeing.
- It is retailers' civic duty to make shopping data available for research benefiting public good.

### **Key recommendations**

- Set expectations upon the retail industry to support sharing of the shopping data for public health research.
- Put the public at the heart of data sharing decisions, e.g., through data donation initiatives.
- Encourage collaboration between retail industry and academic research institutions through appropriate funding schemes.
- Facilitate and endorse community standards and accessible infrastructure for efficient, responsible and transparent data sharing.

## Using shopping data to improve population health

1. **Shopping history data recorded by supermarkets and other retailers is an existing resource about the UK's behaviours and choices.** The general public have a right to access and make use of this resource, yet it currently remains out of reach to them, while predominantly being used for commercial purposes.
2. **Shopping data hold a wealth of information about population health<sup>1 2 3</sup> and wellbeing.** <sup>4 5</sup>
3. Shopping data is an important type of a broader range of digital footprint data sources that play a significant role in social science and health research. Digital footprints and digital phenotypes more broadly, and shopping data specifically, offer great potential to advance population health research. <sup>6 7</sup>
4. Lifestyle choices are one of the major contributors to ill health. For example, 1 in 7 deaths in the UK are attributed to diet and lifestyle choices, yet we still struggle to capture dietary information in a timely manner. Shopping data present a real-time solution to understanding what food is bought for home consumption, at a scale

---

<sup>1</sup> Aiello, L. M. *et al.* (2019). Large-scale and high-resolution analysis of food purchases and health outcomes. *EPJ Data Science*, 8(1), 14. <https://doi.org/10.1140/epjds/s13688-019-0191-y>

<sup>2</sup> Bandy, L. *et al.* (2019). The use of commercial food purchase data for public health nutrition research: A systematic review. *PLoS One*, 14(1), <https://doi.org/10.1371/journal.pone.0210192>

<sup>3</sup> Skatova, A. *et al.* (2019). Those Whose Calorie Consumption Varies Most Eat Most. <https://psyarxiv.com/ah8jp/>

<sup>4</sup> Ljevar, V. *et al.* (2020). Exploration of links between anxiety purchases, deprivation and personality traits. *2020 IEEE International Conference on Big Data*. <https://doi.org/10.1109/BigData50022.2020.9378485>

<sup>5</sup> Dzogang, F. *et al.* (2017). Seasonal variation in collective mood via twitter content and medical purchases. In *International Symposium on Intelligent Data Analysis* (pp. 63-74). Springer, Cham. [https://doi.org/10.1007/978-3-319-68765-0\\_6](https://doi.org/10.1007/978-3-319-68765-0_6)

<sup>6</sup> <https://www.ukri.org/councils/esrc/facilities-and-resources/find-an-esrc-facility-or-resource/>

<sup>7</sup> Jain, S. H. *et al.* (2015). The digital phenotype. *Nature Biotechnology*, 33(5), 462-463. <https://doi.org/10.1038/nbt.3223>

<sup>8</sup> Jenneson, V. L. *et al.* (2021). A systematic review of supermarket automated electronic sales data for population dietary surveillance. *Nutrition Reviews*. <https://doi.org/10.1093/nutrit/nuab089>

<sup>9</sup> <https://www.clocproject.org.uk/>

<sup>10</sup> Dolan, E. H. & Skatova, A. (2021) Ovarian cancer, misdiagnosis and shopping for healthcare products: loyalty card data sharing study. *Webinar for Ovacome*. <https://www.youtube.com/watch?v=XWB5kakyBc>

<sup>11</sup> Birkin, M. *et al.* (2019). Creating a long-term future for big data in obesity research. *International Journal of Obesity*. 43(12), pp. 2587-2592, <https://doi.org/10.1038/s41366-019-0477-y>

<sup>12</sup> Dolan, E. H. (2022) Value of commercial sales data on respiratory death predictions using Model Class Reliance - commercial-data-healthcare-predictions. *Report for NHSX*. [https://github.com/nhsx/commercial-data-healthcare-predictions/blob/main/report/NHSX%20Report\\_Value%20of%20Commercial%20Product%20Sales%20Data%20n%20Healthcare%20Prediction.pdf](https://github.com/nhsx/commercial-data-healthcare-predictions/blob/main/report/NHSX%20Report_Value%20of%20Commercial%20Product%20Sales%20Data%20n%20Healthcare%20Prediction.pdf)

<sup>13</sup> Davies, A. *et al.* (2018). Using machine learning to investigate self-medication purchasing in England via high street retailer loyalty card data. *PLoS One*, 13(11), e0207523. <https://doi.org/10.1371/journal.pone.0207523>

previously not possible.<sup>8</sup> Importantly these “happenstance” shopping data are not subject to the same participation burdens and selection biases that exist within survey data or market research data (e.g., Kantar world panel). **Shopping history data can uncover relationships between lifestyle choices and major health outcomes like cancer<sup>9</sup> <sup>10</sup>, diabetes<sup>11</sup> and mental health** (e.g., change in alcohol consumption patterns through pandemic).

5. **Shopping data can contribute to a better understanding of early symptoms of disease** (e.g., systematic relationships between buying painkillers and Gaviscon when suffering chest pain can be an indicator for cardiovascular problems), or **exacerbation of disease** (e.g., indicators of arthritis or arm nerve injuries through detecting the switch to pre-chopped packs of vegetables) <sup>12</sup>.
6. Through analyses at population scale but with the ability to examine local and neighbourhood levels<sup>13</sup> due to the distribution of the shops across the country, **evidence from shopping data can help plan national and local healthcare more dynamically and with greater sensitivity**. It can also **help to produce tailored recommendations for individuals** (e.g., whether they need to reduce their sugar or alcohol consumption, or how many ready meals they are consuming in comparison to their peers).
7. Shopping data can help to better understand behaviours of hard-to-reach groups who often find themselves excluded from surveys. **Using shopping data** from such groups **can provide evidence to better understand the social determinants of ill health, identify and explain health inequalities**, and provide improved coverage and understanding of vulnerable communities.

### **Recommendations**

8. **Shopping history data has high public value** in many areas including healthcare planning, understanding lifestyle causes of diseases, explaining health inequalities and helping individuals to be healthier. To achieve these targets, **shopping data needs to be made available for research benefiting public good**.

### **Public rights and attitudes to sharing shopping data**

9. The public has the rights to their personal data, including shopping history data recorded by the supermarkets. **Members of the general public have low**

**awareness on what kind of data is collected as a part of their shopping history. The public are also largely unaware of their rights in relation to their shopping data.** Through right to data portability and right to subject access request, UK GDPR mandates that individuals should have control over personal data collected about them, including ability to request the data, authorise sharing it with a third party or authorise a third party (e.g., research organisation) to access their data.

- 10. The general public are largely unaware of the value that shopping data might have for benefiting the public good** including health research and healthcare planning, and for individuals' healthcare pathways (e.g., through improved personalised healthcare) <sup>14 15</sup>.
- 11. Public are unaware of the high level of protection that research organisations provide to safeguard research data.** For example, the Data and Analytics Research Environments UK (DARE UK) programme of UKRI, ADR UK and HDR UK was launched in July 2021 to design and deliver a co-ordinated, national data research infrastructure that demonstrates trustworthiness and supports research at scale for wider public good. <sup>16 17</sup>
- 12. When the value and safeguards of using shopping data for research is explained to the members of the public and when people are asked whether they would donate their data, the public are positive about shopping data being used for social good,** such as public health research<sup>18 19</sup>.

## Recommendations

- 13. It is important to continue informing the public about their rights in relation to shopping history data as well as to raise awareness of the utility of shopping data for improving the UK's health.** This can be done through creating resources to explain the value of shopping history data for research as well as the level of

---

<sup>14</sup> Skatova, A., & Goulding, J. (2019). Psychology of personal data donation. *PloS One*, 14(11), <https://doi.org/10.1371/journal.pone.0224240>

<sup>15</sup> Clarke, H. *et al.* (2021). Understanding barriers to novel data linkages: topic modeling of the results of the LifeInfo survey. *Journal of medical Internet research*, 23(5), e24236. <https://doi.org/10.2196/24236>

<sup>16</sup> [https://lida-data-analytics-team.github.io/laserdocs/docs/laser\\_info/laser.html](https://lida-data-analytics-team.github.io/laserdocs/docs/laser_info/laser.html)

<sup>17</sup> <https://www.data-can.org.uk/health-data/trusted-research-environment/>

<sup>18</sup> Dolan, E. H. *et al.* (2021). Public Attitudes Towards Sharing Loyalty Card Data for Academic Health Research: a Qualitative Study. <https://doi.org/10.21203/rs.3.rs-1103902/v1>

<sup>19</sup> Skatova, A. *et al.* (2019). Attitudes towards transactional data donation and linkage in a longitudinal population study: evidence from the Avon Longitudinal Study of Parents and Children. *Wellcome Open Research* 4. <https://dx.doi.org/10.12688/wellcomeopenres.15557.2>

protection that researchers adhere to (e.g., similar to *Understanding Patients Data*<sup>20</sup> created by Wellcome Trust to explain research with medical records).

- 14. It is important to encourage data donation initiatives and facilitate mechanisms where individuals can share their shopping history data** for important causes such as cancer or diabetes research. This type of data sharing means that people are at the heart of decision making about their data.

### **Potential benefits of effective use and data sharing between research institutions and commercial organisations**

- 15. Ability to respond quickly to system shocks, for example the Covid-19 pandemic.**<sup>21</sup> Shopping data reveal behaviours directly relevant to health outcomes, but also social interactions which too are indicative of our nation's health. Shopping data can be considered 'happenstance' but provide essential information about daily activities which is critical to responding quickly to emergencies. Evidence is clear that we need to be more 'data ready' to respond to emergencies or system shocks like the pandemic and this should include shopping data.<sup>22 23</sup>
- 16. Less wasted time and resources** both for public-funded research institutions and for companies allows for prompt and efficient response to healthcare challenges. For example, the COPI notices that made health data sharing much easier during the pandemic brought great benefits to the public good<sup>24</sup>.
- 17. Effective data sharing will provide reputational gains for retailers.** Sharing shopping data can contribute to the image of the retailer and help consumer retention. Results of research utilising shopping data can be beneficial to business strategies that support good health of the nation, e.g., through shop-level interventions to promote healthy eating.
- 18. Effective data sharing will give people higher control over their data, by allowing to access and authorise access to their data.** This will ensure that the public remain at the heart of controlling the use of their data and empowers them to

---

<sup>20</sup> <https://understandingpatientdata.org.uk/>

<sup>21</sup> Spooner, F. *et al.* (2021). A dynamic microsimulation model for epidemics. *Social Science & Medicine*, 291, 114461. <https://doi.org/10.1016/j.socscimed.2021.114461>

<sup>22</sup> <https://royalsociety.org/news/2021/09/global-pandemic-data-alliance/>

<sup>23</sup> <https://rs-delve.github.io/reports.html>

<sup>24</sup> <https://www.gov.uk/government/publications/coronavirus-covid-19-notification-of-data-controllers-to-share-information>

choose whether it can be used for public good, rather than solely for commercial gain as it is the case now.

19. **Improve existing understanding of populations**, e.g., Office for National Statistics analyses. These routinely collected shopping data have the potential to supplement more traditional survey data resources to provide a more complete picture of shopping behaviours and more broadly lifestyle decisions and choices, in near real time, across the whole population.<sup>25</sup>

## Recommendations

20. **Encourage collaboration between retail industry and researchers, e.g., through specialised funding schemes.**<sup>26 27</sup> Tesco<sup>28</sup> have made some steps to allow their customers to access and share their data with third parties – although this still relies on high individual engagement which is not always practical for research. Others<sup>29</sup> (e.g., Sainsburys'<sup>30</sup>) are collaborating with researchers but there is no systematic approach with each research project considered on case-by-case basis. The funding schemes need to adequately support the resource required from the retail industry to collaborate effectively.
21. **Facilitate development of better infrastructure to host datasets.** The retail industry does not have the data infrastructure to readily share the level of data required to inform health research. This has recently been identified by the independent National Food Strategy<sup>31</sup>, which recommended introducing “mandatory reporting for large food companies” and creating “a National Food System Data programme”, which would help make these data access requests possible. Shopping data is most useful when linked to health outcomes. Research with aggregated anonymised data sets cannot answer all questions but can still progress

---

<sup>25</sup> Morris M. A. *et al.* (2018). Can big data solve a big problem? Reporting the obesity data landscape in line with the Foresight obesity system map. *International Journal of Obesity*. 42 (12) <https://doi.org/10.1038/s41366-018-0184-0>

<sup>26</sup> <https://www.igd.com/charitable-impact/healthy-eating/content-library/article/t/igds-healthy-and-sustainable-diets-project-group/i/28089>

<sup>27</sup> <https://www.theconsumergoodsforum.com/health-wellness/healthier-lives/>

<sup>28</sup> Aiello, L. M. *et al.* (2020). Tesco Grocery 1.0, a large-scale dataset of grocery purchases in London. *Scientific data*, 7(1), 1-11. <https://www.nature.com/articles/s41597-020-0397-7>

<sup>29</sup> <https://www.igd.com/charitable-impact/healthy-eating/content-library/article/t/igds-healthy-and-sustainable-diets-project-group/i/28089>

<sup>30</sup> Clark, S. D. *et al.* (2021). Dietary Patterns Derived from UK Supermarket Transaction Data with Nutrient and Socioeconomic Profiles. *Nutrients*, 13(5), 1481. <https://doi.org/10.3390/nu13051481>

<sup>31</sup> <https://www.nationalfoodstrategy.org/>

understanding of population behaviours<sup>32</sup>. **New infrastructure needs to support individual level data linkage.**

### **Barriers to sharing shopping data for research**

22. Currently, shopping history datasets are mostly locked in industry silos. Below we outline major barriers for sharing these data with **the public, the retail industry, and the academic researchers.**

#### **Barriers faced by the public**

23. **Exercising subject access request and data portability rights is obscured on supermarkets and other retailers' websites.** Individuals do not currently have practical and transparent means to exercise these rights in relation to their shopping data in cases of most supermarkets. Tesco PLC is one example of allowing their customers to exercise right for data portability easily<sup>33</sup>. Specifically, customers can download their data in machine-readable format and/or share their data with a third party.

#### **Barriers faced by the industry**

24. **Supermarkets and other retailers do not have clear incentives to share shopping data for health research.** At the same time perceived reputational risks for the retail industry are high.

25. **Retailers do not have infrastructure to support efficient data sharing.** Even when companies are positive about sharing shopping data for research, the trade-off between costs, time spent, and perceived impact achieved does not presently support a valid business case.

26. **There are no standardised protocols or guidelines for retailers** on how to share the data for health research.

#### **Barriers faced by researchers**

27. **There are no clear standards in the research community of how shopping history data can be shared or/and used.** There are a handful of research projects

---

<sup>32</sup> Fard, N. A. *et al.* (2021). On the interplay between educational attainment and nutrition: a spatially-aware perspective. *EPJ Data Science*, 10(1), 1-21. <https://doi.org/10.1140/epjds/s13688-021-00273-y>

<sup>33</sup> <https://www.tesco.com/data-portability/en-GB/>

that have used shopping history data for health research that can help to draw up a required best practice framework.

28. **There is no framework for governance of shopping history data including how to share it between industry and researchers.** Each health research project using shopping data on a case-by-case basis and largely dependent on personal connections of investigators, and personal enthusiasm to facilitate data sharing.
29. **Legal guidelines on how to share the data between researchers and industry are not well-defined.** For example, it is not clear what form of consent is acceptable (e.g., GDPR consent vs common law consent), and how the identity of the individual needs to be confirmed (e.g., it is not practical to collect copies of IDs as a part of the research process).
30. **Data sharing infrastructure is not supported by most supermarkets and other retailers.** One of the exceptions is Tesco PLC as it allows individuals to share their data with third parties. However, such data sharing comes at high burden to individuals and is not practical for population level research.

## **Recommendations**

31. **There needs to be sector standards on how individuals should be able to exercise their right for data portability and subject access request,** whether they want to access their data for themselves, share it with a third party (e.g., researcher) or authorise third party to access their data. The process needs to be made easy and intuitive.
32. It is vital for the government to set expectations with regards to use and governance of shopping history data. It should communicate that **shopping history data, in part, is a public resource with high value to improve population health.** The government should encourage and reward sector leaders/businesses who have already shared their data for health research or created infrastructure that can facilitate such sharing (e.g., Tesco, Sainsbury's). Retailers are fearful and uncertain of being the first to set up data sharing infrastructure, given a lack of guidelines and best practice standards. If the sharing of shopping data for research in the public interest become a sector norm, it will bring transparency and clarity.

33. **There is a need for community standards for data sharing, data governance and use, including legal and infrastructural aspects.** These need to be endorsed by relevant public bodies (e.g., Centre for Data Ethics and Innovation, Information Commissioner Office) in collaboration with academic expertise in the domain. Such standards will reduce opportunity costs for both industry and researchers. A clear *Information Governance Framework* is required for researchers and industry data scientists who are not necessarily information governance experts. There are examples to draw on from other types of standalone data (e.g., see recommendations in the context of data structure to prevent harms to health from gambling)<sup>34</sup> and with data linkages (e.g., health<sup>35</sup>, administrative).

**Key Recommendations: Support better data governance and collaboration between researchers and industry**

34. It has been nearly 4 years since GDPR came into effect changing consumer rights with regard to the data collected on them. If commercial companies were going to facilitate a streamlined and public-facing mechanism to access this data and share for public health benefit, they would have done so already. In lieu of a ‘bottom up’ pressure to create such mechanisms, **a ‘top down’ governmental approach is needed to provide clear incentives for companies to establish effective mechanisms to support data sharing health research.**
35. Through incorporating shopping data into ‘data readiness’ recommendations and governance procedures for data linkages, researchers and industry will have a framework and guidance to support better collaborations. **Support and approve good practice guidelines that make it clear how various stakeholders (e.g., supermarkets and other retailers, academia) can share these data for research efficiently.**
36. **Reinforce regulations around consumer data and consumer rights,** so individuals can share their data for research. This is possible through strengthening current legislation on rights to data portability and subject access request and by supporting streamlined infrastructure for easy access and sharing of personal data.

---

<sup>34</sup> Lomax N. (2019). Independent repository of gambling industry data – a scoping study. [https://assets.ctfassets.net/j16ev64qyf6l/6CukrZUkyxTKV8923HaTsO/0fa97f492bf0a81d08d2c32966021b82/Independent\\_Repository\\_of\\_gambling\\_industry\\_data.pdf](https://assets.ctfassets.net/j16ev64qyf6l/6CukrZUkyxTKV8923HaTsO/0fa97f492bf0a81d08d2c32966021b82/Independent_Repository_of_gambling_industry_data.pdf)

<sup>35</sup> Crossfield, S. S. *et al.* (2022). A data flow process for confidential data and its application in a health research project. *Plos one*, 17(1), e0262609. <https://pubmed.ncbi.nlm.nih.gov/35061834/>

37. **The public needs to remain at the heart of data sharing and data linkages** and therefore shopping data donation platforms should be established. These should focus on consent, responsible use, and transparency.
38. **Set an expectation for companies to share these shopping datasets for public good as a part of their civic duty.** Companies who collect personal data have a social, legal and moral responsibility to share this co-owned data for public benefit - not merely use it for commercial purposes. Commercial companies need to be made aware of their moral and legal right to share the data by reinforcing the ethical responsibility to use existing resources created by the general public to benefit the general public.

***February 2022***