

*Polish Migrant Essential Workers  
in the UK during COVID-19:  
Survey Data, 2021*

**Technical report**

UKRI/ESRC-funded research project  
on Polish essential workers in the UK  
(ref. ES/V015877/1)

Authors:

Aneta Piekut, Anna Gawlewicz, Kasia Narkowicz,  
Sharon Wright and Paulina Trevena

Sheffield, July 2023

## Contents

1. About the study .....	3
2. Target population .....	3
3. Sampling techniques .....	3
4. Survey fieldwork .....	4
Survey calendar .....	4
5. Sampling via Facebook Advertisement Campaign .....	4
Facebook advertisement specification .....	5
Campaign structure / Ad sets .....	5
Changes in the Facebook ad campaign .....	6
The design of one advertisement .....	7
6. The questionnaire .....	8
Survey settings .....	8
Questionnaire structure .....	8
Survey flow .....	9
Other survey practicalities .....	12
Survey management .....	12
7. Post-fieldwork data preparation .....	12
Data cleaning and quality check .....	12
Data anonymization .....	13
Qualitative data coding .....	13
Correction to question f1 .....	14
Missing data .....	14
Question with multiple options .....	14
8. Data description .....	14
Dataset 1 – <i>Clickers</i> .....	15
Dataset 2 – <i>Respondents</i> .....	15
Survey metadata .....	17
New variables in the <i>Clickers</i> dataset .....	18
New variables in the <i>Respondents</i> dataset .....	19
9. References .....	21
10. Appendix .....	23
Table A1. Polish population in the UK, age by years, in 2019 .....	23
Figure A1. Pictures used in the Facebook ads in a carousel .....	25

## 1. About the study

The survey was prepared as a part of a mixed-methods project “Health, social, economic and cultural impacts of COVID-19 on migrant essential workers in the UK” (Wright et al. 2023). The project run from 6 November 2020 until 5 May 2023. The project collected data through an online survey to map Covid-19 impacts on Polish migrant essential workers (PMEW) in the UK, online interviews with Polish essential workers and expert interviews with key stakeholders providing information and support to migrant workers in the UK.

The project received funding from the Economic and Social Research Council (ESRC), which is part of UK Research and Innovation (UKRI). For more information, please visit the project UKRI website: <https://gtr.ukri.org/projects?ref=ES%2FV015877%2F1#/tabOverview>.

## 2. Target population

The study aim was to find out how the Covid-19 pandemic affected various spheres of life of Polish migrant essential workers in the UK. The survey target population were adult Polish people (aged 18+) who have worked in the essential work sectors in the UK during the Covid-19 pandemic (since March 2020) and lived in the UK for at least three months in 2020. To take part in the study, participants could only have Polish citizenship or had Polish and another citizenship (e.g., British), or used to hold Polish citizenship in the past.

Following UK government guidance, we prepared an operational definition of essential workers (also referred to as ‘key workers’ or ‘frontline workers’):

*As essential work we understand the type of work that is essential for keeping British society and economy running during the pandemic. Hence, essential workers are for example health and care workers, staff maintaining public safety (e.g., police, fire service) or delivering pivotal services (e.g., public transport, childcare for essential workers, food production and sale, product delivery), among others. In the study, we are interested in Polish essential workers in the UK who are either formally (self-)employed or work informally (e.g., cash-in hand work).*

We used the following government guidance regarding who the essential workers were:

- UK Government, 8 January 2020, Children of critical workers and vulnerable children who can access schools or educational settings

<https://www.gov.uk/government/publications/coronavirus-covid-19-maintaining-educational-provision/guidance-for-schools-colleges-and-local-authorities-on-maintaining-educational-provision>

- Scottish Government, 11 May 2020, Coronavirus (COVID-19): prioritisation matrix for key workers to be tested

<https://www.gov.scot/publications/coronavirus-covid-19-access-to-testing/>

## 3. Sampling techniques

The survey participants were recruited through multiple channels in order to diversify the sample and improve its representativeness. An accessible sampling frame containing all ‘essential workers’ or Poles living in the UK, which could be used to randomly sample only Polish essential workers, does not exist. We inquired whether the Yougov online panel could be used in the study, but the company explained that they did not have enough Polish UK

residents in their panel. Hence, we used sampling techniques suitable to recruit for so called 'hard-to-reach' populations.

### **Convenience and snowball sampling**

Our partner organisations (Polish Expats Associations, Fife Migrants Forum, PKAVS Minority Communities Hub and Polish Social and Cultural Association) supported us in survey recruitment. We also recruited via Facebook advertisements, the most popular social media site, since this recruitment method was successfully applied in migration studies before (Pötzschke & Braun 2017; Pötzschke & Weiß 2020) and with other hard-to-reach populations (Borodovsky et al. 2018; Whitaker et al. 2017).

## **4. Survey fieldwork**

The project started in November 2020. Survey design and the fieldwork were the first major tasks of the project and we applied for Ethics Review in December 2020, after a first version of the questionnaire was prepared. January and early February were devoted to improving the flow of the questionnaire, piloting it, translating into Polish language and inserting questions into the online survey software we used - Qualtrics.

### **Survey calendar**

- Survey design
  - Questionnaire and sampling design: November-December 2020
  - Ethical approval obtained: 13 January 2021
  - Pilot study: 25 January – 19 January 2021
  - Translation into Polish language: 5-8 February 2021
  - Imputation of questions into Qualtrics software: 8-12 February
  - Technical testing: 15-19 February 2021
- Survey fieldwork
  - Convenience sampling: 22 February - 12 April 2021
  - Facebook Advertisement Campaign: 27 February - 26 March 2021

## **5. Sampling via Facebook Advertisement Campaign**

Facebook social network site was used to run a dedicated campaign targeting Polish migrants in the UK. The ads were only in Polish language and presented with a carousel of pictures with essential work occupations allowing indirect targeting of PMEW. After clicking on an ad, interested Facebook users were redirected into our Qualtrics survey page.

The Facebook ads were distributed via the official Facebook page of the University of Glasgow, so promoted by a well-known organisation page. Additionally, this webpage was managed at that time by an external marketing company (SRMS), who supported us in setting up the ads and professional management of the campaign.

The initial available budget for the campaign was £1,500, which we spread evenly across the study period, with an initial testing phase for 72 hours to allow the Facebook algorithm to learn where the ad was most effective in generating clicks on the ads and for us to see if we are receiving survey responses. Some additional funds became available during the fieldwork, and we increased the campaign budget to £1,718.

## Facebook advertisement specification

### Targeting specification summary:

- Polish people living in the UK
  - Geolocation: UK
  - Behaviour: Lived in Poland (formerly ex-pats – Poland)
- Aged 18-65 (excluding pension age groups, who are not likely to work anymore)

### Age cohorts

Table A1 in Appendix 1 overviews the age structure of the Polish population in the UK, who were aged from 16 to 64 years in 2019, according to the Annual Population Survey (APS, Polish nationality N=2,271; ONS 2020a). Based on these estimates, Poles in the UK were divided into three age cohorts of similar size, taking into account that the population will be older in 2021:

- aged 18-31 (33.6%),
- aged 32-39 (35.1%),
- 40 years old and more (31.3%).

### UK countries

According to the same data, 85% of Poles aged 16-64 in 2019 lived in England, 10% - Scotland, and the rest in other countries in the UK (ONS 2020a). We wanted to collect larger sample sizes outside England, in order to be able to produce more reliable statistics for all UK countries. As such, we used disproportionate sampling, and the same number of ads set was prepared for England and the rest of the UK combined.

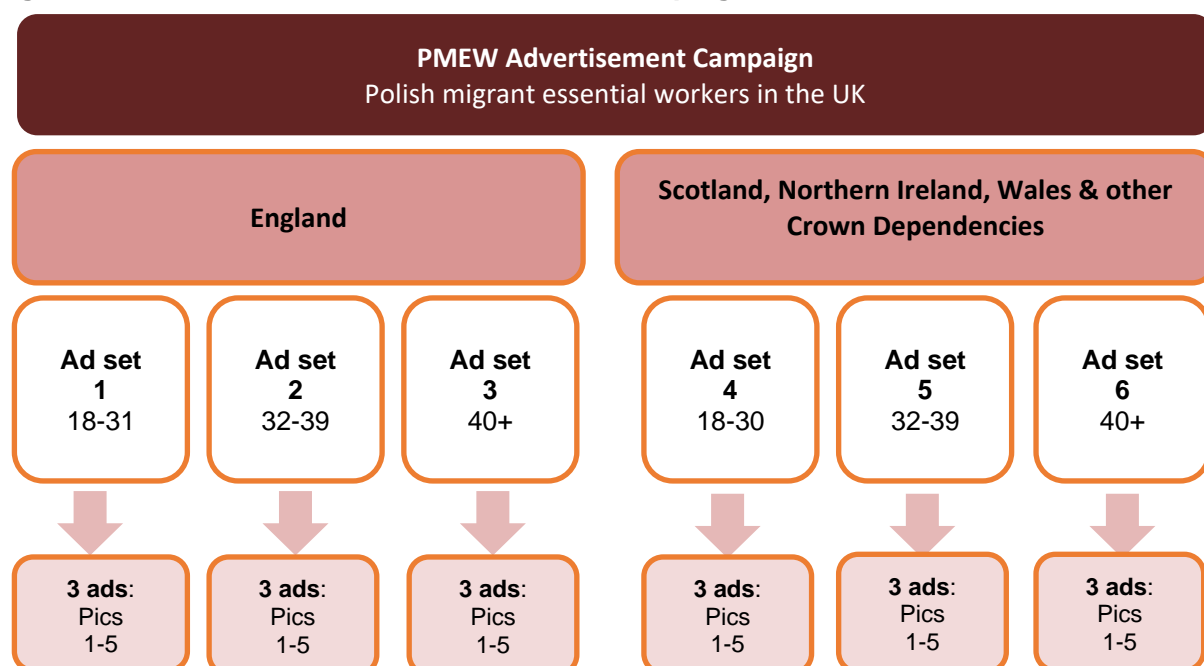
### Ad specifications

- Pricing model: payment per click
- Campaign objective - Traffic (to external website – survey in Qualtrics)
- Pricing model - Cost per click
- Buying type: Auction
- Campaign structure: 6 ad sets: 2 regions (England or the rest of the UK) x 3 age groups

## Campaign structure / Ad sets

The Facebook Advertisement Campaign consisted of three elements: the ads which have a text and a visual element (a picture or a video), an ad set which can hold a few different ads and is used to target specific demographic or interest groups, and the campaign itself. Campaigns, the highest level, are defined by a specific objective, such as raising awareness of an issue, generating traffic to an external website (used here), selling a product or engaging Facebook users via the Meta products: Messenger, Instagram and WhatsApp. Ad sets are also used to define timeframe, budget and ad channels of the campaign (Kühne & Zindel 2020).

**Figure 1. The structure of the advertisement campaign**



The graph in Figure 1 summarises the sampling strategy for the Facebook Ad Campaign:

- Ad sets were defined according to the UK country (2: England / the rest) and 3 age cohorts.
- In total, there were 6 ad sets – 3 broad age cohort groups targeted in England, and the same 3 age cohorts in the rest of the UK.
- Each ad set used had three ads with five pictures in a carousel (see photos numbers in Appendix 2). The Facebook algorithm would decide which one was more effective in attracting clicks and use that one more often. The order of the photos was:
  - Ad 1: 1, 2, 3, 4, 5 (disabled after 3<sup>rd</sup> of March, to attract more male users);
  - Ad 2: 5, 4, 3, 2, 1;
  - Ad 3: 3, 4, 5, 2, 1.

### Changes in the Facebook ad campaign

The initial budget per each ad set (1-6) was £200; so, at the start of the campaign we allocated £1,200 out of available £1,500, leaving £300 as a backup money. After one week we switched off targeting women for 10 days, because about 80% of respondents identified as women. On day 18, we changed the campaign duration from 3 to 4 weeks and distributed the remaining £300 unevenly, by placing more resources on the ad sets outside England, £75 per ad set, and £25 into each ad set in England.

Some additional funds became available in the last days of the campaign, and we hoped to improve sampling in Northern Ireland and Wales, so we placed £72.66 per ad set outside England, but switched off sampling Scotland there. In total we spent £1,718 for ads. Table 1 summarises changes made in the campaign over time.

**Table 1. Summary of changes made in the specifications of the Facebook ad campaign**

Sampling days	Change in campaign settings	Reason for change
Day 1-7	All people sampled meeting the targeting criteria, 6 ad sets x £200	Start of the Facebook ad campaign
Day 8-18	Switched off sampling women	Men underrepresented
Day 19-25	All sampled, campaign extended into 4 weeks, £300 added to the campaign budget	Distributing the remaining funds
Day 26-28	Switched off sampling Scotland, £218 added to the campaign budget	Additional budget available and hoping to boost sample size in Northern Ireland and Wales

### The design of one advertisement

We decided to run the campaign in Polish language only. Running the parallel campaign in English would mean spreading the budget thinly across 12 ad sets. In order to lower the selection bias, we used neutral wording in the text of the ad (Zhang et al. 2020). Below we present text of each ad element. Wording in red is the original ad texts in Polish language, while in blue – a translation into English. Figure 2 presents the same information visually.

#### Our ad elements:


- Profile picture - the University of Glasgow logo
- Profile page: the University of Glasgow.
- Primary text up to 125 characters:
  - Jesteś Polakiem/Polką [PL Flag] i mieszkasz w UK [UK Flag]? Powiedz nam w jaki sposób pandemia Covid-19 wpłynęła na Twoje życie
  - Are you Polish [Flag] and work in the UK [Flag]? Please tell us how the Covid-19 pandemic impacted your life
- Headline up to 40 characters:
  - Zapraszamy do badania
  - Do you live in the UK? Take part in a study
- Link description up to 35 characters:
  - 15-min. ankieta online
  - 15-minute online survey
- Link to the external, promoted page - link to our survey (see Survey distribution)
- 'Learn more' button - project website.

#### Ad pictures

- Photos: a selection of photos of various essential worker occupations: a female health worker, a male female cleaner, a male driver, a male delivery person and a female cashier. See copies of the selected photos in Appendix 2.
- Text embedded in the picture:
  - Jesteś z Polski? Migracja, Covid-19, Praca (wybrane sektory). Wygraj £100 lub £50
  - Are you from Poland? Migration, Covid-19, Work (selected sectors). Win £100 or £50

**Figure 2. One ad design – an example**

**Main text in Polish:**  
‘Are you a Pole [PL flag] and live in the UK [UK flag]? Say how the Covid pandemic affected your life’



**Text embedded in the picture:**  
‘Are you from Poland? Migration, Covid-19, Work (selected sectors). Win £100 or £50’

**Headline**  
‘We invite you to research’  
‘15-minute survey’

**Zapraszamy do badania**  
15 min. ankieta on...  
[LEARN MORE](#)

**Carousel with five pictures** (with the same text)

## 6. The questionnaire

### Survey settings

**Software:** Qualtrics, Sheffield Methods Institute, University of Sheffield account: <https://shefmethods.eu.qualtrics.com/>

**Language:** One questionnaire available in two language versions: Polish (default) and English. Option to switch the language at any point.

### Preview copy of the questionnaire:

[https://shefmethods.eu.qualtrics.com/jfe/preview/previewId/e3006d1c-31a6-4f93-ae43-70a7e6969291/SV\\_3CNCjfj8iNGygIK?Q\\_CHL=preview&Q\\_SurveyVersionID=current](https://shefmethods.eu.qualtrics.com/jfe/preview/previewId/e3006d1c-31a6-4f93-ae43-70a7e6969291/SV_3CNCjfj8iNGygIK?Q_CHL=preview&Q_SurveyVersionID=current)

### Questionnaire structure

The questionnaire started with an introduction, providing detailed information for participants and an option to give a consent to take part in the survey. After agreeing to take part in the survey, respondents were asked a series of questions about their migration status and employment during the pandemic in order to determine their eligibility to participate. Only those who held Polish citizenship, worked in the UK for at least three months in 2020, and worked in one of the essential work sectors were eligible to take part.

Overall, the survey was divided into 10 thematic blocks of questions (X, A-J) and consisted of 40 questions.



## Questionnaire blocks

- X. Study information (1 question / consent)
- A. Screening questions (4 questions)
- B. Employment situation (5 questions)
- C. Work experiences during the Covid-19 pandemic (5 questions, incl. 3 multi-item)
- D. Health impacts of the Covid-19 pandemic (4 questions, incl. 1 multi-item)
- E. Household and caring responsibilities (8 questions, incl. 1 multi-item)
- F. Economic impacts of the Covid-19 pandemic (2 questions)
- G. Migration status and plans (6 questions, incl. 1 multi-item)
- H. Demographic profile (5 questions)
- I. Open question (1 question)
- J. Further Consents (a separate Qualtrics survey – 4 questions)

At the end of the survey respondents had an option to take part in a draw to win two £100 vouchers and four £50 vouchers and leave email addresses to take part in the qualitative interview. Participants were redirected to a separate survey, so no record was retained of how data from the main survey relates to the personal data identifiers (emails). 1,005 participants left emails for a voucher draw and 640 agreed to be contacted about the follow-up qualitative study.

### Preview copy of data collection survey:

[https://shefmethods.eu.qualtrics.com/jfe/preview/previewId/ad888861-651b-477b-83f6-d9d261cdd2f1/SV\\_9ZHBb3NjAsOMSHA?Q\\_CHL=preview&Q\\_SurveyVersionID=current](https://shefmethods.eu.qualtrics.com/jfe/preview/previewId/ad888861-651b-477b-83f6-d9d261cdd2f1/SV_9ZHBb3NjAsOMSHA?Q_CHL=preview&Q_SurveyVersionID=current)

Survey codebook provides an overview of variables generated by questions in sections A-I. Personal data from section J is not published.

## Survey flow

Data collection via the online survey was terminated if participants did not give consent or were not eligible to participate. Out of 36 questions in the sections B-J, 24 questions were asked to all participants. Some remaining questions were asked only to those in employment during the pandemic, self-employed or still living in the UK.

Table 2 overviews survey questions by block, applied survey logics and the question type:

- single answer question – SA – one answer was allowed,
- multiple answer question – MA – more than one answer was allowed,
- question matrix with more than one items – QM – usually with scales,
- open question – OQ – text entry box,
- question in multiple versions – MV – respondents could decide which version of the question to answer.

**Table 2. Survey flow and logics**

Survey block	Question code	Question description	Question type	Who answered (Logics)
X. Introduction	x1	Information about the study	-	All
X. Introduction	x2	Consent Form	SA	All
A. Screening questions	a1	Polish citizenship status	SA	All
A. Screening questions	a2	UK residency	SA	All
A. Screening questions	a3	Number of months spent in the UK in 2020	SA	All
A. Screening questions	a4	Work sector in the UK during Covid-19 (C19) pandemic	SA	All
B. Employment situation	b1a	Current employment	MA	In the UK
B. Employment situation	b1b	Last employment in the UK	MA	Left the UK
B. Employment situation	b2	Employment situation change	MA	All
B. Employment situation	b3	Worry about the C19 pandemic impacts on business	SA	In the UK, self-employed
B. Employment situation	b4	Self-Employment Income Support Scheme	SA	In the UK, self-employed
C. Work experiences	c1	Main work during pandemic	OQ	All
C. Work experiences	c234	Work experience during the C19 pandemic (3 items)	MQ-SA	All
C. Work experiences	c5a	Job satisfaction (5 items)	MQ-SA	In the UK
C. Work experiences	c5b	Job satisfaction – last employment (5 items)	MQ-SA	Left the UK
C. Work experiences	c6	How workplace has been managed during the pandemic (2 items)	MQ-SA	In the UK, employed, self-employed or in unpaid work
D. Health impacts	d1	Covid-19 illness experience	SA	All
D. Health impacts	d2	Intention to get Covid-19 vaccine in the UK	SA	All
D. Health impacts	d3	Wellbeing change during C19 pandemic	SA	All
D. Health impacts	d4	C19 pandemic impacts (2 items)	MQ	All
E. Household and caring	e1_1	Number of adult (including the respondent) living in household	SA	All

E. Household / caring	e1_2	Number of children (under 18) living in household	SA	All
E. Household / caring	e2	Number of rooms in current accommodation	SA	All
E. Household / caring	e3	Access to green space	SA	All
E. Household / caring	e4	Living arrangement change due to C19 pandemic	SA	All
E. Household / caring	e5	Reason of living arrangement change	MA	If yes in e4
E. Household / caring	e6	Care provision (multi-answer)	MA	All
E. Household / caring	e7	Care provision - impacts	MQ-SA	If yes in e6
F. Economic impacts	f1	Covid-19 pandemic financial impact	SA	All
F. Economic impacts	f2	Household applied for benefits	SA	All
G. Migration status & plans	g1	Legal immigration status	SA	All
G. Migration status & plans	g2a	Number of years living permanently in the UK	SA	In the UK
G. Migration status & plans	g2b	Number of years living permanently in the UK (before left)	SA	Left the UK
G. Migration status & plans	g3	Next 5-year plans on living in the UK	SA	In the UK
G. Migration status & plans	g4	Decision to leave the UK	SA	If yes to g3
G. Migration status & plans	g5	C19 pandemic impacts on culture (2 items)	MQ-SA	All
G. Migration status & plans	g6	Support during the C19 pandemic from organisation(s) in the UK	SA	All
H. Demographic profile	h1	Birth year	SA	All
H. Demographic profile	h2	Gender	SA	All
H. Demographic profile	h3	Highest level of education (2 options: Polish and UK)	MV-SA	All
H. Demographic profile	h4	UK country of residency	SA	All
H. Demographic profile	h5	Income (3 options: weekly, monthly, annual)	MV-SA	All
I. Open question	I1	Tell us more about your life during the C19 pandemic	OQ	All

## Other survey practicalities

### Survey options

In order to maintain full participant anonymity (Borodowsky et al. 2018), we deactivated IP data collection in Qualtrics.

Option 'Prevent ballot box stuffing' was selected in Qualtrics, which uses cookies to determine a duplicate respondent.

### URL parameters for passive data collection:

We automatically collected data on through which distribution channel respondents accessed the survey and for Facebook ads which ad set they clicked on to be redirected to the survey. This was done by adding dynamic parameters after a question mark to the survey link and activating embedded data in Qualtrics (with the same names for generated variables).

- Language: Polish or English (parameter name - Q\_Language)
- Distribution channel (Dis)
- For Facebook data: Region (Reg), Age group (Age) and Picture (Pic)

An example ending of the link: ?Q\_Language=PL&Dis=8&Reg=1&Age=1&Pic=1

## Survey management

Project researchers monitored the number of responses by the main socio-demographics characteristics daily. We received a few emails from participants who could not fill in the survey claiming it closed down, and then we temporarily de-activated option 'Prevent ballot box stuffing', so they could access the link again.

## 7. Post-fieldwork data preparation

### Data cleaning and quality check

#### Quality check

Data was checked for duplicates and anomalies, and suspicious cases were removed from the final datasets:

- survey preview and recorded data before/after fieldwork dates: 81,
- responses marked by Qualtrics as spam and potential bot activity: 131,
- potential duplicates: 3 (same answers to all demographics).

#### Survey length

The minimum survey length for the final dataset with valid survey responses is 3.4 minutes, while maximum – 35,649 (25 days).

- 3.6% (n=40) of the sample filled in the survey in less than 5 minutes.
- 5.1% (n=57) of the sample filled in the survey in more than 30 minutes.

Long survey completion time is related to two situations: 1) breaking the survey and returning to it later (e.g., after a few days), or 2) a respondent not moving to the final survey page (with

a 'Thank you' note), but responding to all or most survey questions, and the survey automatically closing on the 12<sup>th</sup> of April.

In case of shorter surveys, it is hard to establish a definitive threshold for survey speeding and limited cognitive engagement with most of the questions. We tested that people who read fast and people not eligible to answer all questions could have filled the questionnaire in about 3 minutes.

## Data anonymization

In order to preserve anonymisation and avoid possible identification of participants after cross-tabulation of various variables, several continuous variables were re-coded into categorical variables after exploring frequency distribution, so small counts are avoided.

### Re-coding numerical variables

- a3 – Number of months lived in the UK in 2020 – grouped into four categories: 0-2, 3-6, 7-10 and 11-12 years.
- e1\_1 – Number of adult household members (including the respondent) – grouped into three categories: 1-3, 4-5 and above 6 people.
- e1\_2 – Number of children under 18 in household – grouped into three categories: none, 1-2, and above 3.
- e2 – Number of rooms (apart from the bathroom and kitchen) – grouped into four categories: none, 1, 2-4, and above 5.

### Re-coding variables measuring years

As the survey was conducted in early 2021, we subtracted year of arrival and birth year from 2020, not 2021. Variables asking about year we anonymised in the following way:

- g2a\_1 – Number of years living permanently in the UK – we subtracted provided year from 2020, and recoded into 0-5, 6-10, 11-15 and 16 and above.
- g2b\_1 – Number of years living in the UK before leaving UK – we subtracted provided year from 2020, and recoded into 0-5 and 6 and above.
- h1 – birth year – birth year was re-coded into 5-year age groups: 18-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64 and above 65.

## Qualitative data coding

- c1 – question asked respondents to describe their occupation in their own words. We used UK Standard Occupational Classification (SOC) developed in 2020 by the Office for National Statistics – SOC 2020 (ONS 2020b) and created two new variables:
  - c1\_soc2020\_major – SOC2020 major categories (9)
  - c1\_soc2020\_submajor – SOC2020 sub-major categories (23)Users of these new variables should note that it has high number of nonresponse (25%) due to vague or too generic wording used by some respondents to describe what they do
- e5\_9\_text – it was short text for 'Other' response option in the question asking about change in the accommodation. It was recoded into the following responses: 1 "Bought own property" 2 "Cannot afford to move out" 3 "Not paying rent" 4 "Someone moved in or out" 5 "Other".

- i1 – responses to the open question asked at the end of the survey were translated into English language and anonymised following the protocol developed for the qualitative part of this study (see Wright et al. 2023).

### Correction to question f1

Question f1 asked about change in respondents' financial situation on a scale from 1 to 5, where one was 'I'm much better off', and 5 – 'I'm much worse off'. Due to technical glitch when editing the questionnaire online, 5 was assigned to response 4 on a scale 'I'm a little worse off'. As such, the response values were recoded in the following way:

- I'm a little worse off – original value 5, recoded into 4.
- I'm much worse off – original value 4, recoded into 5.

### Missing data

We used the following convention to describe missing data:

- -90 – No answer (respondent did not leave any response, skipped the question).
- -97 – Prefer not to say.
- -99 – Not applicable (if not eligible to answer).

'Don't know' option was offered only in questions d2 and g6 and treated as not missing data.

### Question with multiple options

A few questions were asked in multiple versions and we advise to merge them before data analysis.

- b1 and c5 items – were differently worded for people who still lived in the UK and left the UK;
- h3 and h5 – offered respondents choice which version they wanted to answer.

The highest education level was measured using Polish (h3\_1) and British (h3\_2) education levels. However, a few responses provided responses into both options, so we advise to merge them by identifying the highest education level across the two variables.

Salary before tax and deduction in the last 12 months (h5) – there were three options to answer: weekly, monthly and annual. These brackets were based on the Annual Survey of Hours and Earnings for 2020, conducted by the Office for National Statistics (ONS 2021), and decimal points were rounded. We advise to create a new variable by merging categories 1-11 from h5a, h5b and h5c and treating the variable as an ordered, categorical variable.

## 8. Data description

We created two datasets from collected data. Figure 3 provides illustration of various types of participants that our data covered. One dataset contains all responses provided by participants that moved to the second page of the online survey, while the second one – for those who were classified as valid respondents only (provided enough responses for substantive data analysis).

## Dataset 1 – PMEW Clickers

- File name: PMEW\_Clickers\_dataset\_20230725
- File formats:
  - \*.sav (SPSS data format)
  - \*.csv (A Comma Separated Values)
- Number of variables: 115
- Number of observations: 2,061
- Unit of measurement: person
- Codebook: PMEW\_Clickers\_codebook\_20230725 in formats: \*.doc, \*.pdf, \*.odt

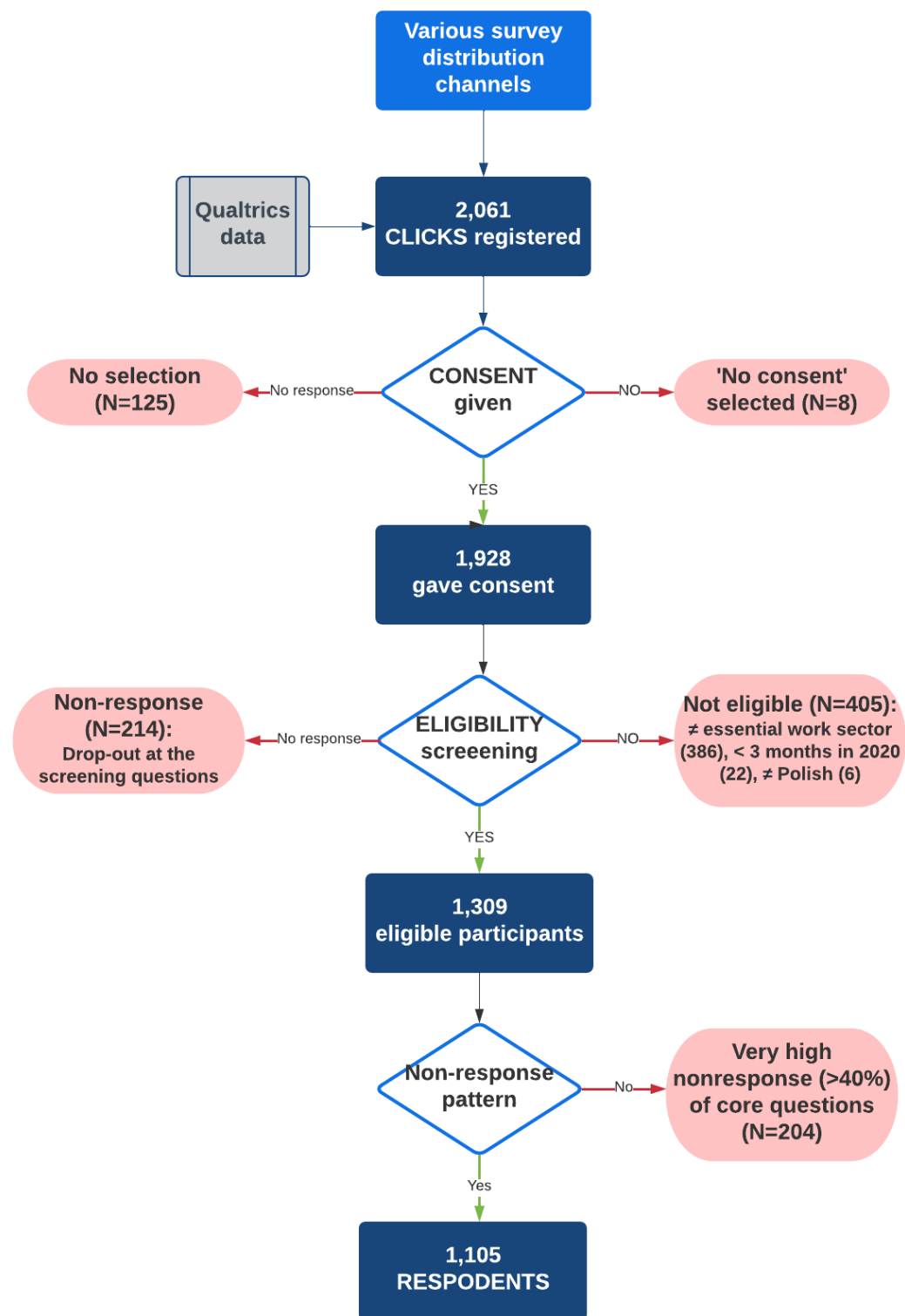
After landing on the online survey page, participants were welcomed by a page with an information about the study. This website did not record the number of visitors. Only when participants decided to move to the second page, it recorded their presence. In total, it was 2,061 'clicks', so we called this dataset 'Clickers'. There were a few more stages, when participants could drop-out from the study, and it was recorded in this data: giving consent, eligibility to study and providing enough responses. As such, this dataset should be used to conduct methodological analyses of survey quality rather than substantive analyses on the studied topics.

## Dataset 2 – PMEW Respondents

- File name: PMEW\_Respondents\_dataset\_20230725
- File format:
  - \*.sav (SPSS data format)
  - \*.csv (A Comma Separated Values)
- Number of variables: 113
- Number of observations: 1,105
- Unit of measurement: person
- Codebook: PMEW\_Respondents\_codebook\_20230725 in formats: \*.doc, \*.pdf, \*.odt

Only participant who responded to at least 60% of core questions (see details below), were counted as the final, valid respondents of the study. It was 1,105 of them, out of which 735 were recruited via the Facebook Ad Campaign. This dataset should be used to conduct substantive analyses on the topics covered by the study.

**Figure 3. Survey sample flowchart**



Below we describe how variables which were not created by collecting data in the questionnaire were generated.



## Survey metadata

### **Responseld**

Qualtrics generated, unique respondent identifier.

### **surveydate1**

Qualtrics generated, survey start date, format: DD-MM-YYYY.

### **surveydate2**

Qualtrics generated, survey end date, format: DD-MM-YYYY.

### **duration\_seconds**

Qualtrics generated, survey duration calculated in seconds based on information on the start and end of the survey.

### **distribution**

Qualtrics generated, embedded data based on dynamic URL parameter, labelled 'Survey distribution channel':

1. Project website
2. Project FB site and Facebook groups
3. Project Twitter
4. Private networks / Friends
5. Respondent via the end of the survey sharable link
6. Organisations / Partners
7. Facebook Ad Campaign

### **language**

Qualtrics generated. Embedded data indicating survey language used by a respondent: Polish (PL) or English (EN-GB).

### **region\_fb**

Qualtrics generated. Respondents recruited via Facebook Ad Campaign only, embedded dynamic URL data indicating sampling region: 'England' and the 'Rest of the UK'.

### **age\_fb**

Qualtrics generated. Respondents recruited via Facebook Ad Campaign only, embedded dynamic URL data indicating sampling age group: '18-31', '32-29' and '40+'.

## **picture**

Qualtrics generated. Respondents recruited via Facebook Ad Campaign only, embedded dynamic URL data indicating picture used in an ad (1-5, see Figure A1).

## **New variables in the *Clickers* dataset**

### **duration\_mins**

Calculated based on Qualtrics generated 'Duration\_\_in\_seconds\_' variable, by dividing it by 60.

### **adset**

Respondents recruited via Facebook Ad Campaign only. Variable based the crosstabulation of on region\_fb and age\_fb variables:

1. England - 18-31
2. England - 32-39
3. England - 40+
4. Rest of the UK - 18-31
5. Rest of the UK - 32-39
6. Rest of the UK - 40+

### **eligibility**

Variable based on responses to question a1, a3 and a4. It was labelled "Survey eligibility", and coded into three response categories:

- 1 - "Eligible to participate" – a respondent held Polish citizenship, lived in the UK in 2020 at least 3 months and worked in essential work sector.
- 2 - "Not eligible to participate" - a respondent did not meet one of the above criteria.
- 3 - "No response (drop out)" - a respondent did not answer
- 99 – Not applicable (If no Consent given)

### **distribution2**

Variable based on 'distribution' variable recording paradata, so participants recruited via Facebook ads are separated from other convenience sampling sources. It was labelled "Survey distribution channel - Facebook Ad / Other" and coded into two response categories:

- 1 - "Facebook Ad Campaign"
- 2 - "Convenience sampling"

### **participant\_type**

Using data from other variables and counting missing responses, we classified all participants into a few categories and labelled this variable "Survey participant type". The categories are:

- 1 "No consent to participate" – no reply or refused to give consent to x2

- 2 - "No response to the eligibility questions" – no response on the page asking about eligibility (a1, a3 and a4)
- 3 - "Not eligible to participate" – did not meet one of the eligibility criteria (a1, a3 and a4)
- 4 - "Non-valid response (>40% nonresponse)" – did not answer 11 or more out of 26 core items
- 5 - "Reluctant respondent (1-40% nonresponse)" – 1-10 out of 26 core items
- 6 - "Engaged respondent (0% nonresponse)" – answered all core survey questions

## **respondent\_type**

Participants were also classified into three groups of respondents: those who could not be classified as a valid respondents, those with some non-response to the core items, and those who answered all 26 core item. The variable was labelled "Survey respondent type", and coded:

- 1 "Non-valid response – a respondent had more than 40% nonresponse to 26 core items, or was not eligible to participate.
- 2 "Reluctant respondent" – a respondent did not answer up to 40% of the core survey items.
- 3 "Engaged respondent" – 0% nonresponse, a respondent answered all core survey items.

## **New variables in the *Respondents* dataset**

### **duration\_mins**

Calculated on the basis of Qualtrics generated Duration\_\_in\_seconds\_ variable, by dividing it by 60.

### **adset**

Respondents recruited via Facebook Ad Campaign only. Variable based the crosstabulation of on region\_fb and age\_fb variables:

- 1. England - 18-31
- 2. England - 32-39
- 3. England - 40+
- 4. Rest of the UK - 18-31
- 5. Rest of the UK - 32-39
- 6. Rest of the UK - 40+

### **weight**

As we oversampled outside England, for analysing the entire dataset and all countries, we recommend applying the weighting variable 'weight'. The variable was developed on the basis of proportions of people with Polish nationality across UK countries as reported in the Annual Population Survey, December 2019 (N=2,271; ONS 2020a). See details in Table 3.

If only data collected via Facebook ads is analysed, we recommend adding a new weight and we provide its values below, in column 6 'weight\_fb'.

**Table 3. Information about the weighting variable**

region	APS 18-64	MEW sample	weight	<i>MEW Fb ads sample</i>	<i>weight_fb</i>
England	85.08%	55.78%	1.53	61.33%	1.39
Scotland	9.4%	33.76%	0.28	25.41%	0.37
Other	5.52%	10.46%	0.53	13.26%	0.42

## **9. Guidelines on working with data**

Both datasets are saved in SPSS .sav format and open data format .csv.

### **Using .sav format data**

To preserve values created for missing data, we recommend opening the datasets in SPSS and saving into a new format, and then opening in a different software. For example, when importing sav data into Stata, the programme automatically replaces values defined as missing with a full stop used for system missing values in Stata.

### **Using .csv format data**

Due to simplified nature of a comma-separated values file only value labels were saved in this format. This dataset should be rather used for initial overview of datasets, but we recommend using the sav version, which can be imported to various type of statistical software, such as R or Stata.

Due to technical issues with saving the file, the qualitative response i1 was removed from the csv version of the data.

## 10. References

- Borodovsky, J. T., Marsch, L. A., & Budney, A. J. (2018). Studying cannabis use behaviors with Facebook and web surveys: methods and insights. *JMIR Public Health and Surveillance*, 4(2), e48.
- Grow, A., Perrotta, D., Del Fava, E., Cimentada, J., Rampazzo, F., Gil-Clavel, S., & Zagheni, E. (2020). Addressing Public Health Emergencies via Facebook Surveys: Advantages, Challenges, and Practical Considerations. *Journal of Medical Internet Research*, 22(12), e20653.
- Iannelli, L., Giglietto, F., Rossi, L., & Zurovac, E. (2020). Facebook digital traces for survey research: Assessing the efficiency and effectiveness of a Facebook Ad-based procedure for recruiting online survey respondents in niche and difficult-to-reach populations. *Social Science Computer Review*, 38(4), 462-476.
- Kühne, S. & Zindel, Z. (2020). *Using Facebook and Instagram to Recruit Web Survey Participants: A Step-by-Step Guide and Application*. Survey Methods: Insights from the Field, Retrieved from <https://surveyinsights.org/?p=13558>.
- Office for National Statistics, ONS, Social Survey Division. (2020a). *Annual Population Survey, January - December, 2019*. [data collection]. 4th Edition. UK Data Service. SN: 8632, <http://doi.org/10.5255/UKDA-SN-8632-4>.
- Office for National Statistics, ONS (2020b). *SOC 2020 Volume 1: structure and descriptions of unit groups*. Available online at: <https://www.ons.gov.uk/methodology/classificationsandstandards/standardoccupationalclassificationsoc/soc2020/soc2020volume1structureanddescriptionsunitgroups>
- Office for National Statistics, ONS (2021, February): *Annual Survey of Hours and Earnings, 2020 provisional edition of this dataset*. Available online at: <https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/earningsandworkinghours/datasets/allemployeesashtable1>.
- Pöttschke, S., & Braun, M. (2017). Migrant sampling using Facebook advertisements: A case study of Polish migrants in four European countries. *Social Science Computer Review*, 35(5), 633-653.
- Pöttschke, S., & Weiß, B. (2020, November 26). Employing Social Networking Sites in migration research: Preliminary findings of the German Emigrants Overseas Online Survey. Preprint: <https://doi.org/10.31219/osf.io/wnc4b>.
- Rosenzweig, L. R., Bergquist, P., Hoffmann Pham, K., Rampazzo, F., & Mildenberger, M. (2020, October 15). Survey sampling in the Global South using Facebook advertisements. Preprint: <https://doi.org/10.31235/osf.io/dka8f>.
- Whitaker, C., Stevelink, S., & Fear, N. (2017). The use of Facebook in recruiting participants for health research purposes: a systematic review. *Journal of Medical Internet Research*, 19(8), e290.
- Wright, S., Gawlewicz, A., Narkowicz, K., Piekut, A. & Trevena, P. (2023). *Polish Migrant Essential Workers in the UK during COVID-19: Qualitative Data, 2021*. [Data Collection]. Colchester, Essex: UK Data Service. Doi: <https://dx.doi.org/10.5255/UKDA-SN-856576>.

Zhang, B., Mildenberger, M., Howe, P. D., Marlon, J., Rosenthal, S. A., & Leiserowitz, A. (2020). Quota sampling using Facebook advertisements. *Political Science Research and Methods*, 8(3), 558-564.

## 11. Appendix

Table A1. Polish population in the UK, age by years, in 2019

Respondent age	Frequency	Percent	Cumulative percent	Age cohort in 2020/21 (+1 year)	Percent
16	27	1.21	1.21	18-31	33.59
17	23	1.02	2.23		
18	32	1.39	3.62		
19	38	1.67	5.29		
20	26	1.14	6.43		
21	24	1.06	7.49		
22	47	2.08	9.57		
23	38	1.66	11.23		
24	38	1.66	12.89		
25	58	2.54	15.43		
26	36	1.6	17.03		
27	48	2.14	19.16		
28	74	3.27	22.43		
29	72	3.15	25.58		
30	67	2.95	28.54		
31	115	5.05	33.59	32-39	35.14
32	97	4.26	37.84		
33	81	3.56	41.41		
34	98	4.3	45.71		
35	112	4.91	50.62		
36	106	4.68	55.30		
37	98	4.32	59.63		
38	96	4.22	63.85		
39	111	4.89	68.74	40+	31.26
40	88	3.9	72.64		
41	74	3.26	75.90		
42	58	2.55	78.45		
43	53	2.34	80.79		
44	51	2.25	83.04		
45	47	2.08	85.12		

46	45	1.99	87.12		
47	34	1.5	88.62		
48	34	1.5	90.12		
49	18	0.8	90.92		
50	20	0.88	91.80		
51	20	0.89	92.69		
52	15	0.66	93.35		
53	19	0.83	94.18		
54	20	0.89	95.07		
55	16	0.71	95.78		
56	14	0.64	96.42		
57	11	0.5	96.92		
58	12	0.51	97.42		
59	17	0.76	98.18		
60	12	0.54	98.73		
61	12	0.53	99.25		
62	7	0.32	99.57		
63	6	0.26	99.83		
64	4	0.17	100.00		
Total	2,271	100.00			100.00

Source: Annual Population Survey 2019: Note: Data weighted.



Figure A1. Pictures used in the Facebook ads in a carrousel

Photo 1



Photo 2



Photo 3



Photo 4



Photo 5

