











STUDY PROTOCOL

Million Migrants study of healthcare and mortality outcomes in non-EU migrants and refugees to England: Analysis protocol for a linked population-based cohort study of 1.5 million migrants [version 1; referees: 2 approved, 2 approved with reservations]

Rachel Burns ¹, Neha Pathak ^{1,2}, Ines Campos-Matos ³, Dominik Zenner^{4,5}, Srinivasa Vittal Katikireddi ⁶, Morris C. Muzyamba³, J. Jaime Miranda ^{7,8}, Ruth Gilbert^{2,9}, Harry Rutter ¹⁰, Lucy Jones¹¹, Elizabeth Williamson ¹², Andrew C. Hayward², Liam Smeeth¹³, Ibrahim Abubakar⁵, Harry Hemingway¹⁴, Robert W. Aldridge ^{1,3}

¹Centre for Public Health Data Science, University College London, London, UK

²Institute of Epidemiology and Healthcare, University College London, London, UK

³Public Health England, London, UK

⁴Migration Health Division, International Organization for Migration, Brussels, Belgium

⁵Institute for Global Health, University College London, London, UK

⁶MRC/CSO Social and Public Health Sciences Unit, Institute of Health and Wellbeing, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, UK

⁷CRONICAS Center of Excellence in Chronic Diseases, Universidad Peruana Cayetano Heredia, Lima, Peru

⁸School of Medicine, Universidad Peruana Cayetano Heredia, Lima, Peru

⁹Administrative Data Research Centre for England, University College London, London, UK

¹⁰Faculty of Humanities and Social Sciences, University of Bath, Bath, UK

¹¹UK programme manager, Doctors of the World, London, UK

¹²Faculty of Epidemiology & Population Health, London School of Hygiene and Tropical Medicine, London, UK

¹³Department of Non-Communicable Diseases Epidemiology, London School of Hygiene and Tropical Medicine, London, UK

¹⁴Institute of Health Informatics Research, Faculty of Population Health Sciences, University College London, London, UK

v1 First published: 17 Jan 2019, 4:4 (<https://doi.org/10.12688/wellcomeopenres.15007.1>)





Latest published: 17 Jan 2019, 4:4 (<https://doi.org/10.12688/wellcomeopenres.15007.1>)





Abstract

Background: In 2017, 15.6% of the people living in England were born abroad, yet we have a limited understanding of their use of health services and subsequent health conditions. This linked population-based cohort study aims to describe the hospital-based healthcare and mortality outcomes of 1.5 million non-European Union (EU) migrants and refugees in England.

Methods and analysis: We will link four data sources: first, non-EU migrant tuberculosis pre-entry screening data; second, refugee pre-entry health assessment data; third, national hospital episode statistics; and fourth, Office of National Statistics death records. Using this linked dataset, we will then generate a population-based cohort to examine hospital-based events and mortality outcomes in England between Jan 1, 2006, and Dec 31, 2017. We will

Open Peer Review

Referee Status:    

	Invited Referees			
	1	2	3	4
version 1 published 17 Jan 2019				
	report	report	report	report




1 **Jessica L. Potter** , Queen Mary University London, UK

compare outcomes across three groups in our analyses: 1) non-EU international migrants, 2) refugees, and 3) general population of England.

Ethics and dissemination: We will obtain approval to use unconsented patient identifiable data from the Secretary of State for Health through the Confidentiality Advisory Group and the National Health Service Research Ethics Committee. After data linkage, we will destroy identifying data and undertake all analyses using the pseudonymised dataset. The results will provide policy makers and civil society with detailed information about the health needs of non-EU international migrants and refugees in England.

Keywords

migration, migrant, refugee, health, healthcare, mortality

- 2 **Matthew Wallace** , Stockholm University Demography Unit (SUDA), Sweden
French National Demographic Institute (INED), France
- 3 **Sarah Rodgers** , Swansea University, UK
University of Liverpool, UK
- 4 **Ursula Trummer** , Center for Health and Migration, Austria

Any reports and responses or comments on the article can be found at the end of the article.

Corresponding author: Robert W. Aldridge (r.aldridge@ucl.ac.uk)

Author roles: **Burns R:** Conceptualization, Methodology, Project Administration, Writing – Original Draft Preparation, Writing – Review & Editing; **Pathak N:** Conceptualization, Methodology, Writing – Review & Editing; **Campos-Matos I:** Conceptualization, Methodology, Writing – Review & Editing; **Zenner D:** Conceptualization, Methodology, Writing – Review & Editing; **Katikireddi SV:** Conceptualization, Methodology, Writing – Review & Editing; **Muzyamba MC:** Conceptualization, Methodology, Writing – Review & Editing; **Miranda JJ:** Conceptualization, Methodology, Writing – Review & Editing; **Gilbert R:** Conceptualization, Methodology, Writing – Review & Editing; **Rutter H:** Conceptualization, Methodology, Writing – Review & Editing; **Jones L:** Conceptualization, Methodology, Writing – Review & Editing; **Williamson E:** Conceptualization, Methodology, Writing – Review & Editing; **Hayward AC:** Conceptualization, Methodology, Writing – Review & Editing; **Smeeth L:** Conceptualization, Methodology, Writing – Review & Editing; **Abubakar I:** Conceptualization, Methodology, Writing – Review & Editing; **Hemingway H:** Conceptualization, Methodology, Writing – Review & Editing; **Aldridge RW:** Conceptualization, Methodology, Project Administration, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: RWA and IA undertook paid consultancy work in support of the Doctors of the World 2017 Observatory report—Falling through the cracks: the failure of universal healthcare coverage in Europe.

Grant information: RWA is supported by a Wellcome Trust Clinical Research Career Development Fellowship (206602/Z/17/Z). NP is funded by a UCL Wellcome Trust Clinical Research Training Fellowship (211162/Z/18/Z). ACH's salary is provided by Central and North West London NHS Community Trust. AS is funded by UCLH Foundation Trust. The views expressed are those of the author(s) and not necessarily those of the NHS, the Wellcome Trust, the NIHR or the Department of Health, International Organisation for Migration or Public Health England. SVK acknowledges funding from a NRS Senior Clinical Fellowship (SCAF/15/02), the Medical Research Council (MC_UU_12017/13 & MC_UU_12017/15) and the Scottish Government Chief Scientist Office (SPHSU13 & SPHSU15). HH is a National Institute for Health Research (NIHR) Senior Investigator. His work is supported by: Health Data Research UK, which is funded by the UK Medical Research Council, Engineering and Physical Sciences Research Council, Economic and Social Research Council, Department of Health and Social Care (England), Chief Scientist Office of the Scottish Government Health and Social Care Directorates, Health and Social Care Research and Development Division (Welsh Government), Public Health Agency (Northern Ireland), British Heart Foundation and Wellcome Trust. The BigData@Heart Consortium, funded by the Innovative Medicines Initiative-2 Joint Undertaking under grant agreement No. 116074. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA; it is chaired, by DE Grobbee and SD Anker, partnering with 20 academic and industry partners and ESC. The National Institute for Health Research University College London Hospitals Biomedical Research Centre. JJM acknowledges having received support from the Alliance for Health Policy and Systems Research (HQHSR1206660), DFID/MRC/Wellcome Global Health Trials (MR/M007405/1), Fogarty International Center (R21TW009982, D71TW010877), Grand Challenges Canada (0335-04), International Development Research Center Canada (106887, 108167), Inter-American Institute for Global Change Research (IAI CRN3036), Medical Research Council (MR/P008984/1, MR/P024408/1, MR/P02386X/1), National Cancer Institute (1P20CA217231), National Heart, Lung and Blood Institute (HHSN268200900033C, 5U01HL114180, 1U01HL134590), National Institute of Mental Health (1U19MH098780), Swiss National Science Foundation (40P740-160366), Wellcome Trust (074833/Z/04/Z, 093541/Z/10/Z, 107435/Z/15/Z, 103994/Z/14/Z, 205177/Z/16/Z) and the World Diabetes Foundation (WDF15-1224).

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2019 Burns R *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution Licence](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Burns R, Pathak N, Campos-Matos I *et al.* Million Migrants study of healthcare and mortality outcomes in non-EU migrants and refugees to England: Analysis protocol for a linked population-based cohort study of 1.5 million migrants [version 1; referees: 2 approved, 2 approved with reservations] Wellcome Open Research 2019, 4:4 (<https://doi.org/10.12688/wellcomeopenres.15007.1>)

First published: 17 Jan 2019, 4:4 (<https://doi.org/10.12688/wellcomeopenres.15007.1>)

Introduction

In 2017, 8.6 million (15.6%) people living in England were born abroad, with 5.3 million (9.2%) born outside of the European Union (EU)¹. However, little is known about how these international migrants use England's National Health Service (NHS) or their subsequent health needs and mortality outcomes. Here we define international migrants as people born outside of England². This may include, for example, people who either have chosen to migrate (e.g. work, study, or join families) or those who may have been forced to migrate due to conflict, persecution or environmental disasters (e.g. refugees and asylum seekers).

A previous systematic review of healthcare usage in Europe showed that usage of accident and emergency (A&E) services by international migrants was high, but that screening and outpatient care usage was low³. Similarly, a recent study from Scotland showed that some ethnic minorities, whether migrants or their offspring, have higher levels of avoidable hospital admissions when compared to the non-migrant white Scottish population⁴. These findings suggest that migrants may be receiving poor quality primary and preventative healthcare or face barriers to accessing health services.

In England, no large scale studies have been conducted to date that were able to examine migrants' usage of hospital-based healthcare services. One primary care based study in England found that individuals who registered with a general practitioner (GP) for the first time when over the age of 15 - used as a proxy for international migrants - had half the rate of hospital admissions than the general population⁵. However, another study found that only one-third of new migrants had registered with a GP, complicating the use of this proxy measure as well as highlighting poor uptake of primary healthcare registration⁶. Other studies in England^{7,8} have used country of birth to examine migration, but did not have information on the date of migration, the country which an individual was migrating from, or the visa category under which they entered England. These studies illustrate the methodological challenges of accurately identifying migrants within national data sources and the lack of available information on their migratory history, thus limiting the interpretations of previous research.

Despite poor access to preventative healthcare, there is evidence that migrants have a mortality advantage compared to host populations in the high-income countries to which they migrate. We recently conducted a systematic review and meta-analysis on the global patterns of mortality data in international migrants⁹. Our review showed that the levels of mortality in international migrants, measured using standardised mortality ratios, were lower compared to the host population in the countries of destination for most disease causes. Two exceptions to these findings were an increase in mortality due to infectious disease and external causes of mortality⁹. As we found very little data on refugees, asylum seekers and other forced migrants, our data is most representative of international migrants in high-income countries who are studying, working or have joined family members. Although our findings supported the idea of the healthy migrant hypothesis - an empirically

observed mortality advantage of migrants relative to the host population¹⁰, there is evidence from migrants residing in England and Wales that suggests that this advantage declines with age^{7,8}. However, these studies were not able to assess whether this advantage also changed with duration of residence in England.

The evaluation of morbidity and mortality outcomes of migrants has been limited by difficulties in their identification in national data sources. To date, we have no data in England linking migrant and refugee health service usage and subsequent health conditions and mortality outcomes. There are existing disparate data sources on the health of international migrants; however, despite their existence, these data have not been integrated or analysed systematically. These data contain information that would enable us to comprehensively evaluate the health needs in migrants and refugees and develop evidence on how to improve access to hospital-based health services and preventative healthcare in England. The Million Migrant study will generate this evidence for the first time by linking records that contain data on the health outcomes for 1.5 million non-EU migrants and refugees.

Aim and objectives

The Million Migrant study will be a population-based cohort study that aims to examine secondary healthcare performance (e.g. quality and accessibility) and mortality in 1.5 million non-EU migrants and refugees in England. There are two main objectives of the study. First, to profile hospital-based healthcare performance by identifying existing health conditions and examining hospital admissions, readmissions and duration of admission of non-EU migrants and refugees compared to the general population in England. Second, to investigate mortality outcomes by health condition for non-EU migrants and refugees in comparison to the general population. This objective will examine whether or not our data replicates the mortality advantage of international migrants found in the literature. The study includes analyses that acknowledge the wider determinants influencing the health of international migrants such as the legal, social, economic and health structures and systems, health service access and support, exposures and behaviours, and epidemiological changes associated with population mobility (Figure 1).

Protocol

Data collection, processing, and linkage

Our study will link four data sources as outlined in Figure 2. First, non-EU migrant tuberculosis pre-entry screening data. This data set contain records on international non-EU migrants resident in a country where tuberculosis is common (40 cases per 100,000 people), and who are planning to come and live in United Kingdom (UK) for more than 6 months¹¹. Individuals in this dataset were screened by the UK pre-entry tuberculosis screening programme between Jan 1, 2006, and Dec 31, 2017. UK pre-entry tuberculosis screening was conducted either by the International Organization for Migration (IOM) or by international clinics recognised by the UK Home Office and quality assessed by Public Health England (PHE). Second, refugee pre-entry health assessment. Refugees undergo a health

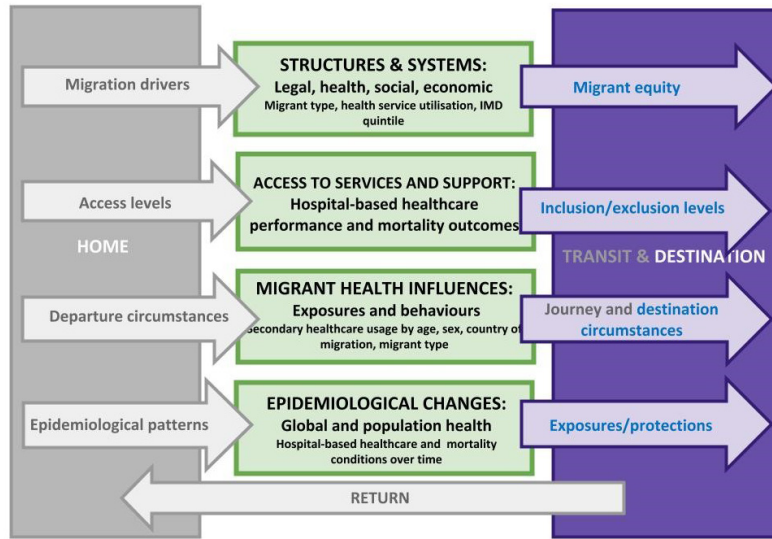
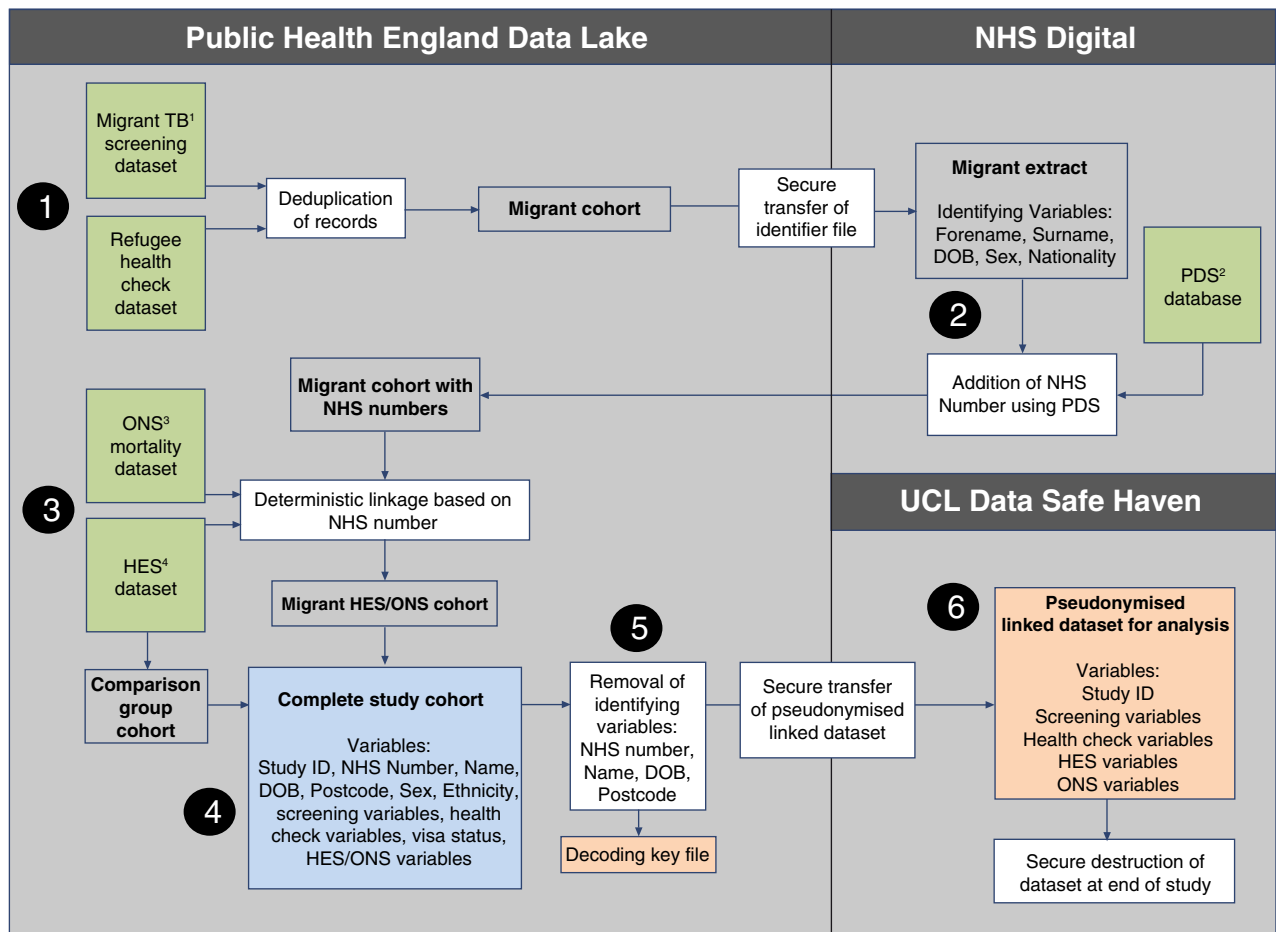


Figure 1. Conceptual framework for influences on migration and health adapted from UCL-Lancet Commission on Migration and Health¹². * The boxes and words in grey will not be measured directly in the Million Migrant study.



¹TB = Tuberculosis; ²PDS = Patient Demographic Service; ³ONS = Office of National Statistics; ⁴HES = Hospital Episode Statistics

Figure 2. Study data flows.

assessment that allows pre-departure information to be shared with local authorities and health services in the UK. Refugee health assessments were conducted by IOM between Mar 1, 2013 and Dec 31, 2017. Our study will therefore include two cohorts of international migrants to the UK - non-EU migrants and refugees that will be linked to the final two datasets from Jan 1, 2006, and Dec 31, 2017. Third, national hospital episode statistics (HES), including hospital admissions and attendances. Fourth, Office of National Statistics (ONS) death records which contain cause of death information on all deaths between Jan 1, 2006, and Dec 31, 2017. We focus on non-EU migrants because of their availability in this dataset, but also because EU migrants are more likely to dual-use health systems and less likely to need adaptation in terms of cultural competence¹³. The analysis will be limited to England as the HES data will only be obtainable for English hospitals.

We will obtain identifying variables (forename, surname, aliases, date of birth, sex, country of origin, country of departure, date of tuberculosis pre-entry screening or refugee pre-entry health assessment, and visa category) from all non-EU migrants and refugees under appropriate legal and ethical approvals. We will then use the Personal Demographics Service (PDS), the national electronic database of NHS patient information such as name, address, and date of birth, to identify and add NHS numbers by matching on these identifying variables to the tuberculous pre-entry screening and refugee health assessment records. Where available, NHS numbers that have already been added by PHE will be integrated. Once the non-EU migrants and refugees have been matched to their NHS number, we will then undertake deterministic linkage using NHS number to identify HES and ONS mortality records. Personal identifiers will then be removed, resulting in a pseudonymised linked dataset which will be securely transferred to UCL for data analysis, as outlined in [Figure 2](#). We will compare matching levels (e.g. whether an individual record is linked to their NHS number or not) across different age, sex, year and age of entry to UK and migrant country of origin groups. The non-EU migrant tuberculosis pre-entry screening dataset might contain multiple duplicate records for some individuals who require repeat tuberculosis screening. Duplicate records will be analysed on the basis of whether they occurred within 12 months of each other or not as per previously defined rules¹⁴ provided in full in [Extended data file 7](#). Non-EU migrant and refugee data are cleaned by the IOM epidemiology unit in coordination with clinics to ensure that records included all results on individuals screened and that any duplicate entries resulting from administrative error were removed or consolidated into one record. Cleaning and consistency checking of the final dataset will be undertaken by examining the distribution of variables, the range of individual variables, and missing data.

Ethics and information governance

To undertake this work, we require access to patient identifiable data without individual consent. We will apply to the Secretary of State for Health through the Confidentiality Advisory Group (CAG) to obtain approval for this work. After data linkage (see earlier section) we will destroy all identifying data and undertake all analyses using the pseudonymised dataset outlined in

[Figure 2](#). We will also seek ethical approval for the study from an NHS Research Ethics Committee. We believe that the benefits of this study for the migrant population outweigh any risks. The primary risk that could be anticipated is a data breach of sensitive information; we are developing a comprehensive data management and data sharing plan to minimise this such that no risks are anticipated. The dataset will be created and analysed by a team with extensive experience of handling large datasets securely, and the organisations involved in this work (UCL and PHE) have extensive Information Security and Governance procedures in place to minimise this risk.

Comparator groups

We will compare three groups in our analyses: 1) non-EU international migrants, 2) refugees, and 3) general population in England. We aim to disaggregate the non-EU migrants and refugees by: 1) age at migration, 2) sex, 3) ethnicity, 4) visa category, 5) country of origin, and 6) date of screening. We will examine each group over the follow up time period. For refugees, we will also examine country of departure as it is often different from the country of origin. Due to their smaller numbers, we will likely examine refugees by World Health Organisation (WHO) region of origin instead of country they migrated from. Final geographical categorisations used for migrant groups will be taken to minimise risk of disclosure and therefore migrants may be grouped into WHO sub-regions. Secondly, we will compare the non-EU migrants and refugees to the general population in England by deprivation level (Index of Multiple Deprivation - IMD - quintile). This will be done using an anonymous HES sample. Here the general population will be composed of mostly England-born residents, along with other types of international migrants who did not partake in one of the two screening programmes. These include EU migrants, international migrants that are not required to get a pre-entry tuberculous screening (e.g. migrants from low-tuberculous countries such as the United States), irregular migrants (e.g. undocumented), and migrants and refugees who arrived before the start of their subsequent screening programmes.

Outcomes

We present a series of outcomes for our healthcare and mortality analyses. We chose outcomes to ensure our analysis is consistent with those used in previous published analyses^{4,7,8}, in addition to outcomes that are of high interest to researchers but that previous studies were not powered to collect. Moreover, we have chosen a range of outcomes that collectively reflect the priorities of health policy makers as well as migrants and refugees who attended our patient engagement workshops on consent process, data linkage and analysis.

Hospital-based healthcare outcomes. We will profile the following hospital-based healthcare outcomes in non-EU migrants and refugees: 1) hospital attendances (inpatient, outpatient, and A&E), 2) hospital admissions (inpatient), 3) duration of hospital admission, and 4) 30 day emergency readmissions. We will explore these four outcomes by sub-conditions where appropriate. The clinical definitions and methodological approaches for each outcome are provided in [Table 1](#). Full clinical definitions of each outcome's subgroup are provided in

Table 2. HES currently uses ICD-10, the 10th version of the international classification of diseases and related health problems, to code for conditions and OPCS-4, the classification of interventions and procedures, to code all interventions and procedures. ICD-10 and OPCS-4 code lists for each are provided in [Extended data file 2–5](#).

Mortality outcomes. We will examine the following mortality outcomes in non-EU migrants and refugees: 1) death and 2) death due to a specific condition. The clinical definition and the methodological approaches for each outcome are provided in [Table 1](#). Full clinical definitions of each outcome's condition are provided in [Table 2](#). ONS datasets use ICD-10 codes to code for the health condition registered at death. ICD-10 code lists for each are provided in [Extended data file 2–Extended data file 5](#). Lastly, we aim to examine if the mortality advantage found in the literature can be replicated with this migrant and refugee population in England.

Entry and exit from cohort

We will include all visa applicants (non-EU migrants) from 101 countries (see [Extended data file 1](#)) with pre-entry tuberculosis

screening in this study and all refugees who underwent pre-entry health assessment.

Individuals will enter the cohort at whichever is the latest of: date at which they were screened for tuberculosis pre-entry screening or refugee pre-entry health assessment. Individuals will be followed up until the earliest of: end of the follow-up period (31st December 2018), emigration, or death. Individuals found to have tuberculosis at pre-entry screening will not enter the cohort as they are not given a certificate of clearance for tuberculosis and therefore their visa process is put on hold until they are treated and a certificate is then produced. Only then will they enter the cohort.

Individuals in our cohort will be at risk of hospital admission or death from the date of entry until the first of the following events: death, emigration, or the end of the follow-up period. Since data were unavailable to indicate whether an individual migrant is living in Scotland/Wales/Northern Ireland (or resettlement to these countries in the case of refugees), or emigration, these events will be accounted for probabilistically by multiple imputation and building on previously described methods^{14,22}.

Table 1. Clinical definition and methodological approach for hospital-based healthcare and mortality outcomes.

Outcome	Clinical Definition	Statistical Definition	Example Statistical Modelling Approach*
Hospital-based healthcare outcomes			
Hospital attendances	Number of hospital attendances in inpatient, outpatient, or A&E.	Binary indicator for attendance with specific condition (yes/no). Time to event defined as entry into cohort date until end of follow up.	Cox proportional hazards model
Hospital admissions	Admission into the hospital as an inpatient.	Binary indicator for specific admission (yes/no). Time to event defined as entry into cohort index attendance until the end of follow up.	Cox proportional hazards model
Duration of hospital admission	Number of days spent in hospital as an inpatient.	Numerical indicator for number of days.	Cox proportional hazards model
30 day emergency readmissions	Emergency admissions to any hospital in England occurring within 30 days of the last, previous discharge from hospital after admission.	Binary indicator for emergency readmission (yes/no) recorded within 30 days of the index admission discharge date. Emergency admissions are defined as those where the admission method is waiting list, booked or planned (11, 12 or 13). To be explored in subgroup of people with an initial hospitalisation.	Logistic regression
Mortality outcomes			
Death from all causes	Deaths in England from any cause.	Binary indicator for death due to a all-causes (yes/no). Deaths will primarily be identified through linkage to ONS deaths registration data, but also through HES (where the Method of Discharge field is coded as "dead" (4) as the latter method may better ascertain information on recent deaths where there is a delay in death registration (e.g. because a coroner's report is required).	Calculation of SMR using ONS death data by age and gender; rates, life expectancy
Death due to a specific condition	Death due to one of the subgroup conditions.	Binary indicator for death due to a specific condition (yes/no). Time to event defined as entry into cohort index attendance until death.	Cox proportional hazards model; calculation of SMR using ONS death data by age and gender; rates, life expectancy

*final statistical modelling approach to be confirmed during analysis stage after review of data and modelling assumptions.

Table 2. Clinical definition for hospital-based healthcare and mortality subgroup outcomes. Extended data can be found at DOI [10.17605/OSF.IO/FUTS4](https://doi.org/10.17605/OSF.IO/FUTS4)¹⁵.

Outcome Subgroups	Clinical Definition
Ambulatory care sensitive (ACS) conditions (Extended data file 4)	Conditions where effective community care can prevent inpatient hospital admission or death.
Amenable conditions	Conditions where hospital admissions or death could be avoided through high quality preventative healthcare ¹⁶ .
Preventable conditions	Conditions where all or most hospital admissions or deaths from a specific cause could be avoided by established medical or public health interventions ¹⁶ .
Avoidable conditions (Extended data file 2 and Extended data file 3)	Conditions that are considered preventable, amendable or both, where each admission or death is only counted once. When cause of admissions or death falls within both the preventable and amenable definition, all admissions or deaths from that cause are counted in both categories when they are presented separately ¹⁶ .
Sexual and reproductive health conditions and treatments (Extended data file 5)	Conditions where hospital admissions or death is for a sexual and reproductive health outcome or treatment. These are split into seven domains based on the Guttmacher-Lancet commission on sexual and reproductive healthcare and rights: abortion, contraception, human immunodeficiency virus and other sexually transmitted infections, maternal and neonatal health, fertility, gender-based violence and reproductive cancers ¹⁷ .
Tuberculosis, HIV, Hepatitis B & C (Extended data file 2 and Extended data file 3)	Conditions where hospital admissions or death is for tuberculosis, HIV, Hepatitis B or Hepatitis C.
Common mental and behavioural disorders (Extended data file 6)	Conditions where hospital admissions or death is for a common mental and behavioural disorder.
Multimorbidity	Co-existence of two or more chronic conditions, each one of which is either: (1) a physical non-communicable disease of long duration, such as a cardiovascular disease or cancer; (2) a mental health condition of long duration, such as a mood disorder or dementia; or (3) an infectious disease of long duration, such as HIV or hepatitis C ¹⁸ .
All causes*	Death due to any cause.
ICD-10 chapter ^{*19}	Death due to a specific conditions, such as infectious disease, disease of the blood, cardiovascular diseases, digestive disease, genitourinary disease, musculoskeletal disease, nervous disease, respiratory disease, endocrine disease, injury or external causes, mental and behavioural, or neoplasms ²⁰ .
Maternal deaths*	Death of a woman while pregnant or within 42 days of termination of pregnancy, irrespective of the duration and site of the pregnancy, from any cause related to or aggravated by the pregnancy or its management but not from accidental or incidental causes ²¹ .

*Subgroup only applies to the mortality outcomes.

Sample size

1,700,000 non-EU migrants who underwent pre-entry tuberculous screening to enter the UK between 2005-2017 will be included. After removal of duplicate records there will be just over 1,500,000 unique individuals. Approximately 10% of migrants move to Scotland and Northern Ireland and will not be linked to HES and ONS data. Therefore, linkage to HES and ONS will be on approximately 1,380,000 non-EU migrants.

[Table 3](#) provides examples of the precision by which prevalence estimates will be estimated and the changes in relative risk detectable between migrant subgroups. The study therefore has sufficient statistical power (80%) to detect changes in common outcomes (e.g. admission due to ambulatory care sensitive condition) and rare outcomes (e.g. mood (affective) disorders), at the 5% significance level.

Analysis plan

Our analysis plan has been designed to meet our two main objectives of profiling hospital-based healthcare and mortality outcomes for non-EU migrants and refugees. To achieve this, we will undertake the analysis in three phases. In the first

phase we will summarise and compare baseline characteristics (see [Table 4](#)) between the non-EU migrant group, the refugee group and the general population in England. With the exception of ethnicity, all baseline characteristics are anticipated to be fully recorded (chronic disease is presumed to be absent unless recorded). Missing values for ethnicity will be analysed grouped as “not recorded”. In the second phase, we will summarise the hospital-based healthcare and mortality outcomes of the three study populations, using the general population of England as a reference. We will estimate the crude association between each of the outcomes and these study population groups. We will then re-estimate the association between each of the outcomes and the study population group after adjusting for characteristics at the time of hospital admission or time of death: age, sex, chronic disease, calendar time period, and reason for hospital admission or cause of death. Finally, an appropriate statistical model (selected on the basis of meeting assumptions such as proportional hazards for Cox regression) will be used to analyse the relationship between the study comparison groups and each of the outcomes. Crude models will be fitted prior to adjustment for “baseline” measurements at or before the index admission. We will write-up the analysis in accordance with the Reporting of

Table 3. Sample size calculations that illustrate the 95% CIs with which the rates for common (ACS admission) and uncommon (Mood (affective) outcomes) can be estimated for migrant subgroups, and the difference between these migrant subgroup rates that can be detected assuming type 1 error = 0.05 and power = 0.8.

Example	Outcome/risk factor			
	Overall rate (95% CIs) per 1000 person years at risk	Subgroup size	95% CIs of rate*	Hazard ratio reduction detectable**
Admissions due to ambulatory care sensitive (ACS) conditions	8.4 (8.4-8.4) ²³	5000 (e.g. Cambodian migrants)	6.1-11.3	0.64
		10000 (e.g. Kenya)	6.7-10.4	0.73
		100000 (e.g. Bangladesh)	7.8-9.0	1.0*
Mood (affective) disorders	0.42 (0.42-0.43) ²⁴	5000	-	-
		10000	0.0-1.0	0.3
		100000	0.38-0.46	1.0*

*This column indicates the confidence interval for the subgroup assuming that the subgroup rate is the same as the general population overall rate (e.g. 8.4 per 1,000 for ACS conditions).

**This column indicates the hazard ratio detectable for the subgroup using the largest group as baseline (e.g. 100,000 Bangladesh) and assuming this baseline group has the same rate as the overall population level (e.g. 8.4 per 1,000 for ACS conditions).

Table 4. Patient characteristics in the time prior to the index admission will be collated as baseline measurements.

Variable	Description
Age at migration	(in years) as recorded at the non-EU migrant pre-entry tuberculous screening or refugee pre-entry health assessment
Sex	as recorded at the non-EU migrant pre-entry tuberculous screening or refugee pre-entry health assessment
Ethnicity	as recorded at the non-EU migrant pre-entry tuberculous screening or refugee pre-entry health assessment
Visa category	as recorded at the non-EU migrant pre-entry tuberculous screening or refugee pre-entry health assessment
Country of origin	as recorded at the non-EU migrant pre-entry tuberculous screening or refugee pre-entry health assessment
Country of departure	as recorded at the refugee pre-entry health assessment
Date of pre-entry tuberculous screening or refugee pre-entry health assessment	as recorded at the non-EU migrant pre-entry tuberculous screening or refugee pre-entry health assessment
Length of time in England prior to index admission/death	(in years) at the given time point will be estimated as (date of admission/death - month & year of pre-entry screening/365.25)
Age at index admission/death	(in years) at a given time point will be estimated as ((date of admission/death - month & year of birth)/365.25) for the index admission or death
ICD-10 chronic disease conditions	obtained from all admissions at the index admission (sub-divided into categories of: <ol style="list-style-type: none"> 1. mental health/behavioural 2. cancer/blood disorders 3. chronic infections 4. respiratory 5. metabolic/endocrine/nutritional 6. renal/genitourinary 7. musculoskeletal/dermatological 8. neurological, 9. cardiovascular

studies Conducted using Observational Routinely-collected Data (RECORD) statement²⁵.

Sensitivity analyses

To determine the robustness of our final results and to quantitatively account for any uncertainty, sensitivity analyses will be conducted to examine the extent to which our findings are affected by changes in methods or values of unmeasured variables. We are uncertain about the length of time non-EU migrants remain in England following their arrival. In a first sensitivity analysis, all migrants will be assumed to stay for one and a half years, the median time of stay for an international migrant, providing a lower estimate of person time at risk. In a second sensitivity analysis, all migrants will be assumed to stay until the end of the study period of 31st December 2017. This is the more conservative assumption, and whilst it unrealistically inflates the denominator, it provides a lower bound for the estimates of incidence and prevalence.

Dissemination of results

At the end of the study we will convene a group of policy makers, non-governmental organizations, migrants and refugees and the public to feedback the results of our study and seek suggestions on ways to take this work forwards. We will also disseminate our findings to relevant policy makers and schemes through a series of regional workshops.

Study status

We are currently in the process of seeking ethical and Confidentiality Advisory Group (CAG) approvals for the study.

Discussion

We describe a novel record linkage study that will use routine data from multiple sources to generate the Million Migrant study. The study has several advantages including the opportunity to accurately identify migrants in UK routine health records and their subsequent health needs. The strengths of this approach are the creation of a highly-powered cohort study that harnesses existing data to uncover health patterns in this often difficult to identify or invisible population. The Million Migrant study aims to improve evidence on hospital-based events and mortality and will better position the scientific community to inform policy makers and civil society with rigorous data about the health of migrants in England.

There are several limitations to our study. We will be unable to include data from primary care due to the lack of a national dataset available for this purpose. As a result, we will be unable to examine any contribution to the health and care of the individuals from primary and community or social care. Designing a new linkage of the million migrant cohort to primary care data would provide a more comprehensive understanding of primary care usage within this population and allow us to identify new opportunities for community-level interventions, and this may become a possibility in the coming years. We also do not include data from all migrant sub-groups. Irregular migrants (e.g. entrants who enter, stay or work in a country without the necessary authorization such as undocumented entrants, failed

asylum seekers, visa overstayers, children born to irregular migrant couples), migrants entering on a temporary visa (e.g. tourist visa), EU and EEA migrants, international migrants from low-incidence tuberculosis countries who subsequently do not have a pre-entry tuberculous screening (e.g. United States of America, Chile, and Egypt), and international migrants who emigrated before the start of either health screening programme will not be captured. As such, the study findings will not be generalisable to these groups. Additionally, although smaller in number, these groups could potentially be included in our randomly generated sample from the general population residing in England. This is important to consider in our interpretation of the findings. Our proposed study will not be able to assess whether the healthcare and mortality outcomes were affected by frequency of travel abroad, health service usage abroad, uncertainty in length of residence in the UK, movement in the UK outside of England to access healthcare, or wider socio-environmental determinants of health. These factors will be later examined through the creation of an electronic longitudinal cohort study using a mobile phone application to collect data on the health of migrants after moving to the UK (part of RWA's Wellcome Trust Fellowship, Public health data science to investigate and improve migrant health in the UK).

To help ensure impact from the work, we have engaged with policy makers, non-governmental organizations, migrants and refugees and the public throughout the design and conduct of the study to ensure relevance to them and prepare a pathway for impact, and will continue to do this through to the end of the project. In designing this study, we held a workshop with international migrants and refugees to understand their views on the consent process, data linkage and analysis. We have also involved national and international policy makers in the design stage.

In England today, nearly 15.6% of the population are international migrants, constituting an important and large group of people. However, there is still a limited understanding of their health needs and use of secondary care in the NHS. This study will fill an important gap in the literature and provide local, regional and national policy makers with detailed information about the health needs of this population. Our results will include information about how and where healthcare services can be improved to prevent hospital admissions, and data on the causes of death in this large and important group of people. Whilst maintaining the highest scientific and ethical standards on the use of existing data, or big data, we will set forth an avenue to advance knowledge and good practices in the field of migration and health, for the UK and internationally.

Data availability

Underlying data

No data is associated with this article.

Extended data

All extended data is publicly available on Open Science Framework: Million Migrants study of healthcare and mortality outcomes in non-EU migrants and refugees to England: Analysis

protocol for a linked population-based cohort study of 1.5 million migrants, <https://doi.org/10.17605/OSF.IO/FUTS4>¹⁵

Files available:

- Extended data 1: Tuberculosis pre-entry screening rollout phases
- Extended data 2: Revised definition of avoidable conditions
- Extended data 3: Avoidable conditions definition for children and young people
- Extended data 4: Definition of Ambulatory Care Sensitive Conditions
- Extended data 5: Definition of sexual and reproductive health outcomes
- Extended data 6: Definition of mental and behavioural disorders
- Extended data 7: Deduplication of records
- Extended data 8: RECORD checklist

Data are available under the terms of the [Creative Commons Zero “No rights reserved” data waiver](#) (CC0 1.0 Public domain dedication).

Reporting guidelines

To review the study’s RECORD checklist, please see [Extended data file 8](#).

Grant information

RWA is supported by a Wellcome Trust Clinical Research Career Development Fellowship (206602/Z/17/Z). NP is funded by a UCL Wellcome Trust Clinical Research Training Fellowship (211162/Z/18/Z). ACH’s salary is provided by Central and North West London NHS Community Trust. AS is funded by UCLH Foundation Trust. The views expressed are those of the author(s) and not necessarily those of the NHS, the

Wellcome Trust, the NIHR or the Department of Health, International Organisation for Migration or Public Health England. SVK acknowledges funding from a NRS Senior Clinical Fellowship (SCAF/15/02), the Medical Research Council (MC_UU_12017/13 & MC_UU_12017/15) and the Scottish Government Chief Scientist Office (SPHSU13 & SPHSU15). HH is a National Institute for Health Research (NIHR) Senior Investigator. His work is supported by: Health Data Research UK, which is funded by the UK Medical Research Council, Engineering and Physical Sciences Research Council, Economic and Social Research Council, Department of Health and Social Care (England), Chief Scientist Office of the Scottish Government Health and Social Care Directorates, Health and Social Care Research and Development Division (Welsh Government), Public Health Agency (Northern Ireland), British Heart Foundation and Wellcome Trust. The BigData@Heart Consortium, funded by the Innovative Medicines Initiative-2 Joint Undertaking under grant agreement No. 116074. This Joint Undertaking receives support from the European Union’s Horizon 2020 research and innovation programme and EFPIA; it is chaired, by DE Grobbee and SD Anker, partnering with 20 academic and industry partners and ESC. The National Institute for Health Research University College London Hospitals Biomedical Research Centre. JJM acknowledges having received support from the Alliance for Health Policy and Systems Research (HQHSR1206660), DFID/MRC/Wellcome Global Health Trials (MR/M007405/1), Fogarty International Center (R21TW009982, D71TW010877), Grand Challenges Canada (0335-04), International Development Research Center Canada (106887, 108167), Inter-American Institute for Global Change Research (IAI CRN3036), Medical Research Council (MR/P008984/1, MR/P024408/1, MR/P02386X/1), National Cancer Institute (1P20CA217231), National Heart, Lung and Blood Institute (HHSN268200900033C, 5U01HL114180, 1UM1HL134590), National Institute of Mental Health (1U19MH098780), Swiss National Science Foundation (40P740-160366), Wellcome Trust (074833/Z/04/Z, 093541/Z/10/Z, 107435/Z/15/Z, 103994/Z/14/Z, 205177/Z/16/Z) and the World Diabetes Foundation (WDF15-1224).

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

1. Office for National Statistics: **Population of the UK by country of birth and nationality: individual country data**. 2018; (accessed Oct 23, 2018). [Reference Source](#)
2. Key Migration Terms: **International Organization for Migration**. 2015; (accessed Dec 3, 2018). [Reference Source](#)
3. Graetz V, Rechel B, Groot W, *et al.*: **Utilization of health care services by migrants in Europe—a systematic literature review**. *Br Med Bull*. 2017; 121(1): 5–18. [PubMed Abstract](#) | [Publisher Full Text](#)
4. Katikireddi SV, Cezard G, Bhopal RS, *et al.*: **Assessment of health care, hospital admissions, and mortality by ethnicity: population-based cohort study of health-system performance in Scotland**. *Lancet Public Health*. 2018; 3(5): e226–36. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
5. Steventon A, Bardsley M: **Use of secondary care in England by international immigrants**. *J Health Serv Res Policy*. 2011; 16(2): 90–4. [PubMed Abstract](#) | [Publisher Full Text](#)
6. Stagg HR, Jones J, Bickler G, *et al.*: **Poor uptake of primary healthcare registration among recent entrants to the UK: a retrospective cohort study**. *BMJ Open*. 2012; 2(4): pii: e001453. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
7. Wallace M, Kulu H: **Low immigrant mortality in England and Wales: a data artefact?** *Soc Sci Med*. 2014; 120: 100–9. [PubMed Abstract](#) | [Publisher Full Text](#)
8. Wallace M, Kulu H: **Mortality among immigrants in England and Wales by major causes of death, 1971-2012: A longitudinal analysis of register-based data**. *Soc Sci Med*. 2015; 147: 209–21. [PubMed Abstract](#) | [Publisher Full Text](#)
9. Aldridge RW, Nellums LB, Bartlett S, *et al.*: **Global patterns of mortality in**

- international migrants: a systematic review and meta-analysis. *Lancet*. In Press.
10. Razum O: **Migrant Mortality, Healthy Migrant Effect**. In: Kirch W, ed. *Encyclopedia of Public Health*. Dordrecht: Springer Netherlands, 2008: 932–5. [PubMed Abstract](#) | [Publisher Full Text](#)
 11. **Tuberculosis screening**. GOV.UK. (accessed July 16, 2018). [Reference Source](#)
 12. Abubakar I, Aldridge RW, Devakumar D, *et al.*: **The UCL-Lancet Commission on Migration and Health: the health of a world on the move**. *Lancet*. 2018; **392**(10164): 2606–2654. [PubMed Abstract](#) | [Publisher Full Text](#)
 13. Villa-Torres L, González-Vázquez T, Fleming PJ, *et al.*: **Transnationalism and health: A systematic literature review on the use of transnationalism in the study of the health practices and behaviors of migrants**. *Soc Sci Med*. 2017; **183**: 70–9. [PubMed Abstract](#) | [Publisher Full Text](#)
 14. Aldridge RW, Zenner D, White PJ, *et al.*: **Tuberculosis in migrants moving from high-incidence to low-incidence countries: a population-based cohort study of 519 955 migrants screened before entry to England, Wales, and Northern Ireland**. *Lancet*. 2016; **388**(10059): 2510–2518. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 15. Burns R: **Million Migrants study of healthcare and mortality outcomes in non-EU migrants and refugees to England: Analysis protocol for a linked population-based cohort study of 1.5 million migrants**. 2019. <http://www.doi.org/10.17605/OSF.IO/FUTS4>
 16. Eddolls J: **Avoidable mortality in England and Wales QMI - Office for National Statistics**. (accessed Dec 3, 2018). [Reference Source](#)
 17. Starrs AM, Ezeh AC, Barker G, *et al.*: **Accelerate progress-sexual and reproductive health and rights for all: report of the Guttmacher-Lancet Commission**. *Lancet*. 2018; **391**(10140): 2642–92. [PubMed Abstract](#) | [Publisher Full Text](#)
 18. **Multimorbidity: a priority for global health research**. The Academy of Medical Sciences. (accessed Dec 3, 2018). [Reference Source](#)
 19. **ICD-10 Version: 2010**. (accessed Dec 3, 2018). [Reference Source](#)
 20. **ICD-10 Version: 2016**. (accessed Nov 14, 2018). [Reference Source](#)
 21. **Maternal mortality**. World Health Organization. (accessed Nov 14, 2018). [Reference Source](#)
 22. Thomas HL, Harris RJ, Muzyamba MC, *et al.*: **Reduction in tuberculosis incidence in the UK from 2011 to 2015: a population-based study**. *Thorax*. 2018; **73**(8): 769–75. [PubMed Abstract](#) | [Publisher Full Text](#)
 23. NHS Digital. CCG Outcomes Indicator Set - September 2018: **2.6 Unplanned hospitalisation for chronic ambulatory care sensitive conditions**. 2018; published online Sept 20. (accessed Nov 14, 2018). [Reference Source](#)
 24. Hospital Admitted Patient Care Activity: **Hospital Admitted Patient Care Activity, 2017-18**. 2018; published online Sept 20. (accessed Nov 14, 2018). [Reference Source](#)
 25. Benchimol EI, Smeeth L, Guttmann A, *et al.*: **The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement**. *PLoS Med*. 2015; **12**(10): e1001885. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Open Peer Review

Current Referee Status:



Version 1

Referee Report 04 March 2019

<https://doi.org/10.21956/wellcomeopenres.16372.r34598>



Ursula Trummer 

Center for Health and Migration, Vienna, Austria

The authors do present an interesting paper on a timely subject. I will not duplicate scope and purpose or data sets used, as this is outlined in detail by the authors.

The paper reads very technical, and addresses an inner circle of scholars. A potential audience outside this inner circle may be lost after the first paragraph or the abstract.

Although it presents and refers to the conceptual framework for influences on migration and health adapted from the UCL Lancet Commission on migration and health, a link from this conceptual framework to the actual analysis is presented only to a marginal degree. In some lines the text presents a rather restricted causal approach of “use of health services and subsequent health conditions” (in the abstract and the introduction paragraph) which may suggest to the reader that the use of health service is the main determinant of migrant’s health conditions.

The analyses include some interesting outcome subgroups, suitable for further secondary analysis e.g. for economic considerations, like ambulatory sensitive conditions (ACS), amendable conditions, preventable conditions. It would be interesting to know if the data will be made available to other researchers for such secondary analysis purposes.

Is the rationale for, and objectives of, the study clearly described?

Yes

Is the study design appropriate for the research question?

Yes

Are sufficient details of the methods provided to allow replication by others?

Yes

Are the datasets clearly presented in a useable and accessible format?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: ethics and economics, migration, health and intergenerational solidarity, social determinants of health, science communication

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Referee Report 15 February 2019

<https://doi.org/10.21956/wellcomeopenres.16372.r34600>



Sarah Rodgers  1,2

¹ Health Data Research UK (HDR-UK), Swansea University, Swansea, UK

² Department of Public Health and Policy, University of Liverpool, Liverpool, UK

The paper sets out a protocol to investigate non-EU migrant and refugee health care utilisation and mortality against that of the host country, the UK. The aim is to investigate rates to understand if there are barriers for migrants accessing healthcare. For instance they mention how only a third of migrants are registered with a GP.

It describes their intention to use routinely collected data for a subset of the total migrant population. Those who receive a refugee health check. And those who are screened for TB because they come from originating countries with high TB rates. It is not possible to analyse all migrants; only those who enter as refugees and those who undergo screening because they originate from countries with high TB rates. This is the basis of their migrant cohort, which, as authors point out, omits migrants from developed countries with few TB cases.

Data linkage

I concentrate here on data linkage aspects because two reviewers have previously covered other aspects. The data linkage relies on the Personal Demographics Service (PDS) dataset. This dataset is key because they will use it to add NHS numbers to the migrant/refugee datasets to add NHS numbers and enable deterministic joining using the NHS number of the migrant cohort to their corresponding records in HES and ONS mortality datasets.

While there is an extended file on the de-duplication of refugee/migrants to create the migrant cohort, there are not additional files on the linkage. I suggest one is added. As it currently stands, there are insufficient details in the methods to allow replication. The datasets they will be using would have some useful additional detail. Their explanation would benefit from details: What is the percentage of NHS numbers added by PHE? What will they do with the cohort members who do not have an NHS number? The completeness of the PDS dataset, and allocation of an NHS number relies on the person having utilised health care services. What is the percentage of the cohort they expect to have an NHS number?

What will they do if the PDS contains only a third of the migrants who have registered with a GP, as highlighted by Stagg et al.¹? I remain sceptical about the PDS having data added through all the sources mentioned on the NHS Digital website (pharmacies, child and mental health, secondary care providers) and believe the main source of data is from the patients themselves when they register with a new GP. Checking thoroughly with NHS Digital at this stage may save time later. Or adding such reassurances to the protocol would help.

Will they simply assume no NHS number means no utilisation? Thereby including these cohort migrants in the no healthcare utilisation denominator? They are using data up until end 2017 and it is likely the rates of NHS numbers populated increase the longer they spend in the UK. Adding time since entry to the UK would also be a useful variable to stratify your comparisons. A young healthy migrant is unlikely to need any healthcare, particularly if they are male, due to lack of screening and reproductive care needs and I note they will use sex in comparing matching levels.

Outcomes and analysis plan

I suggest that unplanned (emergency) admissions are the type that should be avoided ideally, and are also an indicator that there is a lack of preventative care. Can this be added as an additional group for analysis? These then would be separate from the planned admissions that have a prevention goal.

There seems like there is an opportunity to model several outcomes together in a combined model, for instance using multivariate generalized linear mixed models. This would place the relative importance of the different outcomes in the same model and take into account multiple testing corrections they will have to complete otherwise (e.g. Bonferroni). Therefore they should use a smaller significance level to account for multiple hypothesis testing; there are four outcomes for healthcare utilisation, for instance. Justification of the use of Cox regression seems a little confused. They should rephrase as time to event (e.g. hospital attendance) rather than specify the follow up time as the event itself. Their note about confirming statistical modelling approaches following review of data is usual in the case of routine data when the distribution of each outcome variable is unknown. Why not use Cox regression for the all cause death outcome also? Overall their analysis plan seems underspecified.

Sample size

I note that they mention loss of 10% from migration to Scotland and Northern Ireland. Although the proportion moving to Wales will be small, it is incorrect to include Wales because these data are held in the Patient Episode Dataset for Wales (PEDW) and are not included in the HES dataset. This needs adjusting.

Discussion/additional points

I agree with the first reviewer that it is particularly important the authors remove the conflicting statements about the analytical potential to examine the wider determinants of health.

References

1. Stagg HR, Jones J, Bickler G, Abubakar I: Poor uptake of primary healthcare registration among recent entrants to the UK: a retrospective cohort study. *BMJ Open*. 2012; 2 (4). [PubMed Abstract](#) | [Publisher Full Text](#)

Is the rationale for, and objectives of, the study clearly described?

Yes

Is the study design appropriate for the research question?

Yes

Are sufficient details of the methods provided to allow replication by others?

Partly

Are the datasets clearly presented in a useable and accessible format?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: routine data, data linkage, mortality, healthcare utilisation, longitudinal analysis, epidemiology

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Referee Report 07 February 2019

<https://doi.org/10.21956/wellcomeopenres.16372.r34599>



Matthew Wallace  1,2

¹ Stockholm University Demography Unit (SUDA), Stockholm, Sweden

² French National Demographic Institute (INED), Paris, France

Summary

This article summarizes the analysis protocol for a study investigating refugees and migrants originating from non-EU countries with a high prevalence of tuberculosis in England. Specifically, the authors plan to link four data sources to generate a population-based cohort study of around 1,3 million people. The study will examine hospital-based events and all-cause and cause-specific mortality among (specific groups of) migrants in England in the past decade. The study will not be representative of international immigration to England because EU-migrants and non-EU migrants from countries with low tuberculosis prevalence cannot be identified (and are in fact grouped with the English-born as part of the reference population). This is not problematic (definition of the reference group aside) because the biggest value in this study lies in the ability to identify and study such outcomes among *refugees*, so long as the data is packaged in such a way. Previous studies into the health and mortality of refugees in England are lacking because of an inability to identify them in national data sources. However, the authors should recognize that this is not the case for international immigrants in England in general, which is the impression they give in the introduction. Additionally, in some areas of the protocol the authors could be more explicit (e.g. with regards to potential errors induced in how the entries and exits of migrants are captured). Nonetheless, this large-scale study has a great deal of potential and could provide significant insight into health and mortality of refugees and non-EU migrants from countries with high tuberculosis prevalence, in turn contributing to the existing body of literature on the migrant mortality advantage in England.

Introduction

In the introduction, the authors tell the readers that little research has been conducted on mortality among international migrants in England. While this is (I think) accurate for access to health services and health, this is less true for immigrant mortality. See references provided. Together these studies document variation among international immigrants with respect to their all-cause and cause-specific mortality, country or region of origin, generation, age, sex, and duration of residence (among other factors). It is important to give appropriate acknowledgement to previous works so as not to *appear* to be overstating the gap in current literature in England. Acknowledging these works would not lessen the contribution of this study. There are limitations in these studies (including my own) upon which the “Million Migrants study” can build to provide exciting new insight into migrant mortality, health, and health care access

patterns in England.

Additionally, it is important to note that other national data sources exist in which one can identify migrants and examine these outcomes, contrary to that suggested by the authors. For example, see the *Office for National Statistics Longitudinal Study* (LS) which combines data from National Health Service systems, censuses and civil registers. With the LS data, it is possible to identify migrants and examine all-cause and cause-specific mortality, cancer registrations, self-reported health and long-term illness.

The above issues stem from how the study is framed: around all international immigrants in England. Indeed, the introduction defines international migrants in the broadest possible terms. However, the Million Migrants Study excludes a significant proportion of all international immigrants (at least 40%). With the study focusing mainly on refugees and non-EU migrants from high TB prevalence countries, the introduction should explicitly focus on these groups. Then it *would* be accurate to state that there is a lack of studies in England on refugees (I think just one, see: Swerdlow, 1991¹). Similarly, then it *would* be accurate to state that there is a lack of national data sources in England that have the relevant information to be able to identify refugees (which would require information on country of birth, reason for arrival, visa type etc...). I recommend the authors re-frame their introduction around the migrant subgroups they are studying, place their study more specifically in the context of the previous literature on migrant mortality in England¹⁻¹⁴, and be more explicit in their phrasing.

Aims

As alluded to above, the study sample is selective: refugees and non-EU migrants from countries with high tuberculosis prevalence. This is fine and explicitly acknowledged by the authors in their discussion as a limitation. However, it conflicts with a repeatedly-stated objective to investigate whether or not a migrant mortality advantage can be replicated in England (this also taps in to how the introduction is framed). Of course, a migrant mortality advantage has already been observed in England (and Wales). Observing, or indeed not observing, lower mortality among the migrants in this sample relative to the general population would not provide evidence of an overall migrant mortality advantage in England. Suggest the authors qualify such statements to reflect this.

Data collection, processing, and linkage

Are data on EU migrants simply not available or is it a choice by the authors to exclude them? If the data is not available, then the authors can simply state this. Better to do this than to over-generalize about culture and use of health care systems among EU migrants. If it is a restriction of the data, the exclusion does not need to be further justified and cannot be avoided. If, on the other hand, it was a choice, then the authors should consider including EU migrants to provide a broader appeal of this new data source for its potential users.

Will the study cover immigrants arriving at all ages?

Exactly what variables will be available in this data source? It might be nice if the authors were to provide a preliminary list to give some indication of the scope of the data.

Comparator groups

It is not possible to remove other types of migrants who did not undergo screening from the general population? Further, it is important to note that the general population will include children of migrants,

who form a small but growing proportion of England's population. A growing body of research suggest that children of migrants have different mortality patterns to majority populations of high-income host countries too. As a basic approximation, if we take the 6,4% of EU migrants, some proportion of the 9,2% of non-EU migrants who are not subjected to screening, and the 9,2% of the population who are children of migrants (from Eurostat, 2014), then approximately 16% of the general population is formed of individuals who should be excluded (or form other categories). Is there any additional information (country of birth or ethnicity) the authors can use to remove these individuals from the general population?

More information on exactly how the general population is sampled would also be very useful. Knowing this is *as important* as knowing migrants are accurately defined.

Please expand upon the rationale behind comparing migrants, who have just arrived and have very little exposure to conditions in England, to the general population by IMD score.

Entry and exit from cohort

Accurately calculating exposures for mobile populations such as migrants is one of the most challenging tasks to achieve in the studies of migrant populations (see previous studies on under-coverage, over-coverage, censoring bias etc...). With this in mind, it would be informative if the authors devoted a little more space to telling us exactly how immigrant entries and exits are captured in their data and discussing the potential biases incurred from the information they have used.

For example, entries are defined as the latest date of two pre-entry screenings. Could this lead to cases where individuals are considered "at risk" when they are still living in their origin countries? Are delays between pre-entry screening and arrival to England common? How long do they tend to be? Are there people who undertake screening, but never arrive in England? What happens to these individuals? Could they be incorrectly included in the data source?

Some clarification on the way emigrations are imputed would also be helpful (rather than linking to previous studies). Is there any information in the four data sources that could provide some (in)direct information on (r)emigration? These events are unlikely to be negligible among foreign-born and a lack of knowledge about exactly when people leave can contribute to the downward biasing of mortality rates. I appreciate the sensitivity analysis that the authors have done in response to this issue. Nonetheless, a good deal of uncertainty remains around when exactly individuals leave the country.

Sample size

I am not sure if the sample size of the general population is stated anywhere (forgive me if I missed it). We know that the study will contain around 1.3 million migrants. It would make sense to state the general population figure here too to provide overall sample size.

References

1. Swerdlow Aj: Mortality and cancer incidence in Vietnamese refugees in England and Wales: a follow-up study. *Int J Epidemiol.* 1991; **20** (1): 13-9 [PubMed Abstract](#)
2. Guillot M, Khlal M, Elo I, Solignac M, Wallace M: Understanding age variations in the migrant mortality advantage: An international comparative perspective. *PLoS One.* 2018; **13** (6): e0199669 [PubMed Abstract](#) | [Publisher Full Text](#)
3. Harding S: Mortality of migrants from the Caribbean to England and Wales: effect of duration of residence. *Int J Epidemiol.* 2004; **33** (2): 382-6 [PubMed Abstract](#) | [Publisher Full Text](#)

4. Harding S, Rosato M, Teyhan A: Trends for coronary heart disease and stroke mortality among migrants in England and Wales, 1979-2003: slow declines notable for some groups. *Heart*. 2008; **94** (4): 463-70 [PubMed Abstract](#) | [Publisher Full Text](#)
5. Harding S: Mortality of migrants from the Indian subcontinent to England and Wales: effect of duration of residence. *Epidemiology*. 2003; **14** (3): 287-92 [PubMed Abstract](#)
6. Harding S, Balarajan R: Patterns of mortality in second generation Irish living in England and Wales: longitudinal study. *BMJ*. 1996; **312** (7043): 1389-92 [PubMed Abstract](#)
7. Harding S, Balarajan R: Mortality of third generation Irish people living in England and Wales: longitudinal study. *BMJ*. 2001; **322** (7284): 466-7 [PubMed Abstract](#)
8. Scott AP, Timæus IM: Mortality differentials 1991-2005 by self-reported ethnicity: findings from the ONS Longitudinal Study. *J Epidemiol Community Health*. 2013; **67** (9): 743-50 [PubMed Abstract](#) | [Publisher Full Text](#)
9. Wallace M: Adult mortality among the descendants of immigrants in England and Wales: does migrant mortality advantage persist beyond the first generation?. *Journal of Ethnic and Migration Studies*. 2016; **42** (9): 1558-1577 [Publisher Full Text](#)
10. Wallace M, Kulu H: Low immigrant mortality in England and Wales: a data artefact?. *Soc Sci Med*. 2014; **120**: 100-9 [PubMed Abstract](#) | [Publisher Full Text](#)
11. Wallace M, Kulu H: Mortality among immigrants in England and Wales by major causes of death, 1971-2012: A longitudinal analysis of register-based data. *Soc Sci Med*. 2015; **147**: 209-21 [PubMed Abstract](#) | [Publisher Full Text](#)
12. Wallace M, Kulu H: Can the salmon bias effect explain the migrant mortality advantage in England and Wales?. *Population, Space and Place*. 2018; **24** (8). [Publisher Full Text](#)
13. Wild SH, Fischbacher C, Brock A, Griffiths C, Bhopal R: Mortality from all causes and circulatory disease by country of birth in England and Wales 2001-2003. *J Public Health (Oxf)*. 2007; **29** (2): 191-8 [PubMed Abstract](#) | [Publisher Full Text](#)
14. Wild SH, Fischbacher CM, Brock A, Griffiths C, Bhopal R: Mortality from all cancers and lung, colorectal, breast and prostate cancer by country of birth in England and Wales, 2001-2003. *Br J Cancer*. 2006; **94** (7): 1079-85 [PubMed Abstract](#) | [Publisher Full Text](#)

Is the rationale for, and objectives of, the study clearly described?

Partly

Is the study design appropriate for the research question?

Partly

Are sufficient details of the methods provided to allow replication by others?

Yes

Are the datasets clearly presented in a useable and accessible format?

Not applicable

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Mortality, health, international migrants, descendants of migrants, data artefacts, longitudinal data, survival analysis

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Referee Report 25 January 2019

<https://doi.org/10.21956/wellcomeopenres.16372.r34597>



Jessica L. Potter 

Centre for Primary Care and Public Health, Queen Mary University London, London, UK

This article summarises the analysis protocol for a linked population cohort study. The aim of the study is to describe hospital-based healthcare and mortality outcomes among 1.5 million non-EU migrants. This is because immigration status is not routinely collected on a large scale in the UK to allow comparative studies of health outcomes between migrant populations and UK-born populations.

The research team plan to compare outcomes between migrants, refugees and the general population of England. To do this they will collate three sources of data: 1. Data from pre-entry screening of non-EU migrants from high burden TB countries 2. Pre-departure health screening of refugees. 3 National hospital episode statistics. 4. ONS death records.

The authors clearly identify a gap in knowledge with respect to migrant health and therefore this large study could provide extremely novel and valuable insights.

Protocol:

To justify the focus on non-EU migrants the authors state, "We focus on non-EU migrants because of their availability in this dataset, but also because EU migrants are more likely to dual-use health systems and less likely to need adaptation in terms of cultural competence¹³"

- I have reviewed the paper referenced. It is a systematic review which compiled predominantly qualitative studies thus I could not find evidence to support the claim that "EU migrants are more likely" to dual-use health systems. Equally, I think the claim that EU migrants are less likely to need adaptation in terms of cultural competence essentialises the experience of being an EU migrant. Equally, not all aspects of an individuals culture play an equal role in determining how they access care. I suggest that the authors simply make the statement that the focus on non-EU migrants is because of the availability of the data set. Similarly, I think that it should be made very clear that it is not simply a focus on non-EU migrants but non-EU migrants from a very specific cohort of 101 countries, the majority of which are LMIC. Given pre-entry screening for TB is performed in high burden countries and TB burden at country level has been shown to be related to GDP which in turn is related to mortality, does this not introduce a selection bias at the beginning? I realise this is mentioned in the limitations section of the discussion but I think it would be useful to comment on the different characteristics of the groups not included and how these might affect the study outcomes.

I also think it should be made clear that not all refugees are screened pre-departure, some individuals claim asylum on arrival in the UK prior to becoming refugees. Please could the authors make a comment on the representativeness of this cohort of refugees in relation to the wider refugee population in the UK?

Ethics

I note in figure 2 (if I have interpreted it correctly) a plan to share identifiable data between Public Health England and NHS Digital. I wondered whether the authors have sought assurances that such data will not be shared with other parts of government for the purposes of immigration control as has previously occurred through NHS Digital. Beyond this concern the authors have clearly outlined a robust plan of psuedoanonymisation.

Comparator groups

The authors state the general population of England will be "composed of mostly England-born residents". Is country of birth routinely collected to allow this assumption or will the size/ methods of selecting this matched sample ensure they represent the background England population which the authors point out later in the paper consists of 15% of individuals originally born outside the UK?

Analysis plan

The authors state in the introduction "the study includes analyses that acknowledge the wider determinants influencing the health of international migrants such as the legal, social, economic and health structures and systems, health service access and support, exposures and behaviours, and epidemiological changes associated with population mobility" but no details of how this will be done are provided. Then in the discussion the authors state they will be unable to assess the impact of the "wider socio-environmental determinants of health" - please could they therefore clarify the earlier point as I think it is a little mis-leading as it stands.

I cannot comment on the details of the mathematical aspects of the analysis plan as this is outside the range of my expertise.

Is the rationale for, and objectives of, the study clearly described?

Yes

Is the study design appropriate for the research question?

Yes

Are sufficient details of the methods provided to allow replication by others?

Yes

Are the datasets clearly presented in a useable and accessible format?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Tuberculosis, qualitative research methods, migration health

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.
