JMIR Public Health and Surveillance

Impact Factor (2023): 3.5
Volume 11 (2025) ISSN 2369-2960 Editor in Chief: Travis Sanchez, PhD, MPH

Contents

Original Papers

Original Paper

Leveraging Administrative Health Databases to Address Health Challenges in Farming Populations: Scoping Review and Bibliometric Analysis (1975-2024)

Pascal Petit[1], PhD; Nicolas Vuillerme[1,2], PhD, Prof Dr

[1]Laboratoire AGEIS, Université Grenoble Alpes, La Tronche Cedex, France
[2]Institut Universitaire de France, Paris, France

Corresponding Author:
Pascal Petit, PhD
Laboratoire AGEIS
Université Grenoble Alpes
Bâtiment Jean Roget, UFR de Médecine
Domaine de La Merci
La Tronche Cedex, 38706
France
Phone: 33 4 76 63 71 04
Email: pascal.petit@univ-grenoble-alpes.fr

Abstract

Background: Although agricultural health has gained importance, to date, much of the existing research relies on traditional epidemiological approaches that often face limitations related to sample size, geographic scope, temporal coverage, and the range of health events examined. To address these challenges, a complementary approach involves leveraging and reusing data beyond its original purpose. Administrative health databases (AHDs) are increasingly reused in population-based research and digital public health, especially for populations such as farmers, who face distinct environmental risks.

Objective: We aimed to explore the reuse of AHDs in addressing health issues within farming populations by summarizing the current landscape of AHD-based research and identifying key areas of interest, research gaps, and unmet needs.

Methods: We conducted a scoping review and bibliometric analysis using PubMed and Web of Science. Building upon previous reviews of AHD-based public health research, we conducted a comprehensive literature search using 72 terms related to the farming population and AHDs. To identify research hot spots, directions, and gaps, we used keyword frequency, co-occurrence, and thematic mapping. We also explored the bibliometric profile of the farming exposome by mapping keyword co-occurrences between environmental factors and health outcomes.

Results: Between 1975 and April 2024, 296 publications across 118 journals, predominantly from high-income countries, were identified. Nearly one-third of these publications were associated with well-established cohorts, such as Agriculture and Cancer and Agricultural Health Study. The most frequently used AHDs included disease registers (158/296, 53.4%), electronic health records (124/296, 41.9%), insurance claims (106/296, 35.8%), population registers (95/296, 32.1%), and hospital discharge databases (41/296, 13.9%). Fifty (16.9%) of 296 studies involved >1 million participants. Although a broad range of exposure proxies were used, most studies (254/296, 85.8%) relied on broad proxies, which failed to capture the specifics of farming tasks. Research on the farming exposome remains underexplored, with a predominant focus on the specific external exposome, particularly pesticide exposure. A limited range of health events have been examined, primarily cancer, mortality, and injuries.

Conclusions: The increasing use of AHDs holds major potential to advance public health research within farming populations. However, substantial research gaps persist, particularly in low-income regions and among underrepresented farming subgroups, such as women, children, and contingent workers. Emerging issues, including exposure to per- and polyfluoroalkyl substances, biological agents, microbiome, microplastics, and climate change, warrant further research. Major gaps also persist in understanding various health conditions, including cardiovascular, reproductive, ocular, sleep-related, age-related, and autoimmune diseases. Addressing these overlooked areas is essential for comprehending the health risks faced by farming communities and guiding public health policies. Within this context, promoting AHD-based research, in conjunction with other digital data sources (eg, mobile health, social health data, and wearables) and artificial intelligence approaches, represents a promising avenue for future exploration.

(JMIR Public Health Surveill 2025;11:e62939) doi:10.2196/62939

KEYWORDS

farming population; digital public health; digital epidemiology; administrative health database; farming exposome; review; bibliometric analysis; data reuse

Introduction

Background

Public health research seeks to identify and understand the factors that influence population health to effectively prevent diseases and promote health and well-being for all [1,2]. A broad range of environmental determinants can impact health across the life span. One of the core areas of public health research, known as the exposome, investigates how cumulative environmental influences contribute to disease etiology and pathogenesis [3-18]. The exposome, which complements genomic research, refers to the comprehensive examination of all environmental exposures experienced throughout an individual's lifetime, including physical, chemical, biological, psychosocial, and behavioral factors, from conception to death [3-18]. The exposome classically includes 3 overlapping domains: the general external exposome (eg, climate and built environment); the specific external exposome (eg, chemical exposure, lifestyle, and occupations); and the internal exposome (eg, aging, oxidative stress, metabolism, and gut microbiome) [8,14,16,17,19,20]. Understanding the exposome is crucial for enabling both population-wide and precision prevention [3,21-23]. However, fully describing the exposome is challenging due to the vast diversity and the temporal and spatial variability of environmental factors [3]. Public health research in this area requires data on both risk factors and adverse health outcomes to progress effectively [3,14,24,25].

The volume of data collected has grown exponentially as the world becomes increasingly reliant on technology and digitization [26,27]. Data are omnipresent in our everyday lives, leading science toward data-driven research [27,28], in particular in the health field. The digital transformation in health care has enabled unprecedented data availability, collection, storage, and analysis capabilities, leading to a paradigm shift in health care systems, with entire care pathways becoming digitized [29,30]. Health-related data now represent approximately 6% of all digital data globally, a figure that continues to rise [31]. This explosion of data has transformed research, providing new opportunities, especially in public health, to enhance disease understanding and evaluate intervention effectiveness [27,28,32-35]. The integration of digital technologies and digital data in public health has led to the emergence of "digital public health," an evolving field focused on using digital data to achieve public health goals [33,36-41]. Public health research is moving from isolated data systems to more integrated, accessible, and reusable data resources [42]. Reusing data allows researchers to explore various health determinants, including environmental, occupational, behavioral, and organizational factors, fostering a holistic approach to disease prevention and health promotion strategies [14].

Within the digital public health framework, 2 main types of data are being used, namely primary and secondary data. Primary data are tailor-made, designed for a specific purpose, and often used once or repeatedly for the same goal [43-45]. Primary data are the cornerstone of traditional public health policy and decision-making. These data are derived from several types of studies [46-50], in particular observational cohorts (eg, the Framingham Cohort study [51,52]) [46-50,53,54], case-control studies [46-50], cross-sectional surveys (eg, the China Health and Retirement Longitudinal Study [55,56]) [46-50,53], and experimental studies [46-50]. Primary data have many advantages [46-50]. They are rich, of high quality, and are designed to answer specific research questions for public health and epidemiological purposes. Primary data are usually available at the individual level and are derived from studies that control for certain biases. By contrast, they are cumbersome, time-consuming, and costly to set up and maintain [53,54,57]. The representativeness of primary data is also limited in size, geographic scope, and temporal coverage and can erode with time [46-50,53]. Primary data are not free from bias, such as selection, healthy worker, recall, or prevarication biases [53,58].

Unlike traditional public health, digital public health does not rely solely on primary data but takes advantage of the myriad of existing digital data that have not been generated originally for research purposes (ie, secondary data) to overcome some limitations intrinsic to primary data and complement them [28,43,44,53,59-68]. Indeed, some data can have an additional impact when used beyond the context for which they were originally created [68,69]. Secondary data are collected for purposes other than public health or epidemiology and include contextual data (eg, air quality and climate data) [14,24,26,29,70-74], person-generated data (eg, social media, crowdsourcing, and mobile health) [2,24,26,31,43,61,62,73,75-86], synthetic data (eg, digital twin) [87-91], and administrative health databases (AHDs) [26,64,68,81,92-102].

AHD is a broad term encompassing a wide range of routinely collected data on individuals' health and sociodemographic information collected for registration, billing, record keeping, and other administrative purposes [26,64,81,93,95,98,100-102]. For this review, based on previous works [93,95,96,103-105], AHDs included population registers, claims databases, disease registers, electronic health or medical records, and hospital discharge databases that were collected at a local, regional, national, or international level (Table 1) [26,61,62,93,95,96,100,103-110].

Table 1. Definition and characteristics of administrative health databases included in this review.

	Population register	Claims database	Disease register	Electronic health or medical record	Hospital discharge database
Definition	Digital sociodemographic information on the residents of a country	Routinely collected digital information on individual data regarding reimbursement, records of health services, medical procedures, prescriptions, and medical diagnoses	A continuous and exhaustive digital collection of individual data regarding 1 or more health events in a geographically defined population	Systematized digital record of a patient's medical information collected in real time	Digital records of service use with information about patients, their care, and their stay in the hospital
Source	Local or national authorities	Insurance programs or schemes and health care providers	Health care institutions (eg, hospitals)	Hospitals, physicians, health care centers, and institutions	Hospitals
Population	All individuals residing in a country	All individuals covered by an insurance program or scheme	All individuals diagnosed with a specific health event in a population on a geographically defined scale	All patients using the health care system	All patients from a hospital
Purpose or finality	For the administrative purposes of government: to provide reliable information	To store financial and administrative information for medical insurers' and providers' use	For clinical and research purposes: to collect information about people diagnosed with a specific health event	For clinical and billing purposes: to document patients' clinical condition	For billing or accounting purposes
Health event	None	Health events covered by insurance or a health care provider	Specific health events (eg, cancer)	Health events requiring care that are reported in medical records	Health events from hospital admission

AHDs offer many advantages for research. Such data are collected as part of routine administrative processes, reducing additional costs for researchers. Therefore, AHDs offer relatively inexpensive access to a large number of individuals who can be tracked with time for several years, guaranteeing the representativeness of the populations studied [26,54,78,93,95,104,105,111-114]. Data recorded within AHDs are structured, coded in a standardized way, and less affected by participation and recall biases [54,58,95,106,113,115]. AHDs enable the study of rare events and populations underrepresented in studies using only primary data [95,111-113]. AHDs have limitations inherent to their nature, such as the absence of some confounding factors, the limited granularity of certain information, the data complexity, and confidentiality issues [73,78,93,95,115-125].

Rationale

AHDs are increasingly used in population-based health research due to their complementarity with traditional sources of public health and epidemiological data (ie, primary data) [42,59,64,87,93,95,96,126-128]. The reuse of AHDs, referring to their application beyond their original or intended purpose, holds major potential to advance public health and epidemiological research, offering insights that can guide public health decision-making [42,59,64,87,93,95,96,105,126-131]. Although several reviews have previously explored the general use of AHDs in research [42,95,107,129,132-135], others have focused on their application within specific countries [96,136], examined individual AHDs [108,137], or investigated their role in studying specific diseases and adverse health outcomes [44,93,104,138-142]. However, to the best of our knowledge, no study has synthesized how AHDs are reused for epidemiological and public health research within a specific population group.

To address this gap, we conducted a comprehensive scoping review and bibliometric analysis aimed at identifying how AHDs are used to address health issues in a specific population. We selected farming populations as an illustrative example because they present unique health and disease patterns [143-147]. Globally, approximately 27% of the workforce is engaged in occupational farming, and this group is exposed to numerous risk factors (ie, exposomes), including pesticides, biological agents, and limited access to health care [148]. These exposures put them at heightened risk for a wide range of adverse health outcomes [143,145,147,149]. Although agricultural safety and health have become a major public health issue in recent decades, most research on the health of farming populations has relied on traditional epidemiological and community-based studies, which often face limitations in terms of sample size, geographic scope, temporal coverage, and the range of health events examined [145,150,151].

In this context, AHDs offer valuable opportunities to enhance public health and epidemiological research in farming populations by providing broader insights, identifying at-risk subgroups, and informing health services and policy development [152]. The primary objectives of this scoping review were two-fold: (1) to summarize the current state of AHD-based research in farming populations by examining which types of AHDs are used and why, whether AHDs are integrated with other data sources, which farming populations have been studied, and what exposures and health outcomes have been explored and (2) to identify key areas of interest and potential research gaps and unmet needs in this field.

Methods

Overview

This scoping review was conducted and reported according to the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) and evidence maps guidelines (Table S1 in Multimedia Appendix 1) [153] following a single screening approach. The protocol of this study was not registered. A 7-step procedure was used: research question formulation, identifying relevant publications, title review, abstract review, full-text review, data extraction, and data analysis.

To formulate our research question, we followed the Joanna Briggs Institute guidelines, using the population, concept, and context criteria framework [154]. Our population included all individuals engaged in farming and all individuals exposed to farming-related exposures. The concepts included all possible public health and epidemiological research works that involved the study of a health outcome of interest. The context was the use, in any setting, of at least one of the AHDs, as defined in Table 1.

Search Strategy and Selection Criteria

To develop and validate the search strategy, previous reviews that examined the reuse of AHDs for population-based research were identified and refined [93,96,103,104]. Our initial search revealed that electronic health records (EHRs) are often interchangeably referred to as electronic medical records (EMRs). A distinction between EHR and EMR is sometimes made, with EMR describing patients' care from only 1 practice (eg, specific encounters in hospitals), which is contrary to EHR [105]. In that case, EMR serves as a data source for EHR. This distinction was not considered in this paper. In addition, to ensure comprehensiveness, the search terms were broadened by searching for their synonyms. For example, search terms such as "electronic health record," "digital health record," "electronic medical record," "EHR," or "EMR" were used as synonyms for electronic health or medical records. A total of 72 terms pertaining to 2 categories (farming and AHDs) were used (Textbox 1). The search terms were reflective of our research topic and question.

Textbox 1. Search terms.

Farming

- husbandry* OR agriculture* OR farming OR farm* OR agricultural* OR farmworker*

Administrative health databases (combined using AND)

- "health record" OR "health records" OR "digital record" OR "digital records" OR "health administrative register" OR "health administrative registry" OR "health register" OR "health registry" OR "medical register" OR "medical registry" OR "electronic health record" OR "electronic health records" OR "EHR" OR "EMR" OR "electronic medical record" OR "electronic medical records" OR "digital medical record" OR "digital medical records" OR "digital health record" OR "digital health records" OR "health administrative data" OR "health administrative database" OR " health administrative dataset" OR " health administrative datasets" OR "health administrative databases" OR "administrative health data" OR "administrative health database" OR "administrative health dataset" OR "administrative health datasets" OR "administrative health databases" OR "insurance data" OR "insurance database" OR "insurance databases" OR "insurance dataset" OR "insurance claim" OR "insurance claims" OR "cancer registry" OR "cancer register" OR "health insurance" OR "health surveillance program" OR "health surveillance programs" OR "Mutualite Sociale Agricole" OR "MSA" OR "health insurance system" OR "record-linkage" OR "population register" OR "population registry" OR "insurance scheme" OR "social security scheme" OR "hospital discharge" OR "administrative claim" OR "administrative claims" OR "medical claims" OR "medical claim" OR "electronic claim" OR "electronic claims" OR "mortality register" OR "mortality registry" OR "mortality record" OR "mortality records" OR "disease register" OR "disease registry" OR "illness register" OR "illness registry" OR "disorder register" OR "disorder registry"

To develop the eligibility criteria, an initial search of the literature was conducted on PubMed, with a review of the first 100 articles that used AHDs for public health and epidemiological research. In our pilot run, disease and morbidity registers were initially not considered as AHD because they were created for clinical and research purposes [47-50,53,76,155,156]. However, because disease registers contain some information derived from medical records, we decided to consider them as AHD for this review. The eligibility criteria are presented in Textbox 2. The search was restricted to original peer-reviewed records (all types were included) written in English or French but not constrained by the year of publication [93,106,157]. Publications that examined partly farming populations, with, for instance, studies reporting health risks for various sectors of activity, were included.

Textbox 2. Eligibility criteria for selection of publications.

Inclusion criteria for articles

- Data had to originate at least partly from the administrative health database (AHD)
- The study had to pertain at least partly to the farming population
- The study had to relate to public health or epidemiological research
- Original peer-reviewed publications
- Publications in English or French

Exclusion criteria for articles

- Publications not describing the use of an AHD
- Animal or in vitro studies
- Publications not in English or French

The final literature search was conducted on both PubMed and Web of Science Core Collection databases. Regarding the Web of Science Core Collection, a topic search was performed. To reduce the bias induced by daily database changes, all data collection (literature retrieval and data download) was conducted and completed on the same day, that is, April 15, 2024. Titles, abstracts, and full-text publications were screened based on pre-established inclusion and exclusion criteria. The inclusion criteria for each phase of the literature search are provided in Table S1 in Multimedia Appendix 1. When abstracts did not contain enough information about correspondence to inclusion or exclusion criteria, the article was considered for full-text review. Reference lists of included publications were not searched, although they might have also yielded new relevant studies.

Data Collection and Processing

A total of 29 metadata were extracted from each publication included in the scoping review (Table 2).

The data underwent rigorous manual validation, cleaning, and harmonization following a structured 5-step process. First, duplicate items (eg, keywords and institutions) were removed. Second, leading and trailing white spaces were eliminated. Third, items were standardized by converting text to lower case, with only the first letter capitalized. In the fourth step, items were harmonized to either singular or plural forms consistently. Finally, synonyms or terms with similar meanings (eg, “illness” and “disease”) were unified under a single term. For instance, “Pesticide,” “Pesticide exposure,” and “Pesticide use” were standardized to “Pesticide,” while “Pulmonary disease copd,” “Copd,” and “Chronic obstructive pulmonary disease” were unified as “COPD.” For cancer-related keywords, the International Classification of Diseases, eleventh revision, was used to consolidate varied terms (eg, “lung cancer,” “lung cancer risk,” “lung and bronchus cancer,” “lung tumor,” “lung tumour,” “lung neoplasm,” and “basal cell carcinoma of the lung”) into standard categories (eg, lung cancer). Quality appraisals were not performed because they were beyond the aim of this review [106,157].

Table 2. List of metadata of interest to collect from the literature search.

Metadata	Fictional example
Publication year	2024
Publication type	Article
Study name	Project X
Goal of the study	To study the association between farming and health outcome
Study type	Ecological study
Is the study nationwide?	Yes
Digital data used	Insurance claims
Goal of the digital data used	To identify farmers
Is active data used?	Yes
Active data used	Clinical examination
Farming exposure considered	Farming activity and pesticide compounds
Farming activities studied, n	10
Pesticide compounds studied, n	29
Population	Adults
Sex	Female
Participants included, n	100 to 1000
Country	France
Oldest data used (year)	1991
Most recent data used (year)	2020
Data follow-up period (years)	4
Years between the most recent data used and publication year, n	7
Disease or health events	Parkinson disease
Authors' names	Gauthier J
Authors' keywords	Pesticide
Authors' country	France
Authors' institution	Université Grenoble Alpes
Journal	Environmental Health Perspectives
Funding body	MIAI@Grenoble Alpes[a]
Citations, n	14

[a]MIAI@Grenoble Alpes: Multidisciplinary Institute in Artificial Intelligence at the Université Grenoble Alpes.

Data Analysis

To analyze the research directions (ie, hot spots and gaps) on the use of AHDs for public health and epidemiological research in farming populations, a bibliometric approach was conducted [158-160]. This analysis examined the number of publications, countries of publications, most active journals, institutions, authors, funding bodies, subject areas, citations of publications, and keywords of publications. Seven bibliometric metrics were computed, including the h-, g-, m-, and Y-indices; dominance factor; annual growth rate (AGR); and fractionalized frequency (Table S3 in Multimedia Appendix 1). The h-index attempts to measure both the productivity and citation impact of the published body of work of an entity (eg, author, institution, and journal) [161,162]. It refers to the total number of publications by a particular entity with at least the same number of citations. The m-index is calculated by dividing the h-index by the number of years of an entity's productive life (eg, researcher) [161]. The g-index of an entity corresponds to the largest number g such that the top g publications have at least $\geq g^2$ citations together [162]. The Y-index refers to the sum of both the total number of first-authored publications and the total number of corresponding author publications [163]. The dominance factor refers, for a particular researcher, to the proportion of multiauthored publications as a specific author's rank to the total number of multiauthored publications [164]. The fractionalized frequency intends to reflect an author's contribution. The AGR refers to the variable's change in percentage as a year-over-year statistic [165]. The most up-to-date journals' impact factors and ranks were retrieved manually using the Journal Citation Report in April 2024.

Spearman correlations were calculated to examine the association between the number of publications and gross domestic product (GDP); population size [166]; and the total labor force, the number of researchers in research and development (per million people), fertilizer consumption (in both % of fertilizer production and kilograms per hectare of arable land), agricultural land (km^2), agricultural land (% of land area), land under cereal production (hectares), permanent cropland (% of land area), cereal production (metric tons), crop production index, food production index, livestock production index, cereal yield (kilogram per hectare), female individual employment in agriculture (% of female employment), male individual employment in agriculture (% of male employment), employment in agriculture (% of total employment); and agriculture, forestry, and fishing, value added (% of GDP). These country characteristics were obtained from the World Bank. The most recent country characteristic (eg, GDP) was used when available.

Research directions, including hot spots and gaps, were investigated with keyword frequency, co-occurrence (counting of paired keywords), and thematic mapping analyses. Thematic mapping and keyword co-occurrence network are 2 complementary but distinct approaches that serve different purposes and offer different insights. In summary, thematic mapping focuses more on the strategic positioning of research themes within a field, while keyword co-occurrence networks emphasize the relationships and connections between specific keywords in the literature [158,167]. Both methods complement each other and are usually used to provide a more comprehensive understanding of research landscapes. The co-occurrence of 2 keywords was defined by the frequency with which they appear together in publications and was quantified using association strength (AS) or equivalent index, calculated as , where c_{ij} is the number of publications in which keywords *i* and *j* co-occur, while c_i and c_j are the number of publications in which each keyword appears, respectively [158,167]. AS measures how close 2 keywords are to each other. An AS value of 1 indicates keywords always appear together, while 0 indicates they never co-occur. These keyword co-occurrences can be visualized using a co-occurrence network graph, where a vertex or node represents a keyword, the size of the node represents the keyword frequency, and the edge represents the association between 2 keywords [158,167]. On the basis of the keyword co-occurrence network graph, a community detection procedure can be used to identify groups of words highly associated with each other [158,167]. In other words, equivalent keywords based on AS can be grouped together to identify research themes [158,167]. A strategic diagram or thematic map is based on Callon centrality (x-axis) and Callon density (y-axis) [158,167]. Callon centrality measures the degree of interaction of a theme with other themes. It is defined as , where *k* is a keyword belonging to a theme and *h* is a keyword belonging to another theme [158,167]. Callon centrality can be interpreted as an indicator of the importance of a particular topic within the broader research landscape. Callon density measures the internal strength of a theme. It is defined as , where *i* and *j* are keywords belonging to the same theme and *w* is the total number of keywords in a theme [158,167]. Callon density serves as a metric for assessing the progression and maturation of that topic [158,167]. A strategic diagram is divided into 4 quadrants according to Callon centrality and density values, which correspond to 4 types of topics. Hot spots or hot topics are defined by both high density and high centrality values (upper-right quadrant), while basic topics are defined by high centrality but low density values (lower-right quadrant). Peripheral topics are defined by both low centrality and low density values (lower-left quadrant), while niche topics are defined by low centrality and high density values (upper-left quadrant) [158,167].

To focus on agricultural or farming exposome research, a bibliometric profile of the "farming exposome" was constructed, which restricts the exposome concept to environmental exposures specific to farming populations [152,168]. This bibliometric farming exposome picture examined co-occurrences between keywords related to potential risk factors and specific health events (eg, cancers and reproductive disorders).

The bibliometric analysis was conducted and reported according to the preliminary guideline for reporting bibliometric reviews of the biomedical literature (BIBLIO; Table S4 in Multimedia Appendix 1) [169]. All analyses were performed using R software (version 4.3.2; R Foundation for Statistical Computing) for Windows 10 (Microsoft Corporation). The bibliometric analysis was performed using the *bibliometrix* R package (version 4.1.4) [170].

Results

Overview

After excluding 4485 irrelevant records, 296 publications were analyzed (Figure 1). The majority were articles (293/296, 98.9%), with a small number of reviews (2/296, 0.7%) and editorial materials (1/296, 0.3%; Table 3). Only one-third of the publications (107/296, 36.1%) were open access (Table 3 and Figure S1 in Multimedia Appendix 1).

Figure 1. PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) flowchart depicting the literature search and the evaluation process for finding relevant records. The search, conducted on April 15, 2024, in PubMed and Web of Science, had no date restrictions. AHD: administrative health database.

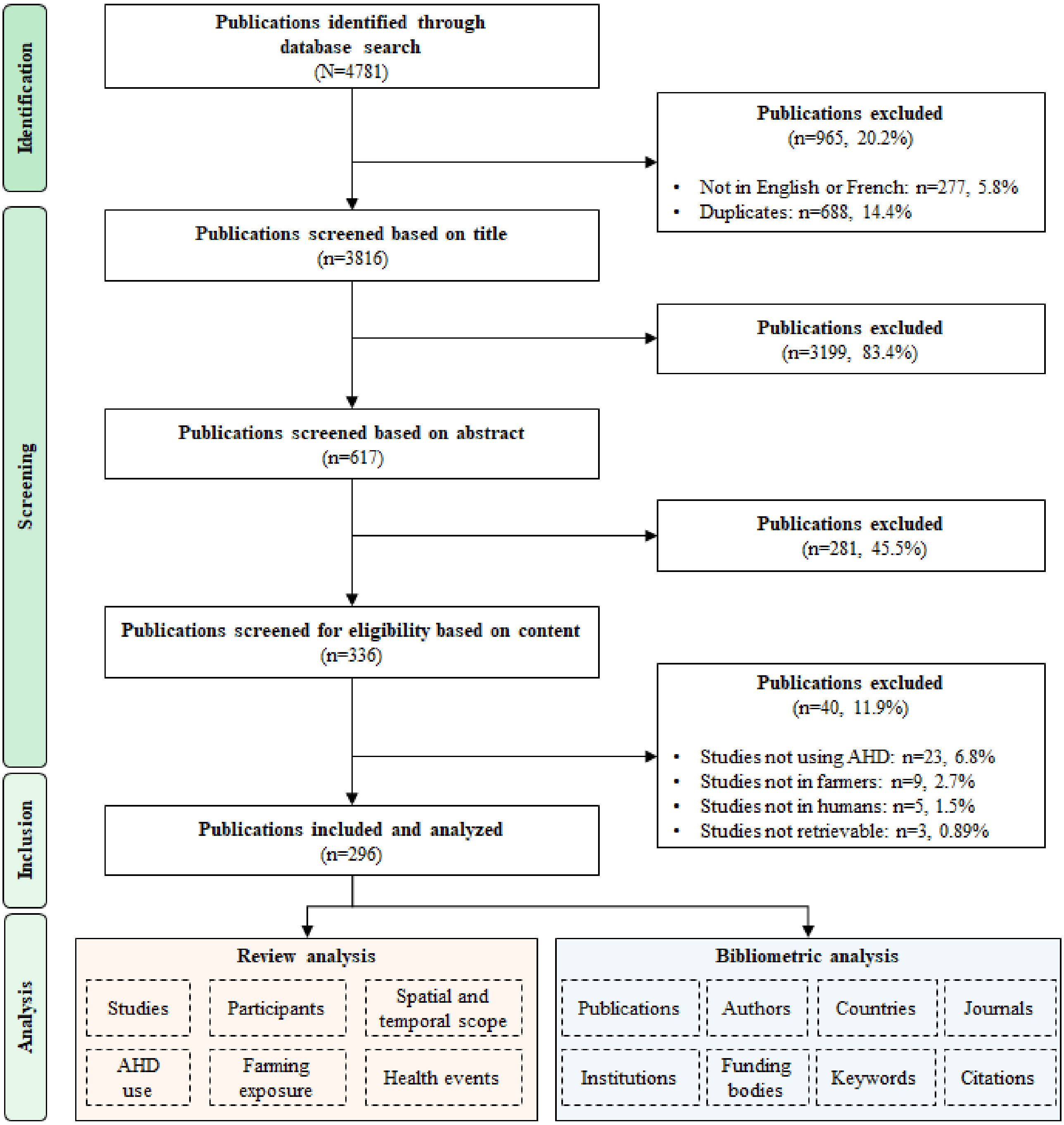


Table 3. Main characteristics of the included publications (N=296).

Description	Results
Timespan	1975 to 2024
Publication type, n (%)	
Article	293 (99)
Review	2 (0.7)
Editorial material	1 (0.3)
Open-access publications, n (%)	107 (36.1)
Document age (y), mean (SD)	14.2 (11.8)
Annual growth rate (%)	5.2
Publication citations	
Total, n[a]	9379
Average citations per publication	31.7
Average citations per year per publications	2.02
References, n	8814
Journals	
Total, n	118
Average number of publications per journal	1.86
Average number of citations per journal	79.5
Authors	
Total, n	1225
Single-author publications, n	4
Author appearances, n	1882
Average number of coauthors per publication	6.36
Average number of publications per author	0.24
International coauthorships (%)	24.3
Author's keywords, n	576
Author's country	
Total, n	34
Average number of publications per country	2.86
Average number of citations per country	436.0
Author's institution	
Total, n	338
Average number of publications per institution	3.11
Average number of citations per institution	101.3
Author's funding body	
Total, n	181
Average number of publications per funding body	2.48
Average number of citations per funding body	77.7

[a]Total, n indicates that the respective parameter has been cited n number of times, as in 296 publications have been cited 9379 times.

The average publication age was 14.2 (SD 11.8) years, ranging from the oldest in 1975 [171] to the most recent in April 2024 [152]. From 1975 onward, there has been a steady increase in publications using AHDs to address health issues in farming populations, with an AGR of 5.2%. Notably, almost one-third of these articles (91/296, 30.7%) were published in the last 5 years, highlighting the rising interest in AHD-based public health research in this population (Figure S2 in Multimedia

Appendix 1). Collectively, the publications received 9379 citations, averaging 31.7 citations per publication. Figure S3 in Multimedia Appendix 1 presents the historical direct citation network. The body of work involved 1225 authors from 338 institutions, with 1882 author appearances and an average of 6 authors per paper (Table 3). Four (1.4%) out of the 296 publications were single-author publications. On average, each paper cited 30 references.

Studies were led by authors from 34 countries, predominantly high-income nations, with 24.3% (72/296) of studies involving multicountry collaborations (Figure S4 in Multimedia Appendix 1). US-based authors contributed the most publications (91/296, 30.7%), followed by authors based in France (71/296, 24%) and Finland (35/296, 11.8%). US authors also had the most citations (3495/9379, 37.2%), with France and Finland ranking second and third, respectively.

Of 296 publications, the 25 (8.4%) most cited ones, appearing in 17 different journals, received between 83 (83/9379, 0.9%) and 485 (485/9379, 5.2%) citations (Table S5 in Multimedia Appendix 1) [150,172-196]. Of these 25 publications, 10 (40%) were published before 2000, another 10 (40%) between 2000 and 2010, and 5 (20%) after 2010. Most of these studies focused on cancer risk (16/25, 64%), while others investigated neurodegenerative disorders (5/25, 20%); respiratory conditions (2/25, 8%); and multiple health outcomes, such as sleep disorders, mental health disorders, and musculoskeletal disorders (2/25, 8%).

Table S6 in Multimedia Appendix 1 provides details on the most productive countries, prolific authors, active journals, institutions, and funding bodies.

Study Characteristics

Table 4 provides an overview of the included publications. Longitudinal study designs were the most common, including retrospective cohorts (129/296, 43.6%) and prospective cohorts (56/296, 18.9%). Case-control studies (62/296, 20.9%), cross-sectional studies (39/296, 13.2%), and ecological studies (17/296, 5.7%) were less common (Multimedia Appendix 2). A few studies (10/296, 3.4%) used multiple study designs [188,194,197-204].

The median follow-up period was 9.5 (IQR 5-17) years. On average, there was a 7-year gap (90% CI 3-14) between the most recent data used and the year of publication, with considerable variation depending on the publication year (Figure 2). The oldest data were from 1801 [205], and the most recent data were from 2022 [206]. Notably, one-third of the data used (98/296, 33.1%) were from before 2000, while nearly three-quarters (214/296, 72.3%) were from before 2015 (Figure S2 in Multimedia Appendix 1). Of 296 studies, only 10 (3.4%) used data from the last 5 years (from 2020), while 80 (27%) used data from the last 10 years (from 2015).

Studies were conducted in all continents, but most participants were from Europe (249/296, 84.1%), followed by North America (85/296, 28.7%), Asia (24/296, 8.1%), Oceania (17/296, 5.7%), Africa (4/296, 1.4%), and Central and South America (4/296, 1.4%). France (70/296, 23.6%) and the United States (67/296, 22.6%) were the most represented countries, followed by Finland (36/296, 12.2%), Sweden (32/296, 10.8%), Denmark (28/296, 9.5%), and Norway (26/296, 8.8%; Figure 3 and Figure S5 in Multimedia Appendix 1). Most studies had a regional or local scope (177/296, 59.8%), in particular, traditional epidemiological studies, such as Agriculture and Cancer (AGRICAN) [207] and Agricultural Health Study (AHS) [172], which used AHDs to either identify potential individuals for inclusion or enrich their cohorts.

Table 4. Characteristics of the included studies (1975 to 2024; N=296).

Characteristic	Values
Research goal, n (%)	
Study the association between farming and a health event	156 (52.7)
Study the association between individual characteristics and a health event	131 (44.3)
Other research goals	9 (3)
Study design, n (%)	
Retrospective cohort	129 (43.6)
Case-control study	62 (20.9)
Prospective cohort	56 (18.9)
Cross-sectional study	39 (13.2)
Ecological study	17 (5.7)
Multiple designs	10 (3.4)
Review	2 (0.7)
Perspective	1 (0.3)
Geographic scope, n (%)	
Nationwide	117 (39.5)
Regional or local	176 (59.5)
Temporal scope (y)	
Follow-up period, median (IQR)	9.50 (5-17)
Follow-up period, mean (SD)	12.8 (14.0)
Gap between the latest data used and publication year, median (IQR)	7.21 (5-9)
Gap between the latest data used and publication year, mean (SD)	7.21 (4.67)
Population, n (%)	
Adult	265 (89.5)
Adult and child	19 (6.4)
Child	8 (2.7)
Not reported	1 (0.3)
Sex, n (%)	
Female	130 (43.9)
Male	169 (57.1)
Female and male	188 (63.5)
Not specified	108 (36.5)
Participants, n (%)	
>1,000,000	50 (16.9)
100,001 to 1,000,000	53 (17.9)
10,001 to 100,000	65 (22)
1001 to 10,000	67 (22.6)
101 to 1000	47 (15.9)
10 to 100	8 (2.7)
Not reported	3 (1)
AHD [a] type, n (%)	
Disease register	158 (53.4)
Electronic health or medical record	124 (41.9)

Characteristic	Values
Insurance claim	106 (35.8)
Population register	95 (32.1)
Hospital discharge databases	41 (13.9)
AHD use, n (%)	
Obtain information on sociodemographics	272 (91.9)
Obtain information on a health event	269 (90.9)
Identify a farmer	147 (49.7)
Identify an individual	140 (47.3)
Obtain information on occupations	117 (39.5)
Exposure assessment	57 (19.3)
Obtain information on a farming activity	43 (14.5)
Other uses	14 (4.7)

[a]AHD: administrative health database.

Figure 2. Number of years between the most recent data used and publication for all included articles (1975-2024). Points refer to the average number of years or gap between the most recent data used and publication (x-axis) for each publication year (y-axis). Error bars refer to the 90% CI of the number of years between the most recent data used and publication.

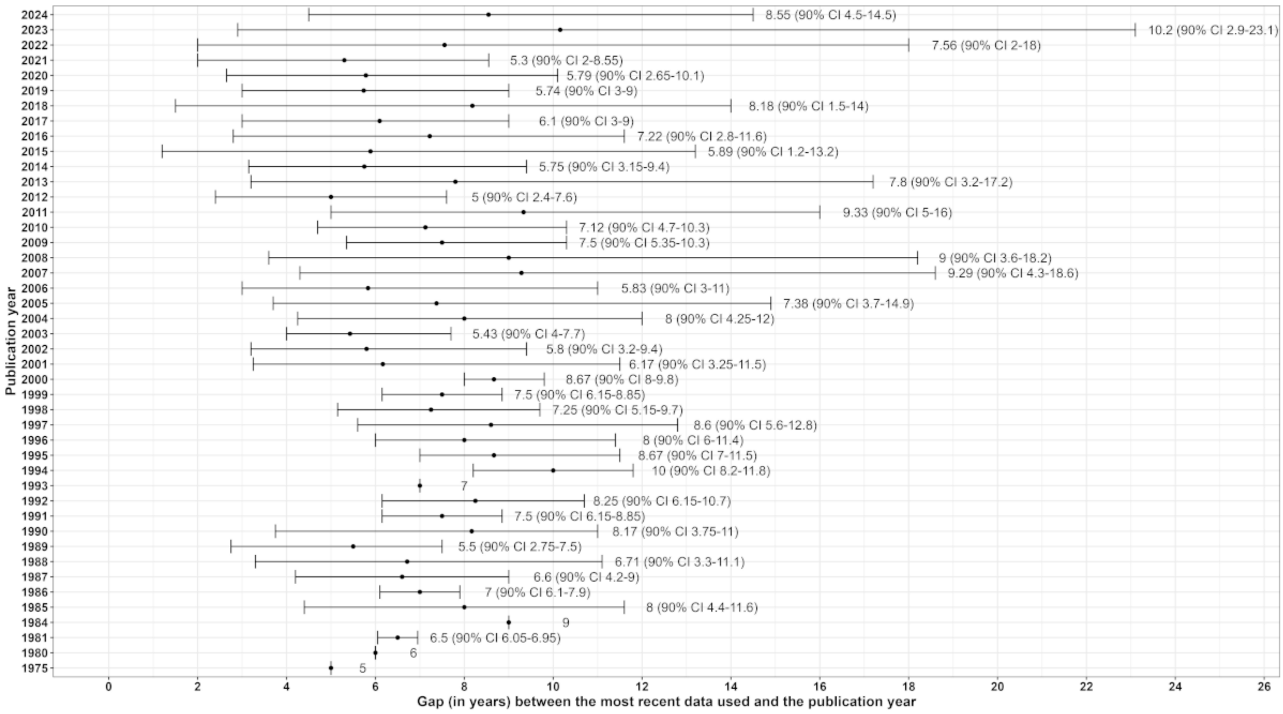
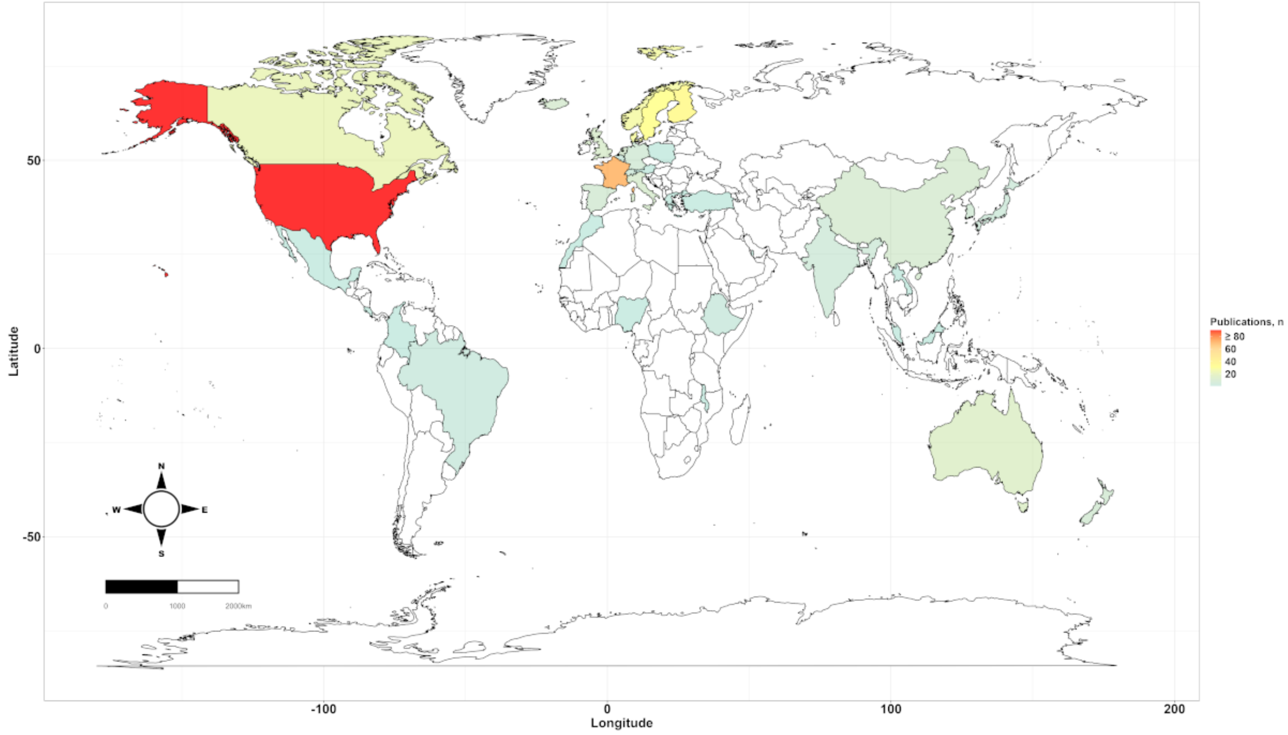


Figure 3. World map of the number of publications per country of the farming population studied between 1975 and 2024.



Most studies included 1001 to 10,000 participants (67/296, 22.6%), followed by studies with 10,001 to 100,000 participants (65/296, 22%) and 100,001 to 1 million participants (53/296, 17.9%; Table 4). Larger studies (>1 million participants) accounted for 16.9% (50/296) of the included publications. Smaller studies, with 100 to 1000 participants, were less common (47/296, 15.9%), and very few (8/296, 2.7%) had <100 participants. Most studies included adult participants (284/296, 95.9%). Of 296 studies, 169 (57.1%) examined male individuals, 130 (43.9%) examined female individuals, and 188 (63.5%) examined both sexes, but 108 (36.5%) did not specify the participants' sex.

More than half of the studies (156/296, 52.7%) aimed to explore the relationship between farming activities (eg, dairy farming) and health events, while 131 studies (44.3%) focused on individual characteristics, such as occupation, age, sex, and socioeconomic status (Figure S6 in Multimedia Appendix 1). Among those studies examining individual characteristics, farming was often considered broadly and compared to other occupations (95/131, 72.5%). Conversely, in studies investigating health outcomes specifically related to farming activities, agriculture was treated as a broad category in only 27.6% (43/156) of the cases. Most studies (277/296, 93.6%) used the general population or other nonfarming groups as the reference category without differentiating farmers by job role (eg, farm managers vs farm workers). Descriptive statistics and multivariable regression were the most commonly used methods. Notably, only 2 studies (2/296, 0.7%) incorporated artificial intelligence (AI) in their analysis [208,209].

Few studies investigated health outcomes in farmers' family members or nonfarmers exposed to farming. Of 296 studies, only 3 (1%) focused on health events in farmers' partners [177,210,211], 5 (1.7%) on farmers' children [179,212-215], and 6 (2%) on nonfarmers exposed to farming-related risks [209,216-221]. There were 11 (3.7%) studies that explored health risks in migrant workers.

Some publications reported findings from the same cohorts (Figure S7 in Multimedia Appendix 1). The 10 most prolific cohorts included France-based AGRICAN (18/296, 6.1%) [207], the US-based AHS (17/296, 5.7%) [172], Nordic Occupational Cancer Study (NOCCA; 12/296, 4.1%) from Nordic countries (Finland, Denmark, Norway, Sweden, and Iceland) [189], France-based Tracking and Monitoring Occupational Risks in Agriculture (TRACTOR; 7/296, 2.4%) [222], and Cancer in the Norwegian Agricultural Population (7/296, 2.4%) [182] cohorts. Other notable cohorts included the US-based United Farm Workers of America (6/296, 2.0%) [223], France-based BALISTIC (5/296, 1.7%) [224], the international (29 countries) consortium agricultural cohort (AGRICOH; 4/296, 1.4%) [150,225], AIRBAg (4/296, 1.4%) from France [226], and the US-based National Agricultural Workers Survey (3/296, 1%) [227]. Among these top 10 cohorts, only NOCCA, United Farm Workers of America, and TRACTOR exclusively used AHDs.

AHD Use

There was high heterogeneity in the coding systems used and the granularity of the information available regarding health events (outcomes), population, and exposure determinants, depending on the AHD and study considered. Regardless of the publication reviewed, AHDs and other datasets were never reported as adhering to the findable, accessible, interoperable, and reusable (FAIR) data principles [228-230]. In addition, none of them could be considered as FAIR data because, with a few exceptions [222], most AHDs were not precisely described, and data availability statements were rare. Furthermore, mainly due to privacy concerns, AHDs were not available for open and free access.

The most commonly used AHDs were disease registers, used in more than half of the studies (158/296, 53.4%), followed by electronic health or medical records (124/296, 41.9%), insurance claims (106/296, 35.8%), population registers (95/296, 32.1%), and hospital discharge databases (41/296, 13.9%; Table 4). Among disease registers, cancer (120/158, 75.9) and mortality registers (75/158, 47.5%) were the most frequently used (Figure S8 in Multimedia Appendix 1 and Multimedia Appendix 2). Nearly one-third of the studies (91/296, 30.7%) relied on a single AHD, with disease registers being the most common (38/91, 42%), followed by insurance claims (29/91, 32%), electronic health or medical records (18/91, 20%), population registers (5/91, 5%), and hospital discharge databases (1/91, 1%). Other types of digital data were used less frequently, including pesticide registration records (13/296, 4.4%), job-exposure matrices (JEMs; 12/296, 4.1%), crop-exposure matrices (11/296, 3.7%), pesticide use records (8/296, 2.7%), climate data (7/296, 2.4%), and air quality data (2/296, 0.7%). While contextual data were sometimes used (9/296, 3.0%), person-generated data, smart agriculture data, and omics were never used.

The AHDs and other digital data were primarily used to obtain sociodemographic information (272/296, 91.9%) and health event data (269/296, 90.9%). They were also used to identify farmers (147/296, 49.7%) or individuals (140/296, 47.3%), gather occupational information (117/296, 39.5%), assess exposure (57/296, 19.3%), obtain data on farming activities (43/296, 14.5%), or track climate conditions (7/296, 2.4%).

Nearly two-thirds of the studies (181/296, 61.1%) relied exclusively on digital data (AHDs or other), while more than one-third (112/296, 37.8%) incorporated self-reported information/active data (requiring active participant involvement) as part of epidemiological cohorts. A total of 111 (37.5%) out of 296 studies used participant-completed questionnaires (paper or electronic) to gather sociodemographic data and confounding factors (98/296, 33.1%), assess exposure (96/296, 32.4%), or collect health information (83/296, 28%). Some information was obtained through interviews (44/296, 14.9%) or clinical examinations (32/296, 10.8%). Biological monitoring (24/296, 8.1%) and airborne monitoring (2/296, 0.7%) were sometimes used, whereas no study reported dermal monitoring (Figure S9 in Multimedia Appendix 1).

Among all the AHDs used, the Mutualité Sociale Agricole (MSA) is a singularity. To the best of our knowledge, it is the only AHD specifically dedicated to the entire farming population of a country. Indeed, MSA is the French national insurance scheme that covers the entire farming workforce (5% of the overall French population) [115,128]. MSA was used in 60 studies (60/296, 20.3%). These studies were often part of cohorts with multiple publications, such as AGRICAN (18/60, 30%), TRACTOR (7/60, 12%), BALISTIC (5/60, 8%), AIRBAg (4/60, 7%), Aging Multidisciplinary Investigation (2/60, 3%) [151], BM3R (2/60, 3%) [231], FERMA (risk factors of the rural environment and allergic and respiratory disease; 1/60, 2%) [232], and Phytoner (1/60, 2%) [233]. Of these, TRACTOR was the only cohort using exclusively MSA data [222].

Farming Exposure

A variety of exposure proxies were used to assess farming-related exposure. The most common proxy was a job title, which generally referred to whether the individual was a farmer (184/296, 62.2%). Other proxies included specific farming activities, such as dairy or crop farming (111/296, 37.5%), general pesticide exposure (yes or no; 62/296, 20.9%), and exposure to specific pesticide compounds (eg, glyphosate or paraquat) or pesticide classes (eg, insecticides; 51/296, 17.2%; Figure S10 in Multimedia Appendix 1 and Multimedia Appendix 2). The number of farming activities studied ranged from just 1 [226] to 78 [222], with an average of 8 farming activities per study. Similarly, the number of pesticide compounds assessed ranged from 1 [234] to 943 [235], with an average of 42 pesticides per study. Only 1 study investigated the mixture effect of exposure to multiple pesticide combinations on human health [236]. Investigations into other chemical exposures were also limited, with only 2 papers each addressing silica exposure [237,238] and air pollution [194,217] (2/296, 0.7%). Notably, no studies examined exposure to per- and polyfluoroalkyl substances or micro- and nanoplastics. Research on the broader farming exposome was rare (5/296, 1.7%) and typically used farming activities as proxies [152].

Of 296 studies, few explored exposure to physical agents, with 5 studies (1.7%) focusing on radiation [187,218,239-241]. No studies investigated the effects of climate change on farmers' health. Exposure to biological agents was rarely studied as well, with just 3 (1%) out of 296 papers addressing mycotoxins [241-243]. Finally, only 3 studies (1%) examined psychological factors related to farming exposure [244-246].

Health Events

The most frequently studied health events were cancer (142/296, 48%), followed by mortality (44/296, 14.9%), injuries (38/296, 12.8%), workplace accidents (32/296, 10.8%), respiratory disorders (30/296, 10.1%), neurodegenerative diseases (28/296, 9.5%), and mental health issues (26/296, 8.8%; Figure S11 in Multimedia Appendix 1 and Multimedia Appendix 2). Less frequently studied conditions included cardiovascular diseases (16/296, 5.4%), autoimmune disorders (11/296, 3.7%), musculoskeletal disorders (11/296, 3.7%), reproductive disorders (3/296, 1.0%), sleep disorders (1/296, 0.3%), and frailty (1/296, 0.3%). Notably, no studies explored the farming microbiome.

Among cancers, lung cancer was the most commonly investigated cancer (43/142, 30.3%), followed by prostate cancer (38/142, 26.8%), leukemia (37/142, 26.1%), colorectal cancer (35/142, 34.6%), multiple myeloma (35/142, 34.6%), non-Hodgkin lymphoma (35/142, 34.6%), bladder cancer (31/142, 21.8%), and brain cancer (31/142, 21.8%; Figure S12 in Multimedia Appendix 1). Respiratory disorders were primarily focused on asthma (15/30, 50%) and COPD (chronic obstructive pulmonary disease; 14/30, 47%). Parkinson disease was the most studied neurodegenerative condition (16/28, 57%), followed by multiple sclerosis (6/28, 21%). Fewer publications examined Alzheimer disease (2/28, 7%) and amyotrophic lateral sclerosis (2/28, 7%; Figure S13 in Multimedia Appendix 1). In the mental health field, suicide (12/26, 46%) and depression

(8/26, 31%) were the most investigated issues (Figure S14 in Multimedia Appendix 1).

Keyword Analysis

Overview

Following an initial extraction of 1259 authors' keywords, manual harmonization was performed. Duplicate keywords were removed through singular or plural standardization (130/1259, 10.3%) and synonym unification and grouping of cancer-related terms (553/1259, 43.9%), yielding a final set of 576 (45.8%) harmonized keywords, which were all used in subsequent analyses to prevent selection bias.

On average, each publication included 8.90 keywords (90% CI 0-17), although 35 (11.8%) out of 296 publications lacked any keywords, in line with the journal guidelines. Keyword analysis confirmed prior findings regarding farming exposure and health outcomes. It also provided deeper insights into emerging research hot spots, directions, and gaps.

Of the total 576 keywords, 301 (52.3%) appeared only once, while 68 (11.8%) were mentioned at least 10 times. More frequently used keywords included 39 that appeared at least 20 times (39/576, 6.8%) and 11 that featured in at least 50 publications (11/576, 1.9%). The 50 most frequently used keywords were mentioned in at least 17 (5.7%) out of 296 publications, while the top 10 appeared in at least 51 publications (17.2%; Table S7 in Multimedia Appendix 1). The most frequently cited keyword was "cancer" (150/296, 50.7%), followed by "mortality" (96/296, 32.4%), "pesticide" (88/296, 29.7%), "occupation" (82/296, 27.7%), "farmer" (77/296, 26.0%), "agriculture" (74/296, 25%), "exposure" (57/296, 19.3%), and "epidemiology" (57/296, 19.3%).

In terms of overall citations, "cancer" (5766/9379, 61.5%), "pesticide" (3569/9379, 38.1%), and "mortality" (3097/9379, 33%) were the most cited keywords. During the past decade, the frequency of the top 5 keywords has drastically increased (Figure S15 in Multimedia Appendix 1). Notably, keywords such as "cancer," "mortality," "occupation," "pesticide," "agriculture," and "farmer" have been consistently present in publications spanning at least 30 years (not necessarily consecutively; Figure S16 in Multimedia Appendix 1). In the last decade, emerging keywords, such as "big data," "administrative health database," "dust," and "BMI," have gained prominence (Figure S17 in Multimedia Appendix 1).

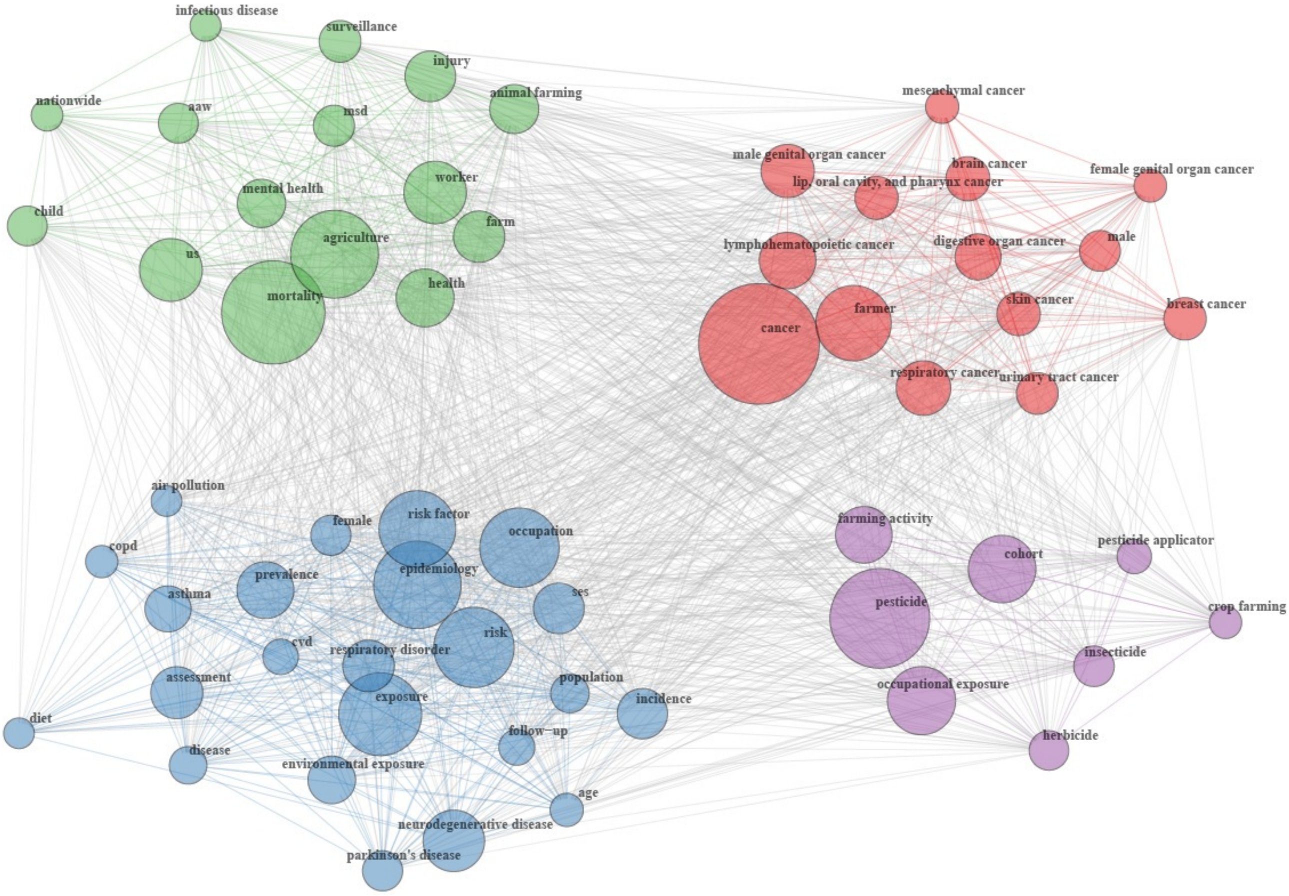
Keyword Co-Occurrence

A keyword co-occurrence network illustrating the frequency of keyword co-mentions in publications was constructed, thereby revealing relationships and conceptual connections (Figure 4). In this network, nodes or vertices represent keywords, with their sizes indicating frequency, while edges denote co-occurrences. The network's density and arrangement reveal topic interconnectivity, with larger vertices representing more frequently mentioned keywords. The network visualization helps identify clusters of related topics and highlights core research areas.

Using a community detection algorithm (spin-glass model with simulated annealing), 4 distinct clusters or communities of keywords were identified. Each cluster groups keywords that are often mentioned together, with stronger internal associations and weaker connections to keywords in other clusters.

The most frequently used keywords for each cluster were "cancer" (red cluster), "pesticide" (purple cluster), "mortality" (green cluster), and "exposure" (blue cluster). The red cluster highlights associations between various types of cancer, reflecting the fact that studies investigating cancer risks often examine multiple types of cancer. The green cluster links "mortality" with terms such as "mental health," "injury," and "animal farming," explained by the association between workplace accidents, mental health issues (eg, suicide), animal farming, and mortality. In the purple cluster, "pesticide" connects with "occupational exposure" and "farming activity," emphasizing that pesticide exposure is primarily studied in occupational settings across different types of farming. The blue cluster connects "exposure" to terms such as "neurodegenerative disease," "respiratory disorder," "cardiovascular disorder," "risk factor," "air pollution," "age," and "diet," indicating the study of various risk factors in relation to several health events. These clusters highlight current research hot spots that focus on 4 main interconnected themes: the associations between risk factors, pesticide exposure, farming activities, and a range of diseases.

Figure 4. Keyword co-occurrence network of the 296 articles published between 1975 and 2024. Each vertex or node represents a keyword, while edges represent the co-occurrence between keywords. Two keywords are connected when they co-occur in the same publication, and the size of each vertex indicates the frequency of a keyword: larger vertices represent more frequently mentioned keywords. Keywords with the same color (cluster) represent a research area. AAW: workplace accident; COPD: chronic obstructive pulmonary disease; CVD: cardiovascular disorder; MSD: musculoskeletal disorder; SES: socioeconomic status.



Thematic Mapping: Research Hot Spots

Figure 5 presents a thematic map that illustrates current research directions. Thematic mapping visualizes the relationship between research themes or topics, enabling the identification of directions, emerging areas, and gaps in the literature. The result is a strategic diagram that shows how themes relate to each other and their relevance within a specific field. The graph is divided into 4 quadrants, categorizing topics based on their relevance (x-axis, Callon centrality) and maturity (y-axis, Callon density) within the broader research landscape. Each circle represents a theme or topic (ie, a cluster of equivalent keywords), with the circle size corresponding to the frequency of the keywords associated with that theme.

The upper-right quadrant represents "hot topics," which are both highly relevant and mature in the research landscape. Four key hot topics drive AHD-based public health research in farming populations. These include 1 topic focused on cancer research; another on respiratory disorders; and a third encompassing neurodegenerative diseases, workplace accidents, injuries, and mental health issues. The final hot topic involves large-scale studies in France and Europe using big data and insurance claims.

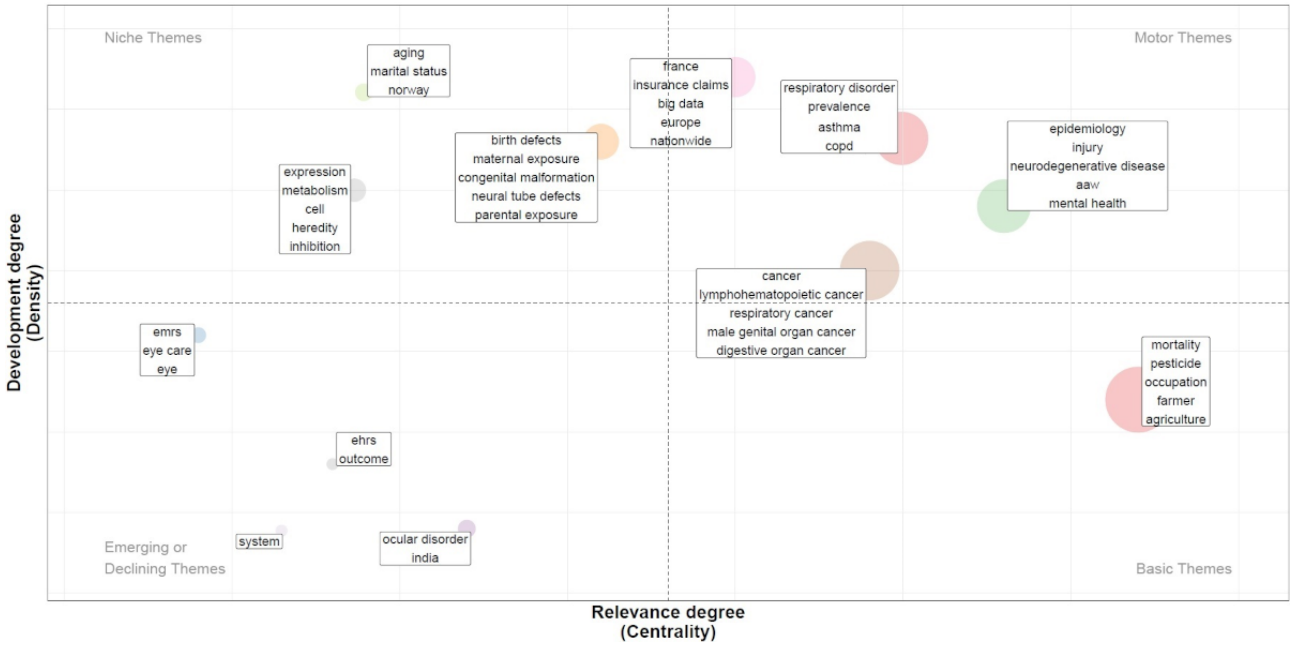
The lower-right quadrant contains "basic topics," which are relevant but not yet mature in the research landscape. Only 1 such theme emerged: research related to pesticide exposure, mortality, and farming.

In the upper-left quadrant, "niche themes" refer to mature research topics that have not yet achieved full relevance. Three niche themes were identified: the first involves studies examining aging and research conducted in Norway; the second focuses on reproductive disorders and parental exposure, a theme poised to potentially evolve into a hot topic; and the final niche theme covers genetics and metabolism.

Finally, the lower-left quadrant contains "peripheral topics," which represent either emerging or declining themes with low relevance and maturity. Four peripheral topics were observed, of which 2 (50%) were primarily centered on research on ocular disorders, 1 (25%) on the use of electronic health or medical records, and 1 (25%) on studies conducted in India.

This thematic map helps highlight both well-established and emerging areas of research, as well as gaps that may be ripe for future investigation.

Figure 5. Thematic mapping: research hot spots based on keywords from the 296 articles published between 1975 and 2024. The graph shows how themes relate to each other and their relevance within a specific field. This graph is divided into 4 quadrants, categorizing topics based on their relevance (x-axis and Callon centrality) and maturity (y-axis and Callon density) within the broader research landscape. Each circle represents a theme or topic (ie, a cluster of equivalent keywords), with the circle size corresponding to the frequency of the keywords associated with that theme. AAW: accident at work; COPD: chronic obstructive pulmonary disease; EHR: electronic health record; EMR: electronic medical record.



Bibliometric Farming Exposome

To identify research directions and gaps in the farming exposome literature, a bibliometric keyword co-occurrence analysis was conducted to explore the farming exposome by examining the co-occurrence between keywords associated with potential risk factors and specific health outcomes. This analysis was restricted to exposome-related and health event–related keywords. Of 576 keywords, 130 (22.6%) were related to the exposome, among which 93 (16.1%) were related to the specific external exposome (eg, pesticide), 19 (3.3%) to the general external exposome (eg, climate), and 18 (3.1%) to the internal exposome (eg, oxidative stress). Furthermore, there were 70 (12.2%) health event–related keywords (eg, brain cancer).

The results of this analysis are provided in Tables 5 and 6 and Multimedia Appendix 3, with each cell representing the percentage of occurrences of an exposome-related keyword (eg, air pollution) in all publications mentioning a specific health event–related keyword (eg, Alzheimer disease). For example, a value of 33.3 indicates that an exposome-related keyword appeared in 33.3% of all publications mentioning a specified health event–related keyword. To facilitate interpretation and ease the reading of Tables 5 and 6, exposome-related keywords were categorized into 19 groups (eg, chemical agent) and health event–related keywords into 20 groups.

Table 5. Co-occurrence between keywords related to internal exposomes and health event categories among the articles published between 1975 and 2024. Each cell refers to the number of times (%) a keyword related to an exposome category (eg, chemical agent) was mentioned among all publications in which a keyword related to a health event category (eg, cancer) appeared (N=296). Please note that the absolute value for each row is provided in parentheses with the row header and remains the same for all the parameters in that row.

Health event, n	Internal exposome (%)										
	Age	Sex	BMI	BP[a]	Heredity	Ethnicity	Hormone	Menopause	Metabolism	OS[b]	Inflammation
Cardiovascular disease (n=17)	5.88	11.8	5.88	5.88	0	5.88	0	0	0	0	0
Work-related disease (n=5)	0	20	0	0	0	0	0	0	0	0	0
Autoimmune disease (IBD[c], RA[d], vasculitis, and NR[e]; n=9)	0	0	11.1	0	0	0	0	0	0	0	0
Cancer (n=150)	3.62	27.5	2.17	0	0.73	1.45	3.62	0.73	2.9	0.73	0
Dental health (n=2)	0	0	0	0	0	0	0	0	0	0	0
Ocular disorder (n=6)	16.7	0	0	0	0	0	0	0	0	0	0
Frailty (n=2)	50	0	50	0	0	0	0	0	0	0	0
Anemia (n=1)	100	100	0	0	0	0	0	0	100	0	0
Infectious disease (malaria, Lyme disease, tuberculosis, toxoplasmosis, and NR; n=14)	0	7.14	0	0	0	7.14	0	0	0	0	0
Injury (including workplace accident and disability; n=40)	4.26	8.51	2.13	0	0	2.13	0	2.13	0	0	0
Chronic kidney disease (n=3)	0	0	0	0	0	0	0	0	0	0	0
Mental health disorder (depression, suicide, and NR; n=25)	9.52	4.76	4.76	0	0	0	0	0	0	0	0
Metabolic disorder (diabetes, dysthyroidism, and NR; n=9)	11.1	0	11.1	11.1	0	11.1	0	0	0	0	0
Mortality (n=75)	4.17	17.7	0	0	1.04	0	1.04	2.08	0	0	0
Musculoskeletal disorder (arthritis, low-back pain, and NR; n=14)	7.14	14.3	7.14	0	0	0	0	0	0	0	0
Neurodegenerative disease (AD[f], ALS[g], MND[h], MS[i], PD[j], and NR; n=33)	6.06	0	0	0	3.03	6.06	0	0	12.1	6.06	0
Sensory impairment (n=1)	0	0	0	0	0	0	0	0	0	0	0
Reproductive disorder (birth defects, infertility, spontaneous abortion, and NR; n=24)	0	28.6	14.3	0	0	7.14	7.14	0	7.14	0	0
Respiratory disorder (allergy, asthma, COPD[k], pneumonia, sarcoidosis, and NR; n=39)	5.88	11.8	8.82	2.94	0	5.88	0	0	5.88	0	2.94
Skin disorder (dermatitis and NR; n=2)	0	50	0	0	0	0	0	0	0	0	0

[a]BP: blood pressure.

[b]OS: oxidative stress.

[c]IBD: inflammatory bowel disease.

[d]RA: rheumatoid arthritis.

[e]NR: not reported.

[f]AD: Alzheimer disease.

[g]ALS: amyotrophic lateral sclerosis.

[h]MND: motor neuron disease

[i]MS: multiple sclerosis.

[j]PD: Parkinson disease.

[k]COPD: chronic obstructive pulmonary disease.

Distinct keyword exposome profiles were developed for each health event–related keyword (Figures S18-S43 in Multimedia Appendix 1), as illustrated in Figure 6 for mental health disorders. Most exposome-related keywords associated with keywords related to mental health disorders pertained to the type of occupations as well as chemical, lifestyle, socioeconomic, and psychological factors. Cancer-related keywords were associated mostly with keywords related to the internal (sex) and specific external exposome (chemical agents, lifestyle, and type of occupations). Autoimmune disease–related keywords co-occurred mostly with external exposome–related keywords (chemical agents, lifestyle, type of occupations, and socioeconomic factors). Neurodegenerative disease–related keywords were associated mostly with keywords related to the specific external exposome (lifestyle, chemical agents, and type of occupations). Reproductive disorders co-occurred mostly with internal (sex and BMI) and specific external exposome–related keywords (chemical agents and type of occupations). Keywords related to both musculoskeletal disorder and injury were associated with keywords from all 3 exposome components, in particular sex, type of occupations, lifestyle, biomechanical factors, chemical agents, and psychological factors. Infectious disease–related keywords co-occurred with specific external exposome–related keywords (biological agents and type of occupations), while respiratory disorder–related keywords were associated mostly with internal (sex) and specific external exposome–related keywords (lifestyle, chemical and biological agents, and type of occupations). Cardiovascular disorder–related keywords were associated with keywords from all 3 exposome components, in particular the sex, type of occupations, lifestyle, chemical agents, and socioeconomic and psychological factors.

Table 6. Co-occurrence between keywords related to specific external and general external exposomes and health event categories among the articles published between 1975 and 2024. Each cell refers to the number of times (%) a keyword related to an exposome category (eg, chemical agent) was mentioned among all publications in which a keyword related to a health event category (eg, cancer) appeared (N=296). Please note that the absolute value for each row is provided in parentheses with the row header and remains the same for all the parameters in that row.

Health event, n	Specific external exposome (%)					General external exposome (%)		
	Lifestyle	CA[a]	BA[b]	BF[c]	Occupation	PA[d]	SF[e]	PF[f]
Cardiovascular disease (n=17)	11.8	29.4	0	11.8	23.5	0	29.4	17.6
Work-related disease (n=5)	0	20	20	20	40	0	20	0
Autoimmune disease (IBD[g], RA[h], vasculitis, and NR[i]; n=9)	22.2	44.4	0	0	55.6	0	22.2	0
Cancer (n=150)	17.4	51.4	7.25	1.45	58.7	5.8	8.7	0
Dental health (n=2)	50	0	0	0	0	0	50	0
Ocular disorder (n=6)	16.7	16.7	0	0	0	0	0	16.7
Frailty (n=2)	100	50	0	0	0	0	0	50
Anemia (n=1)	0	0	0	0	0	0	0	0
Infectious disease (malaria, Lyme disease, tuberculosis, toxoplasmosis, and NR; n=14)	21.4	21.4	50	7.14	28.6	0	14.3	7.14
Injury (including workplace accident and disability; n=40)	12.8	14.9	2.13	14.9	23.4	2.13	4.26	6.38
Chronic kidney disease (n=3)	0	100	66.7	0	0	33.3	0	0
Mental health disorder (depression, suicide, and NR; n=25)	28.6	23.8	0	4.76	33.3	4.76	23.8	28.6
Metabolic disorder (diabetes, dysthyroidism, and NR; n=9)	33.3	44.4	0	11.1	44.4	11.1	0	22.2
Mortality (n=75)	17.7	55.2	8.33	4.17	54.2	5.21	17.7	2.08
Musculoskeletal disorder (arthritis, low-back pain, and NR; n=14)	14.3	21.4	7.14	14.3	42.9	7.14	0	21.4
Neurodegenerative disease (AD[j], ALS[k], MND[l], MS[m], PD[n], and NR; n=33)	21.2	48.5	6.06	0	36.4	0	9.09	3.03
Sensory impairment (n=1)	0	0	0	0	100	0	0	0
Reproductive disorder (birth defects, infertility, spontaneous abortion, and NR; n=24)	7.14	35.7	7.14	0	64.3	0	7.14	0
Respiratory disorder (allergy, asthma, COPD[o], pneumonia, sarcoidosis, and NR; n=39)	47.1	35.3	11.8	2.94	35.3	0	8.82	5.88
Skin disorder (dermatitis and NR; n=2)	50	50	0	0	50	0	0	0

[a]CA: chemical agent.

[b]BA: biological agent.

[c]BF: biomechanical factor.

[d]PA: physical agent.

[e]SF: socioeconomic factor.

[f]PF: psychological factor.

[g]IBD: inflammatory bowel disease.

[h]RA: rheumatoid arthritis.

[i]NR: not reported.

[j]AD: Alzheimer disease.

[k]ALS: amyotrophic lateral sclerosis.

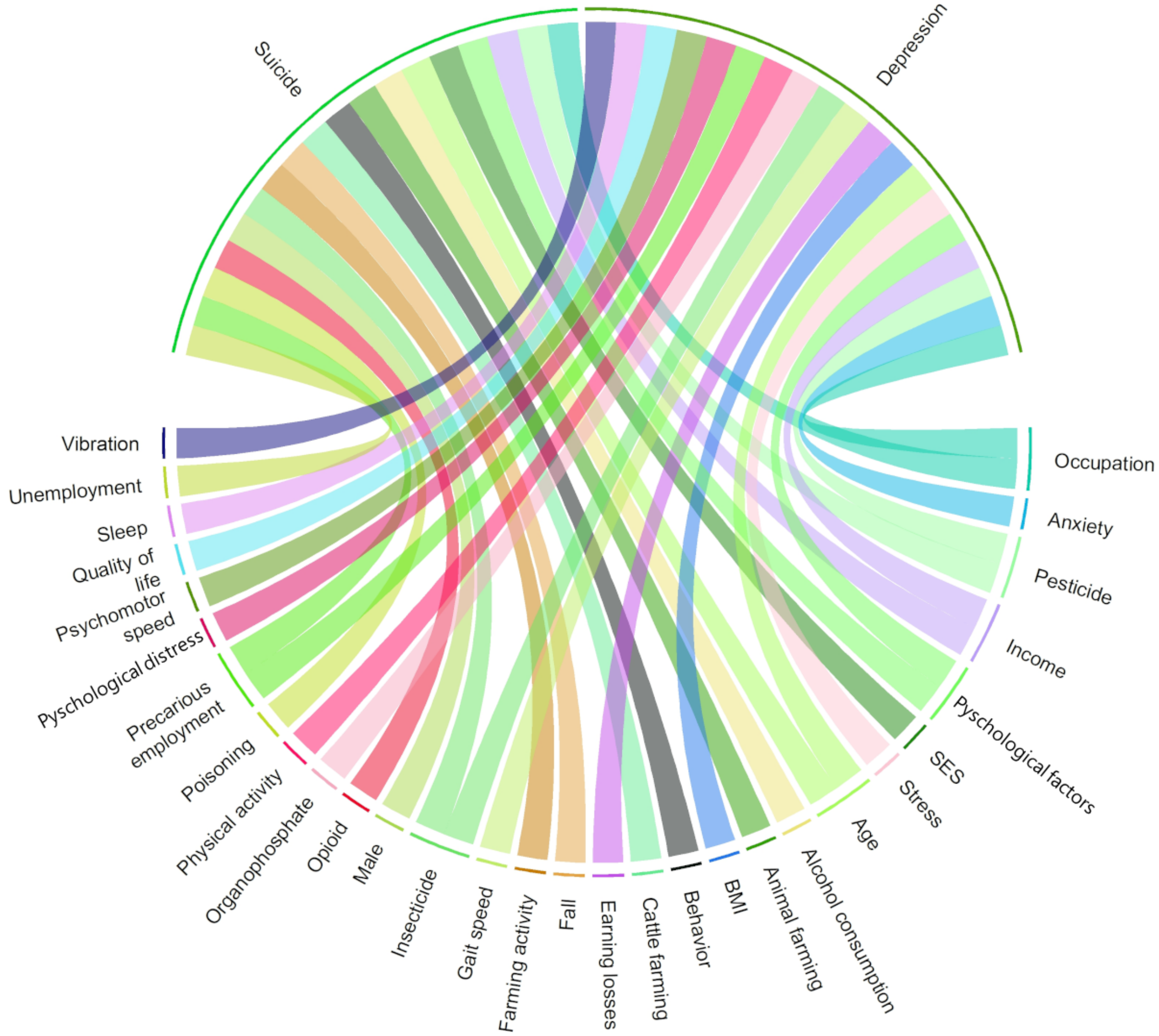
[l]MND: motor neuron disease

[m]MS: multiple sclerosis.

[n]PD: Parkinson disease.

[o]COPD: chronic obstructive pulmonary disease.

Figure 6. Chord diagram of keyword co-occurrence between potential risk factors and mental health disorder keywords from the 296 articles published between 1975 and 2024. Disease-related keywords are displayed on the top half of the chord diagram, while exposome-related keywords are displayed on the bottom half. Each chord or link indicates that an exposome-related keyword was mentioned with a disease-related keyword (co-occurrence) at least once in the same publication. The chord color differs from one exposome-related keyword to another. SES: socioeconomic status.



Discussion

Principal Findings

This review provides the first comprehensive and objective synthesis of research on the use of AHDs to address health issues in farming populations. It identifies major contributors, key publications, and existing research gaps while also suggesting future directions for leveraging AHDs to study health issues in farming populations. Overall, findings indicate that only a small part of the exposome and a limited range of health events have been examined within farming populations through the reuse of AHDs.

Current Directions in AHD Use for Public Health Research in Farming Populations

Research using AHDs in farming populations has been predominantly conducted in developed countries [150,225], with the United States [172,223,227]; France [207,222,224]; Canada [145,186,195]; and Scandinavian nations [182,189], including Denmark, Finland, Norway, and Sweden, leading the field. This dominance is linked to considerable funding from these regions and international collaborations. Scandinavian countries are particularly advanced in AHD use, offering databases that are highly complete, accessible, and well-integrated into public health research. AHDs from Denmark, Sweden, Canada, and France also provide comprehensive data on a patient's digital trajectory within their respective health systems [93,98,104,108,113,115,128,247]. France stands out further, with an AHD dedicated specifically to the entire farming population (MSA). This may explain the frequency of large-scale and long-term studies from these countries, some of which included >100,000 participants. However, many studies still had a regional focus, partly due to the use of AHDs by traditional epidemiological studies such as AGRICAN [207] and AHS [172], which rely on limited resources [47-50,53,54,57]. These studies often used AHDs to identify farming populations for inclusion or to supplement cohort data. The international AGRICOH consortium, initiated by the US National Cancer Institute and the International Agency for Research on Cancer, includes 11 (38%) of the 29 cohorts identified in this review [150,225]. However, several cohorts in AGRICOH were not identified, potentially due to lack of publications, language barriers, or limited use of AHDs. There were many publications associated with these well-established cohorts, for which many of the most prolific authors were working [172,182,189,207,222-227].

The most frequently used AHDs in farming population health research were disease registers, followed by electronic health or medical records and insurance claims. More than two-thirds of the studies used disease registers, in particular, cancer and mortality registers. This is not surprising because disease registers are created for clinical and research purposes with a continuous, exhaustive, and optimal digital collection of individual data regarding ≥1 health event in a geographically defined population [53,105]. The coding systems and the granularity of information related to health outcomes, populations, and exposure determinants varied widely across studies. Most studies (291/296, 98.3%) used AHDs to collect sociodemographic and health event information [222].

There was no consensus on the best methods or proxies to assess farming exposure. A variety of exposure proxies and determinants were used across studies, with indirect methods being the most common. Many studies (237/296, 80.1%) dichotomized proxy, for example, classifying individuals as farmers or nonfarmers or as pesticide-exposed versus nonexposed. In nearly two-thirds of the included studies, job title (ie, "being a farmer") served as the primary exposure proxy. About one-third of studies took into account specific farming activities (eg, dairy farming and crop farming) to reflect the diversity of farming practices. This approach is a valuable proxy for agricultural exposure, offering a broader representation of the farming exposome, which involves multiple stressors beyond just pesticides [147,152,188,201,248,249]. Farming activity information was often derived from digitalized data, such as agricultural censuses or self-reported data from mandatory insurance enrollments [152]. Many studies (111/296, 37.5%) combined AHDs with self-reported data (eg, questionnaires) [172,207], which allowed for more comprehensive data collection but tended to restrict the scope to regional studies due to resource constraints. These studies typically yielded high-quality data, with more potential confounders considered compared to studies relying solely on AHDs. Most studies (68/111, 61.3%) using self-reported data focused on single exposures, mainly pesticides, with only 1 study addressing multiexposure to pesticides [236]. Biological monitoring and airborne chemical sampling were rarely conducted, likely due to practical and financial constraints and the short half-lives of most pesticides [250]. Dermal chemical monitoring has not been reported, even though it is the main exposure route for pesticides [251]. The high number of studies investigating exposure to pesticides may be explained by the fact that AHS focuses on pesticide applicators and their spouses [172] and because many pesticides have adverse health effects on humans, such as neurotoxicity or endocrine disruption [252-256]. Beyond pesticides, farmers face exposure to other chemicals [257], such as air pollution; micro- and nanoplastics [258-263]; biological agents (eg, endotoxins and zoonoses) [264-266]; physical agents (eg, UV radiation, noise, and vibrations) [187,267]; biomechanical factors (eg, repetitive movements, heavy load, and working posture) [198,268,269]; and psychosocial stressors [270-274], which have been less studied than pesticides. Despite these multiple exposures, the broader farming exposome remains understudied.

In addition to AHDs, some studies integrated other secondary data, such as climate data [187,275], air quality data [194,217], JEMs [212,276,277], or crop-exposure matrices [278]. JEMs provide exposure level estimates for various chemicals and stressors based on job categories [250,279]. Although JEMs can provide valuable exposure data, they often lack the specificity of individual-level data, making it difficult to account for task-specific risks, temporal variations, and the inclusion of specific worker subgroups such as female individuals [250,279-282]. The lack of a universal standard for JEMs further complicates their application, which may explain why many studies still rely on self-reported data for more accurate exposure assessment despite the risk of recall bias [250,279-282].

The health outcomes studied were predominantly cancer [145,150,151,171-175,177-179,182,183,185,189,190,196,207,210-212,223], mortality [173,186,194,195,202,205], workplace injuries [198,208], respiratory disorders (eg, asthma and COPD) [151,180,181,213,224,226], neurodegenerative diseases (eg, Parkinson disease) [111,151,176,184,187,188,193,197,201,248,249,283], and mental health issues [151,244,273], which represent focal points within the research field. This is not surprising because many well-established cohorts centered on cancer research, in particular AGRICAN [207], Cancer in the Norwegian Agricultural Population [182], and NOCCA [189]. In addition, arsenic and inorganic arsenic compounds are classified as carcinogenic to humans by the International Agency for Research on Cancer [252,253,284], while malathion, glyphosate, diazinon, dichlorodiphenyltrichloroethane, and occupational exposures in spraying and application of nonarsenical insecticides are classified as probably carcinogenic to humans (group 2A), and several other pesticides are ranked as possibly carcinogenic to humans (group 2B), such as tetrachlorvinphos and parathion. Regarding mortality, it is often cancer and suicide mortalities that are investigated [186,195,202,244]. Furthermore, several pesticides are neurotoxic [252,255], but existing studies focused mainly on Parkinson disease and multiple sclerosis, with a paucity of data on Alzheimer disease and other neurodegenerative diseases [147,283]. In contrast, certain areas, such as cardiovascular diseases [151,194,203,227,285,286], autoimmune disorders [168,219,237], musculoskeletal disorders [192,204,245,287,288], reproductive disorders [242,289,290], sleep disorders [191], aging-related conditions [151,291], hearing impairment [267,292], and the microbiome [8,293-298], remain underexplored, despite their potential relevance to farming populations.

Challenges of Reusing AHD for Public Health Research in Farming Populations

Each AHD presents unique advantages and limitations. For example, large sample sizes and a large number of available health events are strengths, while generalizability and the absence of key confounders are challenges [64,93,95,105,115]. Access to AHDs is frequently restricted by a variety of factors, including governance and technical barriers, such as language, data structure, interoperability, and coding systems. Additional challenges stem from the type of AHD (eg, insurance claims or cancer registers), inadequate documentation (eg, absence of a data dictionary), limited accessibility due to costs or conditions, and jurisdictional and legal constraints [30,62,64,81,113,115]. Identifying the optimal AHD for a given research question is also complex, especially when considering the heterogeneity in coding systems and country-specific data structures. In countries such as Scandinavia, Canada, and France, individual identifiers facilitate data linkage across multiple AHDs, enhancing research opportunities [93,104,113,115,128,247]. However, many AHDs are not ready for research and require significant processing, cleaning, and understanding before they can be analyzed [93,105,113,222,299]. Another major challenge is the long lag between data access, analysis, and research publication. On average, studies used data that were 7 years old at the time of publication, largely due to delays in data access, administrative approvals, and the need to prepare complex datasets for analysis [223,300]. For instance, the TRACTOR project took 2 years to clean and prepare its dataset for research use [222]. These delays are compounded by the time required to conduct statistical analyses and prepare manuscripts for publication, as well as review and publication times (delay from submission to acceptance and from acceptance to publication). Another limitation of AHDs is the lack of detailed exposure data. AHDs rarely include exposure information due to their administrative focus, requiring researchers to supplement with additional data sources, such as JEMs or self-reported data. When exposure information is recorded in AHDs, it is often too generalized, typically only reflecting broad job classifications, such as farming, without specifying detailed activities or stressors. There are some exceptions, such as MSA data, which capture a wider range of specific farming activities (eg, dairy farming) [222]. However, exposure to specific stressors (eg, chemical compounds) is largely absent from AHDs.

The reference populations used in farming studies vary, which precludes direct comparisons and limits the generalizability of the findings. For example, AGRICAN used the general population as a reference [207], while TRACTOR used a farming population [152,168]. Furthermore, studies differ in their focus on specific farming populations, such as the entire agricultural workforce [207], farm managers [152], or pesticide applicators [172], which may lead to distinct exposure profiles that influence health outcomes because these farming populations have different socioeconomic status, experiences, and behaviors. Hence, to avoid or lessen bias, some studies focused on specific farming populations [152,172]. Moreover, the scope of farming populations included in studies is often limited, omitting subgroups, such as farm families, nearby residents, or consumers exposed to agricultural products, which limits the broader application of the findings. In addition, farming practices can vary significantly between countries and studies, and there is no international standardized classification for farming activities. In some cases, farming categories are derived from legal or administrative sources, as seen in the MSA data [152,168,283]. This lack of standardization limits the comparability and generalizability of findings across studies. In addition, the generalizability of the findings to other countries when using AHDs may be limited because of the differences in health care systems [93].

There are several well-known limitations of AHDs that complicate the investigation of health outcomes [301]. Health events captured in AHDs are typically limited to those requiring medical attention, which may not reflect the true incidence of diseases. In addition, the level of detail varies across diseases, even within the same AHD [92]. Although diagnostic accuracy is generally higher in disease registers, these are often geographically limited and cover only a subset of health conditions. For example, in France, cancer registers only cover 23 (24%) out of 96 administrative regions [155]. In addition, certain conditions, such as depression, are not covered by any registers. Identifying health outcomes in AHDs often requires complex algorithms that combine data from multiple sources, such as drug reimbursements, disease declarations, or medical procedures [104,105,128,152,302-304]. In addition, inconsistencies in case definitions and algorithms across studies and countries hinder the ability to compare and pool risk estimates [104,302,304]. Some AHDs also lack critical clinical information, such as laboratory results and genetic data [115,128,305], and the recorded diagnosis or treatment date may not correspond to the actual onset of the disease. Furthermore, diagnosis codes are not always indicative of a confirmed diagnosis.

Emerging Opportunities and Research Needs

While AHDs are well-used in certain countries, there are underexplored opportunities in regions such as the United Kingdom, where AHDs exist but are underused for research [113]. For low- and middle-income countries, the development and access to AHDs remain limited, and international support is needed to expand this research infrastructure. As already reported by Habib et al [306], there is a notable lack of sex-specific data, even though occupational exposures and health outcomes can vary significantly between sexes due to genetic, physiological, psychological, and behavioral factors [307-311]. Future research should address these disparities to provide a more comprehensive understanding of health risks in farming communities. Although there are inherent delays in using AHDs due to the time required for data generation, consolidation, and access, we advocate for the continued publication of studies, even those using older data. Historical data remain vital for better understanding long-term health patterns, particularly for diseases such as cancer, where tumor initiation can span decades [312,313]. Editors should encourage the publication of studies using older datasets, especially when addressing long-term health outcomes (eg, cancer and neurodegenerative diseases) or when recent data are not available [312].

None of the analyzed AHDs fully adhere to the FAIR principles, possibly because most were developed before the establishment of these principles in 2016 [228-230]. Moreover, the assessment of FAIR compliance of AHDs relied solely on information presented in publications, which may not provide a comprehensive evaluation. Nevertheless, there is a critical need to advocate for the integration of FAIR principles within AHDs to enhance public health research [228-230]. Currently, the landscape is favorable for data reuse, particularly with initiatives such as the forthcoming European Health Data Space [314-316]. Data reuse extends beyond mere access; it encompasses data discovery, a fundamental aspect of FAIR principles that involves recognizing the existence of databases [228-230]. To facilitate this, the creation of data catalogs is essential [228]. Numerous data repositories, such as Re3Data [317], Zenodo [318], CANUE [319], Figshare [320], "Epidémiologie – France" [321], data.gouv [322], Dataverse [320], or Data Europa [320], already exist. In addition, specialized multidisciplinary open-access and peer-reviewed journals such as *Scientific Data* and *Data in Brief* publish datasets [318]. A dataset search can also be conducted using the Google (Google LLC) platform [323]. However, the documentation and access conditions for datasets can highly vary across inventories, complicating the selection process for researchers. The absence of indicators or scores for data reusability further hampers efforts to identify the most suitable datasets for specific research questions [45,69]. To our knowledge, no comprehensive catalog of AHDs currently exists to date. A web-based inventory of AHDs, modeled after existing resources, such as OccupationalCohorts.net [324], OccupationalExposureTools.net [325], and Toxicological and Exposure Database Inventory [10], could greatly enhance research endeavors. The motivation for analyzing AHDs often stems from the data they contain. Consequently, as data availability increases, researchers will be better positioned to formulate research questions and engage in a parallel process of "datagraphy" or "datagraphic search" akin to traditional bibliographic research. The objective of datagraphy would be to determine which datasets are best suited for addressing specific research questions, highlighting the need for accessible catalogs to support this goal.

There is also an opportunity to integrate other secondary data, such as person-generated data (eg, mobile health, social media, digital footprints, and wearable sensors); contextual data (eg, climate and air quality); and smart agriculture data [2,83,84,326-328]. These datasets, largely untapped in farming population research, could provide new insights into health outcomes and environmental exposures [101].

Nationwide studies using big data were a hot spot. AI, such as machine learning, is particularly useful for analyzing big data and holds substantial promise for future research [329-331], particularly for identifying predictors of health outcomes in farming populations [332]. To date, AI has been underused, with only 2 studies using it, 1 identifying occupational injuries in agriculture [208] and 1 reviewing the development of chronic kidney disease risk prediction models [209]. Incorporating AI, along with cohort enrichment and interdisciplinary expert interpretation, could open new avenues for research.

Many studies continue to examine agriculture as a broad category, highlighting the need for more detailed investigations into specific farming activities and tasks [147,152,188,201,248,249]. Our analysis reveals major research gaps in understanding environmental and occupational exposures among farming populations, particularly with regard to emerging concerns such as per- and polyfluoroalkyl substances, biological agents, micro- and nanoplastics, and the impact of climate change. Climate change is a critical issue for agriculture, as it may drive shifts in pests, diseases, and farming practices [274,333-339]. Parental exposure appears to be a theme that will soon become a hot topic. Furthermore, research is needed to explore the farming exposome, particularly focusing on the "mixture effects" of multiple simultaneous exposures [340,341]. Omics data, which have not been used in farming population studies to date, represent a promising avenue for future research because genetics and metabolism were found to be a niche theme. Omics data refer to the large-scale datasets generated from various omics technologies that analyze biological molecules (eg, genomics, transcriptomics, proteomics, and metabolomics), which provide comprehensive insights into different biological layers and processes [11,342,343].

To enhance the characterization of farming exposome research using keyword analysis, there is a pressing need for standardized keyword reporting. We advocate for the development of a standardized approach to reporting keywords in scientific journals, including defining a minimum set of information (eg, study type, health outcome, population studied, data sources, and positive, negative, or null associations) and adopting a list of standardized terms. Although challenging, this approach would improve literature searches, make data more comparable and FAIR [228-230], and lead to more efficient, frugal (less time and energy spent to identify relevant information), and accurate synthesis of the scientific literature, such as in reviews and bibliometric analyses.

The prominence of topics such as cancer, neurodegenerative diseases, mortality, injuries, and mental health issues underscores the need for targeted prevention strategies. The thematic map analysis indicates that reproductive disorders (eg, birth defects, endometriosis, and infertility) are on the verge of becoming a central research focus. Emerging and understudied health conditions, including ocular disorders, autoimmune diseases (eg, inflammatory bowel disease and rheumatoid arthritis), sleep disorders (eg, sleep apnea), cardiovascular diseases, and musculoskeletal conditions (eg, low-back pain), warrant increased attention and further research. Aging-related health issues, such as frailty, also represent promising avenues for future research, particularly given the growing aging population and associated health care challenges [24].

Limitations

The findings of this review should be considered in view of their limitations. Because of time and resource constraints, a single screening approach was used, with only 1 author (PP) conducting the review and bibliometric analysis. While single screening is an efficient use of time and resources, there is a higher risk of missing relevant studies than when using dual screening [344,345]. However, when completed by an experienced reviewer familiar with the research topic, the proportion of missed studies is limited and estimated to be around 3% [344]. Therefore, we cannot exclude the possibility that some studies may have been missed. Nevertheless, we are confident that none of these methodological limitations would change the overall conclusions of this work. Our restriction on articles published in English and French may have inadvertently excluded potentially relevant publications. We cannot exclude the possibility that publications using AHDs for addressing health issues in farming populations may have been missed if there was no mention of AHD in the publications' titles and abstracts. However, it is important to mention that our search strategy was similar to recent reviews that specifically examined the use of AHDs for population-based research [93,96,103,104]. We further broadened our search by including synonyms to improve the comprehensiveness of our literature search. Some details and specificities on the AHDs and other digital data used may be limited because only information reported in each study was used. Shortcomings inherent to bibliometric analysis cannot be excluded. Some authors may have duplicate names, and namesakes could exist. This limitation could not be prevented as a unique author identifier (eg, Open Researcher and Contributor ID number) was not available. Self-citation could not be identified.

While our keyword analysis helped map the farming exposome in AHD-based public health research, this profile is incomplete and potentially biased. Because our review focused on AHD-based studies, we likely missed relevant epidemiological studies not using AHDs, leaving gaps in our understanding of the complete farming exposome across public health. In addition, the variability in keyword reporting practices across journals introduced bias into our keyword analysis. Some journals limit the number of keywords, and the lack of standardized keyword ontologies adds further variability. To mitigate this bias, we manually harmonized the keywords (eg, the use of 1 unique term for a given entity). While this approach is time-consuming, it allows for a more accurate analysis. For instance, if this approach was not performed, the same entity could be designated by various terms that would have been considered separate entities or terms, potentially resulting in underestimates (eg, in the number of publications). Despite these challenges, the findings from our scoping review were consistent with the keyword analysis.

Notwithstanding the aforementioned limitations, most of which are inherent to all scoping reviews and bibliometric analyses [93,96,103,104,158-160], we are confident that our findings can provide a comprehensive picture of what has been published until now (the current state of research and general directions) regarding the use of AHDs for addressing health issues in farming populations. This study may lay the groundwork for researchers to quickly identify research priorities and emerging research directions investigating health issues in farming populations using AHDs.

Conclusions

Technological advancements have greatly increased the volume of research data available, positioning AHDs as critical resources for population-based public health studies [41]. Our review underscores the broad public health implications of AHDs, providing actionable insights for researchers, physicians, and policy makers (Textbox 3). Addressing the identified research gaps is crucial to comprehensively understanding health risks in farming populations.

The insights derived from AHDs can inform meaningful recommendations for policy makers and guide future research directions, ultimately aiding health services and health policy development. Our findings underscore the necessity of comprehensive, interdisciplinary approaches to better understand and mitigate the health risks encountered by farming populations. Such efforts will improve data comparability and research quality while also supporting the formulation of targeted prevention strategies. This, in turn, will enhance health outcomes for farming populations and promote the sustainability of agriculture in an increasingly dynamic environment. The findings from this review offer insights that are not only relevant to farming populations but also potentially generalizable to other populations.

Textbox 3. Take-home messages.

Farming population

Research focusing on low- and middle-income countries, as well as on underrepresented subgroups within farming communities (eg, women, children, and contingent workers), remains insufficiently developed. These areas warrant further investigation to ensure more comprehensive insights.

Administrative health database (AHD) use

The use of AHDs in public health research among farming populations is expanding, offering major potential to enhance epidemiological studies and inform public health decisions. Promoting AHD-based research alongside the integration of other secondary data and artificial intelligence approaches represents a promising direction for future exploration. There is also a need to promote findable, accessible, interoperable, and reusable principles. Creating an AHD catalog or inventory could be a solution that would allow researchers to conduct a “datagraphic search” akin to traditional bibliographic research.

Farming exposure

Published studies on farming-related exposures often rely on broad proxies, such as job titles, neglecting to capture the nuances of specific agricultural tasks. While pesticide exposure remains a predominant research focus, emerging concerns, such as per- and polyfluoroalkyl substances, biological agents, micro or nanoplastics, and the effects of climate change, require urgent attention. The farming exposome remains underexplored despite its potential to uncover important associations between risk factors and a diverse range of health outcomes.

Health outcomes

Cancer, respiratory diseases, neurodegenerative disorders, and mental health issues are among the most frequently studied health outcomes in farming populations. However, significant gaps exist in understanding other critical conditions, such as cardiovascular diseases, reproductive disorders, ocular conditions, autoimmune diseases, musculoskeletal disorders, age-related health issues, and microbiome impacts. Addressing these overlooked areas is essential for a more complete understanding of the health risks faced by farming communities.

Acknowledgments

This work was partly funded by MIAI@Grenoble Alpes (Multidisciplinary Institute in Artificial Intelligence at Grenoble Alpes; grant ANR-19-P3IA-0003, 2019) and the French National Research Agency within the framework of the Investissements d’avenir program (grants ANR - 10 - AIRT - 05, 2012; ANR - 15 - IDEX - 02, 2016). The funding sources had no role in the study design and conduct of the study; collection, management, analysis, and interpretation of data; writing of the report; preparation, review, or approval of the paper; or decision to submit the paper for publication. The authors were not precluded from accessing data in the study, and all authors took final responsibility for the decision to submit for publication.

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author upon reasonable request and also in the multimedia appendices.

Authors' Contributions

PP was involved in the conceptualization, methodology, software selection, validation, formal analysis, investigation, data curation, writing the original draft and reviewing and editing, visualization, project administration, and funding acquisition of the study. NV was involved in the conceptualization, reviewing and editing, resource collection, and funding acquisition of the study.

All authors meet all 4 criteria for authorship in the International Committee of Medical Journal Editors recommendations. All authors affirm that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any potential discrepancies from the study as originally planned (and, if relevant, registered) have been explained.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary materials.
[DOCX File , 8778 KB - publichealth_v11i1e62939_app1.docx]

Multimedia Appendix 2
Publications included and analyzed.
[XLSX File (Microsoft Excel File), 80 KB - publichealth_v11i1e62939_app2.xlsx]

Multimedia Appendix 3
Co-occurrence between exposome-related keywords and health event-related keywords.
[XLSX File (Microsoft Excel File), 39 KB - publichealth_v11i1e62939_app3.xlsx]

Multimedia Appendix 4
PRISMA-ScR checklist.
[PDF File (Adobe PDF File), 186 KB - publichealth_v11i1e62939_app4.pdf]

Multimedia Appendix 5
BIBLIO checklist for reporting the bibliometric reviews of the biomedical literature.
[PDF File (Adobe PDF File), 270 KB - publichealth_v11i1e62939_app5.pdf]

References

1. Azari R, Borisch B. What is public health? A scoping review. Arch Public Health 2023 May 10;81(1):86 [FREE Full text] [doi: 10.1186/s13690-023-01091-6] [Medline: 37165370]
2. Velmovitsky PE, Kirolos M, Alencar P, Leatherdale S, Cowan D, Morita PP. Leveraging mHealth technologies for public health. JMIR Public Health Surveill 2024 Sep 12;10:e49719 [FREE Full text] [doi: 10.2196/49719] [Medline: 39265164]
3. Barouki R, Lévi Y. Rapport 24-02. L'exposome, une contribution majeure pour la prévention. Bull Acad Natl Med 2024 May;208(5):523-535. [doi: 10.1016/j.banm.2024.03.004]
4. Barouki R, Audouze K, Becker C, Blaha L, Coumoul X, Karakitsios S, et al. The exposome and toxicology: a win-win collaboration. Toxicol Sci 2022 Feb 28;186(1):1-11 [FREE Full text] [doi: 10.1093/toxsci/kfab149] [Medline: 34878125]
5. Bennett B, Workman T, Smith MN, Griffith WC, Thompson B, Faustman EM. Characterizing the neurodevelopmental pesticide exposome in a children's agricultural cohort. Int J Environ Res Public Health 2020 Feb 25;17(5):1479 [FREE Full text] [doi: 10.3390/ijerph17051479] [Medline: 32106530]
6. Goodrich JA, Wang H, Jia Q, Stratakis N, Zhao Y, Maitre L, et al. Integrating Multi-Omics with environmental data for precision health: a novel analytic framework and case study on prenatal mercury induced childhood fatty liver disease. Environ Int 2024 Aug;190:108930 [FREE Full text] [doi: 10.1016/j.envint.2024.108930] [Medline: 39128376]
7. Haddad N, Andrianou XD, Makris KC. A scoping review on the characteristics of human exposome studies. Curr Pollution Rep 2019 Nov 28;5(4):378-393. [doi: 10.1007/S40726-019-00130-7]
8. Merra G, Gualtieri P, La Placa G, Frank G, Della Morte D, De Lorenzo A, et al. The relationship between exposome and microbiome. Microorganisms 2024 Jul 08;12(7):1386 [FREE Full text] [doi: 10.3390/microorganisms12071386] [Medline: 39065154]
9. Pandics T, Major D, Fazekas-Pongor V, Szarvas Z, Peterfi A, Mukli P, et al. Exposome and unhealthy aging: environmental drivers from air pollution to occupational exposures. Geroscience 2023 Dec 09;45(6):3381-3408 [FREE Full text] [doi: 10.1007/s11357-023-00913-3] [Medline: 37688657]
10. Petit P. Toxicological and exposure database inventory: a review. Int J Hyg Environ Health 2022 Sep;246:114055. [doi: 10.1016/j.ijheh.2022.114055] [Medline: 36279787]
11. Price EJ, Vitale CM, Miller GW, David A, Barouki R, Audouze K, et al. Merging the exposome into an integrated framework for "omics" sciences. iScience 2022 Mar 18;25(3):103976 [FREE Full text] [doi: 10.1016/j.isci.2022.103976] [Medline: 35310334]
12. Santos S, Maitre L, Warembourg C, Agier L, Richiardi L, Basagaña X, et al. Applying the exposome concept in birth cohort research: a review of statistical approaches. Eur J Epidemiol 2020 Mar 27;35(3):193-204 [FREE Full text] [doi: 10.1007/s10654-020-00625-4] [Medline: 32221742]
13. Siroux V, Agier L, Slama R. The exposome concept: a challenge and a potential driver for environmental health research. Eur Respir Rev 2016 Jun 31;25(140):124-129 [FREE Full text] [doi: 10.1183/16000617.0034-2016] [Medline: 27246588]
14. Tagliaferro S, Maio S, Pirona F, Stanisci I, Sarno G, Silvi P, EarlyFOOD, HEALS EXHES, et al. Assessing external exposome by implementing an environmental data management system using open data. Sci Rep 2024 Jul 26;14(1):17142 [FREE Full text] [doi: 10.1038/s41598-024-62924-0] [Medline: 39060268]
15. Vineis P, Barouki R. The exposome as the science of social-to-biological transitions. Environ Int 2022 Jul;165:107312 [FREE Full text] [doi: 10.1016/j.envint.2022.107312] [Medline: 35635963]
16. Wei X, Huang Z, Jiang L, Li Y, Zhang X, Leng Y, et al. Charting the landscape of the environmental exposome. Imeta 2022 Dec 02;1(4):e50. [doi: 10.1002/imt2.50] [Medline: 38867899]
17. Wild CP. The exposome: from concept to utility. Int J Epidemiol 2012 Feb;41(1):24-32. [doi: 10.1093/ije/dyr236] [Medline: 22296988]
18. Zhang P, Carlsten C, Chaleckis R, Hanhineva K, Huang M, Isobe T, et al. Defining the scope of exposome studies and research needs from a multidisciplinary perspective. Environ Sci Technol Lett 2021 Oct 12;8(10):839-852 [FREE Full text] [doi: 10.1021/acs.estlett.1c00648] [Medline: 34660833]

19. Ibanez A, Melloni L, Świeboda P, Hynes W, Ikiz B, Ayadi R, et al. Neuroecological links of the exposome and one health. Neuron 2024 Jun 19;112(12):1905-1910 [FREE Full text] [doi: 10.1016/j.neuron.2024.04.016] [Medline: 38723637]
20. Liang KH, Colombijn JM, Verhaar MC, Ghannoum M, Timmermans EJ, Vernooij RW. The general external exposome and the development or progression of chronic kidney disease: a systematic review and meta-analyses. Environ Pollut 2024 Oct 01;358:124509 [FREE Full text] [doi: 10.1016/j.envpol.2024.124509] [Medline: 38968981]
21. Canfell OJ, Littlewood R, Burton-Jones A, Sullivan C. Digital health and precision prevention: shifting from disease-centred care to consumer-centred health. Aust Health Rev 2022 Jun 10;46(3):279-283. [doi: 10.1071/AH21063] [Medline: 34882538]
22. Mess F, Blaschke S, Schick TS, Friedrich J. Precision prevention in worksite health-a scoping review on research trends and gaps. PLoS One 2024 Jun 10;19(6):e0304951 [FREE Full text] [doi: 10.1371/journal.pone.0304951] [Medline: 38857277]
23. Wang Y, Zhu M, Ma H, Shen H. Polygenic risk scores: the future of cancer risk prediction, screening, and precision prevention. Med Rev (2021) 2021 Dec;1(2):129-149 [FREE Full text] [doi: 10.1515/mr-2021-0025] [Medline: 37724297]
24. Kalia V, Belsky DW, Baccarelli AA, Miller GW. An exposomic framework to uncover environmental drivers of aging. Exposome 2022;2(1):osac002 [FREE Full text] [doi: 10.1093/exposome/osac002] [Medline: 35295547]
25. Chung MK, House JS, Akhtari FS, Makris KC, Langston MA, Islam KT, Members of the Exposomics Consortium. Decoding the exposome: data science methodologies and implications in exposome-wide association studies (ExWASs). Exposome 2024;4(1):osae001 [FREE Full text] [doi: 10.1093/exposome/osae001] [Medline: 38344436]
26. Andreu-Perez J, Poon CC, Merrifield RD, Wong ST, Yang GZ. Big data for health. IEEE J Biomed Health Inform 2015 Jul;19(4):1193-1208 [FREE Full text] [doi: 10.1109/JBHI.2015.2450362] [Medline: 26173222]
27. Li C, Huang M. Environmental sustainability in the age of big data: opportunities and challenges for business and industry. Environ Sci Pollut Res Int 2023 Dec 03;30(56):119001-119015. [doi: 10.1007/s11356-023-30301-5] [Medline: 37919492]
28. Pastorino R, De Vito C, Migliara G, Glocker K, Binenbaum I, Ricciardi W, et al. Benefits and challenges of Big Data in healthcare: an overview of the European initiatives. Eur J Public Health 2019 Oct 01;29(Supplement_3):23-27 [FREE Full text] [doi: 10.1093/eurpub/ckz168] [Medline: 31738444]
29. Flahault A, Geissbuhler A, Guessous I, Guérin P, Bolon I, Salathé M, et al. Precision global health in the digital age. Swiss Med Wkly 2017 Apr 07;147(1314):w14423 [FREE Full text] [doi: 10.4414/smw.2017.14423] [Medline: 28421566]
30. Manaouil C, Chamot S, Petit P. Le médecin confronté à l'IA (Intelligence artificielle): Éthique et responsabilité. Méd Droit 2024 Jun;2024(186):50-66. [doi: 10.1016/j.meddro.2024.02.001]
31. Samuel G, Lucassen AM. The environmental sustainability of data-driven health research: a scoping review. Digit Health 2022 Jul 05;8:20552076221111297 [FREE Full text] [doi: 10.1177/20552076221111297] [Medline: 35847526]
32. Adam NR, Wieder R, Ghosh D. Data science, learning, and applications to biomedical and health sciences. Ann N Y Acad Sci 2017 Jan 25;1387(1):5-11. [doi: 10.1111/nyas.13309] [Medline: 28122121]
33. Iyamu I, Xu AX, Gómez-Ramírez O, Ablona A, Chang H, Mckee G, et al. Defining digital public health and the role of digitization, digitalization, and digital transformation: scoping review. JMIR Public Health Surveill 2021 Nov 26;7(11):e30399 [FREE Full text] [doi: 10.2196/30399] [Medline: 34842555]
34. Mooney SJ, Pejaver V. Big data in public health: terminology, machine learning, and privacy. Annu Rev Public Health 2018 Apr 01;39:95-112 [FREE Full text] [doi: 10.1146/annurev-publhealth-040617-014208] [Medline: 29261408]
35. Wesson P, Hswen Y, Valdes G, Stojanovski K, Handley MA. Risks and opportunities to ensure equity in the application of big data research in public health. Annu Rev Public Health 2022 Apr 05;43(1):59-78 [FREE Full text] [doi: 10.1146/annurev-publhealth-051920-110928] [Medline: 34871504]
36. Iyamu I, Gómez-Ramírez O, Xu AX, Chang H, Watt S, Mckee G, et al. Challenges in the development of digital public health interventions and mapped solutions: findings from a scoping review. Digit Health 2022;8:20552076221102255. [doi: 10.1177/20552076221102255] [Medline: 35656283]
37. Iyamu I, McKee G, Haag D, Gilbert M. Defining the role of digital public health in the evolving digital health landscape: policy and practice implications in Canada. Health Promot Chronic Dis Prev Can 2024 Feb;44(2):66-69 [FREE Full text] [doi: 10.24095/hpcdp.44.2.04] [Medline: 38353941]
38. Khan S, Kandukuri DK, Subramaniyan EU, MohanaSundaram A. Harnessing the untapped potential of digital twin technology in digital public health interventions. Explor Digit Health Technol 2023 Nov 1;5(4):11-16. [doi: 10.37349/edht.2023.00003]
39. Maaß L, Pan CC, Freye M. Mapping digital public health interventions among existing digital technologies and internet-based interventions to maintain and improve population health in practice: protocol for a scoping review. JMIR Res Protoc 2022 Mar 31;11(3):e33404 [FREE Full text] [doi: 10.2196/33404] [Medline: 35357321]
40. Odone A, Buttigieg S, Ricciardi W, Azzopardi-Muscat N, Staines A. Public health digitalization in Europe. Eur J Public Health 2019 Oct 01;29(Supplement_3):28-35 [FREE Full text] [doi: 10.1093/eurpub/ckz161] [Medline: 31738441]
41. Wienert J, Jahnel T, Maaß L. What are digital public health interventions? First steps toward a definition and an intervention classification framework. J Med Internet Res 2022 Jun 28;24(6):e31921 [FREE Full text] [doi: 10.2196/31921] [Medline: 35763320]
42. Burgun A, Bernal-Delgado E, Kuchinke W, van Staa T, Cunningham J, Lettieri E, et al. Health data for public health: towards new ways of combining data sources to support research efforts in Europe. Yearb Med Inform 2017 Aug 19;26(01):235-240. [doi: 10.1055/S-0037-1606509]

43. Canfell OJ, Davidson K, Sullivan C, Eakin E, Burton-Jones A. Data sources for precision public health of obesity: a scoping review, evidence map and use case in Queensland, Australia. BMC Public Health 2022 Mar 24;22(1):584 [FREE Full text] [doi: 10.1186/s12889-022-12939-x] [Medline: 35331189]
44. Canfell OJ, Davidson K, Woods L, Sullivan C, Cocoros NM, Klompas M, et al. Precision public health for non-communicable diseases: an emerging strategic roadmap and multinational use cases. Front. Public Health 2022 Apr 8;10:854525 [FREE Full text] [doi: 10.3389/fpubh.2022.854525] [Medline: 35462850]
45. Sielemann K, Hafner A, Pucker B. The reuse of public datasets in the life sciences: potential risks and rewards. PeerJ 2020;8:e9954 [FREE Full text] [doi: 10.7717/peerj.9954] [Medline: 33024631]
46. Deal JA, Betz J, Lin FR, Reed NS. Interpreting results from epidemiologic studies. Semin Hear 2021 Feb 15;42(1):3-9 [FREE Full text] [doi: 10.1055/s-0041-1725996] [Medline: 33883787]
47. DiPietro NA. Methods in epidemiology: observational study designs. Pharmacotherapy 2010 Oct 06;30(10):973-984. [doi: 10.1592/phco.30.10.973] [Medline: 20874034]
48. Cancer epidemiology: principles and methods. International Agency for Research on Cancer. 1999. URL: https://publications.iarc.fr/_publications/media/download/4106/a72983d2524d630f6ad0539f747e4780447f1ae3.pdf [accessed 2024-04-29]
49. Munnangi S, Boktor SW. Epidemiology of study design. StatPearls. 2014. URL: https://www.ncbi.nlm.nih.gov/books/NBK470342/ [accessed 2024-04-29]
50. Röhrig B, du Prel JB, Wachtlin D, Blettner M. Types of study in medical research: part 3 of a series on evaluation of scientific publications. Dtsch Arztebl Int 2009 Apr;106(15):262-268 [FREE Full text] [doi: 10.3238/arztebl.2009.0262] [Medline: 19547627]
51. Himali JJ, Baril AA, Cavuoto MG, Yiallourou S, Wiedner CD, Himali D, et al. Association between slow-wave sleep loss and incident dementia. JAMA Neurol 2023 Dec 01;80(12):1326-1333. [doi: 10.1001/jamaneurol.2023.3889] [Medline: 37902739]
52. Mahmood SS, Levy D, Vasan RS, Wang TJ. The Framingham Heart Study and the epidemiology of cardiovascular disease: a historical perspective. Lancet 2014 Mar 15;383(9921):999-1008 [FREE Full text] [doi: 10.1016/S0140-6736(13)61752-3] [Medline: 24084292]
53. Salines G, De LC. Les cohortes : intérêt, role et position de l'InVS. In: Saint-Maurice (Fra) : Institut de veille sanitaire, janvier 2010, 20 p. Princeton, NJ: Princeton University Press; 2018.
54. Zins M, Goldberg M. L’apport des bases de données d’origine administrative aux cohortes épidémiologiques : l’exemple de la cohorte Constances. Statistique et Société 2014;2(2):43-48 [FREE Full text]
55. Chen W, Wang X, Chen J, You C, Ma L, Zhang W, et al. Household air pollution, adherence to a healthy lifestyle, and risk of cardiometabolic multimorbidity: results from the China health and retirement longitudinal study. Sci Total Environ 2023 Jan 10;855:158896. [doi: 10.1016/j.scitotenv.2022.158896] [Medline: 36150596]
56. Zhao Y, Hu Y, Smith JP, Strauss J, Yang G. Cohort profile: the China Health and Retirement Longitudinal Study (CHARLS). Int J Epidemiol 2014 Feb 12;43(1):61-68 [FREE Full text] [doi: 10.1093/ije/dys203] [Medline: 23243115]
57. Kesse-Guyot E, Assmann K, Andreeva V, Castetbon K, Méjean C, Touvier M, et al. Metadata correction of: lessons learned from methodological validation research in e-epidemiology. JMIR Public Health Surveill 2016 Oct 31;2(2):e160 [FREE Full text] [doi: 10.2196/publichealth.6876] [Medline: 27797322]
58. Quantin C. Bases de données médicoadministratives et épidémiologie : intérêt, limites. Revue d'Épidémiologie et de Santé Publique 2008 Sep;56(5):S254. [doi: 10.1016/j.respe.2008.06.011]
59. Babik KR, Shockey TM, Moore LL, Wurzelbacher SJ. Standardizing industrial hygiene data collection forms used by workers' compensation insurers. J Occup Environ Hyg 2018 Sep 04;15(9):676-685 [FREE Full text] [doi: 10.1080/15459624.2018.1490022] [Medline: 29985777]
60. Bosward R, Braunack-Mayer A, Frost E, Carter S. Mapping precision public health definitions, terminology and applications: a scoping review protocol. BMJ Open 2022 Feb 23;12(2):e058069 [FREE Full text] [doi: 10.1136/bmjopen-2021-058069] [Medline: 35197357]
61. Canfell OJ, Kodiyattu Z, Eakin E, Burton-Jones A, Wong I, Macaulay C, et al. Real-world data for precision public health of noncommunicable diseases: a scoping review. BMC Public Health 2022 Nov 24;22(1):2166 [FREE Full text] [doi: 10.1186/s12889-022-14452-7] [Medline: 36434553]
62. Chao K, Sarker MN, Ali I, Firdaus RB, Azman A, Shaed MM. Big data-driven public health policy making: potential for the healthcare industry. Heliyon 2023 Sep;9(9):e19681 [FREE Full text] [doi: 10.1016/j.heliyon.2023.e19681] [Medline: 37809720]
63. Dolley S. Big Data's role in precision public health. Front Public Health 2018 Mar 7;6:68 [FREE Full text] [doi: 10.3389/fpubh.2018.00068] [Medline: 29594091]
64. Doyle CM, Lix LM, Hemmelgarn BR, Paterson JM, Renoux C. Data variability across Canadian administrative health databases: differences in content, coding, and completeness. Pharmacoepidemiol Drug Saf 2020 Jan 10;29 Suppl 1(S1):68-77. [doi: 10.1002/pds.4889] [Medline: 31507029]
65. Gallini A, Moisan F, Maura G, Carcaillon-Bentata L, Leray E, Haesebaert J, et al. [Identification of neurodegenerative diseases in administrative databases in France: a systematic review of the literature]. Rev Epidemiol Sante Publique 2017 Oct;65 Suppl 4:S183-S197. [doi: 10.1016/j.respe.2017.01.115] [Medline: 28341166]

66. Khoury MJ, Iademarco MF, Riley WT. Precision public health for the era of precision medicine. Am J Prev Med 2016 Mar;50(3):398-401 [FREE Full text] [doi: 10.1016/j.amepre.2015.08.031] [Medline: 26547538]
67. McCullough LE, Maliniak ML, Amin AB, Baker JM, Baliashvili D, Barberio J, et al. Epidemiology beyond its limits. Sci Adv 2022 Jun 10;8(23):eabn3328 [FREE Full text] [doi: 10.1126/sciadv.abn3328] [Medline: 35675391]
68. Näher AF, Vorisek CN, Klopfenstein SA, Lehne M, Thun S, Alsalamah S, et al. Secondary data for global health digitalisation. Lancet Digit Health 2023 Feb;5(2):e93-101 [FREE Full text] [doi: 10.1016/S2589-7500(22)00195-9] [Medline: 36707190]
69. Koesten L, Vougiouklis P, Simperl E, Groth P. Dataset reuse: toward translating principles to practice. Patterns (N Y) 2020 Nov 13;1(8):100136 [FREE Full text] [doi: 10.1016/j.patter.2020.100136] [Medline: 33294873]
70. Hu H, Galea S, Rosella L, Henry D. Big data and population health: focusing on the health impacts of the social, physical, and economic environment. Epidemiology 2017 Nov;28(6):759-762. [doi: 10.1097/EDE.0000000000000711] [Medline: 28682850]
71. Shaw RJ, Harron KL, Pescarini JM, Pinto Junior EP, Allik M, Siroky AN, et al. Biases arising from linked administrative data for epidemiological research: a conceptual framework from registration to analyses. Eur J Epidemiol 2022 Dec 05;37(12):1215-1224 [FREE Full text] [doi: 10.1007/s10654-022-00934-w] [Medline: 36333542]
72. Subramanian M, Wojtusciszyn A, Favre L, Boughorbel S, Shan J, Letaief KB, et al. Precision medicine in the era of artificial intelligence: implications in chronic disease management. J Transl Med 2020 Dec 09;18(1):472 [FREE Full text] [doi: 10.1186/s12967-020-02658-5] [Medline: 33298113]
73. Velmovitsky PE, Bevilacqua T, Alencar P, Cowan D, Morita PP. Convergence of precision medicine and public health into precision public health: toward a big data perspective. Front Public Health 2021 Apr 6;9:561873 [FREE Full text] [doi: 10.3389/fpubh.2021.561873] [Medline: 33889555]
74. Zhang H, Hu H, Diller M, Hogan WR, Prosperi M, Guo Y, et al. Semantic standards of external exposome data. Environ Res 2021 Jun;197:111185 [FREE Full text] [doi: 10.1016/j.envres.2021.111185] [Medline: 33901445]
75. Daniels K, Bonnechère B. Harnessing digital health interventions to bridge the gap in prevention for older adults. Front Public Health 2023 Jan 8;11:1281923 [FREE Full text] [doi: 10.3389/fpubh.2023.1281923] [Medline: 38259780]
76. Liu F, Panagiotakos D. Real-world data: a brief review of the methods, applications, challenges and opportunities. BMC Med Res Methodol 2022 Nov 05;22(1):287 [FREE Full text] [doi: 10.1186/s12874-022-01768-6] [Medline: 36335315]
77. Maaß L, Freye M, Pan CC, Dassow HH, Niess J, Jahnel T. The definitions of health apps and medical apps from the perspective of public health and law: qualitative analysis of an interdisciplinary literature overview. JMIR Mhealth Uhealth 2022 Oct 31;10(10):e37980 [FREE Full text] [doi: 10.2196/37980] [Medline: 36315221]
78. Mooney SJ, Garber MD. Sampling and sampling frames in big data epidemiology. Curr Epidemiol Rep 2019 Mar 2;6(1):14-22 [FREE Full text] [doi: 10.1007/s40471-019-0179-y] [Medline: 31360626]
79. Pearson AL, Tribby C, Brown CD, Yang JA, Pfeiffer K, Jankowska MM. Systematic review of best practices for GPS data usage, processing, and linkage in health, exposure science and environmental context research. BMJ Open 2024 Feb 01;14(2):e077036 [FREE Full text] [doi: 10.1136/bmjopen-2023-077036] [Medline: 38307539]
80. Prosperi M, Min JS, Bian J, Modave F. Big data hurdles in precision medicine and precision public health. BMC Med Inform Decis Mak 2018 Dec 29;18(1):139 [FREE Full text] [doi: 10.1186/s12911-018-0719-2] [Medline: 30594159]
81. Safran C. Update on data reuse in health care. Yearb Med Inform 2017 Aug;26(1):24-27 [FREE Full text] [doi: 10.15265/IY-2017-013] [Medline: 29063535]
82. Severo M, Feredj A, Romele A. Soft data and public policy: can social media offer alternatives to official statistics in urban policymaking? Policy Internet 2016 Jul 19;8(3):354-372. [doi: 10.1002/poi3.127]
83. Shakeri Hossein Abad Z, Butler GP, Thompson W, Lee J. Physical activity, sedentary behavior, and sleep on Twitter: multicountry and fully labeled public data set for digital public health surveillance research. JMIR Public Health Surveill 2022 Feb 14;8(2):e32355 [FREE Full text] [doi: 10.2196/32355] [Medline: 35156938]
84. Singh T, Roberts K, Cohen T, Cobb N, Wang J, Fujimoto K, et al. Social media as a research tool (SMaaRT) for risky behavior analytics: methodological review. JMIR Public Health Surveill 2020 Nov 30;6(4):e21660 [FREE Full text] [doi: 10.2196/21660] [Medline: 33252345]
85. Zins M, Cuggia M, Goldberg M. Les données de santé en France - Abondantes mais complexes. Med Sci (Paris) 2021 Feb 16;37(2):179-184 [FREE Full text] [doi: 10.1051/medsci/2021001] [Medline: 33591261]
86. Motahari-Nezhad H, Fgaier M, Mahdi Abid M, Péntek M, Gulácsi L, Zrubka Z. Digital biomarker-based studies: scoping review of systematic reviews. JMIR Mhealth Uhealth 2022 Oct 24;10(10):e35722 [FREE Full text] [doi: 10.2196/35722] [Medline: 36279171]
87. Ayilara OF, Platt RW, Dahl M, Coulombe J, Ginestet PG, Chateau D, et al. Generating synthetic data from administrative health records for drug safety and effectiveness studies. Int J Popul Data Sci 2023 Nov 27;8(1):2176 [FREE Full text] [doi: 10.23889/ijpds.v8i1.2176] [Medline: 38414538]
88. Fadel M, Petot J, Gourraud PA, Descatha A. Flexibility of a large blindly synthetized avatar database for occupational research: example from the CONSTANCES cohort for stroke and knee pain. PLoS One 2024 Jul 31;19(7):e0308063 [FREE Full text] [doi: 10.1371/journal.pone.0308063] [Medline: 39083487]
89. Giuffrè M, Shung DL. Harnessing the power of synthetic data in healthcare: innovation, application, and privacy. NPJ Digit Med 2023 Oct 09;6(1):186 [FREE Full text] [doi: 10.1038/s41746-023-00927-3] [Medline: 37813960]

90. Kamel Boulos MN, Zhang P. Digital twins: from personalised medicine to precision public health. J Pers Med 2021 Jul 29;11(8):745 [FREE Full text] [doi: 10.3390/jpm11080745] [Medline: 34442389]
91. Vallevik VB, Babic A, Marshall SE, Elvatun S, Brøgger HB, Alagaratnam S, et al. Can I trust my fake data - a comprehensive quality assessment framework for synthetic tabular data in healthcare. Int J Med Inform 2024 May;185:105413 [FREE Full text] [doi: 10.1016/j.ijmedinf.2024.105413] [Medline: 38493547]
92. Chamot S, Manaouil C, Fantoni S, Frimat P, Al-Salameh A, Desailloud R, et al. Système national des données de santé (SNDS) : perspectives en santé au travail et environnementale. Arch Mal Prof Environ 2023 Nov;84(6):101875. [doi: 10.1016/j.admp.2023.101875]
93. Cohen S, Gilutz H, Marelli AJ, Iserin L, Benis A, Bonnet D, et al. Administrative health databases for addressing emerging issues in adults with CHD: a systematic review. Cardiol Young 2018 Jun 29;28(6):844-853. [doi: 10.1017/S1047951118000446] [Medline: 29704902]
94. Fraser HS, Mugisha M, Bacher I, Ngenzi JL, Seebregts C, Umubyeyi A, et al. Factors influencing data quality in electronic health record systems in 50 health facilities in Rwanda and the role of clinical alerts: cross-sectional observational study. JMIR Public Health Surveill 2024 Jul 03;10:e49127 [FREE Full text] [doi: 10.2196/49127] [Medline: 38959048]
95. Gavrielov-Yusim N, Friger M. Use of administrative medical databases in population-based research. J Epidemiol Community Health 2014 Mar;68(3):283-287. [doi: 10.1136/jech-2013-202744] [Medline: 24248997]
96. Hinds A, Lix LM, Smith M, Quan H, Sanmartin C. Quality of administrative health databases in Canada: a scoping review. Can J Public Health 2016 Jun 27;107(1):e56-e61 [FREE Full text] [doi: 10.17269/cjph.107.5244] [Medline: 27348111]
97. Howell CR, Zhang L, Clay OJ, Dutton G, Horton T, Mugavero MJ, et al. Social determinants of health phenotypes and cardiometabolic condition prevalence among patients in a large academic health system: latent class analysis. JMIR Public Health Surveill 2024 Aug 07;10:e53371 [FREE Full text] [doi: 10.2196/53371] [Medline: 39113389]
98. Jones KH, Heys S, Tingay KS, Jackson P, Dibben C. The good, the bad, the clunky: improving the use of administrative data for research. Int J Popul Data Sci 2018 Dec 04;4(1):587 [FREE Full text] [doi: 10.23889/ijpds.v4i1.587] [Medline: 32935024]
99. Namulanda G, Qualters J, Vaidyanathan A, Roberts E, Richardson M, Fraser A, et al. Electronic health record case studies to advance environmental public health tracking. J Biomed Inform 2018 Mar;79:98-104 [FREE Full text] [doi: 10.1016/j.jbi.2018.02.012] [Medline: 29476967]
100. Nicholls SG, Quach P, von Elm E, Guttmann A, Moher D, Petersen I, et al. The REporting of studies conducted using observational routinely-collected health data (RECORD) statement: methods for arriving at consensus and developing reporting guidelines. PLoS One 2015;10(5):e0125620 [FREE Full text] [doi: 10.1371/journal.pone.0125620] [Medline: 25965407]
101. Silverwood R, Rajah N, Calderwood L, De Stavola B, Harron K, Ploubidis G. Examining the quality and population representativeness of linked survey and administrative data: guidance and illustration using linked 1958 National Child Development Study and Hospital Episode Statistics data. Int J Popul Data Sci 2024 Jan 09;9(1):2137 [FREE Full text] [doi: 10.23889/ijpds.v9i1.2137] [Medline: 38425790]
102. Zuiderwijk A, Türk BO, Brazier F. Identifying the most important facilitators of open research data sharing and reuse in Epidemiology: a mixed-methods study. PLoS One 2024 Feb 8;19(2):e0297969 [FREE Full text] [doi: 10.1371/journal.pone.0297969] [Medline: 38330007]
103. Hobensack M, Song J, Scharp D, Bowles KH, Topaz M. Machine learning applied to electronic health record data in home healthcare: a scoping review. Int J Med Inform 2023 Feb;170:104978 [FREE Full text] [doi: 10.1016/j.ijmedinf.2022.104978] [Medline: 36592572]
104. Tassi MF, le Meur N, Stéfic K, Grammatico-Guillon L. Performance of French medico-administrative databases in epidemiology of infectious diseases: a scoping review. Front Public Health 2023 May 12;11:1161550 [FREE Full text] [doi: 10.3389/fpubh.2023.1161550] [Medline: 37250067]
105. Yoshida S, Takeuchi M, Tanaka-Mizuno S, Mizuno K, Nakashima M, Fukasawa T, et al. Clinical epidemiology and pharmacoepidemiology studies with real-world databases. Proc Jpn Acad Ser B Phys Biol Sci 2022;98(10):517-528 [FREE Full text] [doi: 10.2183/pjab.98.026] [Medline: 36504194]
106. Howren A, Aviña-Zubieta JA, Puyat JH, Esdaile JM, Da Costa D, De Vera MA. Defining depression and anxiety in individuals with rheumatic diseases using administrative health databases: a systematic review. Arthritis Care Res (Hoboken) 2020 Feb 09;72(2):243-255. [doi: 10.1002/acr.24048] [Medline: 31421021]
107. Kim MK, Rouphael C, McMichael J, Welch N, Dasarathy S. Challenges in and opportunities for electronic health record-based data analysis and interpretation. Gut Liver 2024 Mar 15;18(2):201-208 [FREE Full text] [doi: 10.5009/gnl230272] [Medline: 37905424]
108. Lee S, Xu Y, D Apos Souza AG, Martin EA, Doktorchik C, Zhang Z, et al. Unlocking the potential of electronic health records for health research. Int J Popul Data Sci 2020 Jan 30;5(1):1123 [FREE Full text] [doi: 10.23889/ijpds.v5i1.1123] [Medline: 32935049]
109. Pletscher F, Mändli Lerch K, Glinz D. Willingness to share anonymised routinely collected clinical health data in Switzerland: a cross-sectional survey. Swiss Med Wkly 2022 Jun 20;152(2324):w30182 [FREE Full text] [doi: 10.4414/smw.2022.w30182] [Medline: 35752970]

110. Yuan H, Ali MS, Brouwer ES, Girman CJ, Guo JJ, Lund JL, ISPE Comparative Effectiveness Research Special Interest Group. Real-world evidence: what it is and what it can tell us according to the International Society for Pharmacoepidemiology (ISPE) Comparative Effectiveness Research (CER) Special Interest Group (SIG). Clin Pharmacol Ther 2018 Aug 07;104(2):239-241. [doi: 10.1002/cpt.1086] [Medline: 29733448]
111. Andrew A, Zhou J, Gui J, Harrison A, Shi X, Li M, et al. Pesticides applied to crops and amyotrophic lateral sclerosis risk in the U.S. Neurotoxicology 2021 Dec;87:128-135 [FREE Full text] [doi: 10.1016/j.neuro.2021.09.004] [Medline: 34562505]
112. Chazard E, Ficheur G, Caron A, Lamer A, Labreuche J, Cuggia M, et al. Secondary use of healthcare structured data: the challenge of domain-knowledge based extraction of features. Stud Health Technol Inform 2018;255:15-19. [Medline: 30306898]
113. Grath-Lone LM, Jay MA, Blackburn R, Gordon E, Zylbersztejn A, Wiljaars L, et al. What makes administrative data "research-ready"? A systematic review and thematic analysis of published literature. Int J Popul Data Sci 2022;7(1):1718 [FREE Full text] [doi: 10.23889/ijpds.v6i1.1718] [Medline: 35520099]
114. Hashimoto RE, Brodt ED, Skelly AC, Dettori JR. Administrative database studies: goldmine or goose chase? Evid Based Spine Care J 2014 Oct 24;5(2):74-76 [FREE Full text] [doi: 10.1055/s-0034-1390027] [Medline: 25278880]
115. Bezin J, Duong M, Lassalle R, Droz C, Pariente A, Blin P, et al. The national healthcare system claims databases in France, SNIIRAM and EGB: powerful tools for pharmacoepidemiology. Pharmacoepidemiol Drug Saf 2017 Aug 24;26(8):954-962. [doi: 10.1002/pds.4233] [Medline: 28544284]
116. Al-Sahab B, Leviton A, Loddenkemper T, Paneth N, Zhang B. Biases in electronic health records data for generating real-world evidence: an overview. J Healthc Inform Res 2024 Mar 14;8(1):121-139. [doi: 10.1007/s41666-023-00153-2] [Medline: 38273982]
117. Bastarache L, Brown JS, Cimino JJ, Dorr DA, Embi PJ, Payne PR, et al. Developing real-world evidence from real-world data: transforming raw data into analytical datasets. Learn Health Syst 2022 Jan 14;6(1):e10293 [FREE Full text] [doi: 10.1002/lrh2.10293] [Medline: 35036557]
118. Fang J, Kapral MK, Richards J, Robertson A, Stamplecoski M, Silver FL. The Registry of Canadian Stroke Network : an evolving methodology. Acta Neurol Taiwan 2011 Jun;20(2):77-84. [Medline: 21739386]
119. Fillinger S, de la Garza L, Peltzer A, Kohlbacher O, Nahnsen S. Challenges of big data integration in the life sciences. Anal Bioanal Chem 2019 Oct 28;411(26):6791-6800. [doi: 10.1007/s00216-019-02074-9] [Medline: 31463515]
120. Fonteneau L, Le Meur N, Cohen-Akenine A, Pessel C, Brouard C, Delon F, et al. [The use of administrative health databases in infectious disease epidemiology and public health]. Rev Epidemiol Sante Publique 2017 Oct;65 Suppl 4:S174-S182 [FREE Full text] [doi: 10.1016/j.respe.2017.03.131] [Medline: 28624133]
121. Silenou BC, Avalos M, Helmer C, Berr C, Pariente A, Jacqmin-Gadda H. Health administrative data enrichment using cohort information: comparative evaluation of methods by simulation and application to real data. PLoS One 2019 Jan 31;14(1):e0211118 [FREE Full text] [doi: 10.1371/journal.pone.0211118] [Medline: 30703112]
122. Wang SV, Schneeweiss S. Data checks before registering study protocols for health care database analyses. JAMA 2024 May 07;331(17):1445-1446. [doi: 10.1001/jama.2024.2988] [Medline: 38587830]
123. Yamana H, Yamasaki K, Jo T, Yatsuhashi H, Yasunaga H. A descriptive analysis of acute viral hepatitis using a database with electronic medical records and claims data. Ann Clin Epidemiol 2023;5(4):107-112 [FREE Full text] [doi: 10.37737/ace.23014] [Medline: 38504953]
124. Gianfrancesco MA, Goldstein ND. A narrative review on the validity of electronic health record-based research in epidemiology. BMC Med Res Methodol 2021 Oct 27;21(1):234 [FREE Full text] [doi: 10.1186/s12874-021-01416-5] [Medline: 34706667]
125. Denisov NS, Kamenskikh EM, Fedorova OS. Trends in population-based studies: molecular and digital epidemiology (review). Sovrem Tekhnologii Med 2022 Aug 29;14(4):60-70 [FREE Full text] [doi: 10.17691/stm2022.14.4.07] [Medline: 37179982]
126. Hamm NC, Hamad AF, Wall-Wieler E, Roos LL, Plana-Ripoll O, Lix LM. Multigenerational health research using population-based linked databases: an international review. Int J Popul Data Sci 2021 Oct 07;6(1):1686 [FREE Full text] [doi: 10.23889/ijpds.v6i1.1686] [Medline: 34734126]
127. Poulain R, Tarek MF, Beheshti R. Improving fairness in AI models on electronic health records: the case for federated learning methods. FAccT 23 (2023) 2023 Jun;2023:1599-1608 [FREE Full text] [doi: 10.1145/3593013.3594102] [Medline: 37990734]
128. Tuppin P, Rudant J, Constantinou P, Gastaldi-Ménager C, Rachas A, de Roquefeuil L, et al. Value of a national administrative database to guide public decisions: from the système national d'information interrégimes de l'Assurance Maladie (SNIIRAM) to the système national des données de santé (SNDS) in France. Rev Epidemiol Sante Publique 2017 Oct;65 Suppl 4:S149-S167. [doi: 10.1016/j.respe.2017.05.004] [Medline: 28756037]
129. Casey JA, Schwartz BS, Stewart WF, Adler NE. Using electronic health records for population health research: a review of methods and applications. Annu Rev Public Health 2016;37:61-81 [FREE Full text] [doi: 10.1146/annurev-publhealth-032315-021353] [Medline: 26667605]

130. Chan JM, Carroll MW, Smyth M, Hamilton Z, Evans D, McGrail K, et al. Comparing health administrative and clinical registry data: trends in incidence and prevalence of pediatric inflammatory bowel disease in British Columbia. Clin Epidemiol 2021;13:81-90 [FREE Full text] [doi: 10.2147/CLEP.S292546] [Medline: 33603489]
131. Chowdhury T, Hemmelgarn B. Evidence-based decision-making 6: utilization of administrative databases for health services research. In: Parfrey PS, Barrett BJ, editors. Clinical Epidemiology: Practice and Methods. Cham, Switzerland: Springer; 2015:469-484.
132. Chishtie J, Sapiro N, Wiebe N, Rabatach L, Lorenzetti D, Leung AA, et al. Use of epic electronic health record system for health care research: scoping review. J Med Internet Res 2023 Dec 15;25:e51003 [FREE Full text] [doi: 10.2196/51003] [Medline: 38100185]
133. Jutte DP, Roos LL, Brownell MD. Administrative record linkage as a tool for public health research. Annu Rev Public Health 2011;32:91-108. [doi: 10.1146/annurev-publhealth-031210-100700] [Medline: 21219160]
134. Mazzali C, Duca P. Use of administrative data in healthcare research. Intern Emerg Med 2015 Jun 25;10(4):517-524. [doi: 10.1007/s11739-015-1213-9] [Medline: 25711312]
135. McPheeters ML, Sathe NA, Jerome RN, Carnahan RM. Methods for systematic reviews of administrative database studies capturing health outcomes of interest. Vaccine 2013 Dec 30;31 Suppl 10:K2-K6. [doi: 10.1016/j.vaccine.2013.06.048] [Medline: 24331070]
136. Sato S, Yasunaga H. A review of studies using Japanese nationwide administrative claims databases. Ann Clin Epidemiol 2023;5(2):58-64 [FREE Full text] [doi: 10.37737/ace.23008] [Medline: 38505730]
137. Mues KE, Liede A, Liu J, Wetmore J, Zaha R, Bradbury BD, et al. Use of the Medicare database in epidemiologic and health services research: a valuable source of real-world evidence on the older and disabled populations in the US. Clin Epidemiol 2017;9:267-277 [FREE Full text] [doi: 10.2147/CLEP.S105613] [Medline: 28533698]
138. Bernstein CN. Large registry epidemiology in IBD. Inflamm Bowel Dis 2017 Nov;23(11):1941-1949. [doi: 10.1097/MIB.0000000000001279] [Medline: 28991858]
139. Bloom PP, Tapper EB. The use of administrative data to investigate the population burden of hepatic encephalopathy. J Clin Med 2020 Nov 10;9(11):3620 [FREE Full text] [doi: 10.3390/jcm9113620] [Medline: 33182743]
140. Hudson M, Tascilar K, Suissa S. Comparative effectiveness research with administrative health data in rheumatoid arthritis. Nat Rev Rheumatol 2016 Jun 15;12(6):358-366. [doi: 10.1038/nrrheum.2016.34] [Medline: 27080692]
141. Mitchell E, Waring T, Ahern E, O'Donovan D, O'Reilly D, Bradley DT. Predictors and consequences of homelessness in whole-population observational studies that used administrative data: a systematic review. BMC Public Health 2023 Aug 24;23(1):1610 [FREE Full text] [doi: 10.1186/s12889-023-16503-z] [Medline: 37612701]
142. Ulrich EH, So G, Zappitelli M, Chanchlani R. A review on the application and limitations of administrative health care data for the study of acute kidney injury epidemiology and outcomes in children. Front Pediatr 2021 Oct 27;9:742888 [FREE Full text] [doi: 10.3389/fped.2021.742888] [Medline: 34778133]
143. Curl CL, Spivak M, Phinney R, Montrose L. Synthetic pesticides and health in vulnerable populations: agricultural workers. Curr Environ Health Rep 2020 Mar 20;7(1):13-29 [FREE Full text] [doi: 10.1007/s40572-020-00266-5] [Medline: 31960353]
144. Demos K, Sazakli E, Jelastopulu E, Charokopos N, Ellul J, Leotsinidis M. Does farming have an effect on health status? A comparison study in west Greece. Int J Environ Res Public Health 2013 Feb 26;10(3):776-792 [FREE Full text] [doi: 10.3390/ijerph10030776] [Medline: 23442558]
145. Kachuri L, Harris MA, MacLeod JS, Tjepkema M, Peters PA, Demers PA. Cancer risks in a population-based study of 70,570 agricultural workers: results from the Canadian census health and Environment cohort (CanCHEC). BMC Cancer 2017 May 19;17(1):343 [FREE Full text] [doi: 10.1186/s12885-017-3346-x] [Medline: 28525996]
146. Ramos AK, Girdžiūtė L, Starič J, Rautianinen RH. Identifying "vulnerable agricultural populations" at risk for occupational injuries and illnesses: a European perspective. J Agromedicine 2021 Jul 27;26(3):340-345. [doi: 10.1080/1059924X.2020.1771498] [Medline: 32458756]
147. Sturm ET, Castro C, Mendez-Colmenares A, Duffy J, Burzynska AZ, Stallones L, et al. Risk factors for brain health in agricultural work: a systematic review. Int J Environ Res Public Health 2022 Mar 13;19(6):3373 [FREE Full text] [doi: 10.3390/ijerph19063373] [Medline: 35329061]
148. Guindo Y, Parent ME, Richard H, Luce D, Barul C. Expert-based assessment of chemical and physical exposures, and organizational factors, in past agricultural jobs. Environ Res 2024 Dec 15;263(Pt 3):120238 [FREE Full text] [doi: 10.1016/j.envres.2024.120238] [Medline: 39461702]
149. Komarek AM, De Pinto A, Smith VH. A review of types of risks in agriculture: what we know and what we need to know. Agric Syst 2020 Feb;178:102738. [doi: 10.1016/j.agsy.2019.102738]
150. Leon ME, Schinasi LH, Lebailly P, Beane Freeman LE, Nordby KC, Ferro G, et al. Pesticide use and risk of non-Hodgkin lymphoid malignancies in agricultural cohorts from France, Norway and the USA: a pooled analysis from the AGRICOH consortium. Int J Epidemiol 2019 Oct 01;48(5):1519-1535 [FREE Full text] [doi: 10.1093/ije/dyz017] [Medline: 30880337]
151. Pérès K, Matharan F, Allard M, Amieva H, Baldi I, Barberger-Gateau P, et al. Health and aging in elderly farmers: the AMI cohort. BMC Public Health 2012 Jul 27;12(1):558 [FREE Full text] [doi: 10.1186/1471-2458-12-558] [Medline: 22839764]

152. Petit P, Chamot S, Al-Salameh A, Cancé C, Desailloud R, Bonneterre V. Farming activity and risk of treated thyroid disorders: insights from the TRACTOR project, a nationwide cohort study. Environ Res 2024 May 15;249:118458 [FREE Full text] [doi: 10.1016/j.envres.2024.118458] [Medline: 38365059]
153. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
154. The Joanna Briggs Institute Reviewers' Manual 2015 methodology for JBI scoping reviews. The Joanna Briggs Institute. 2015. URL: https://reben.com.br/revista/wp-content/uploads/2020/10/Scoping.pdf [accessed 2024-04-03]
155. Beydon M, Pinto S, De Rycke Y, Fautrel B, Mariette X, Seror R, et al. Risk of cancer for patients with rheumatoid arthritis versus general population: a national claims database cohort study. Lancet Reg Health Eur 2023 Dec;35:100768 [FREE Full text] [doi: 10.1016/j.lanepe.2023.100768] [Medline: 37954220]
156. Moysan M. Des archives à la base de données : valeur(s) des données en épidémiologie. Hist Méd Santé 2022 Dec 15(22):151-164 [FREE Full text] [doi: 10.4000/hms.6274]
157. Syed R, Eden R, Makasi T, Chukwudi I, Mamudu A, Kamalpour M, et al. Digital health data quality issues: systematic review. J Med Internet Res 2023 Mar 31;25:e42615 [FREE Full text] [doi: 10.2196/42615] [Medline: 37000497]
158. Cobo MJ, López-Herrera AG, Herrera-Viedma E, Herrera F. An approach for detecting, quantifying, and visualizing the evolution of a research field: a practical application to the Fuzzy Sets Theory field. J Informetr 2011 Jan;5(1):146-166. [doi: 10.1016/j.joi.2010.10.002]
159. Donthu N, Kumar S, Mukherjee D, Pandey N, Lim WM. How to conduct a bibliometric analysis: an overview and guidelines. J Bus Res 2021 Sep;133:285-296. [doi: 10.1016/j.jbusres.2021.04.070]
160. Moral-Muñoz JA, Herrera-Viedma E, Santisteban-Espejo A, Cobo MJ. Software tools for conducting bibliometric analysis in science: an up-to-date review. Prof Inf J 2020 Jan 19;29(1) [FREE Full text] [doi: 10.3145/epi.2020.ene.03]
161. Bienert IR, Oliveira RC, Andrade PB, Caramori CA. Bibliometric indexes, databases and impact factors in cardiology. Rev Bras Cir Cardiovasc 2015;30(2):254-259 [FREE Full text] [doi: 10.5935/1678-9741.20150019] [Medline: 26107458]
162. Bihari A, Tripathi S, Deepak A. A review on h-index and its alternative indices. J Inf Sci 2021 May 31;49(3):624-665. [doi: 10.1177/01655515211014478]
163. Shao Y, Chien TW, Jang FL. The use of radar plots with the Yk-index to identify which authors contributed the most to the journal of Medicine in 2020 and 2021: a bibliometric analysis. Medicine (Baltimore) 2022 Nov 11;101(45):e31033 [FREE Full text] [doi: 10.1097/MD.0000000000031033] [Medline: 36397440]
164. Kumar S, Kumar S. Trends of collaborative research in journals of oilseeds research (India), 1993-2004. Indian J Agric Library Inf Services 2008;24:90.
165. Wu T, Duan Y, Zhang T, Tian W, Liu H, Deng Y. Research trends in the application of artificial intelligence in oncology: a bibliometric and network visualization study. Front Biosci (Landmark Ed) 2022 Aug 31;27(9):254 [FREE Full text] [doi: 10.31083/j.fbl2709254] [Medline: 36224012]
166. Wang B, Zhao P. Worldwide research productivity in the field of back pain: a bibliometric analysis. Medicine (Baltimore) 2018 Oct;97(40):e11566 [FREE Full text] [doi: 10.1097/MD.0000000000011566] [Medline: 30290588]
167. Aria M, Cuccurullo C, D'Aniello L, Misuraca M, Spano M. Thematic analysis as a new culturomic tool: the social media coverage on COVID-19 pandemic in Italy. Sustainability 2022 Mar 20;14(6):3643. [doi: 10.3390/su14063643]
168. Petit P, Leroyer A, Chamot S, Fumery M, Bonneterre V. Farming activities and risk of inflammatory bowel disease: a French nationwide population-based cohort study. J Crohns Colitis 2024 Sep 03;18(9):1415-1429. [doi: 10.1093/ecco-jcc/jjae050] [Medline: 38605515]
169. Montazeri A, Mohammadi S, M Hesari P, Ghaemi M, Riazi H, Sheikhi-Mobarakeh Z. Preliminary guideline for reporting bibliometric reviews of the biomedical literature (BIBLIO): a minimum requirements. Syst Rev 2023 Dec 15;12(1):239 [FREE Full text] [doi: 10.1186/s13643-023-02410-2] [Medline: 38102710]
170. Aria M, Cuccurullo C. bibliometrix : an R-tool for comprehensive science mapping analysis. J Informetr 2017 Nov;11(4):959-975. [doi: 10.1016/j.joi.2017.08.007]
171. Spitzer WO, Hill GB, Chambers LW, Helliwell BE, Murphy HB. The occupation of fishing as a risk factor in cancer of the lip. N Engl J Med 1975 Aug 28;293(9):419-424. [doi: 10.1056/nejm197508282930903]
172. Alavanja MC, Samanic C, Dosemeci M, Lubin J, Tarone R, Lynch CF, et al. Use of agricultural pesticides and prostate cancer risk in the Agricultural Health Study cohort. Am J Epidemiol 2003 May 01;157(9):800-814. [doi: 10.1093/aje/kwg040] [Medline: 12727674]
173. Alavanja MC, Dosemeci M, Samanic C, Lubin J, Lynch CF, Knott C, et al. Pesticides and lung cancer risk in the agricultural health study cohort. Am J Epidemiol 2004 Nov 01;160(9):876-885. [doi: 10.1093/aje/kwh290] [Medline: 15496540]
174. Andersen A, Barlow L, Engeland A, Kjaerheim K, Lynge E, Pukkala E. Work-related cancer in the Nordic countries. Scand J Work Environ Health 1999;25 Suppl 2:1-116 [FREE Full text] [Medline: 10507118]
175. Andreotti G, Koutros S, Hofmann JN, Sandler DP, Lubin JH, Lynch CF, et al. Glyphosate use and cancer incidence in the agricultural health study. J Natl Cancer Inst 2018 May 01;110(5):509-516 [FREE Full text] [doi: 10.1093/jnci/djx233] [Medline: 29136183]

176. Elbaz A, Clavel J, Rathouz PJ, Moisan F, Galanaud J, Delemotte B, et al. Professional exposure to pesticides and Parkinson disease. Ann Neurol 2009 Oct;66(4):494-504. [doi: 10.1002/ana.21717] [Medline: 19847896]
177. Engel LS, Hill DA, Hoppin JA, Lubin JH, Lynch CF, Pierce J, et al. Pesticide use and breast cancer risk among farmers' wives in the agricultural health study. Am J Epidemiol 2005 Jan 15;161(2):121-135. [doi: 10.1093/aje/kwi022] [Medline: 15632262]
178. Eriksson M, Karlsson M. Occupational and other environmental factors and multiple myeloma: a population based case-control study. Br J Ind Med 1992 Feb 01;49(2):95-103 [FREE Full text] [doi: 10.1136/oem.49.2.95] [Medline: 1536825]
179. Hemminki K, Saloniemi I, Salonen T, Partanen T, Vainio H. Childhood cancer and parental occupation in Finland. J Epidemiol Community Health 1981 Mar 01;35(1):11-15 [FREE Full text] [doi: 10.1136/jech.35.1.11] [Medline: 7264527]
180. Karjalainen A, Kurppa K, Martikainen R, Klaukka T, Karjalainen J. Work is related to a substantial portion of adult-onset asthma incidence in the Finnish population. Am J Respir Crit Care Med 2001 Aug 15;164(4):565-568. [doi: 10.1164/ajrccm.164.4.2012146] [Medline: 11520716]
181. Karjalainen A, Martikainen R, Klaukka T, Saarinen K, Uitti J. Risk of asthma among Finnish patients with occupational rhinitis. Chest 2003 Jan;123(1):283-288. [doi: 10.1378/chest.123.1.283] [Medline: 12527633]
182. Kristensen P, Andersen A, Irgens LM, Laake P, Bye AS. Incidence and risk factors of cancer among men and women in Norwegian agriculture. Scand J Work Environ Health 1996 Feb;22(1):14-26 [FREE Full text] [doi: 10.5271/sjweh.104] [Medline: 8685669]
183. Lee WJ, Sandler DP, Blair A, Samanic C, Cross AJ, Alavanja MC. Pesticide use and colorectal cancer risk in the Agricultural Health Study. Int J Cancer 2007 Jul 15;121(2):339-346 [FREE Full text] [doi: 10.1002/ijc.22635] [Medline: 17390374]
184. Levecque C, Elbaz A, Clavel J, Richard F, Vidal JS, Amouyel P, et al. Association between Parkinson's disease and polymorphisms in the nNOS and iNOS genes in a community-based case-control study. Hum Mol Genet 2003 Jan 01;12(1):79-86. [doi: 10.1093/hmg/ddg009] [Medline: 12490535]
185. Mills PK, Yang R. Prostate cancer risk in California farm workers. J Occup Environ Med 2003 Mar;45(3):249-258. [doi: 10.1097/01.jom.0000058339.05741.0c] [Medline: 12661182]
186. Morrison H, Savitz D, Semenciw R, Hulka B, Mao Y, Morison D, et al. Farming and prostate cancer mortality. Am J Epidemiol 1993 Feb 01;137(3):270-280. [doi: 10.1093/oxfordjournals.aje.a116674] [Medline: 8452135]
187. Orton SM, Wald L, Confavreux C, Vukusic S, Krohn J, Ramagopalan S, et al. Association of UV radiation with multiple sclerosis prevalence and sex ratio in France. Neurology 2011 Feb;76(5):425-431. [doi: 10.1212/wnl.0b013e31820a0a9f]
188. Pouchieu C, Piel C, Carles C, Gruber A, Helmer C, Tual S, et al. Pesticide use in agriculture and Parkinson's disease in the AGRICAN cohort study. Int J Epidemiol 2018 Feb 01;47(1):299-310. [doi: 10.1093/ije/dyx225] [Medline: 29136149]
189. Pukkala E, Martinsen JI, Lynge E, Gunnarsdottir HK, Sparén P, Tryggvadottir L, et al. Occupation and cancer - follow-up of 15 million people in five Nordic countries. Acta Oncol 2009 Jul 21;48(5):646-790. [doi: 10.1080/02841860902913546] [Medline: 19925375]
190. Reif J, Pearce N, Fraser J. Cancer risks in New Zealand farmers. Int J Epidemiol 1989 Dec;18(4):768-774. [doi: 10.1093/ije/18.4.768] [Medline: 2621012]
191. Stokes L, Stark A, Marshall E, Narang A. Neurotoxicity among pesticide applicators exposed to organophosphates. Occup Environ Med 1995 Oct 01;52(10):648-653 [FREE Full text] [doi: 10.1136/oem.52.10.648] [Medline: 7489054]
192. Vingård E, Alfredsson L, Goldie I, Hogstedt C. Occupation and osteoarthrosis of the hip and knee: a register-based cohort study. Int J Epidemiol 1991 Dec;20(4):1025-1031. [doi: 10.1093/ije/20.4.1025] [Medline: 1800399]
193. Vukusic S, Van Bockstael V, Gosselin S, Confavreux C. Regional variations in the prevalence of multiple sclerosis in French farmers. J Neurol Neurosurg Psychiatry 2007 Jul 18;78(7):707-709 [FREE Full text] [doi: 10.1136/jnnp.2006.101196] [Medline: 17299020]
194. Weichenthal S, Villeneuve PJ, Burnett RT, van Donkelaar A, Martin RV, Jones RR, et al. Long-term exposure to fine particulate matter: association with nonaccidental and cardiovascular mortality in the agricultural health study cohort. Environ Health Perspect 2014 Jun;122(6):609-615 [FREE Full text] [doi: 10.1289/ehp.1307277] [Medline: 24633320]
195. Wigle DT, Semenciw RM, Wilkins K, Riedel D, Ritter L, Morrison HI, et al. Mortality study of Canadian male farm operators: non-Hodgkin's lymphoma mortality and agricultural practices in Saskatchewan. J Natl Cancer Inst 1990 Apr 04;82(7):575-582. [doi: 10.1093/jnci/82.7.575] [Medline: 2313734]
196. Wiklund K, Dich J, Holm LE, Eklund G. Risk of cancer in pesticide applicators in Swedish agriculture. Br J Ind Med 1989 Nov 01;46(11):809-814 [FREE Full text] [doi: 10.1136/oem.46.11.809] [Medline: 2590647]
197. Horwitz H, Ahlgren B, Nærum E. Effect of occupation on risk of developing MS: an insurance cohort study. BMJ Open 2013 Jun 20;3(6):e002894 [FREE Full text] [doi: 10.1136/bmjopen-2013-002894] [Medline: 23794592]
198. Karttunen JP, Rautiainen RH, Quendler E. Gender division of farm work and occupational injuries. J Agric Saf Health 2019 Jul 16;25(3):117-127. [doi: 10.13031/jash.13177] [Medline: 32429612]
199. Keller JE, Howe HL. Risk factors for lung cancer among nonsmoking Illinois residents. Environ Res 1993 Jan;60(1):1-11. [doi: 10.1006/enrs.1993.1001] [Medline: 8432260]
200. Marant Micallef C, Shield KD, Vignat J, Baldi I, Charbotel B, Fervers B, et al. Cancers in France in 2015 attributable to occupational exposures. Int J Hyg Environ Health 2019 Jan;222(1):22-29. [doi: 10.1016/j.ijheh.2018.07.015] [Medline: 30174219]

201. Moisan F, Spinosi J, Dupupet JL, Delabre L, Mazurie J, Goldberg M, et al. The relation between type of farming and prevalence of Parkinson's disease among agricultural workers in five French districts. Mov Disord 2011 Feb 01;26(2):271-279 [FREE Full text] [doi: 10.1002/mds.23370] [Medline: 21412834]
202. Toch-Marquardt M, Menvielle G, Eikemo TA, Kulhánová I, Kulik MC, Bopp M, Euro-GBD-SE consortium. Occupational class inequalities in all-cause and cause-specific mortality among middle-aged men in 14 European populations during the early 2000s. PLoS One 2014 Sep 30;9(9):e108072 [FREE Full text] [doi: 10.1371/journal.pone.0108072] [Medline: 25268702]
203. Tüchsen F, Andersen O, Costa G, Filakti H, Marmot MG. Occupation and ischemic heart disease in the European Community: a comparative study of occupations at potential high risk. Am J Ind Med 1996 Oct;30(4):407-414. [doi: 10.1002/(SICI)1097-0274(199610)30:4<407::AID-AJIM5>3.0.CO;2-S] [Medline: 8892545]
204. Wang CC, Lin GL, Lin YJ, Chen WL, Wu WT. Occupational health surveillance and detection of emerging occupational diseases among Taiwan farmers, through analysis of national-based farmers' and medico-administrative databases. Am J Ind Med 2023 Jan 17;66(1):85-93. [doi: 10.1002/ajim.23443] [Medline: 36398405]
205. Langholz PL, Sommerseth HL. Old-age mortality and social class in northern Norway in the first half of the twentieth century. Scand Econ Hist Rev 2023 Sep 13;72(2):153-170. [doi: 10.1080/03585522.2023.2255599]
206. Das AV, Bagga B, Joseph J. Clinical profile and demographic distribution of keratitis: an electronic medical record-driven data analytics from an eye care network in India. Ocul Immunol Inflamm 2024 Jul 21;32(5):722-726. [doi: 10.1080/09273948.2023.2199331] [Medline: 37084281]
207. Lemarchand C, Tual S, Levêque-Morlais N, Perrier S, Belot A, Velten M, AGRICAN group. Cancer incidence in the AGRICAN cohort study (2005-2011). Cancer Epidemiol 2017 Aug;49:175-185 [FREE Full text] [doi: 10.1016/j.canep.2017.06.003] [Medline: 28709062]
208. Scott E, Hirabayashi L, Levenstein A, Krupa N, Jenkins P. The development of a machine learning algorithm to identify occupational injuries in agriculture using pre-hospital care reports. Health Inf Sci Syst 2021 Dec 29;9(1):31 [FREE Full text] [doi: 10.1007/s13755-021-00161-9] [Medline: 34422257]
209. Wilke RA, Qamar M, Lupu RA, Gu S, Zhao J. Chronic kidney disease in agricultural communities. Am J Med 2019 Oct;132(10):e727-e732 [FREE Full text] [doi: 10.1016/j.amjmed.2019.03.036] [Medline: 30998912]
210. Engel LS, Werder E, Satagopan J, Blair A, Hoppin JA, Koutros S, et al. Insecticide use and breast cancer risk among farmers' wives in the agricultural health study. Environ Health Perspect 2017 Sep 06;125(9):097002 [FREE Full text] [doi: 10.1289/EHP1295] [Medline: 28934092]
211. Louis LM, Lerro CC, Friesen MC, Andreotti G, Koutros S, Sandler DP, et al. A prospective study of cancer risk among agricultural health study farm spouses associated with personal use of organochlorine insecticides. Environ Health 2017 Sep 06;16(1):95 [FREE Full text] [doi: 10.1186/s12940-017-0298-1] [Medline: 28874165]
212. Elbaek Pedersen J, Hansen J. Risk of breast cancer in daughters of agricultural workers in Denmark. Environ Res 2024 Jan 01;240(Pt 1):117374. [doi: 10.1016/j.envres.2023.117374] [Medline: 37866542]
213. Kasten-Arias C, Hodes T, Marino M, Kaufmann J, Lucas JA, Estela Vasquez Guzman C, et al. Healthcare utilization for asthma exacerbation among children of migrant and seasonal farmworkers. Prev Med Rep 2024 Feb;38:102598 [FREE Full text] [doi: 10.1016/j.pmedr.2024.102598] [Medline: 38283959]
214. Mengist Dessie A, Fenta Feleke S, Getaye Workie S, Getinet Abebe T, Mossu Chanie Y, Kassa Yalew A. Prevalence of skin disease and its associated factors among primary schoolchildren: a cross-sectional study from a Northern Ethiopian town. Clin Cosmet Investig Dermatol 2022;15:791-801 [FREE Full text] [doi: 10.2147/CCID.S361051] [Medline: 35521561]
215. Mills PK, Zahm SH. Organophosphate pesticide residues in urine of farmworkers and their children in Fresno County, California. Am J Ind Med 2001;40(5):571-577. [doi: 10.1002/ajim.10007.abs]
216. Andrew N, Beare R, Ravipati T, Parker E, Snowdon D, Naude K, et al. Developing a linked electronic health record derived data platform to support research into healthy ageing. Int J Popul Data Sci 2023 Jun 12;8(1):2129 [FREE Full text] [doi: 10.23889/ijpds.v8i1.2129] [Medline: 37670961]
217. Contiero P, Borgini A, Bertoldi M, Abita A, Cuffari G, Tomao P, et al. An epidemiological study to investigate links between atmospheric pollution from farming and SARS-CoV-2 mortality. Int J Environ Res Public Health 2022 Apr 12;19(8):4637 [FREE Full text] [doi: 10.3390/ijerph19084637] [Medline: 35457503]
218. Lotterman A, Baliatsas C, de Rooij MM, Huss A, Jacobs J, Dückers M, et al. Increased risk of pneumonia amongst residents living near goat farms in different livestock-dense regions in the Netherlands. PLoS One 2023 Jul 5;18(7):e0286972 [FREE Full text] [doi: 10.1371/journal.pone.0286972] [Medline: 37405987]
219. Nerich V, Monnet E, Weill A, Vallier N, Vanbockstael V, Auleley G, et al. Fine-scale geographic variations of inflammatory bowel disease in France: correlation with socioeconomic and house equipment variables. Inflamm Bowel Dis 2010 May;16(5):813-821. [doi: 10.1002/ibd.21122] [Medline: 19902547]
220. Nilles JD, Lim D, Boyer MP, Wilson BD, Betar RA, Showalter HA, et al. The occurrence of bone and joint cancers and their association with rural living and radon exposure in Iowa. Environ Geochem Health 2023 Mar 05;45(3):925-940 [FREE Full text] [doi: 10.1007/s10653-022-01261-5] [Medline: 35381949]

221. Yzermans CJ, Moleman YP, Spreeuwenberg P, Nielen MM, Dückers ML, Smit LA, et al. Risk of pneumonia in the vicinity of goat farms: a comparative assessment of temporal variation based on longitudinal health data. Pneumonia (Nathan) 2023 Sep 05;15(1):13 [FREE Full text] [doi: 10.1186/s41479-023-00115-7] [Medline: 37667350]
222. Petit P, Bosson-Rieutort D, Maugard C, Gondard E, Ozenfant D, Joubert N, et al. The TRACTOR project: TRACking and MoniToring occupational risks in agriculture using French insurance health data (MSA). Ann Work Expo Health 2022 Mar 15;66(3):402-411. [doi: 10.1093/annweh/wxab083] [Medline: 34562080]
223. Mills PK, Kwong S. Cancer incidence in the United Farmworkers of America (UFW), 1987–1997. Am J Ind Med 2001;40(5):596-603. [doi: 10.1002/ajim.1125.abs]
224. Guillien A, Laurent L, Soumagne T, Puyraveau M, Laplante J, Andujar P, et al. Anxiety and depression among dairy farmers: the impact of COPD. Int J Chron Obstruct Pulmon Dis 2018;13:1-9 [FREE Full text] [doi: 10.2147/COPD.S143883] [Medline: 29296078]
225. A consortium of agricultural cohort studies. AGRICOH. URL: https://agricoh.iarc.fr/ [accessed 2024-05-17]
226. Jouneau S, Pinault M, Gouyet T, Brinchault G, Guillot S, Viel J, GERPRA. [AIRBAg study: preliminary results after one year of screening for COPD in dairy farmers]. Rev Mal Respir 2015 Mar;32(3):240-248. [doi: 10.1016/j.rmr.2014.03.014] [Medline: 25847201]
227. Soto S, Yoder AM, Nuño T, Aceves B, Sepulveda R, Rosales CB. Health conditions among farmworkers in the southwest: an analysis of the National Agricultural Workers survey. Front Public Health 2022 Nov 3;10:962085 [FREE Full text] [doi: 10.3389/fpubh.2022.962085] [Medline: 36407981]
228. Contaxis N, Clark J, Dellureficio A, Gonzales S, Mannheimer S, Oxley PR, et al. Ten simple rules for improving research data discovery. PLoS Comput Biol 2022 Feb 10;18(2):e1009768 [FREE Full text] [doi: 10.1371/journal.pcbi.1009768] [Medline: 35143479]
229. Vesteghem C, Brøndum RF, Sønderkær M, Sommer M, Schmitz A, Bødker JS, et al. Implementing the FAIR Data Principles in precision oncology: review of supporting initiatives. Brief Bioinform 2020 May 21;21(3):936-945 [FREE Full text] [doi: 10.1093/bib/bbz044] [Medline: 31263868]
230. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 2016 Mar 15;3(1):160018 [FREE Full text] [doi: 10.1038/sdata.2016.18] [Medline: 26978244]
231. Roux P, Guillien A, Soumagne T, Ritter O, Laplante J, Travers C, et al. Smoking habits in French farmers: a cross-sectional study. BMC Public Health 2017 Feb 04;17(1):166 [FREE Full text] [doi: 10.1186/s12889-017-4030-4] [Medline: 28160776]
232. Audi C, Baïz N, Maesano CN, Ramousse O, Reboulleau D, Magnan A, et al. Serum cytokine levels related to exposure to volatile organic compounds and PM2.5 in dwellings and workplaces in French farmers - a mechanism to explain nonsmoking COPD. Int J Chron Obstruct Pulmon Dis 2017;12:1363-1374 [FREE Full text] [doi: 10.2147/COPD.S117866] [Medline: 28503065]
233. Baldi I, Gruber A, Rondeau V, Lebailly P, Brochard P, Fabrigoule C. Neurobehavioral effects of long-term exposure to pesticides: results from the 4-year follow-up of the PHYTONER study. Occup Environ Med 2011 Feb 22;68(2):108-115. [doi: 10.1136/oem.2009.047811] [Medline: 21097948]
234. Christensen CH, Platz EA, Andreotti G, Blair A, Hoppin JA, Koutros S, et al. Coumaphos exposure and incident cancer among male participants in the Agricultural Health Study (AHS). Environ Health Perspect 2010 Jan;118(1):92-96 [FREE Full text] [doi: 10.1289/ehp.0800446] [Medline: 20056581]
235. Achard P, Maugard C, Cancé C, Spinosi J, Ozenfant D, Maître A, et al. Medico-administrative data combined with agricultural practices data to retrospectively estimate pesticide use by agricultural workers. J Expo Sci Environ Epidemiol 2020 Jul 04;30(4):743-755. [doi: 10.1038/s41370-019-0166-x] [Medline: 31484997]
236. Hohenadel K, Harris SA, McLaughlin JR, Spinelli JJ, Pahwa P, Dosman JA, et al. Exposure to multiple pesticides and risk of non-Hodgkin lymphoma in men from six Canadian provinces. Int J Environ Res Public Health 2011 Jun 21;8(6):2320-2330 [FREE Full text] [doi: 10.3390/ijerph8062320] [Medline: 21776232]
237. Chung EY, Risi D, Holt JL, Lonergan M, Kotwal S, Yong K, et al. Retrospective study on the epidemiology of antineutrophil cytoplasmic autoantibodies-associated vasculitis in two Australian health districts. Intern Med J 2022 Apr 13;52(4):605-613. [doi: 10.1111/imj.15098] [Medline: 33040456]
238. Li X, Sundquist J, Sundquist K. Socioeconomic and occupational risk factors for rheumatoid arthritis: a nationwide study based on hospitalizations in Sweden. J Rheumatol 2008 Jun;35(6):986-991. [Medline: 18464310]
239. Apalla Z, Lallas A, Sotiriou E, Lazaridou E, Vakirlis E, Trakatelli M, et al. Farmers develop more aggressive histologic subtypes of basal cell carcinoma. Experience from a tertiary hospital in Northern Greece. J Eur Acad Dermatol Venereol 2016 Apr 17;30 Suppl 3(S3):17-20. [doi: 10.1111/jdv.13605] [Medline: 26995018]
240. Lope V, Pollán M, Gustavsson P, Plato N, Pérez-Gómez B, Aragonés N, et al. Occupation and thyroid cancer risk in Sweden. J Occup Environ Med 2005 Sep;47(9):948-957. [doi: 10.1097/01.jom.0000169564.21523.5d] [Medline: 16155480]
241. Nordby KC, Andersen A, Kristensen P. Incidence of lip cancer in the male Norwegian agricultural population. Cancer Causes Control 2004 Aug;15(6):619-626. [doi: 10.1023/b:caco.0000036169.90864.e2]

242. Kristensen P, Andersen A, Irgens LM. Hormone-dependent cancer and adverse reproductive outcomes in farmers' families--effects of climatic conditions favoring fungal growth in grain. Scand J Work Environ Health 2000 Aug;26(4):331-337 [FREE Full text] [doi: 10.5271/sjweh.550] [Medline: 10994799]
243. Døssing M, Petersen KT, Vyberg M, Olsen JH. Liver cancer among employees in Denmark. Am J Ind Med 1997 Sep;32(3):248-254. [doi: 10.1002/(sici)1097-0274(199709)32:3<248::aid-ajim10>3.0.co;2-v] [Medline: 9219654]
244. Klingelschmidt J, Chastang JF, Khireddine-Medouni I, Chérié-Challine L, Niedhammer I. [Occupational factors associated with suicide among French employees from the special agricultural social security scheme (MSA) working between 2007 and 2013]. Rev Epidemiol Sante Publique 2020 Feb;68(1):1-8. [doi: 10.1016/j.respe.2019.09.008] [Medline: 31843361]
245. Manninen P, Heliövaara M, Riihimäki H, Mäkelä P. Does psychological distress predict disability? Int J Epidemiol 1997 Oct;26(5):1063-1070. [doi: 10.1093/ije/26.5.1063] [Medline: 9363529]
246. Zöller B, Li X, Sundquist J, Sundquist K. Socioeconomic and occupational risk factors for venous thromboembolism in Sweden: a nationwide epidemiological study. Thromb Res 2012 May;129(5):577-582. [doi: 10.1016/j.thromres.2011.07.050] [Medline: 21868069]
247. Lucyk K, Lu M, Sajobi T, Quan H. Administrative health data in Canada: lessons from history. BMC Med Inform Decis Mak 2015 Aug 19;15(1):69 [FREE Full text] [doi: 10.1186/s12911-015-0196-9] [Medline: 26286712]
248. Kab S, Spinosi J, Chaperon L, Dugravot A, Singh-Manoux A, Moisan F, et al. Agricultural activities and the incidence of Parkinson's disease in the general French population. Eur J Epidemiol 2017 Mar 9;32(3):203-216. [doi: 10.1007/s10654-017-0229-z] [Medline: 28185034]
249. Perrin L, Spinosi J, Chaperon L, Kab S, Moisan F, Ebaz A. Pesticides expenditures by farming type and incidence of Parkinson disease in farmers: a French nationwide study. Environ Res 2021 Jun;197:111161. [doi: 10.1016/j.envres.2021.111161] [Medline: 33887276]
250. Carles C, Bouvier G, Lebailly P, Baldi I. Use of job-exposure matrices to estimate occupational exposure to pesticides: a review. J Expo Sci Environ Epidemiol 2017 Mar;27(2):125-140. [doi: 10.1038/jes.2016.25] [Medline: 27189257]
251. Tudi M, Li H, Li H, Wang L, Lyu J, Yang L, et al. Exposure routes and health risks associated with pesticide application. Toxics 2022 Jun 19;10(6):335 [FREE Full text] [doi: 10.3390/toxics10060335] [Medline: 35736943]
252. Ahmad MF, Ahmad FA, Alsayegh AA, Zeyaullah M, AlShahrani AM, Muzammil K, et al. Pesticides impacts on human health and the environment with their mechanisms of action and possible countermeasures. Heliyon 2024 Apr 15;10(7):e29128 [FREE Full text] [doi: 10.1016/j.heliyon.2024.e29128] [Medline: 38623208]
253. Cavalier H, Trasande L, Porta M. Exposures to pesticides and risk of cancer: evaluation of recent epidemiological evidence in humans and paths forward. Int J Cancer 2023 Mar 01;152(5):879-912 [FREE Full text] [doi: 10.1002/ijc.34300] [Medline: 36134639]
254. Cresto N, Forner-Piquer I, Baig A, Chatterjee M, Perroy J, Goracci J, et al. Pesticides at brain borders: impact on the blood-brain barrier, neuroinflammation, and neurological risk trajectories. Chemosphere 2023 May;324:138251. [doi: 10.1016/j.chemosphere.2023.138251] [Medline: 36878369]
255. Richardson JR, Fitsanakis V, Westerink RH, Kanthasamy AG. Neurotoxicity of pesticides. Acta Neuropathol 2019 Sep 13;138(3):343-362 [FREE Full text] [doi: 10.1007/s00401-019-02033-9] [Medline: 31197504]
256. Sun J, Fang R, Wang H, Xu D, Yang J, Huang X, et al. A review of environmental metabolism disrupting chemicals and effect biomarkers associating disease risks: where exposomics meets metabolomics. Environ Int 2022 Jan;158:106941 [FREE Full text] [doi: 10.1016/j.envint.2021.106941] [Medline: 34689039]
257. Blair A, Freeman LB. Epidemiologic studies in agricultural populations: observations and future directions. J Agromedicine 2009 May 07;14(2):125-131 [FREE Full text] [doi: 10.1080/10599240902779436] [Medline: 19437268]
258. Carvalho JG, Augusto HC, Ferraz R, Delerue-Matos C, Fernandes VC. Micro(nano)plastic and related chemicals: emerging contaminants in environment, food and health impacts. Toxics 2024 Oct 20;12(10):762 [FREE Full text] [doi: 10.3390/toxics12100762] [Medline: 39453182]
259. Clarke K, Manrique A, Sabo-Attwood T, Coker ES. A narrative review of occupational air pollution and respiratory health in farmworkers. Int J Environ Res Public Health 2021 Apr 13;18(8):4097 [FREE Full text] [doi: 10.3390/ijerph18084097] [Medline: 33924663]
260. Domingo NG, Balasubramanian S, Thakrar SK, Clark MA, Adams PJ, Marshall JD, et al. Air quality-related health damages of food. Proc Natl Acad Sci U S A 2021 May 18;118(20):e2013637118 [FREE Full text] [doi: 10.1073/pnas.2013637118] [Medline: 33972419]
261. Hao Y, Sun H, Zeng X, Dong G, Kronzucker HJ, Min J, et al. Smallholder vegetable farming produces more soil microplastics pollution than large-scale farming. Environ Pollut 2023 Jan 15;317:120805. [doi: 10.1016/j.envpol.2022.120805] [Medline: 36470457]
262. Ng EL, Huerta Lwanga E, Eldridge SM, Johnston P, Hu H, Geissen V, et al. An overview of microplastic and nanoplastic pollution in agroecosystems. Sci Total Environ 2018 Jun 15;627:1377-1388. [doi: 10.1016/j.scitotenv.2018.01.341] [Medline: 30857101]
263. van Kersen W, Bossers A, de Steenhuijsen Piters WA, de Rooij MM, Bonten M, Fluit AC, et al. Air pollution from livestock farms and the oropharyngeal microbiome of COPD patients and controls. Environ Int 2022 Nov;169:107497 [FREE Full text] [doi: 10.1016/j.envint.2022.107497] [Medline: 36088872]

264. Ganter M. Zoonotic risks from small ruminants. Vet Microbiol 2015 Dec 14;181(1-2):53-65. [doi: 10.1016/j.vetmic.2015.07.015] [Medline: 26275853]
265. Golomazou E, Mamedova S, Eslahi AV, Karanis P. Cryptosporidium and agriculture: a review. Sci Total Environ 2024 Mar 15;916:170057. [doi: 10.1016/j.scitotenv.2024.170057] [Medline: 38242460]
266. McDaniel CJ, Cardwell DM, Moeller Jr RB, Gray GC. Humans and cattle: a review of bovine zoonoses. Vector Borne Zoonotic Dis 2014 Jan;14(1):1-19 [FREE Full text] [doi: 10.1089/vbz.2012.1164] [Medline: 24341911]
267. Arora K, Xu L, Bhagianadh D. Dementia and cognitive decline in older adulthood: are agricultural workers at greater risk? J Gerontol B Psychol Sci Soc Sci 2021 Sep 13;76(8):1629-1643 [FREE Full text] [doi: 10.1093/geronb/gbab005] [Medline: 33406265]
268. Douphrate DI, Rosecrance JC, Reynolds SJ, Stallones L, Gilkey DP. Tractor-related injuries: an analysis of workers' compensation data. J Agromedicine 2009 May 07;14(2):198-205. [doi: 10.1080/10599240902773215] [Medline: 19437278]
269. Dos Anjos Magri C, Garófallo Garcia R, Binotto E, Duarte da Silva Lima N, de Alencar Nääs I, Sgavioli S, et al. Occupational risk factors in health of broiler-farm workers: a systematic review. Arch Environ Occup Health 2021 Oct 15;76(8):482-493. [doi: 10.1080/19338244.2020.1832036] [Medline: 33054688]
270. Barbosa Junior M, Sokulski CC, Salvador R, Pinheiro E, de Francisco AC, Trojan F. What kills the agricultural worker? A systematic review on suicide. Rural Remote Health 2021 Aug;21(3):6067 [FREE Full text] [doi: 10.22605/RRH6067] [Medline: 34384223]
271. Garrett-Wright D, Malin C, Jones MS. Mental health in farming communities. J Psychosoc Nurs Ment Health Serv 2023 Oct;61(10):39-43. [doi: 10.3928/02793695-20230424-02] [Medline: 37134283]
272. Lunner Kolstrup C, Kallioniemi M, Lundqvist P, Kymäläinen HR, Stallones L, Brumby S. International perspectives on psychosocial working conditions, mental health, and stress of dairy farm operators. J Agromedicine 2013 Jul 03;18(3):244-255. [doi: 10.1080/1059924X.2013.796903] [Medline: 23844791]
273. Petit P, Gandon G, Dubuc M, Vuillerme N, Bonneterre V. Agricultural activities and risk of treatment for depressive disorders among the entire French agricultural workforce: the TRACTOR project, a nationwide retrospective cohort study. Lancet Reg Health Eur 2023 Aug;31:100674 [FREE Full text] [doi: 10.1016/j.lanepe.2023.100674] [Medline: 37408876]
274. Daghagh Yazd S, Wheeler SA, Zuo A. Key risk factors affecting farmers' mental health: a systematic review. Int J Environ Res Public Health 2019 Dec 02;16(23):4849 [FREE Full text] [doi: 10.3390/ijerph16234849] [Medline: 31810320]
275. Nordby K, Andersen A, Irgens LM, Kristensen P. Indicators of mancozeb exposure in relation to thyroid cancer and neural tube defects in farmers' families. Scand J Work Environ Health 2005 Apr;31(2):89-96 [FREE Full text] [doi: 10.5271/sjweh.855] [Medline: 15864902]
276. Kauppinen T, Uuksulainen S, Saalo A, Mäkinen I, Pukkala E. Use of the Finnish Information System on Occupational Exposure (FINJEM) in epidemiologic, surveillance, and other applications. Ann Occup Hyg 2014 Apr;58(3):380-396. [doi: 10.1093/annhyg/met074] [Medline: 24401793]
277. Shim YK, Mlynarek SP, van Wijngaarden E. Parental exposure to pesticides and childhood brain cancer: U.S. Atlantic coast childhood brain cancer study. Environ Health Perspect 2009 Jun;117(6):1002-1006 [FREE Full text] [doi: 10.1289/ehp.0800209] [Medline: 19590697]
278. Baldi I, Carles C, Blanc-Lapierre A, Fabbro-Peray P, Druet-Cabanac M, Boutet-Robinet E, PESTIMAT Group. A French crop-exposure matrix for use in epidemiological studies on pesticides: PESTIMAT. J Expo Sci Environ Epidemiol 2017 Jan 23;27(1):56-63. [doi: 10.1038/jes.2015.72] [Medline: 26696463]
279. Oltramare C, Mediouni Z, Shoman Y, Hopf NB, Graczyk H, Berthet A. Determinants of pesticide exposure in occupational studies: a meta-analysis. Toxics 2023 Jul 18;11(7):623 [FREE Full text] [doi: 10.3390/toxics11070623] [Medline: 37505588]
280. Burstyn I, Lavoué J, Van Tongeren M. Aggregation of exposure level and probability into a single metric in job-exposure matrices creates bias. Ann Occup Hyg 2012 Nov;56(9):1038-1050. [doi: 10.1093/annhyg/mes031] [Medline: 22986426]
281. Fadel M, Evanoff BA, Andersen JH, d'Errico A, Dale AM, Leclerc A, et al. Not just a research method: if used with caution, can job-exposure matrices be a useful tool in the practice of occupational medicine and public health? Scand J Work Environ Health 2020 Sep 01;46(5):552-553 [FREE Full text] [doi: 10.5271/sjweh.3900] [Medline: 32367143]
282. Peters S, Vienneau D, Sampri A, Turner MC, Castaño-Vinyals G, Bugge M, et al. Occupational exposure assessment tools in Europe: a comprehensive inventory overview. Ann Work Expo Health 2022 Jun 06;66(5):671-686 [FREE Full text] [doi: 10.1093/annweh/wxab110] [Medline: 34935027]
283. Petit P, Gondard E, Gandon G, Moreaud O, Sauvée M, Bonneterre V. Agricultural activities and risk of Alzheimer's disease: the TRACTOR project, a nationwide retrospective cohort study. Eur J Epidemiol 2024 Mar 10;39(3):271-287 [FREE Full text] [doi: 10.1007/s10654-023-01079-0] [Medline: 38195954]
284. List of classifications. International Agency for Research on Cancer. URL: https://monographs.iarc.who.int/list-of-classifications [accessed 2024-10-30]
285. Desai B, Sahni S, Jordan H, Sahni R, Reinbeau R, Nguyen A, et al. Risk of cardiovascular disease in male farmers over the age of 45: a review of literature. Cureus 2022 May;14(5):e24642 [FREE Full text] [doi: 10.7759/cureus.24642] [Medline: 35663675]

286. Kim H, Jung W, Jung S, Cho S, Jung I, Song H, et al. Is farming a risk occupation for cardio-cerebrovascular diseases? A scoping review on cardio-cerebrovascular disease risk in farmers. J Prev Med Public Health 2024 Aug 23;57(6):521-529 [FREE Full text] [doi: 10.3961/jpmph.24.302] [Medline: 39438011]
287. Hulkkonen S, Shiri R, Auvinen J, Miettunen J, Karppinen J, Ryhänen J. Risk factors of hospitalization for carpal tunnel syndrome among the general working population. Scand J Work Environ Health 2020 Jan 01;46(1):43-49 [FREE Full text] [doi: 10.5271/sjweh.3835] [Medline: 31168633]
288. Kim J, Youn K, Park J. Risk factors for musculoskeletal disorders in Korean farmers: survey on occupational diseases in 2020 and 2022. Healthcare (Basel) 2024 Oct 11;12(20):2026 [FREE Full text] [doi: 10.3390/healthcare12202026] [Medline: 39451441]
289. Hemminki K, Niemi ML, Saloniemi I, Vainio H, Hemminki E. Spontaneous abortions by occupation and social class in Finland. Int J Epidemiol 1980 Jun;9(2):149-153. [doi: 10.1093/ije/9.2.149] [Medline: 7409966]
290. Strand K, Wergeland E, Bjerkedal T. Fertility patterns according to occupational grouping in Norway, 1989. Scand J Soc Med 1996 Mar 01;24(1):50-54. [doi: 10.1177/140349489602400108] [Medline: 8740876]
291. Liu Y, Huang L, Hu F, Zhang X. Investigating frailty, polypharmacy, malnutrition, chronic conditions, and quality of life in older adults: large population-based study. JMIR Public Health Surveill 2024 Oct 11;10:e50617 [FREE Full text] [doi: 10.2196/50617] [Medline: 39145920]
292. Lie A, Skogstad M, Johannessen HA, Tynes T, Mehlum IS, Nordby K, et al. Occupational noise exposure and hearing: a systematic review. Int Arch Occup Environ Health 2016 Apr 7;89(3):351-372 [FREE Full text] [doi: 10.1007/s00420-015-1083-5] [Medline: 26249711]
293. Kraemer JG, Aebi S, Oppliger A, Hilty M. The Indoor-Air Microbiota of pig farms drives the composition of the pig farmers' nasal microbiota in a season-dependent and farm-specific manner. Appl Environ Microbiol 2019 May 01;85(9):e03038 [FREE Full text] [doi: 10.1128/AEM.03038-18] [Medline: 30824439]
294. Mucci N, Tommasi E, Chiarelli A, Lulli LG, Traversini V, Galea RP, et al. WORKbiota: a systematic review about the effects of occupational exposure on microbiota and workers' health. Int J Environ Res Public Health 2022 Jan 18;19(3):1043 [FREE Full text] [doi: 10.3390/ijerph19031043] [Medline: 35162072]
295. Shukla SK, Ye Z, Sandberg S, Reyes I, Fritsche TR, Keifer M. The nasal microbiota of dairy farmers is more complex than oral microbiota, reflects occupational exposure, and provides competition for staphylococci. PLoS One 2017 Aug 29;12(8):e0183898 [FREE Full text] [doi: 10.1371/journal.pone.0183898] [Medline: 28850578]
296. Sun J, Liao XP, D'Souza AW, Boolchandani M, Li S, Cheng K, et al. Environmental remodeling of human gut microbiota and antibiotic resistome in livestock farms. Nat Commun 2020 Mar 18;11(1):1427 [FREE Full text] [doi: 10.1038/s41467-020-15222-y] [Medline: 32188862]
297. Vilela C, Araújo B, Soares-Guedes C, Caridade-Silva R, Martins-Macedo J, Teixeira C, et al. From the gut to the brain: is microbiota a new paradigm in Parkinson’s disease treatment? Cells 2024 Apr 30;13(9):770 [FREE Full text] [doi: 10.3390/cells13090770] [Medline: 38727306]
298. Gois MF, Fernández-Pato A, Huss A, Gacesa R, Wijmenga C, Weersma RK, et al. Impact of occupational pesticide exposure on the human gut microbiome. Front Microbiol 2023 Aug 10;14:1223120 [FREE Full text] [doi: 10.3389/fmicb.2023.1223120] [Medline: 37637104]
299. Christen P, Schnell R. Thirty-three myths and misconceptions about population data: from data capture and processing to linkage. Int J Popul Data Sci 2023 Jan 31;8(1):2115 [FREE Full text] [doi: 10.23889/ijpds.v8i1.2115] [Medline: 37636835]
300. Dattani N, Hardelid P, Davey J, Gilbert R. Accessing electronic administrative health data for research takes time. Arch Dis Child 2013 May 28;98(5):391-392. [doi: 10.1136/archdischild-2013-303730] [Medline: 23448863]
301. Verheij RA, Curcin V, Delaney BC, McGilchrist MM. Possible sources of bias in primary care electronic health record data use and reuse. J Med Internet Res 2018 May 29;20(5):e185 [FREE Full text] [doi: 10.2196/jmir.9134] [Medline: 29844010]
302. Bousquet P, Caillet P, Coeuret-Pellicer M, Goulard H, Kudjawu Y, Le Bihan C, et al. [Using cancer case identification algorithms in medico-administrative databases: literature review and first results from the REDSIAM Tumors group based on breast, colon, and lung cancer]. Rev Epidemiol Sante Publique 2017 Oct;65 Suppl 4:S236-S242. [doi: 10.1016/j.respe.2017.04.057] [Medline: 28576379]
303. Goldberg M, Carton M, Doussin A, Fagot-Campagna A, Heyndrickx E, Lemaitre M, et al. [The REDSIAM network]. Rev Epidemiol Sante Publique 2017 Oct;65 Suppl 4:S144-S148. [doi: 10.1016/j.respe.2017.06.001] [Medline: 28844426]
304. Lipscombe LL, Hwee J, Webster L, Shah BR, Booth GL, Tu K. Identifying diabetes cases from administrative data: a population-based validation study. BMC Health Serv Res 2018 May 02;18(1):316 [FREE Full text] [doi: 10.1186/s12913-018-3148-0] [Medline: 29720153]
305. Caputo M, Pecere A, Sarro A, Mele C, Ucciero A, Pagano L, et al. Incidence and prevalence of hyperthyroidism: a population-based study in the Piedmont Region, Italy. Endocrine 2020 Jul 13;69(1):107-112. [doi: 10.1007/s12020-020-02222-7] [Medline: 32056093]
306. Habib RR, Hojeij S, Elzein K. Gender in occupational health research of farmworkers: a systematic review. Am J Ind Med 2014 Dec 16;57(12):1344-1367 [FREE Full text] [doi: 10.1002/ajim.22375] [Medline: 25227724]

307. Biswas A, Harbin S, Irvin E, Johnston H, Begum M, Tiong M, et al. Sex and gender differences in occupational hazard exposures: a scoping review of the recent literature. Curr Environ Health Rep 2021 Dec 27;8(4):267-280 [FREE Full text] [doi: 10.1007/s40572-021-00330-8] [Medline: 34839446]
308. Biswas A, Tiong M, Irvin E, Zhai G, Sinkins M, Johnston H, et al. Gender and sex differences in occupation-specific infectious diseases: a systematic review. Occup Environ Med 2024 Aug 29;81(8):425-432 [FREE Full text] [doi: 10.1136/oemed-2024-109451] [Medline: 39168602]
309. Comfort N, Re DB. Sex-specific neurotoxic effects of organophosphate pesticides across the life course. Curr Environ Health Rep 2017 Dec 23;4(4):392-404 [FREE Full text] [doi: 10.1007/s40572-017-0171-y] [Medline: 29063415]
310. Wang W, Jin J, He R, Gong H. Gender differences in pesticide use knowledge, risk awareness and practices in Chinese farmers. Sci Total Environ 2017 Jul 15;590-591:22-28. [doi: 10.1016/j.scitotenv.2017.03.053] [Medline: 28285132]
311. Xu L, Xu X, Wu X, Kuang H, Xu C. Sex-dependent environmental health risk analysis of flupyradifurone. Environ Sci Technol 2022 Feb 01;56(3):1841-1853. [doi: 10.1021/acs.est.1c07726] [Medline: 35041393]
312. Ketchen Jr DJ, Roccapriore AY, Connelly BL. Using old data: when is it appropriate? J Manag 2023 Jun 04;49(8):2541-2548. [doi: 10.1177/01492063231177785]
313. Vallot C. Epigenomic tumor evolution under the spotlight: the promises of single-cell approaches. C R Biol 2022 May 11;345(1):11-16 [FREE Full text] [doi: 10.5802/crbiol.75] [Medline: 35787617]
314. Horgan D, Hajduch M, Vrana M, Soderberg J, Hughes N, Omar MI, et al. European health data space-an opportunity now to grasp the future of data-driven healthcare. Healthcare (Basel) 2022 Aug 26;10(9):1629 [FREE Full text] [doi: 10.3390/healthcare10091629] [Medline: 36141241]
315. Martani A, Geneviève LD, Wangmo T, Maurer J, Crameri K, Erard F, et al. Sensing the (digital) pulse. Future steps for improving the secondary use of data for research in Switzerland. Digit Health 2023 Apr 20;9:20552076231169826 [FREE Full text] [doi: 10.1177/20552076231169826] [Medline: 37113255]
316. Raab R, Küderle A, Zakreuskaya A, Stern AD, Klucken J, Kaissis G, et al. Federated electronic health records for the European Health Data Space. Lancet Digit Health 2023 Nov;5(11):e840-e847 [FREE Full text] [doi: 10.1016/S2589-7500(23)00156-5] [Medline: 37741765]
317. Pampel H, Weisweiler NL, Strecker D, Witt M, Vierkant P, Elger K, et al. re3data - indexing the global research data repository landscape since 2012. Sci Data 2023 Aug 29;10(1):571 [FREE Full text] [doi: 10.1038/s41597-023-02462-y] [Medline: 37644024]
318. Suhr B, Dungl J, Stocker A. Search, reuse and sharing of research data in materials science and engineering-a qualitative interview study. PLoS One 2020 Sep 15;15(9):e0239216 [FREE Full text] [doi: 10.1371/journal.pone.0239216] [Medline: 32931508]
319. Noaeen M, Doiron C, Syer J, Brook J. Advancing population health through open environmental data platforms. Curr Top Behav Neurosci 2024;68:297-323. [doi: 10.1007/7854_2024_512] [Medline: 39112811]
320. Dul J, van Raaij E, Caputo A. Advancing scientific inquiry through data reuse: necessary condition analysis with archival data. Strateg Change 2023 Nov 02;33(1):35-40. [doi: 10.1002/jsc.2562]
321. Catalogue des bases de données individuelles en santé en France. INSERM. 2020. URL: https://epidemiologie-france.aviesan.fr/ [accessed 2024-10-30]
322. Plateforme ouverte des données publiques françaises. République Française. URL: https://www.data.gouv.fr/fr/ [accessed 2024-10-30]
323. Dataset search. Google. URL: https://datasetsearch.research.google.com/ [accessed 2024-10-30]
324. Kogevinas M, Schlünssen V, Mehlum IS, Turner MC. The OMEGA-NET international inventory of occupational cohorts. Ann Work Expo Health 2020 Jul 01;64(6):565-568 [FREE Full text] [doi: 10.1093/annweh/wxaa039] [Medline: 32556221]
325. Peters S. Although a valuable method in occupational epidemiology, job-exposure -matrices are no magic fix. Scand J Work Environ Health 2020 May 01;46(3):231-234 [FREE Full text] [doi: 10.5271/sjweh.3894] [Medline: 32356897]
326. Nittas V, Puhan MA, von Wyl V. Toward a working definition of eCohort studies in health research: narrative literature review. JMIR Public Health Surveill 2021 Jan 21;7(1):e24588 [FREE Full text] [doi: 10.2196/24588] [Medline: 33475521]
327. Singh AV, Bansod G, Mahajan M, Dietrich P, Singh SP, Rav K, et al. Digital transformation in toxicology: improving communication and efficiency in risk assessment. ACS Omega 2023 Jun 20;8(24):21377-21390. [doi: 10.1021/acsomega.3c00596] [Medline: 37360489]
328. Soussi A, Zero E, Sacile R, Trinchero D, Fossa M. Smart sensors and smart data for precision agriculture: a review. Sensors (Basel) 2024 Apr 21;24(8):2647 [FREE Full text] [doi: 10.3390/s24082647] [Medline: 38676264]
329. Bukhtiyarova O, Abderrazak A, Chiu Y, Sparano S, Simard M, Sirois C. Major areas of interest of artificial intelligence research applied to health care administrative data: a scoping review. Front Pharmacol 2022 Jul 18;13:944516 [FREE Full text] [doi: 10.3389/fphar.2022.944516] [Medline: 35924057]
330. Caruana A, Bandara M, Musial K, Catchpoole D, Kennedy PJ. Machine learning for administrative health records: a systematic review of techniques and applications. Artif Intell Med 2023 Oct;144:102642. [doi: 10.1016/j.artmed.2023.102642] [Medline: 37783537]

331. Leal Neto O, Von Wyl V. Digital transformation of public health for noncommunicable diseases: narrative viewpoint of challenges and opportunities. JMIR Public Health Surveill 2024 Jan 25;10:e49575 [FREE Full text] [doi: 10.2196/49575] [Medline: 38271097]
332. Wang J, Kharrat FG, Gariépy G, Gagné C, Pelletier J, Massamba VK, et al. Predicting the population risk of suicide using routinely collected health administrative data in Quebec, Canada: model-based synthetic estimation study. JMIR Public Health Surveill 2024 Jun 28;10:e52773 [FREE Full text] [doi: 10.2196/52773] [Medline: 38941610]
333. Ben-Ari T, Boé J, Ciais P, Lecerf R, Van der Velde M, Makowski D. Causes and implications of the unforeseen 2016 extreme yield loss in the breadbasket of France. Nat Commun 2018 Apr 24;9(1):1627 [FREE Full text] [doi: 10.1038/s41467-018-04087-x] [Medline: 29691405]
334. Bhawra J, Elsahli N, Patel J. Applying digital technology to understand human experiences of climate change impacts on food security and mental health: scoping review. JMIR Public Health Surveill 2024 Jul 23;10:e54064 [FREE Full text] [doi: 10.2196/54064] [Medline: 39042453]
335. Comi M, Becot F, Bendixsen C. Automation, climate change, and the future of farm work: cross-disciplinary lessons for studying dynamic changes in agricultural health and safety. Int J Environ Res Public Health 2023 Mar 08;20(6):4778 [FREE Full text] [doi: 10.3390/ijerph20064778] [Medline: 36981685]
336. Ebi KL, Vanos J, Baldwin JW, Bell JE, Hondula DM, Errett NA, et al. Extreme weather and climate change: population health and health system implications. Annu Rev Public Health 2021 Apr 01;42:293-315 [FREE Full text] [doi: 10.1146/annurev-publhealth-012420-105026] [Medline: 33406378]
337. El Khayat M, Halwani DA, Hneiny L, Alameddine I, Haidar MA, Habib RR. Impacts of climate change and heat stress on farmworkers' health: a scoping review. Front Public Health 2022 Feb 8;10:782811 [FREE Full text] [doi: 10.3389/fpubh.2022.782811] [Medline: 35211437]
338. Shahzad A, Ullah S, Dar AA, Sardar MF, Mehmood T, Tufail MA, et al. Nexus on climate change: agriculture and possible solution to cope future climate change stresses. Environ Sci Pollut Res 2021 Jan 29;28(12):14211-14232. [doi: 10.1007/S11356-021-12649-8]
339. Gibb K, Beckman S, Vergara XP, Heinzerling A, Harrison R. Extreme heat and occupational health risks. Annu Rev Public Health 2024 May;45(1):315-335 [FREE Full text] [doi: 10.1146/annurev-publhealth-060222-034715] [Medline: 38166501]
340. Lefèvre-Arbogast S, Chaker J, Mercier F, Barouki R, Coumoul X, Miller GW, et al. Assessing the contribution of the chemical exposome to neurodegenerative disease. Nat Neurosci 2024 May 29;27(5):812-821 [FREE Full text] [doi: 10.1038/s41593-024-01627-1] [Medline: 38684891]
341. Shan L, Heusinkveld HJ, Paul KC, Hughes S, Darweesh SK, Bloem BR, et al. Towards improved screening of toxins for Parkinson's risk. NPJ Parkinsons Dis 2023 Dec 19;9(1):169 [FREE Full text] [doi: 10.1038/s41531-023-00615-9] [Medline: 38114496]
342. Abdelzaher H, Tawfik SM, Nour A, Abdelkader S, Elbalkiny ST, Abdelkader M, et al. Climate change, human health, and the exposome: utilizing OMIC technologies to navigate an era of uncertainty. Front Public Health 2022 Sep 21;10:973000 [FREE Full text] [doi: 10.3389/fpubh.2022.973000] [Medline: 36211706]
343. DeBord DG, Carreón T, Lentz TJ, Middendorf PJ, Hoover MD, Schulte PA. Use of the "exposome" in the practice of epidemiology: a primer on -Omic technologies. Am J Epidemiol 2016 Aug 15;184(4):302-314 [FREE Full text] [doi: 10.1093/aje/kwv325] [Medline: 27519539]
344. Waffenschmidt S, Knelangen M, Sieben W, Bühn S, Pieper D. Single screening versus conventional double screening for study selection in systematic reviews: a methodological systematic review. BMC Med Res Methodol 2019 Jun 28;19(1):132 [FREE Full text] [doi: 10.1186/s12874-019-0782-0] [Medline: 31253092]
345. Waffenschmidt S, Sieben W, Jakubeit T, Knelangen M, Overesch I, Bühn S, et al. Increasing the efficiency of study selection for systematic reviews using prioritization tools and a single-screening approach. Syst Rev 2023 Sep 14;12(1):161 [FREE Full text] [doi: 10.1186/s13643-023-02334-x] [Medline: 37705060]

Abbreviations

AGR: annual growth rate
AGRICAN: Agriculture and Cancer
AGRICOH: agricultural cohort
AHD: administrative health database
AHS: Agricultural Health Study
AI: artificial intelligence
AMI: Aging Multidisciplinary Investigation
ANR: French National Research Agency (Agence Nationale de la Recherche in French)
AS: association strength
BIBLIO: preliminary guideline for reporting bibliometric reviews of the biomedical literature
COPD: chronic obstructive pulmonary disease
EHR: electronic health record

EMR: electronic medical record
FAIR: findable, accessible, interoperable, and reusable
FERMA: risk factors of the rural environment and the allergic and respiratory disease
GDP: gross domestic product
JEM: job-exposure matrix
MIAI: Multidisciplinary Institute in Artificial Intelligence
MSA: Mutualité Sociale Agricole
NOCCA: Nordic Occupational Cancer Study
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
TRACTOR: Tracking and Monitoring Occupational Risks in Agriculture

Edited by A Mavragani; submitted 05.06.24; peer-reviewed by G Wu, F Clerc; comments to author 30.09.24; revised version received 08.10.24; accepted 07.11.24; published 09.01.25.

Please cite as:
Petit P, Vuillerme N
Leveraging Administrative Health Databases to Address Health Challenges in Farming Populations: Scoping Review and Bibliometric Analysis (1975-2024)
JMIR Public Health Surveill 2025;11:e62939
URL: https://publichealth.jmir.org/2025/1/e62939
doi:10.2196/62939
PMID:

Trends in Prescription of Stimulants and Narcoleptic Drugs in Switzerland: Longitudinal Health Insurance Claims Analysis for the Years 2014-2021

Tamara Scharf[1,2,3*], MD; Carola A Huber[4,5*], MPH, PhD; Markus Näpflin[4*]; Zhongxing Zhang[3,6*], PhD; Ramin Khatami[3,6,7*], Prof Dr Med

[1]Graduate School of Health Sciences, University of Bern, Bern, Switzerland
[2]Institute of Primary Health Care (BIHAM), University of Bern, Bern, Switzerland
[3]Centre of Sleep Medicine and Epileptology Barmelweid, Klinik Barmelweid AG, Aargau, Switzerland
[4]Department of Health Sciences, Helsana Group, Zürich, Switzerland
[5]Institute of Primary Care, University of Zürich, University Hospital Zürich, Zürich, Switzerland
[6]Barmelweid Academy, Klinik Barmelweid AG, Barmelweid, Switzerland
[7]Department of Neurology, Inselspital Bern University Hospital, University of Bern, Bern, Switzerland
[*]all authors contributed equally

Corresponding Author:
Ramin Khatami, Prof Dr Med
Centre of Sleep Medicine and Epileptology Barmelweid, Klinik Barmelweid AG, Aargau, Switzerland

Abstract

Background: Stimulants are potent treatments for central hypersomnolence disorders or attention-deficit/hyperactivity disorders/attention deficit disorders but concerns have been raised about their potential negative consequences and their increasing prescription rates.

Objective: We aimed to describe stimulant prescription trends in Switzerland from 2014 to 2021. Second, we aimed to analyze the characteristics of individuals who received stimulant prescriptions in 2021 and investigate the link between stimulant prescriptions and hospitalization rates in 2021, using hospitalization as a potential indicator of adverse health outcomes.

Methods: Longitudinal and cross-sectional data from a large Swiss health care insurance were analyzed from all insureds older than 6 years. The results were extrapolated to the Swiss general population. We identified prescriptions for methylphenidate, lisdexamfetamine, modafinil, and sodium oxybate and calculated prevalences of each drug prescription over the period from 2014 to 2021. For 2021 we provide detailed information on the prescribers and evaluate the association of stimulant prescription and the number and duration of hospitalization using logistic regression models.

Results: We observed increasing prescription rates of all stimulants in all age groups from 2014 to 2021 (0.55% to 0.81%, 43,848 to 66,113 insureds with a prescription). In 2021, 37.1% (28,057 prescriptions) of the medications were prescribed by psychiatrists, followed by 36.1% (n=27,323) prescribed by general practitioners and 1% (n=748) by neurologists. Only sodium oxybate, which is highly specific for narcolepsy treatment, was most frequently prescribed by neurologists (27.8%, 37 prescriptions). Comorbid psychiatric disorders were common in patients receiving stimulants. Patients hospitalized in a psychiatric institution were 5.3 times (odds ratio 5.3, 95% CI 4.63 - 6.08, *P*<.001) more likely to have a stimulant prescription than those without hospitalization. There were no significant associations between stimulant prescription and the total length of inpatient stay (odds ratio 1, 95% CI 1 - 1, *P*=.13).

Conclusions: The prescription of stimulant medication in Switzerland increased slightly but continuously over years, but at lower rates compared to the estimated prevalence of central hypersomnolence disorders and attention-deficit/hyperactivity disorders/attention deficit disorders. Most stimulants are prescribed by psychiatrists, closely followed by general practitioners. The increased odds for hospitalization to psychiatric institutions for stimulant receivers reflects the severity of disease and the higher psychiatric comorbidities in these patients.

(JMIR Public Health Surveill 2025;11:e53957) doi:10.2196/53957

KEYWORDS

prescription trends; claims data; cross-sectional data; narcolepsy; prescribers; prescribing practices; medical care; stimulants; stimulant medication

Introduction

Stimulants such as methylphenidate, lisdexamfetamine, and modafinil are highly potent pharmacologic treatment options for hypersomnolence disorders, including narcolepsy or attention-deficit/hyperactivity disorders/attention deficit disorders (ADHDs/ADDs).

ADHDs/ADDs are some of the most common diagnosed psychiatric disorders worldwide with a prevalence of 5.3% worldwide in the years of 1978 to 2010 among people aged 18 years or younger [1]. Prevalence can be different due to varying diagnostic methods per country. In Switzerland, a prevalence of 5.2% was found in children aged 7 to 17 years and in adult men a prevalence of 4% [2,3].

Prescriptions of stimulant drugs have been increasing at various rates over the past decades, with methylphenidate showing an 8.2 fold increase from 1996 to 2013 [4,5]. In the United States, amphetamine and methylphenidate increased from 5.6% to 6.1% in adults aged 20 years or older between 2014 and 2019 [4,5]. In New Zealand, 1.06% of adolescents received stimulants in 2016, an increase of 41.3% from 0.75% in 2011 [6]. Denmark observed a trend in stimulant prescriptions rising from 0.31 per 1000 person-years in 1996 to 7.29 per 1000 person-years in 2010 [7]. In Switzerland, data on stimulant prescription rates is scarce. A Swiss study from 2015 found a lifetime prevalence of stimulants and other substances enhancing cognitive abilities of about 1.4% in employees and students [8]. No sufficient data have, however, been collected in a nationwide study.

Stimulant medication is used for diseases often manifesting during childhood or adolescence and in many cases long-term pharmacological treatment throughout adulthood is needed. Prescription of stimulant agents in this age group is therefore of special interest to balance the need of medication and potential risk by over prescription or under prescription. Prescription of stimulants in young age groups increased from 0.02% to 0.26% over time in Asia, Australia, Europe, and North America in children and 0.003% to 1.48% in adults [9]. Among Swiss school children between 2002 and 2005, methylphenidate prescription increased from 0.74% to 1.02% for children aged 5 to 14 years [10].

This rapid prescription increase may indicate over prescription or even misuse. Misuse of stimulants is common, with up to 17% prevalence in US college students according to a meta-analysis. Misuse can lead to a range of negative consequences such as decreased appetite, insomnia, and increases in heart rate and blood pressure with increased long-term cardiovascular risk and possibly lead to increased hospitalization rates [11-13].

Properly identifying current stimulant prescription rates and discovering the prescription patterns or circumstances of their prescription (prescriber, package size, and comorbidities) may help to further identify alarming prescription increases and potential misuse in Switzerland, and the possible causes.

We therefore had 2 objectives: first, we aim to describe the rate of stimulant prescriptions in Switzerland from 2014 to 2021, focusing on both minors and adults. We hypothesize that the rate of stimulant prescriptions in Switzerland, similar to international trends, has increased over the last decade, with the highest prescription rates being recorded in 2021.

Second, we aimed at analyzing the characteristics of individuals who received stimulant prescriptions in 2021. This analysis will include factors such as comorbidities, age, and prescription details such as package size and health care providers most frequently issuing these prescriptions. Our hypotheses are that recipients likely mirror the characteristics of ADHDs/ADDs and central hypersomnolence disorders, most are minors with few comorbidities, and the likelihood of receiving a prescription decreases with age. Specialists, particularly psychiatrists and neurologists, are expected to be the primary prescribers.

Lastly, this study aims to investigate the link between stimulant prescriptions and hospitalization rates in 2021, using hospitalization as a potential indicator of adverse health outcomes. Given the expected increase in stimulant prescriptions until 2021, the hypothesis is that individuals prescribed stimulants may have a higher risk of hospitalization, particularly if prescription rates are on the rise.

Methods

Study Design

This study is a longitudinal and cross-sectional analysis of the Helsana health care insurance data of around 1.5 million people in Switzerland insured over the period of 2014 to 2021. Helsana belongs to a group of the biggest insurance companies in Switzerland and insures 14% of the Swiss population, with insureds in 26/26 cantons of Switzerland. Data describes general information on the insured persons and all their invoices for health services directed to the insurance. These invoices are representative of all health care costs of the insureds, except for the costs that were not sent as invoices to the insurance and paid by the insureds themselves (ie, over the counter drug costs and dental costs). We decided to provide more detailed descriptive statistics for data in 2021, since it was the most recent data available to us and presumably with the highest rate of stimulant prescription.

Identification of Drugs

We identified the drug invoices through their Anatomical Therapeutic Chemical code, which classifies chemical substances based on their therapeutical properties. The identification was performed for the following drugs: methylphenidate (N06BA04), lisdexamfetamine (N06BA12), modafinil (N06BA07), and sodium oxybate (N07XX04) which is a specific medication for narcolepsy treatment and used to estimate the treatment prevalence of narcolepsy patients in the dataset. Pitolisant (N07XX11) was identified too as a stimulant with specific use for narcolepsy but was excluded from further analysis, as it was only authorized for use in 2020 but had neglectable low prescription rates. We also identified the Swissmedic code of the medications, which is specific not only for the chemical substance but also the producer of the medication and package size. These drugs are only accessible through prescription by a medical professional and reimbursed by the insurance company. Overlapping prescriptions of the 4 drugs was defined as 1 prescription of one of the 4 drugs invoiced with at least one of the other drugs once or multiple times during the year of 2021.

Variables

The dataset consisted of all insureds aged ≥6 years with information on their age, sex, region of language, and region of residence. We categorized 5 age groups (6 ‐ 17 y, 18 ‐ 35 y, 36 ‐ 65 y, 66 ‐ 75 y, and 76+ y). We divided the insureds` residential regions into "rural," "intermediate," and "urban" subgroups according to the Swiss federal office for statistics.

The insureds had various health care plans including standard care and managed care models (eg family physician model). These health insurance plans were identified and categorized into standard care and managed care (ie, the combination of telemedical care and general practitioner [GP] care).

Chronic health condition status was identified by substance prescriptions related to chronic diseases. This was carried out according to approaches developed in previous research on the dataset [14]. We classified 22 different chronic conditions and categorized them into psychiatric, cardiologic, rheumatologic, respiratory comorbidities, or all other.

All invoices for hospitalization (ie, hospitals of all sizes providing acute care and psychiatric clinics) and the length of stay were included in our analysis.

Based on the medical prescriber who issued the invoices, several prescriber categories were defined: GP, psychiatrist, neurologist, other specialists (combining all other prescribers, such as nonspecific group practices, cardiologists, pulmonologists, rheumatologists, etc). Only health care personnel in Switzerland are allowed to prescribe medication. We further grouped them into "only prescriber" of the medication when a describer prescribed all medication exclusively for single individuals, ">50% prescriber" meaning more than 50% of the prescriptions for single individuals were invoiced by the prescriber, and "rest" with all other prescription proportions.

Within the prescriber categories (inpatient psychiatry, inpatient acute, and rehabilitation, nursing) we differentiated between inpatient (during a hospital stay) and outpatient invoices.

Statistical Analysis

For our first objective we provide descriptive statistics for prescription trends among different age groups across the years 2014 and 2021 by identifying individuals in the dataset with at least one prescription of the predefined drugs. To obtain representative data for Swiss population we extrapolated these data by current residency numbers and populations statistics of the Federal Statistical Office.

For our second objective we restricted the descriptive statistics to the year 2021 and provide detailed information on age, sex, region of residence, health insurance status, comorbidities, prescribers, and package size. Logistic regression models were performed to evaluate the association of at least one stimulant prescription versus no prescription (ie, the dependent variable is prescription yes or no), number of hospitalization, and length of stay, adjusted for age, sex, region of residence, health insurance status, and number of chronic diseases for the year 2021.

All data management, graphic generation, and analysis was performed with the statistics program R (version 4.2.1; R Foundation).

Data Availability

The authors were permitted access to the data by collaboration with the insurance companies research team. The datasets generated and analyzed during this study are available from the corresponding author on reasonable request. AI was not used in any way in data generation, analysis, and presentation of results.

Ethical Considerations

According to ethical and legal regulations in Switzerland no ethical approval or patient consent was needed for this study, as all data complied with privacy regulations and personal data protection, data was anonymized when presented to the research team. The Swiss Human Research Act (REQ-2017 ‐ 00280) did not apply to this project. The exploratory statistical analyses of the feasibility test complied with the Swiss Federal Law on data protection. All data were anonymized and deidentified prior to the performed analysis to protect the privacy of patients, physicians, and hospitals. According to the national ethical and legal regulation, an ethical approval was not needed because the data were pre-existing and deidentified. Since data was anonymized, no consent of patients was required.

Results

Trend of Stimulant Prescription Per Year

As baseline we refer to the prescription period in 2014. Between baseline and 2021, on average 14% of Swiss people were at any time insured with the Helsana Group.

In the year 2014, 0.55% (42,848 insureds) of insured people of any age received at least 1 stimulant agent (or stimulant prescription). This number increased up to 0.8% (66,113 insureds) in the year of 2021. The largest growing percentage of stimulus prescriptions were in the youngest age group with an increase of prescription of 0.6 percentage points (17,972, 1.8% to 24,982, 2.4%) between 2014 and 2021 (Table 1).

Between 2014 and 2021, 32,2418 packages of methylphenidate were the most prescribed stimulant, followed by 46,074 packages of lisdexamfetamine, 8797 packages of modafinil, and 3115 packages of sodium oxybate. We found a prescription increase of all identified medications in younger and middle age groups over the years of 2014 to 2021, extrapolated by the Swiss population. The increase in prescription was steady in age groups aged 36 ‐ 65 years, whereas in other age groups prescription stagnated from 2018 to 2020, with a steep increase from 2020 to 2021.

Methylphenidate prescription increased overall in all age groups with a steady increase in insureds aged 36 ‐ 65 years. All other age groups experienced a steep increase after 2020. Only insureds aged 66 ‐ 75 years and 76+ years ever experienced a smaller prescription rate than at baseline in 2014, with a drop to 89% (664/746) in 2017 (Figure 1).

Lisdexamfetamine prescription was low at baseline and increased steadily in all age groups with great increase from 0 prescriptions at baseline to 39 in insureds aged 60 - 65 years. Smaller increase in prescription was seen in age groups aged 66 - 75 years and 76+ years (Figure 1).

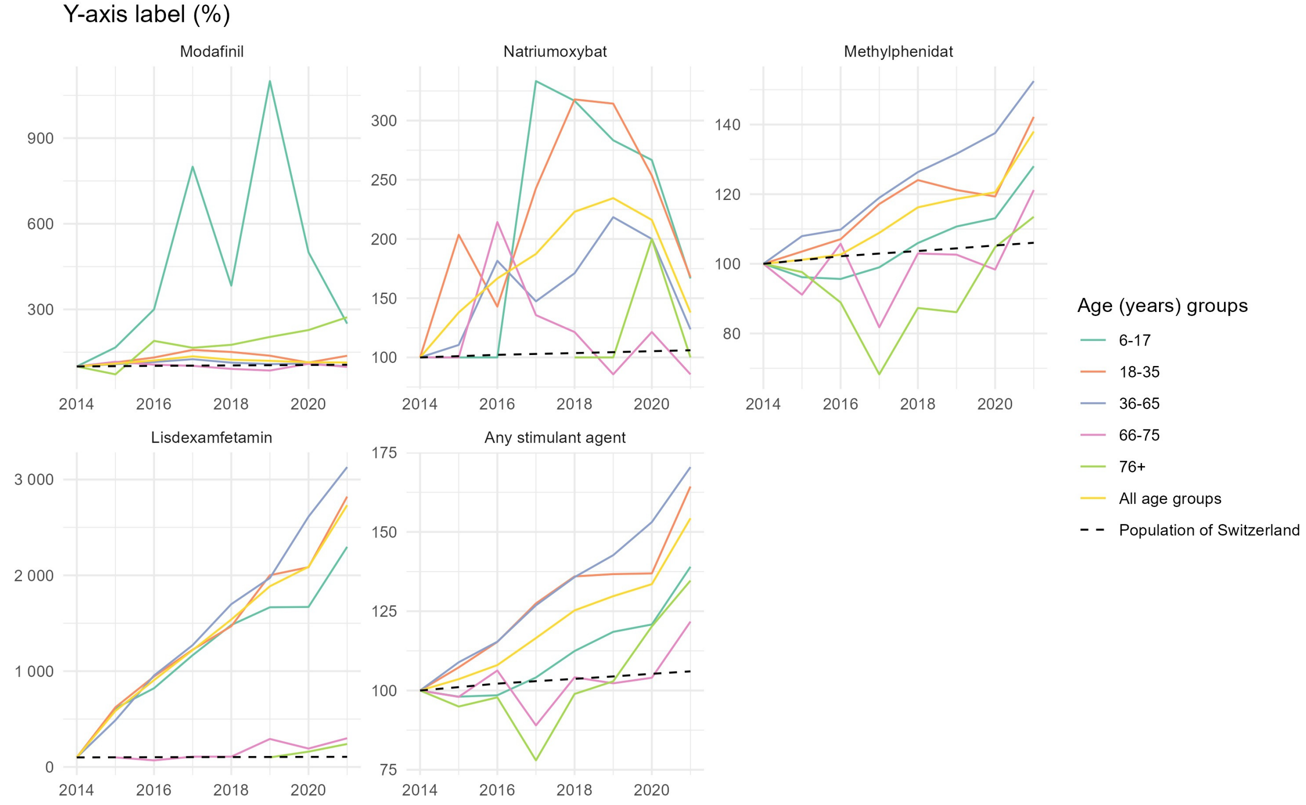
Prescription rates of modafinil—only prescribed to few—increased the most in the youngest age group (6 - 17 y). The prescribing trend in this age group strongly fluctuated. From its highest peak in 2019 rates decreased from around 1100% (66/6) to slightly less than 300% (15/6) prescription compared to the baseline in 2014. In 2021, modafinil was most frequently prescribed in age groups aged 6 - 17 years and 76+ years. All other age groups had only a moderate increase in modafinil prescriptions over the years 2014 to 2021.

Sodium oxybate overall increased the most in insureds aged 18 - 35 years and 6 - 17 years from 100% to 175% in 2021. Prescription rates for sodium oxybate only decreased overall in insureds aged 66 - 75 years. Age groups aged 36 - 65 years and 76+ experienced a similar prescription trend over the years with overall increase but declining prescription rates after 2020.

Table . Proportions of Swiss insureds with at least 1 stimulant agent or other narcolepsy treatment prescription within age groups for each year between 2014 and 2021.

Age (years)	2014, n (%)	2015, n (%)	2016, n (%)	2017, n (%)	2018, n (%)	2019, n (%)	2020, n (%)	2021, n (%)	*P* value (chi-square test for trend in proportions)
6 - 17	17,972 (1.8)	17,624 (1.8)	17,702 (1.8)	18,711 (1.9)	20,209 (2)	21,296 (2.1)	21,714 (2.1)	24,982 (2.4)	<.001
18 - 35	13,145 (0.68)	14,103 (0.72)	15,159 (0.77)	16,755 (0.85)	17,874 (0.91)	17,969 (0.92)	17,998 (0.92)	21,599 (1.10)	<.001
36 - 65	10,708 (0.31)	11,661 (0.34)	12,355 (0.35)	13,584 (0.38)	14,537 (0.41)	15,277 (0.43)	16,394 (0.45)	18,251 (0.50)	<.001
66 - 75	746 (0.1)	731 (0.1)	793 (0.1)	664 (0.1)	777 (0.1)	763 (0.1)	776 (0.1)	908 (0.1)	.10
76+	277 (0)	263 (0)	271 (0)	216 (0)	274 (0)	285 (0)	333 (0)	373 (0.10)	.05
All age groups	42,848 (0.55)	44,382 (0.57)	46,280 (0.59)	49,931 (0.63)	53,671 (0.67)	55,590 (0.69)	57,216 (0.70)	66,113 (0.81)	<.001

Figure 1. Trends of stimulant prescription from 2014 to 2021 per age group and active ingredient (indexed, base year=2014).



Factors Associated With Prescription in 2021

Characteristics of Stimulant Users

Most stimulant receivers in the year of 2021 were male representing 61% (42,803/70,396) of our population. Only modafinil was more often prescribed to women than to men with 55% (970/1776) female receivers. The highest proportion of prescriptions was provided to people living in urban areas with a proportion of 67% (46,968/70,396) of all stimulants compared to intermediate and rural area residents. Managed care was the preferred health care plan for patients receiving stimulants with 72% (50,513/70,396) receiving. A total of 45% (798/1776) of all modafinil receivers had 3+ chronic illnesses, 55% (775/1399) of those had psychiatric comorbidities, followed by cardiological and rheumatological diseases. A total of 42% (49/121) of sodium oxybate users had no comorbidities. Most common chronic illness in sodium oxybate users was psychiatric (43/72, 60%) or cardiologic (31/72, 43%). Additionally, half of the methylphenidate users had comorbidities (28,619/57,128, 50%, 28,509 had no chronic illness identified) with psychological (14,396/28,619, 50%) and other chronic conditions (7034/28,619, 25%) as the most common identified chronic diseases. Similar results were found for lisdexamfetamine users, that is, 43% (4908/11,371) users had no comorbidities and psychological (3556/6463, 55%) and other (1989/6463, 31%) comorbidities were the most common chronic diseases (Table 2).

Table . Characteristics of Swiss insureds receiving any stimulant prescription in the year 2021.

Characteristic	Modafinil (n=1776)	Sodium oxybate (n=121)	Methylphenidate (n=57,128)	Lisdexamfetamine (n=11,371)
Sex, n (%)				
Male	806 (45.4)	63 (52.1)	35,038 (61.3)	6896 (60.7)
Female	970 (54.6)	58 (48)	22,090 (38.7)	4475 (39.4)
Age (years)				
Median (IQR)	45 (33-6)	38 (29-5)	22 (14-4)	26 (16-4)
Mean (SD)	46 (16)	40 (18)	27 (16)	28 (14)
Age (years, in groups), n (%)				
6 - 17	16 (0.9)	10 (8.3)	22,948 (40.2)	3355 (29.5)
18 - 35	568 (32)	47 (38.8)	18,002 (31.5)	4401 (38.7)
36 - 65	976 (55)	48 (39.7)	15,157 (26.5)	3565 (31.4)
66 - 75	139 (7.8)	12 (9.9)	742 (1.3)	38 (0.3)
76+	77 (4.3)	4 (3.3)	279 (0.5)	12 (0.1)
Region of residence, n (%)				
Urban	1060 (59.7)	80 (66.1)	38,196 (66.9)	7632 (67.1)
Intermediate	449 (25.3)	29 (24)	11,554 (20.2)	2422 (21.3)
Rural	267 (15)	12 (9.9)	7378 (12.9)	1317 (11.6)
Health insurance status, n (%)				
Managed care	1064 (59.9)	82 (67.8)	41,178 (72.1)	8189 (72)
Standard care	712 (40.1)	39 (32.2)	15,950 (27.9)	3182 (28)
Comorbidities [14]**, n (%)**				
0	377 (21.2)	49 (40.5)	28,509 (49.9)	4908 (43.2)
1	340 (19.1)	27 (22.3)	12,617 (22.1)	2601 (22.9)
2	261 (14.7)	25 (21.7)	6678 (11.7)	1619 (14.2)
3+	798 (44.9)	20 (16.5)	9324 (16.3)	2243 (19.7)
Most frequent, n (%)	psyd:[a] 775 (43.6)	psyd: 43 (36)	psyd: 14,396 (25.2)	psyd: 3556 (31.3)
2nd most frequent, n (%)	card:[b] 592 (33.3)	card: 31 (26)	ther:[c] 7034 (12.3)	ther: 1989 (17.5)
3rd most frequent, n (%)	rheu:[d] 430 (24.2)	rheu: 17 (14)	rheu: 6588 (11.5)	resp:[e] 1435 (12.6)

[a] psyd: psychiatric.
[b] card: cardiological.
[c] ther: other.
[d] rheu: rheumatological.
[e] resp: respiratory.

Package Size

All medication was predominantly prescribed more than once within a year, with ≥5 packages prescribed in 40.4% (715/1768) of all modafinil prescriptions, 95.8% (115/120) of all sodium oxybate prescriptions, 47.5% (27,124/57,093) of all methylphenidate prescriptions, and 59.1% (6727/11,392) of all lisdexamfetamine prescriptions (Table 3).

Table . Number of Swiss insureds and their stimulant prescriptions by number of packages and by prescriber (profession of the physician) in the year 2021.

			At least 1 stimulant agent (total)	At least 1 modafinil use	At least 1 sodium oxybate use	At least 1 methylphenidate use	At least 1 lisdexamfetamin use
Total patients (N)			66,113	1768	120	57,093	11,392
Number of packages, n (%)							
	1		12,235 (17.4)	536 (30.3)	0 (0)	9984 (17.5)	1715 (15.1)
	2		9301 (13.2)	228 (12.9)	5 (4.2)	7968 (14)	1099 (9.6)
	3		7435 (10.6)	113 (6.4)	0 (0)	6300 (11)	1022 (9)
	4		6721 (9.6)	176 (10)	0 (0)	5716 (10)	829 (7.3)
	≥5		34,681 (49.3)	715 (40.4)	115 (95.8)	27,124 (47.5)	6727 (59.1)
Package sizes, median (IQR)[a]							
	1		—[b]	30 (30-90)	—	30 (30-100)	30 (30-30)
	2		—	90 (60-90)	—	50 (30-100)	30 (30-30)
	3		—	90 (67.5 - 90)	—	50 (30-83)	30 (30-30)
	4		—	90 (90-90)	—	50 (35-72)	30 (30-30)
	≥5		—	90 (90-90)	—	45 (30-60)	30 (30-30)
Prescriber of the issued prescriptions, n (%)							
	General practitioner (GP)		27,323 (41.3)	639 (36.1)	10 (8.3)	24,127 (42.3)	3529 (31)
		Only	22,578 (82.6)	476 (74.5)	10 (100)	19,521 (80.9)	1762 (49.9)
		>50	1911 (7)	63 (9.9)	0 (0)	1802 (7.5)	717 (20.3)
		Rest	2834 (10.4)	100 (15.6)	0 (0)	2804 (11.6)	1050 (29.8)
	Psychiatrist		28,057 (42.4)	336 (19)	5 (4.2)	23,553 (41.3)	6222 (54.6)
		Only	23,267 (82.9)	220 (65.5)	0 (0)	18,167 (77.1)	3204 (51.5)
		>50	2247 (8)	31 (9.2)	5 (100)	2260 (9.6)	1308 (21)
		Rest	2543 (9.1)	85 (25.3)	0 (0)	3126 (13.3)	1710 (27.5)
	Neurologist		748 (1.1)	181 (10.2)	37 (30.8)	545 (1)	63 (0.6)
		Only	558 (74.6)	97 (53.6)	11 (29.7)	384 (70.5)	26 (41.3)
		>50	99 (13.2)	31 (17.1)	20 (54.1)	48 (8.8)	25 (39.7)
		Rest	91 (12.2)	53 (29.3)	6 (16.2)	113 (20.7)	12 (19)
	Other specialists		7962 (12)	286 (16.2)	15 (12.5)	6680 (11.7)	1300 (11.4)
		Only	4599 (57.8)	176 (61.5)	10 (66.7)	3782 (56.6)	437 (33.6)
		>50	955 (12)	28 (9.8)	5 (33.3)	801 (12)	234 (18)
		Rest	2408 (30.2)	82 (28.7)	0 (0)	2097 (31.4)	629 (48.4)
	Inpatient psychiatry		6000 (9.1)	70 (4)	6 (5)	5027 (8.8)	1234 (10.8)
		Only	3429 (57.2)	27 (38.6)	0 (0)	2857 (56.8)	362 (29.3)
		>50	898 (15)	11 (15.7)	0 (0)	756 (15)	214 (17.3)
		Rest	1673 (27.9)	32 (45.7)	6 (100)	1414 (28.1)	658 (53.3)
	Inpatient acute, rehabilitation, and nursing		5549 (8.4)	496 (28.1)	60 (50)	4251 (7.4)	1110 (9.7)
		Only	3312 (59.7)	284 (57.3)	11 (18.3)	2376 (55.9)	415 (37.4)
		>50	756 (13.6)	47 (9.5)	32 (53.3)	699 (16.4)	136 (12.3)
		Rest	1481 (26.7)	165 (33.3)	17 (28.3)	1176 (27.7)	559 (50.4)

[a]Median (IQR) package size (number of units) per number of prescribed packages. Not shown for sodium oxybate, since this substance is a liquid.

[b]Not applicable.

Prescribers

In 2021, psychiatrists were the most frequent prescribers with 42.4% (28,057/66,172) of all the prescribed medications. Lisdexamfetamine (6222/11,396, 54.6%), followed by methylphenidate with 41.3% (23,553/57,038) were mainly prescribed by them. By contrast modafinil and sodium oxybate were rarely prescribed by psychiatrists as 19% and 4.2% (336/1769 and 5/120), respectively. If chosen as a prescriber, psychiatrists are often the only source of prescription for 82.9% (23,267/28,057) insured individuals of all medication of 2021.

The second most frequent prescribers were GPs, with a similar high proportion of 41.3% (27,323/66,172) of all drug prescriptions. More specifically 42.3% (24,127/57,038) methylphenidate was prescribed by GPs followed by modafinil (639/1769, 36.1%), lisdexamfetamine (3529/11,396, 31%), and smaller proportions for sodium oxybate (10/120, 8.3%). GPs were most often the exclusive prescribers of the medications (22,578/27,323, 82.6%). Only 10.4% (2834/27,323) of prescriptions by GPs shared the prescribing job with other medical specialists.

Neurologists were rarely prescribers of stimulants or narcolepsy treatments, as only 1.1% (748/66,172) of all prescriptions were invoiced by them. Only sodium oxybate was the most frequent (37/120, 30.8%) medication prescribed by neurologists. If chosen as the prescriber, they are often the only source (558/748, 74.6%) from which individuals received the prescriptions in 2021 (Table 3).

Concerning invoices handed in by hospitals or psychiatric clinics or rehabilitation facilities, they only made up a small part of overall stimulant invoices. An exception is sodium oxybate, of which 45.1% (60/133) of invoices are issued by an acute clinic, rehabilitation clinic, or nursing home (Table 3).

Association of Stimulant Use and Outcomes

We found an association with patients receiving a stimulant or narcolepsy treatment prescription and increased hospitalizations in a psychiatric facility (odds ratio [OR] 5.30, 95% CI 4.63 - 6.08, $P<.001$). In contrast, there was a negative association between stimulant prescription and hospitalization in an acute medical care facility (OR 0.77, 95% CI 0.73 - 0.82, $P<.001$). There were no significant associations between stimulant prescription and the total length of inpatient stay (OR 1, 95% CI 1 - 1, $P=.13$; Table 4).

Table . Regression model of predicting the outcomes of hospitalization and length of stay in Swiss insureds, who received stimulant prescription in the year 2021.

Characteristic		Odds ratio (95% CI)	*P* value
Sex			
	Male	—[a] (—)	—
	Female	0.89 (0.86 - 0.92)	<.001
Age (years, in groups)			
	6-17	1.17 (1.13-1.21)	<.001
	18 - 35	— (—)	—
	36 - 65	0.62 (0.60 - 0.64)	<.001
	66 - 75	0.31 (0.27 - 0.34)	<.001
	76+	0.24 (0.20 - 0.28)	<.001
Region of residence			
	Urban	— (—)	—
	Intermediate	1.09 (1.05 - 1.13)	<.001
	Rural	0.97 (0.94 - 1.02)	.2
Health insurance status			
	Standard	— (—)	—
	Managed care	0.69 (0.67 - 0.71)	<.001
Number of comorbidities		1.56 (1.54 - 1.59)	<.001
Total inpatient length of stay		1 (1 - 1)	.13
Hospitalization acute (yes or no)		0.77 (0.73 - 0.82)	<.001
Hospitalization psychiatry (yes or no)		5.3 (4.63 - 6.08)	<.001

[a]Not applicable.

When only focusing on methylphenidate or lisdexamfetamine compared to a balanced sample of nonstimulant users, we found a positive association between their prescription and hospitalization in a psychiatric facility (OR 6.85, 95% CI 5.89 - 7.99, *P*<.001). No significant association was found between the prescription rate and the total inpatient length of stay (OR 1, 95% CI 1 - 1, *P*=.1). Hospitalization in an acute medical care facility was less likely (OR 0.76, 95% CI 0.71 - 0.8, *P*<.001) to prescribe methylphenidate or lisdexamfetamine (Table 5).

Table . Regression models of predicting the outcomes of hospitalization and length of stay in Swiss insureds who received stimulant prescriptions in the year 2021. Prescriptions of methylphenidate or lisdexamfetamine are shown in the left columns, prescriptions of modafinil or sodium oxybate are shown in the right columns.

Characteristic		Methylphenidate- or lisdexamfetamine-users versus nonusers		Modafinil- or sodium oxybate-users versus nonusers	
		OR[a] (95% CI)	*P* value	OR[a] (95% CI)	*P* value
Sex					
	Male	— (—)	—	— (—)	—
	Female	0.9 (0.87 - 0.93)	<.001	0.84 (0.72 - 0.97)	.02
Age (years, in groups)					
	6-17	1.17 (1.13-1.21)	<.001	1.17 (0.73-1.88)	.5
	18 - 35	— (—)	—	— (—)	—
	36 - 65	0.63 (0.6 - 0.65)	<.001	0.57 (0.48 - 0.68)	<.001
	66 - 75	0.31 (0.28 - 0.36)	<.001	0.25 (0.19 - 0.34)	<.001
	76+	0.26 (0.21 - 0.31)	<.001	0.15 (0.09 - 0.23)	<.001
Region of residence					
	Urban	— (—)	—	— (—)	—
	Intermediate	1.11 (1.07 - 1.15)	<.001	0.96 (0.8 - 1.15)	.7
	Rural	0.95 (0.91 - 0.99)	.019	1.11 (0.89 - 1.38)	.3
Health insurance status					
	Standard	— (—)	—	— (—)	—
	Managed care	0.69 (0.67 - 0.71)	<.001	0.58 (0.5 - 0.68)	<.001
Number of comorbidities		1.55 (1.53 - 1.57)	<.001	1.62 (1.54 - 1.71)	<.001
Total inpatient length of stay		1 (1-1)	.10	1 (1-1)	.8
Hospitalization acute (yes or no)		0.76 (0.71 - 0.8)	<.001	1.23 (0.96 - 1.56)	.10
Hospitalization psychiatry (yes or no)		6.85 (5.89 - 7.99)	<.001	2.15 (1.17 - 4.23)	.02

[a]Odds ratio

[b]Not applicable.

When only focusing on modafinil or sodium oxybate compared to a balanced sample of nonstimulant users, we found a positive association between their prescription and hospitalization in a psychiatric facility (OR of 2.15, 95% CI 1.17-4.23, *P*=.02). We found no significant association between total inpatient length of stay (OR 1, 95% CI 1 - 1, *P*=.80) and hospitalization in an acute medical care facility (OR 1.23, 95% CI 0.96 - 1.56, *P*=.10; Table 5).

Discussion

Summary

We found an increasing trend of stimulant and narcoleptic drug prescriptions in Switzerland from 0.6% to 0.8% over the years 2014 to 2021. Most stimulants are prescribed continuously with more than 5 packages in 1 year per insured to underaged individuals with no comorbidities. Psychiatrists and GPs are often the prescribers of stimulants, much more frequently than neurologists.

Context in Research

Compared with data on global stimulant prescription rates, our results are notably lower, showing a smaller increase in prescription rates than in the United States (5.6% - 6.1%), New Zealand (0.75% - 1.06%), and Denmark (0.03% - 0.73%) [5-7]. Given the ADHDs/ADDs prevalence rates of 4% in adult men and 5.3% in children in Switzerland [2,3], compared to a worldwide prevalence of between 5% and 11.4% [1,15,16], it is reasonable to interpret the lower Swiss prescription rates as either an indication of underprescription or as lower stimulant misuse rates in Switzerland [15]. Even when taking all stimulants together, current prescription rates do not reach the prevalence rate of ADHDs/ADDs in Switzerland. The fact that

ADHDs/ADDs are also treated nonpharmacologically is another argument for assuming that our result of low prescription prevalence is lower than the disease prevalence. Unfortunately, there are no data that quantifies the extent of drug or nonpharmacological treatment for ADHDs/ADDs and thus could help define the normal gap between prescription rate and disease prevalence. Our distribution of prescription rates corresponds to the disease distribution in different age groups, with prescriptions and prevalence of ADHDs/ADDs being higher in the younger age groups [17]. The distribution of prescriptions by gender reflects the current disease prevalence of ADHDs/ADDs, as we found a slightly lower proportion of females than males, in line with another summary by Thapar and Cooper [17]

Stimulants were predominantly prescribed in urban areas. A higher population compared to rural areas, a higher density of prescribing physicians, and a higher number of hospitals specialized in diagnosis and treatment of these diseases in urban settings may account for this predominant prescription pattern. Vice versa underprescription in rural areas could be due to reduced access to adequate health care services.

Methylphenidate was by far the most prescribed stimulant agent, followed by lisdexamfetamine, a new stimulant showing promising results for treatment of ADHDs/ADDs [18]. The prescription characteristics were very similar between the two stimulants, with lisdexamfetamine more often prescribed to older adults than methylphenidate. This shows that lisdexamfetamine is possibly used as a second medication after methylphenidate was prescribed in the young and did not give continuous results while the patients aged.

Prescriptions were most frequently issued by psychiatrists. Since methylphenidate and lisdexamfetamine are standard treatment for the psychiatric diseases ADHDs and ADDs, this prescription pattern makes sense from a health provider perspective and is confirmed by other findings about prescribers of stimulants [19].

Prescription characteristics of sodium oxybate should reflect its specificity for narcolepsy because it is not indicated for any other disease. Here we find most of the prescriptions in middle aged groups and fewer in children and teenagers, with a nearly even distribution between women and men. This grouping does not match with the expected narcolepsy features of young patients with a possible second peak in the late forties [20]. The lack of prescription in young patients is best explained by a missed or severely delayed diagnosis of narcolepsy, which is in line with the recently published delayed diagnosis for Switzerland and other European countries [21,22]. Surprisingly, neurologists who diagnose and treat narcolepsy, are rarely the patients' prescription source, even for highly specific and not easy to handle medication, such as sodium oxybate. Multiple reasons may account for this prescription practice, among them is the lower barrier for receiving an appointment with GPs before prescriptions expires, compared to neurologists.

We found no significant change in the odds of at least one stimulant prescription in patients hospitalized in an acute hospital but a significant increase in the odds of patients hospitalized in a psychiatric facility. This reflects the fact that ADHDs/ADDs are often overlapping or comorbid with other psychiatric diseases which can lead to hospitalization, such as addiction, disruptive disorders, anxiety disorders, or bipolar disorders [23-26], and by the severity of disease.

We assume that our data is representative for Switzerland, since we analyzed claims data extrapolated to the entire population from one of the biggest insurance companies in Switzerland, with a nearly equal distribution across the country. As stimulant agents are only accessed through prescription, we were able to register all invoices for the medication in question in real world; therefore, our data minimize sampling bias and recall bias that frequently influence the accuracy and reliability of retrospective studies.

Limitations

This study is an analysis of health insurance claims data, which does not contain any information on the clinical reason of why a medication was indicated. We therefore could not distinguish between prescriptions according to current treatment guidelines and prescriptions of pharmacological treatment for diseases without proper diagnosis.

We identified prescribers by Zahlsteller register (registered number for medical personel allowed to bill insurance companies) number and medication prescription pattern, which may lead in some cases to misclassification, as some physicians share Zahlsteller register numbers in group practices and GPs go sometimes through additional training as psychiatrists or neurologists.

Conclusion

The prescription of stimulants and sodium oxybate in Switzerland increased slightly but continuously over the past years, but at lower rates compared to the estimated prevalence of central hypersomnolence disorders and ADHDs/ADDs. Most stimulants are prescribed by psychiatrists, closely followed by GPs. The increased odds for hospitalization to psychiatric institutions for stimulant receivers reflects the severity of disease and the higher psychiatric comorbidities in these patients.

Conflicts of Interest

None declared.

References

1. Polanczyk GV, Willcutt EG, Salum GA, Kieling C, Rohde LA. ADHD prevalence estimates across three decades: an updated systematic review and meta-regression analysis. Int J Epidemiol 2014 Apr;43(2):434-442. [doi: 10.1093/ije/dyt261] [Medline: 24464188]

2. Steinhausen HC, Metzke CW, Meier M, Kannenberg R. Prevalence of child and adolescent psychiatric disorders: the Zürich epidemiological study. Acta Psychiatr Scand 1998 Oct;98(4):262-271. [doi: 10.1111/j.1600-0447.1998.tb10082.x] [Medline: 9821446]
3. Estévez N, Eich-Höchli D, Dey M, Gmel G, Studer J, Mohler-Kuo M. Prevalence of and associated factors for adult attention deficit hyperactivity disorder in young Swiss men. PLoS ONE 2014;9(2):e89298. [doi: 10.1371/journal.pone.0089298] [Medline: 24586672]
4. Safer DJ. Recent trends in stimulant usage. J Atten Disord 2016 Jun;20(6):471-477. [doi: 10.1177/1087054715605915] [Medline: 26486603]
5. Board AR, Guy G, Jones CM, Hoots B. Trends in stimulant dispensing by age, sex, state of residence, and prescriber specialty - United States, 2014-2019. Drug Alcohol Depend 2020 Dec 1;217:108297. [doi: 10.1016/j.drugalcdep.2020.108297] [Medline: 32961454]
6. Barczyk ZA, Rucklidge JJ, Eggleston M, Mulder RT. Psychotropic medication prescription rates and trends for New Zealand children and adolescents 2008-2016. J Child Adolesc Psychopharmacol 2020 Mar;30(2):87-96. [doi: 10.1089/cap.2019.0032] [Medline: 31633377]
7. Steinhausen HC, Bisgaard C. Nationwide time trends in dispensed prescriptions of psychotropic medication for children and adolescents in Denmark. Acta Psychiatr Scand 2014 Mar;129(3):221-231. [doi: 10.1111/acps.12155] [Medline: 23738593]
8. Maier LJ, Haug S, Schaub MP. Prevalence of and motives for pharmacological neuroenhancement in Switzerland--results from a national Internet panel. Addiction 2016 Feb;111(2):280-295. [doi: 10.1111/add.13059] [Medline: 26189457]
9. Raman SR, Man KKC, Bahmanyar S, et al. Trends in attention-deficit hyperactivity disorder medication use: a retrospective observational study using population-based databases. Lancet Psychiatry 2018 Oct;5(10):824-835. [doi: 10.1016/S2215-0366(18)30293-1] [Medline: 30220514]
10. Gumy C, Huissoud T, Dubois-Arber F. Prevalence of methylphenidate prescription among school-aged children in a Swiss population: increase in the number of prescriptions in the Swiss Canton of Vaud, from 2002 to 2005, and changes in patient demographics. J Atten Disord 2010 Nov;14(3):267-272. [doi: 10.1177/1087054709356386] [Medline: 20484708]
11. Benson K, Flory K, Humphreys KL, Lee SS. Misuse of stimulant medication among college students: a comprehensive review and meta-analysis. Clin Child Fam Psychol Rev 2015 Mar;18(1):50-76. [doi: 10.1007/s10567-014-0177-z] [Medline: 25575768]
12. Charach A, Ickowicz A, Schachar R. Stimulant treatment over five years: adherence, effectiveness, and adverse effects. J Am Acad Child Adolesc Psychiatry 2004 May;43(5):559-567. [doi: 10.1097/00004583-200405000-00009]
13. Lakhan SE, Kirchgessner A. Prescription stimulants in individuals with and without attention deficit hyperactivity disorder: misuse, cognitive impact, and adverse effects. Brain Behav 2012 Sep;2(5):661-677. [doi: 10.1002/brb3.78] [Medline: 23139911]
14. Huber CA, Szucs TD, Rapold R, Reich O. Identifying patients with chronic conditions using pharmacy data in Switzerland: an updated mapping approach to the classification of medications. BMC Public Health 2013 Oct 30;13(1):1-10. [doi: 10.1186/1471-2458-13-1030] [Medline: 24172142]
15. Danielson ML, Claussen AH, Bitsko RH, et al. ADHD prevalence among U.S. children and adolescents in 2022: diagnosis, severity, co-occurring disorders, and treatment. J Clin Child Adolesc Psychol 2024;53(3):343-360. [doi: 10.1080/15374416.2024.2335625] [Medline: 38778436]
16. Thomas R, Sanders S, Doust J, Beller E, Glasziou P. Prevalence of attention-deficit/hyperactivity disorder: a systematic review and meta-analysis. Pediatrics 2015 Apr;135(4):e994-1001. [doi: 10.1542/peds.2014-3482] [Medline: 25733754]
17. Thapar A, Cooper M. Attention deficit hyperactivity disorder. The Lancet 2016 Mar;387(10024):1240-1250. [doi: 10.1016/S0140-6736(15)00238-X]
18. Roskell NS, Setyawan J, Zimovetz EA, Hodgkins P. Systematic evidence synthesis of treatments for ADHD in children and adolescents: indirect treatment comparisons of lisdexamfetamine with methylphenidate and atomoxetine. Curr Med Res Opin 2014 Aug;30(8):1673-1685. [doi: 10.1185/03007995.2014.904772] [Medline: 24627974]
19. Karanges EA, Stephenson CP, McGregor IS. Longitudinal trends in the dispensing of psychotropic medications in Australia from 2009–2012: focus on children, adolescents and prescriber specialty. Aust N Z J Psychiatry 2014 Oct;48(10):917-931. [doi: 10.1177/0004867414538675] [Medline: 24927734]
20. Dauvilliers Y, Montplaisir J, Molinari N, et al. Age at onset of narcolepsy in two large populations of patients in France and Quebec. Neurology (ECronicon) 2001 Dec 11;57(11):2029-2033. [doi: 10.1212/wnl.57.11.2029] [Medline: 11739821]
21. Oberle D, Drechsel-Bäuerle U, Schmidtmann I, Mayer G, Keller-Stanislawski B. Incidence of narcolepsy in Germany. Sleep 2015 Oct 1;38(10):1619-1628. [doi: 10.5665/sleep.5060] [Medline: 25902804]
22. Zhang Z, Dauvilliers Y, Plazzi G, et al. Idling for decades: a european study on risk factors associated with the delay before a narcolepsy diagnosis. Nat Sci Sleep 2022;14:1031-1047. [doi: 10.2147/NSS.S359980] [Medline: 35669411]
23. van de Glind G, Van Emmerik ‐ van Oortmerssen K, Carpentier PJ, et al. The International ADHD in Substance Use Disorders Prevalence (IASP) study: background, methods and study population. Int J Methods Psych Res 2013 Sep;22(3):232-244 [FREE Full text] [doi: 10.1002/mpr.1397]

24. Strålin P, Hetta J. First episode psychosis and comorbid ADHD, autism and intellectual disability. Eur Psychiatry 2019 Jan;55:18-22. [doi: 10.1016/j.eurpsy.2018.09.007] [Medline: 30384107]
25. Jensen PS, Hinshaw SP, Kraemer HC, et al. ADHD comorbidity findings from the MTA study: comparing comorbid subgroups. J Am Acad Child Adolesc Psychiatry 2001 Feb;40(2):147-158. [doi: 10.1097/00004583-200102000-00009] [Medline: 11211363]
26. Marangoni C, De Chiara L, Faedda GL. Bipolar disorder and ADHD: comorbidity and diagnostic distinctions. Curr Psychiatry Rep 2015 Aug;17(8):604. [doi: 10.1007/s11920-015-0604-y] [Medline: 26084666]

Abbreviations

ADHD/ADD: attention-deficit/hyperactivity disorder/attention deficit disorder
GP: general practitioner
OR: odds ratio

Edited by A Mavragani, T Sanchez; submitted 25.10.23; peer-reviewed by H Demir, K Karimova; revised version received 29.09.24; accepted 30.09.24; published 07.01.25.

Please cite as:
Scharf T, Huber CA, Näpflin M, Zhang Z, Khatami R
Trends in Prescription of Stimulants and Narcoleptic Drugs in Switzerland: Longitudinal Health Insurance Claims Analysis for the Years 2014-2021
JMIR Public Health Surveill 2025;11:e53957
URL: https://publichealth.jmir.org/2025/1/e53957
doi:10.2196/53957

Population Size Estimation of Men Who Have Sex With Men in Low- and Middle-Income Countries: Google Trends Analysis

Carly M Malburg[1,2*], MPH; Steve Gutreuter[1*], PhD; Horacio Ruiseñor-Escudero[1*], MD, PhD; Abu Abdul-Quader[1*], PhD; Wolfgang Hladik[1*], MD, PhD

[1]Division of Global HIV and TB, Centers for Disease Control and Prevention, 1600 Clifton Rd, Atlanta, GA, United States

[2]CDC Global Health Fellowship Program, Public Health Institute, Oakland, CA, United States

[*]all authors contributed equally

Corresponding Author:
Carly M Malburg, MPH
Division of Global HIV and TB, Centers for Disease Control and Prevention, 1600 Clifton Rd, Atlanta, GA, United States

Abstract

Background: Population size estimation (PSE) for key populations is needed to inform HIV programming and policy.

Objective: This study aimed to examine the utility of applying a recently proposed method using Google Trend (GT) internet search data to generate PSE (Google Trends Population Size Estimate [GTPSE]) for men who have sex with men (MSM) in 54 countries in Africa, Asia, the Americas, and Europe.

Methods: We examined GT relative search volumes (representing the relative internet search frequency of specific search terms) for “porn” and, as a comparator term, “gay porn” for the year 2020. We assumed “porn” represents “men” (denominator) while “gay porn” represents a subset of “MSM” (numerator) in each county, resulting in a proportional size estimate for MSM. We multiplied the proportional GTPSE values with the countries’ male adult population (15 - 49 years) to obtain absolute size estimates. Separately, we produced subnational MSM PSE limited to countries’ (commercial) capitals. Using linear regression analysis, we examined the effect of countries’ levels of urbanization, internet penetration, criminalization of homosexuality, and stigma on national GTPSE results. We conducted a sensitivity analysis in a subset of countries (n=14) examining the effect of alternative English search terms, different language search terms (Spanish, French, and Swahili), and alternative search years (2019 and 2021).

Results: One country was excluded from our analysis as no GT data could be obtained. Of the remaining 53 countries, all national GTPSE values exceeded the World Health Organization’s recommended minimum PSE threshold of 1% (range 1.2% - 7.5%). For 44 out of 49 (89.8%) of the countries, GTPSE results were higher than Joint United Nations Programme on HIV/AIDS (UNAIDS) Key Population Atlas values but largely consistent with the regional UNAIDS Global AIDS Monitoring results. Substantial heterogeneity across same-region countries was evident in GTPSE although smaller than those based on Key Population Atlas data. Subnational GTPSE values were obtained in 51 out of 53 (96%) countries; all subnational GTPSE values exceeded 1% but often did not match or exceed the corresponding countries’ national estimates. None of the covariates examined had a substantial effect on the GTPSE values (R^2 values 0.01 - 0.28). Alternative (English) search terms in 12 out of 14 (85%) countries produced GTPSE>1%. Using non-English language terms often produced markedly lower same-country GTPSE values compared with English with 10 out of 14 (71%) countries showing national GTPSE exceeding 1%. GTPSE used search data from 2019 and 2021, yielding results similar to those of the reference year 2020. Due to a lack of absolute search volume data, credibility intervals could not be computed. The validity of key assumptions, especially who (males and females) searches for porn and gay porn, could not be assessed.

Conclusions: GTPSE for MSM provides a simple, fast, essentially cost-free method. Limitations that impact the certainty of our estimates include a lack of validation of key assumptions and an inability to assign credibility intervals. GTPSE for MSM may provide an additional data source, especially for estimating national-level PSE.

(JMIR Public Health Surveill 2025;11:e58630) doi:10.2196/58630

KEYWORDS

population size estimation; men who have sex with men; MSM; PSE; google trends; HIV; AIDS; programming and policy; internet; porn; gay porn; male adult; geriatric; linear regression; homosexuality; sensitivity analysis; World Health Organization; WHO; epidemiology

Introduction

The Joint United Nations Programme on HIV/AIDS (UNAIDS) estimated that in 2022, about 39 million people were living with HIV worldwide [1]. HIV burden is higher among men who have sex with men (MSM), people who inject drugs, sex workers, and transgender persons, which together are often described as key populations (KP) [1]. KPs and their paying or nonpaying sexual partners may account for 70% of new HIV infections worldwide, with an estimated 80% of new HIV infections outside sub-Saharan Africa (SSA) and 55% of all new infections within SSA [1,2].

Key population size estimation (PSE) is needed to estimate the number of individuals belonging to a KP in a given geographical area [3,4]. PSEs provide the denominator values to inform KP programming and policy [5]. However, PSE is a difficult field and its methods often lack rigor in design or implementation, and the many methods available reflect the lack of an acceptable gold standard [3,6]. Challenges to PSE include lack of sampling frames, mobility, and nondisclosure of KP-defining behaviors [3,4]. Further, most PSE methods produce local estimates whereas national PSE estimates are often obtained through "expert opinion," simple projection, or modeling and less often through national-level empirical data such as direct survey questions or the network scale-up method, both used in general population-based surveys [6,7]. Direct survey questions about KP-defining traits experience reporting bias and require a major effort unless they can be added to an already planned general population survey. The frequent lack of reliable national-level PSE constitutes an even larger challenge compared with the availability of local PSE and complicates national, regional, and global HIV estimation work [3,8-10].

The rise of the internet facilitates web-based activities to improve public health, including in the field of digital epidemiology and infoveillance [11]. Recently, a new PSE method using Google Trends (GT) internet search data was proposed in a proof of concept paper by Card et al [12] GT is a free cloud-based app that displays the relative frequency of user-specified Google search terms as trends across time and user-selected geographical areas [12-14]. Card et al [12] used GT and Canadian census data to estimate the local PSE of MSM in urban and rural locations throughout Canada. Card et al [12] related search terms presumed to be representative of MSM ("gay porn") to that presumed to be representative of the general (male) population ("porn"). By relating these 2 sets of values, Card et al [12] estimated the relative size of MSM in these Canadian towns. To date, no other published PSE exists using this method.

The literature on pornography consumption by sex and sexual orientation is limited and often the MSM population is not represented. However, a major porn website reported that about a third of its visitors globally in 2021 were reportedly women [15,16]. Further, women, regardless of sexual orientation, may also watch gay porn, possibly in substantial numbers [17]. Beyond this, we found no meaningful gray literature or peer-reviewed articles about internet pornography consumption in low- and middle-income countries (LMICs) or pornography consumption by MSM in LMICs. We are also not aware of (gray) literature about the proportion of heterosexual and homosexual men searching Google for (gay) porn in LMICs.

We expanded the literature search to include high-income settings. A study conducted in the United States reported that more men than women consume pornography (92%:68%, respectively) over the span of a year [18]. The study did not report the type of pornography consumed or disaggregate male respondents by sexual orientation or practice [18]. A separate study from Norway with a sample of some 2300 male and female participants suggested that more men than women consume some pornography (94% of men and 68% of women) [19]. However, only 5% (n=106) of participants identified as gay/lesbian/bisexual, no breakdown of sexual orientation by sex was given, and no information on the type of pornography consumed by participants was available [19].

The aim of this study was to examine the potential utility of using GT data to obtain MSM PSE in selected LMICs.

Methods

Preliminary Literature Search

A nonsystematic literature search was conducted to better understand the behavior of pornography consumption of the general population and sexual minorities, by sex, as well as the relative frequencies with which these populations search for (gay) porn in general (via Google) or by directly accessing specific porn sites.

Selection of Countries

We analyzed GT data for a selected set of 54 countries that receive support from the US President's Emergency Plan for AIDS Relief, the US Government's initiative to support global HIV responses, for which information on MSM PSE has been sought [2,20]. These countries are located in SSA (n=29), Asia (n=13), the Americas (n=11), and Ukraine (Tables 1 and 2).

Table . National men who have sex with men (MSM) population size estimation (PSE) for US President's Emergency Plan for AIDS Relief supported countries (n=53) using Google Trends (GT) data for the year 2020[a,b].

Region and country		GT (number of MSM), n	GT, %	UNAIDS[c] GAM[d] regional %, median (IQR)[e]	UNAIDS KP[f] Atlas, %
East Africa[g]				1.67	
	Burundi	48,500	1.77		0.34
	Ethiopia	365,000	1.28		—[h]
	Kenya	276,000	1.99		0.24
	Rwanda	51,300	1.54		0.15
	Tanzania	243,000	1.73		0.35
	Uganda	154,000	1.47		0.23
Southern Africa[i]				1.67	
	Angola	106,000	1.44		—
	Botswana	13,000	2.12		0.43
	Eswatini	4500	1.57		1.38
	Lesotho	10,000	1.71		1.05
	Malawi	52,500	1.16		0.94
	Mozambique	134,000	1.87		0.22
	Namibia	16,500	2.60		—
	South Africa	393,000	2.46		1.94
	Zambia	51,800	1.18		0.15
	Zimbabwe	53,000	1.64		0.71
West Central Africa[j]				1.28 (IQR 0.45 - 1.50)	
	Benin	34,000	1.18		0.20
	Burkina Faso	88,000	1.80		0.07
	Cameroon	148,000	2.29		0.11
	Cote d'Ivoire	166,000	2.68		0.90
	DRC	33,000	1.66		0.98
	Ghana	112,000	1.40		0.69
	Liberia	22,600	1.83		6.04
	Mali	70,500	1.55		0.09
	Nigeria	614,000	1.26		0.49
	Senegal	73,600	1.94		1.38
	Sierra Leone	27,000	1.36		0.16
	Togo	53,100	2.65		0.30
Asia[k]				1.63 (IQR 0.26 - 3.10)	
	Burma	664,000	4.53		1.72
	Cambodia	258,000	5.67		1.93
	India	6,460,000	1.18		0.06
	Indonesia	1,180,000	1.61		1.03
	Kazakhstan	137,000	2.99		1.35
	Kyrgyz Rep.	53,000	3.10		0.99
	Lao PDR	53,000	2.73		2.96

Region and country		GT (number of MSM), n	GT, %	UNAIDS[c] GAM[d] regional %, median (IQR)[e]	UNAIDS KP[f] Atlas, %
	Nepal	83,000	1.19		0.86
	PNG	31,000	1.30		1.58
	Tajikistan	52,000	2.14		0
	Thailand	215,000	1.25		3.08
	Philippines	1,260,000	4.27		2.33
	Viet Nam	1,953,000	7.46		0.98
Europe					
	Ukraine	366,000	3.48	2.11 (IQR 1.75 - 2.49)	1.71
Caribbean[l]				2.71	
	Dominican Rep.	124,000	4.26		4.90
	Guyana	8200	3.60		1.45
	Haiti	108,000	3.60		1.03
	Jamaica	24,000	2.91		5.15
	Trinidad and Tobago	11,000	3.04		—
Central and South America[m]				3.37	
	Brazil	2,960,000	5.18		3.50
	El Salvador	85,000	5.20		3.31
	Guatemala	245,000	5.09		2.42
	Honduras	147,000	5.32		1.48
	Nicaragua	114,000	6.32		1.97
	Panama	81,000	7.23		2.65

[a]These estimates are for descriptive purposes only, to examine issues related to the potential utility of the method proposed by Card et al [12]. They represent the MSM population national population size estimates (percentage of MSM) for the year 2020. The percentage of MSM was calculated by taking the average relative search volume score produced by Google Trends for "gay porn" and dividing it by the average relative search volume score produced by Google Trends for "porn." MSM population size estimate (number of MSM) was calculated by taking the percentage of MSM population size estimate and dividing it by the total male population (ages 15 - 49 years). Key populations (KPs) Atlas percentage of MSM population size estimate was calculated by dividing the absolute MSM population size estimate taken from the United Nations Programme on HIV/AIDS (UNAIDS) KPs Atlas dashboard by the total adult male population (ages 15 - 49 years), and then multiplying by 100. The absolute value difference was calculated by subtracting the GT absolute MSM population size estimate value from the KPs Atlas MSM population size estimate absolute value. All absolute values under 10,000 are rounded to the nearest 100. All other absolute values are rounded to the nearest 1000. UNAIDS Global AIDS Monitoring system (GAM) values are regional values transcribed from the UNAIDS open-source Spectrum 6 guide. The countries used to create these regions and respective values may not be in full alignment with the countries included in the population size estimate analysis, therefore direct 1:1 comparisons should not be made. Max:Min ratio: The ratio based on the largest and smallest PSE % value in each region.

[b] Absolute values are not provided as Google Trends does not provide absolute search frequency values.

[c]UNAIDS: United Nations Programme on HIV/AIDS.

[d]GAM: Global AIDS Monitoring system.

[e]IQR values were included for available regions. Regions without an IQR listed did not have one available.

[f]KP: key population.

[g]Max:Min ratio: 1.6 (GT) and 2.3 (UNAIDS KP).

[h]Not available (data missing for the country).

[i]Max:Min ratio: 2.2 (GT) and 12.9 (UNAIDS KP).

[j]Max:Min ratio: 2.3 (GT) and 86.3 (UNAIDS KP).

[k]Max:Min ratio: 6.3 (GT) and 51.3 (UNAIDS KP).

[l]Max:Min ratio: 1.5 (GT) and 5 (UNAIDS KP).

[m]Max:Min ratio: 1.2 (GT) and 2.4 (UNAIDS KP).

Table . Regional median Google Trends Population Size Estimate, United Nations Programme on HIV/AIDS (UNAIDS) Global AIDS Monitoring system (GAM), and key populations (KP) Atlas for men who have sex with men (MSM) populations for the year 2020.[a]

Region	Median regional percentage MSM population size estimation[b]		
	GT[c], %	UNAIDS GAM, %	UNAIDS KP Atlas, %
Eastern Africa	1.64	1.67	0.24
Southern Africa	1.68	1.67	0.83
West Central Africa	1.73	1.28	0.40
Asia	2.86	1.63	—[d]
Europe	2.86	2.11	1.47
Caribbean	3.60	2.71	3.17
Central & South America	5.26	3.37	2.54

[a]Absolute values are not provided as Google Trends does not provide absolute search frequency values.

[b]Google Trends (GT) and KP Atlas regional estimates only include estimates from included countries with available data (Table 1). UNAIDS GAM data separate regions differently and include countries that vary from our GT or the KP Atlas regional data: UNAIDS GAM includes eastern and southern Africa in 1 estimate and separates Asia and Europe into 2 estimates (1.63% for Asia and the Pacific, 2.11% for Eastern Europe and Central Asia). Region names were not adjusted in the above table to align with GAM data.

[c]GT: Google Trends.

[d]Not available.

Ethical Considerations

No ethics or review board approval or informed consent was obtained or applicable for this work. All data used in this paper are anonymous, aggregate, and publicly available and sourced.

GT-Based Population Size Estimation

GT provides results based on exact search terms, unlike the "topical" search results that Google's main search engine provides. GT does not provide absolute search frequency values; instead, GT offers relative search volume (RSV) values across time (eg, 52 wk) in a specified space (eg, Kenya), ie, it normalizes search frequencies for specific search terms (eg, porn) to a range from 0 to 100, where a search term's maximum frequency (for the specified geographic area and during the specified time frame) is set at 100 and 0 reflects no search for that term [11,13,14]. Importantly, GT allows users to add "comparator" terms (eg, gay porn) next to the main term (eg, porn); the RSV values for such comparator terms are normalized against the main term's RSV values [13,21]. For the purpose of PSE calculation, the main term "porn" may represent all men whereas the comparator term "gay porn" may be viewed as a subset of men who represent the subpopulation of gay men or MSM. To generate an MSM PSE from the RSV values we divide the comparator RSV value (gay porn) by the larger same-time, same-place RSV value (porn).

National Size Estimates

PSE data collection was carried out through GT's application [13]. We applied this analytic approach for the year 2020 using "porn" and "gay porn" as the main and comparator search terms for each of the 54 countries. The time period for data collection was set as the year 2020, the most recent year for which we could obtain all necessary data for this analysis. Weekly RSV values for "porn" and "gay porn" for the year 2020 were exported, summed, and proportional size estimates obtained. For example, for Botswana, the average of the weekly RSV values for "porn" was 78.3, the corresponding average for "gay porn" was 1.66 and the proportional PSE was therefore calculated as 1.66/78.3=2.1%. This was repeated for all countries. We then calculated the absolute Google Trends Population Size Estimate (GTPSE) by multiplying the proportional GTPSE by the total male population aged 15 - 49 years in each country, the most used age range for KPs. The sizes for countries' 15 - 49 year-old male general population in 2020 were obtained through Spectrum (version 6.1, Avenir Health).

Local Size Estimates

GT data can be restricted to subnational areas. Separately from national estimates, for each country, we also attempted to obtain local GTPSE for the political (or, if different, commercial) capital city. Where data were unavailable for the political or commercial capital city, we used data from the district that contained the capital city. The calculation to obtain relative GTPSE was then the same as for the national level. We did not produce absolute subnational GTPSE.

Consistency of GTPSE Results With WHO-Recommended Minimum Estimate

We assessed whether the GTPSE results met the World Health Organization (WHO) and UNAIDS recommendation that national MSM PSE should represent at least 1% of the general adult male population [22,23].

Comparability

We compared the country-level GTPSE against 2 reference data sources used by UNAIDS: the KP Atlas database and the Global AIDS Monitoring system (GAM) [22,24,25]. The KP Atlas database stores countries' self-reported absolute MSM size estimates using a wide range of PSE methods, often projected up to national scale from local estimates, with primary data collected over different periods of time. Proportional KP Atlas

PSE values were computed by dividing the absolute MSM PSE values from the KP Atlas over the male general population (15 ‐ 49 years). UNAIDS’ GAM is a global data warehousing system that informs policy and facilitates monitoring, including KP size estimates. Using GAM data, UNAIDS curated a table with regional relative MSM PSE (median and IQR) deemed reasonable.

Covariates Potentially Affecting GTPSE

Overview

We examined the potential effect of select covariates on the relative GTPSE values by performing regression analysis for each covariate. The country-specific covariates we examined included internet penetration, urbanization, stigma, and criminalization of homosexuality. The covariate data are provided in Table S1 in Multimedia Appendix 1; these data were not used to adjust GTPSE values.

Internet Penetration and Urbanization

Internet penetration data were extracted from the World Development Indicators database through the World Bank and the Internet World Statistics database, indicating the percentage of each country’s total population with access to the internet. Urbanization data were obtained from the World Development Indicators database through the World Bank, indicating the percent of the total population in each country considered urban [26,27].

Stigma

Country-level stigma values were extracted from the Global Acceptance Index [28]. This index was developed using computer modeling informed by responses to questions that measure attitudes toward lesbian, gay, bisexual, transgender, or intersex people from 11 different global surveys to create a stigma score in 175 countries toward lesbian, gay, bisexual, transgender, or intersex persons. The system scores countries on a scale of 1 to 10; higher scores indicate less stigma [28].

Criminalization

The State-Sponsored Homophobia International Lesbian, Gay, Bisexual, Trans, and Intersex Association report was used to evaluate the effects of criminalization of homosexual orientation or behavior on GTPSE [28]. The report classifies countries based on their level of legal protection or criminalization of sexual orientation and same-sex sexual acts. These classifications, ranging from most severe to most protected, include the death penalty, up to lifelong imprisonment, up to 8 years imprisonment, de facto criminalization, no criminalization or legal protections, limited protections, employment protections, broad protections, and constitutional protections. We converted these classifications into a quantitative ranking ranging from +4 to −4. The most severe classification (death penalty) was assigned the rank value “+4” and descended to the least severe/most protective classification (constitutional protection) with a rank value “−4.”

Sensitivity Analysis

Using a subset (n=14) of the 53 countries we performed 3 sensitivity analyses at the national level. The 14 countries were randomly selected among countries with prominent languages being French, Spanish, or Swahili. The first sensitivity analysis probed the effect of select non-English search languages. The 14 countries comprised 4 using Swahili (Kenya, Tanzania, Uganda, Democratic Republic of Congo [DRC]), 5 using French (Cote d’Ivoire, Senegal, Cameroon, Mali, Haiti), and 5 using Spanish (Dominican Republic, Panama, El Salvador, Nicaragua, Honduras) as their national/dominant language. We generated GTPSE using search terms in Swahili (“ngono” and “ngono za mashoga”), French (“porno” and “porno gay”), and Spanish (“porno” and “porno gay”) and compared them to the original relative GTPSE values. Using the same 14 countries, the second sensitivity analysis probed the effect of different search terms in English on GTPSE, that is, “sex,” ”gay sex” as well as “sex,” ”anal sex” and compared them to the original GTPSE (porn and gay porn). The third sensitivity analysis probed the effect of using different calendar years, ie, (2019 [pre-COVID] and 2021) and compared them to the original 2020 GTPSE values, using the original English language search terms.

Results

GTPSE and Comparability

Of the 54 countries examined, 1 (South Sudan), was omitted for lack of RSV values. All remaining 53 countries had GTPSE exceeding 1% (Table 1), similar to GAM values (all exceeding 1% as well) and compared with KP Atlas values where 24 out of 53 (45%) countries showed values above 1%. GTPSE ranged from 1.16% to 7.46% (median 1.99%, IQR 1.54% ‐ 3.48%), compared with 0.06% to 6.04% (median 0.99%, IQR 0.34 ‐ 1.93%) in the KP Atlas, and 1.38% to 2.82% in GAM regions. In 48 out of 53 (91%) countries, relative GTPSE exceeded estimates in the KP Atlas values; KP Atlas values were larger in 5 countries (DRC, Liberia, Lao People’s Democratic Republic [PDR], Thailand, and Jamaica). Absolute differences between GTPSE and KP Atlas ranged from −312,900 (Thailand) to 6,221,800 (India). Table 2 displays regional median GTPSE, ranging from 1.64% (East Africa) to 5.26% (Central/South America), larger in all regions than the corresponding KP Atlas values and largely similar to GAM values in most regions. Table 1 also displays the ratios between the largest and smallest country-level %PSE for each region, separately for GT and KP Atlas values. While substantial variability is seen in all regions and for both data sources (GT and KP Atlas), in all regions the observed heterogeneity was consistently higher for KP Atlas values compared with GT values.

Local GTPSE pertaining to political or commercial capitals or the larger sub-national areas encompassing these are displayed in Table 3. We could obtain local estimates for 51 out of 53 (96%) countries’ capital cities; GT did not provide data for Nairobi (Kenya) and Kathmandu (Nepal). Among the 51 cities with estimates, the GTPSE ranged from 0% to 13% (median 2.2%); most cities’ estimates (44/51, 86%) exceeded 1%. Five cities yielded noncredible GTPSE values of 0%, including Bujumbura (Burundi), Dodoma (encompassing Dar es Salaam, Tanzania), Ouagadougou (Burkina Faso), Monrovia (Liberia), and Vientiane (Laos PDR). Of the 44 subnational GTPSE with values >1%, 18 (41%) were below the same-country national GTPSE.

Table . Reported local men who have sex with men (MSM) Google Trends Population Size Estimate (GTPSE) (n=53) in the year 2020.[a]

Region and country		Local area[b]	Relative national GTPSE, %	Relative local GTPSE, %	Absolute percentage difference national and local GTPSE, %
East Africa					
	Burundi	Bujumbura	1.77	0	−1.77
	Ethiopia	Addis Ababa	1.28	1.30	0.02
	Kenya	Nairobi	1.99	—[c]	—
	Rwanda	Kigali	1.54	1.70	0.16
	Tanzania	Dodoma	1.73	0	−1.73
	Uganda	Kampala	1.47	1.60	0.13
Southern Africa					
	Angola	Luanda	1.44	2.04	0.60
	Botswana	Gaborone	2.12	0	−2.12
	Eswatini	Mbabane	1.57	1.48	−0.09
	Lesotho	Maseru	1.71	1.01	−0.70
	Malawi	Lilongwe	1.16	2.24	1.08
	Mozambique	Maputo	1.87	2.04	0.17
	Namibia	Windhoek	2.60	2.53	−0.07
	South Africa	Johannesburg (Gauteng)	2.46	0.99	−1.47
	Zambia	Lusaka	1.18	1.56	0.38
	Zimbabwe	Harare	1.64	1.56	−0.08
West Central Africa					
	Benin	Littoral (Cotonou)	1.18	4.11	2.93
	Burkina Faso	Centre (Ouagadougou)	1.80	0	−1.80
	Cameroon	Littoral (Douala)	2.29	2.47	0.18
	Cote d’Ivoire	Abidjan	2.68	1.01	−1.67
	DRC	Kinshasa	1.66	2.04	0.38
	Ghana	Accra	1.40	1.42	0.02
	Liberia	Monrovia	1.83	0	−1.83
	Mali	Bamako	1.55	2.93	1.38
	Nigeria	Abuja (Federal Capital Terriorty)	1.26	1.44	0.18
	Senegal	Dakar	1.94	2.85	0.91
	Sierra Leone	Freetown	1.36	1.01	−0.35
	Togo	Lome	2.65	2.04	−0.61
Asia					
	Burma	Yangon (Yangon Region)	4.53	4.79	0.26
	Cambodia	Phnom Penh	5.67	5.43	−0.24
	India	New Delhi (Uttar Pradesh)	1.18	1.15	−0.03
	Indonesia	Jakarta	1.61	2.20	0.59
	Kazakhstan	Almaty (Almaty Region)	2.99	5.52	2.53

Region and country		Local area[b]	Relative national GTPSE, %	Relative local GTPSE, %	Absolute percentage difference national and local GTPSE, %
	Kyrgyz Rep.	Bishkek	3.10	3.09	−0.01
	Lao PDR	Vientiane	2.73	0	−2.73
	Nepal	Katmandu/Kantipur	1.19	—	—
	PNG	Port Moresby	1.30	1.01	−0.29
	Tajikistan	Dushanbe	2.14	1.01	−1.13
	Thailand	Bangkok	1.25	3.24	1.99
	Philippines	Manila	4.27	5.51	1.24
	Viet Nam	Hanoi	7.46	4.56	−2.90
Europe					
	Ukraine	Kyiv	3.48	4.14	0.66
Caribbean					
	Dominican Rep.	Santo Domingo	4.26	3.99	−0.27
	Guyana	Georgetown	3.60	3.09	−0.51
	Haiti	Port-au-Prince	3.60	3.09	−0.51
	Jamaica	Kingston (St. Andrew Parish)	2.91	3.38	0.47
	Trinidad and Tobago	Port of Spain	3.04	13	9.96
Central and South America					
	Brazil	São Paulo (State of São Paulo)	5.18	5.92	0.74
	El Salvador	San Salvador	5.20	5.73	0.53
	Guatemala	Guatemala City (Guatemala Department)	5.09	4.91	−0.18
	Honduras	Tegucigalpa (Comayagua)	5.32	8.13	2.81
	Nicaragua	Managua	6.32	5.31	−1.01
	Panama	Panama City	7.23	7.44	0.21

[a]Absolute values are not provided as Google Trends does not provide absolute search frequency values.

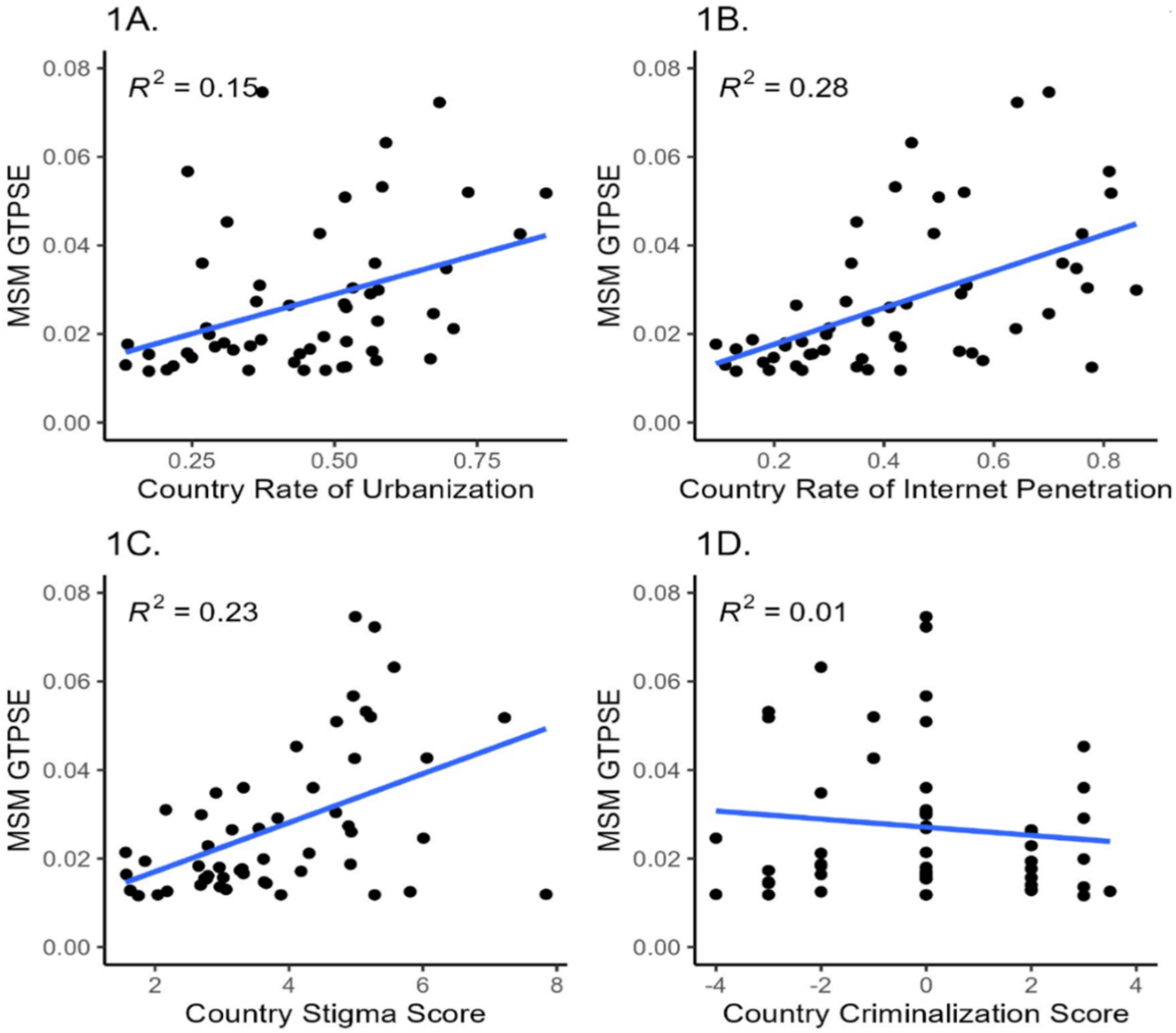
[b]Local MSM GTPSE for 53 countries for the year 2020 was calculated by restricting the geographic entity to the desired capital city or commercial hub. Where Google Trends (GT) did not provide data for a given city, we substituted the place name with the largest city by population or by district that had data available in GT. This is noted by listing what was available in GT in parenthesis next to the capital city. Kenya and Nepal were excluded from this analysis due to insufficient regional data available in GT.

[c]Not available (data missing for that country).

Effect of Covariates

Figure 1A-D displays the correlations between national-level GTPSE and urbanization, internet penetration, stigma, and criminalization. Coefficients ranged from 0.01 (criminalization) to 0.28 (internet penetration).

Figure 1. The linear relationship between the Google Trends national population size estimates and the rate of urbanization in each country (n=53). (A) The linear relationship between the Google Trends national population size estimates and the rate of urbanization in each country (n=53). (B) The linear relationship between the Google Trends national population size estimates and the rate of internet penetration in each country (n=53). (C) The linear relationship between the Google Trends national population size estimates and the level of stigma against LGBTQ+ persons in each country (n=53). (D) The linear relationship between the Google Trends national population size estimates and the degree of criminalization against men who have sex with men population in each country (n=53). LGBTQ+: lesbian, gay, bisexual, transgender, queer, and other identities; MSM: men who have sex with men; GTPSE: Google Trends Population Size Estimate.



Sensitivity Analysis

Table 4 displays how the GTPSE generated from the alternative search terms compares to the original search term GTPSE. In most countries "Porn/Gay Porn" produced higher PSE values compared with "sex/anal sex" (13/14, 93%) as well as compared with "sex/gay sex" (12/14, 86%). For "sex/gay sex," all 14 countries produced estimates exceeding 1%. For "sex/anal sex", 3 out of 14 (21%) countries did not produce estimates reaching the 1% threshold, including Mali for which zero search results were reported for "anal sex."

Table . Sensitivity analysis using alternative search terms in Google Trends to calculate national population size estimations (PSEs) for select US President's Emergency Plan for AIDS Relief countries (n=53) in 2020.[a]

	Original GTPSE[b]	SA alternate search term GTPSE[c]			
	Porn/gay porn PSE	Sex/gay sex PSE	Absolute percentage difference	Sex/anal sex PSE	Absolute percentage difference
Country, %					
Kenya	1.99	1.37	0.62	1.37	0.62
Tanzania	1.73	1.46	0.27	3.54	−1.81
Uganda	1.47	1.38	0.09	1.26	0.21
DRC	1.66	1.55	0.11	1.15	0.51
Cameroon	2.29	1.28	1.01	0.90	1.39
Mali	1.55	1.65	−0.10	0	1.55
Cote d'Ivoire	2.68	1.90	0.78	1.74	0.94
Senegal	1.94	1.50	0.44	0.88	1.06
Haiti	3.60	2.60	1	2.83	0.77
Dominican Rep.	4.26	3.36	0.90	1.83	2.43
Panama	7.23	5.17	2.06	3.71	3.52
El Salvador	5.20	5.34	−0.14	4.19	1.01
Nicaragua	6.32	7.10	−0.78	4.82	1.50
Honduras	5.32	4.85	0.47	2.96	2.36
Median (IQR)	2.49 (1.78-4.97)	1.78 (1.47-4.48)	0.45 (0.10-0.87)	1.79 (1.18-3.40)	1.03 (0.66-1.54)

[a]Absolute values are not provided as Google Trends does not provide absolute search frequency values.

[b]GTPSE: Google Trends Population Size Estimate.

[c]Alternative search terms were chosen based on words that represented the general male population and men who have sex with men subset population in each country (n=53) in the year 2020.

Table 5 shows how GTPSE was generated using alternative language terms compared with the original GT search terms. For Swahili, only 1 country yielded a PSE in that language. All countries using French (n=5), or Spanish (n=5) search terms yielded estimates, all exceeding 1%. All alternative language estimates were lower than the original "porn/gay porn" PSE values.

Table . Sensitivity analysis using alternate national language searches in Google Trends to calculate national population size estimation for select US President's Emergency Plan for AIDS Relief countries (n=14) in 2020.[a]

Language and country		Original GTPSE[b] (English), %	Alternate language term GTPSE, %[c]	Absolute percentage difference, %
Swahili				
	Kenya	1.99	0	1.99
	Tanzania	1.73	0.52	1.21
	Uganda	1.47	0	1.47
	DRC	1.66	0	1.66
French				
	Cameroon	2.29	1.36	0.93
	Mali	1.55	1.07	0.48
	Cote d'Ivoire	2.68	1.35	1.33
	Senegal	1.94	1.28	0.66
	Haiti	3.60	2.23	1.37
Spanish				
	Dominican Rep.	4.26	2.56	1.70
	Panama	7.23	5.14	2.09
	El Salvador	5.20	4.36	0.84
	Nicaragua	6.32	4.13	2.19
	Honduras	5.32	4.07	1.25

[a]Absolute values are not provided as Google Trends does not provide absolute search frequency values.

[b]GTPSE: Google Trends Population Size Estimate.

[c]Alternative language search terms included "ngono/ngono za mashoga" (Swahili), "porno/porno gay" (French), "porno/porno gay" (Spanish).

Table 6 displays how GTPSE generated for alternative years (2019 and 2021) compared with the original 2020 GT searches. All 14 countries in both years produced estimates exceeding 1%. No large discrepancies in PSE between the years were observed; 13 out of 14 in 2019 values were larger than the 2020 values whereas the 2021 values were largely similar to the 2020 values.

Table . Sensitivity analysis for men who have sex with men population size estimates for select US President's Emergency Plan for AIDS Relief supported countries (n=14) using Google Trends data in years 2019 and 2021 compared with the year 2020.[a, b]

	2019 PSE[c], %	2020 PSE, %	2021 PSE, %
Kenya	2.37	1.99	1.99
Tanzania	1.96	1.73	1.85
Uganda	1.73	1.47	1.69
DRC	1.95	1.66	1.60
Cameroon	2.70	2.29	2.25
Mali	2.30	1.55	2.17
Cote d'Ivoire	2.52	2.68	2.23
Senegal	2.54	1.94	1.90
Haiti	4.33	3.60	2.92
Dominican Republic	4.91	4.26	4.34
Panama	9.36	7.23	6.74
El Salvador	6.19	5.20	4.77
Nicaragua	7.31	6.32	4.93
Honduras	6.79	5.32	5.51

[a]2019 and 2021 values were computed in the same way as the reference 2020 estimates.

[b]Absolute values are not provided as Google Trends does not provide absolute search frequency values.

[c]PSE: population size estimation.

Discussion

Principal Findings

Our analysis suggests that national-level MSM GTPSE is feasible in almost all countries. Importantly, all estimates appeared plausible, that is, they exceeded the WHO/UNAIDS suggested minimum threshold of 1%. Heterogeneity of GTPSE across same-region countries was pronounced within all regions yet smaller than the ratios based on the UNAIDS KP Atlas values which contained numerous PSE values well below the 1% threshold.

Our analysis draws on several strengths. We successfully applied the GTPSE method to many low and middle-income countries, suggesting that GTPSE appears to have wide geographic applicability. We compared the values against 2 PSE data sources at UNAIDS, assessed the potential effect of various covariates on GTPSE values, and conducted a sensitivity analysis with varying English search terms, non-English search languages, and different calendar years. Google is the dominant search engine in all countries covered in this analysis, with a market share ranging between 84% and 99% (data shown in Table S1 in Multimedia Appendix 1) [27]. Although no absolute search volume data were available to us, searches for "porn" globally were among the top 20 search terms in 2023 with about 65 million searches globally each month according to one source [29] although this is still well behind the largest porn site-specific searches. GTPSE may emerge as another example of digital public health and epidemiology that includes real-time surveillance of disease outbreaks [30], assessing the impact of global public health days [31], informing health and health policy research [32], or understanding spatiotemporal patterns of dry eye disease [33].

While most local estimates were plausible (>1%), 14% (n=7) did not reach the WHO/UNAIDS minimum threshold, and 2 more locations did not produce a GTPSE value at all due to lack of GT data and how GT organized the subnational data despite some of the affected cities' large population sizes. This is not an uncommon finding, as other PSE methods in active use typically do not meet the WHO/UNAIDs minimum threshold. For a few other country or commercial capital cities with no direct GT data available, such as Johannesburg (South Africa), we could obtain a subnational estimate using the larger district or province within which the city (eg, Johannesburg and Pretoria) are located. This may limit the utility and comparability of such local estimates. About one-third of the local (relative) estimates did not reach or exceed the same country national level estimates, somewhat contrary to our expectation that rural-to-urban migration among MSM may be more pronounced than that of other men and so yielding higher GTPSE values [9]. In Card et al's [12] study on Canadian towns and cities the estimates ranged from 2% to 4% compared with 0% to 13% among the local estimates, whereas the Canadian national estimate was 2.8% compared with 1.2% - 7.5% across all countries we examined. While not a limitation, it is worth noting that weekly RSV data varied widely (data not shown), confirming the recommendation to use GT data for size estimation only over longer time periods, such as a full calendar year.

Limitations

Like most PSE methods, GTPSE has limitations. In particular, the assumptions underlying the GTPSE method deserve close scrutiny: straight men only search for porn, MSM only search for gay porn, MSM and straight men search for (gay) porn in equal proportions, and women do not search for (gay) porn at all or do not affect the generated GTPSE for MSM. Violations of these assumptions will result in bias if they affect RSV for porn and gay porn to differing extents, hence altering the proportion of porn searches that are directed at gay porn. While the literature from LMIC settings on this topic is very sparse, reports and literature from high-income settings suggest that gay porn is also consumed by heterosexual men and women, suggesting that some bias may be present. Complicating speculations about the magnitude and direction of bias is the fact that specific porn websites' user statistics may not accurately reflect searches for (gay) porn on Google. Women's search behavior on Google regarding gay porn may increase or decrease the GTPSE estimates depending on the frequency relative to searches for just porn.

Regrettably, Google does not provide access to its algorithm generating the RSV data nor can users filter GT searches by age or gender. An inherent limitation in using GT data includes the lack of deduplication in the search data (although repeated searches by the same user within a short time period are not counted multiple times by Google) and the lack of absolute search volume data. Not having access to the absolute search volume data impedes the computation of uncertainty intervals (which in most national settings may be expected to be small due to the large search volumes involved). However, absolute search volume information may eventually be made available by Google and is already offered to some extent by select third-party companies. Absolute search volume data may also inform the choice of search language and even search terms and may facilitate composite GTPSE metrics by incorporating multiple GTPSE metrics stemming from different language search terms. Restricting GTPSE-relevant data to male users may further refine GTPSE values by excluding female users, a limitation our analysis could not overcome. VPN (virtual private network) also has the potential to introduce errors if users select a country other than their place of residence. The adoption of VPN may vary considerably across time and by country, and, among US President's Emergency Plan for AIDS Relief countries. According to one industry website in 2020, VPN was highest in Ukraine (7.9%) and lowest in Kenya (0.5%) [34]. Taken together, these limitations constitute a major source of uncertainty about the bias and precision of GTPSE. For that reason, GPTSE should be regarded as an approximate reference value. Clearly, they do not attain the rigor or transparency of statistically principled estimation from accurately measured data, which the currently best available PSE methods do offer. Additionally, GTPSE may not be feasible for a few countries, perhaps due to poor or little data availability on search terms and frequencies.

GTPSE seems infeasible for size estimation among transgender persons, sex workers, or people who inject drugs. Unlike (gay) porn, where the search is about a web-based product (visual depictions of porn), searches for sex work or clients, transgenderism, or injecting drug use are not directly tied to the internet, and may exhibit a more variable search terminology, and may lack fitting "denominator" search terms (analogous to "porn").

Overall, the GTPSEs often were substantially higher than the KP Atlas estimates but were more closely aligned with the reported GAM regional estimates. The KP Atlas estimates are based on a broad range of PSE methods typically generating local PSE that may or may not be projected to national scale, or summed or averaged across multiple localities, and may refer to various time points (calendar years) and various age ranges. Many KP Atlas based MSM PSE were implausibly low (<1%), suggesting that substantial differences to GTPSE may often be due to KP Atlas underestimates. The regional GAM estimates are based on a more curated database of PSE after excluding estimates with subpar quality and hence of perhaps more trustworthy quality [22]. However, GAM regions do not exactly overlap with the regions we used for GTPSE and the KP Atlas estimates.

The national MSM GTPSE values were robust against varying levels of urbanization, internet penetration, stigma, and criminalization or protection of homosexuality, negating the need for adjustment and increasing comparability across different settings. The largest influence was seen with internet penetration which can be expected to increase over time. In the sensitivity analysis, the largest differences to the original GTPSE values were seen using alternate English language search terms. Among the 14 examined countries, almost half (43%) of the alternate estimates were below 1% and hence considered implausibly low. This indicates that search term selection is important, especially for comparison across time and space. Further exploration may be warranted to evaluate if country or region-specific English or non-English slang terms may produce plausible estimates; however, the limited sensitivity analysis suggests that "Porn/Gay Porn" may be dependable and consistently produces plausible values. The use of similar search terms in French, Spanish, and Swahili yielded universally lower results; Swahili, not a nationally dominant language in most countries, appears particularly unsuitable as it frequently produced 0% PSE values. As most countries display prominent non-English language use, countries may want to consider using the predominant language (used for web searches) when applying this method while considering any language's geographic scope in-country. The results also appeared robust across time (two years affected by the COVID pandemic plus 1-year pre-COVID) as the 2 adjacent years produced plausible and (same country) consistent results. The lack of uncertainty intervals however impeded a more meaningful interpretation of the results from the sensitivity analyses.

Conclusions

Generating national-level PSEs for KPs is challenging for many countries. GTPSE is a simple method with the potential to address this problem efficiently without the need of additional resources. However, the lack of validation of key assumptions and the inability to generate credibility intervals suggest important uncertainty regarding the accuracy and precision of the estimates. Additional research, such as expanding or building on our sensitivity and covariate analysis, to address or better understand these limitations may further improve the quality and utility of GTPSE for MSM in LMICs.

Acknowledgments

We wish to acknowledge all authors who contributed to this paper. WH and AAQ contributed to the study concept and design. All authors substantially contributed to the study's conduct, or to data analysis and interpretation. All authors contributed to writing, reviewing, or editing the manuscript. We also wish to acknowledge all colleagues within the Center for Disease Control and Prevention who offered insights and alternative perspectives during the analysis and writing of this paper.

Disclaimer

The findings and conclusions in this report are those of the authors and do not necessarily represent the views of the funding agencies. This research publication has been supported by the President's Emergency Plan for AIDS Relief (PEPFAR) through the Centers for Disease Control and Prevention (CDC) under the terms of project number 0900f3eb820619c0. This publication was also supported by Cooperative Agreement number NU2GGH002093 between the Centers for Disease Control and Prevention and the Public Health Institute.

Data Availability

All data generated or analyzed during this study are publicly accessible data (with the exception that only aggregate data can be obtained through Google Trends). Links to where the data can be accessed are included in this published article and its supplementary information files. Further, all relative search volume data points can be reproduced through Google Trends.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary Table 1.
[XLSX File, 18 KB - publichealth_v11i1e58630_app1.xlsx]

References

1. Global HIV & AIDS statistics — fact sheet. UNAIDS. URL: https://www.unaids.org/en/resources/fact-sheet [accessed 2024-12-31]
2. Indicators and questions for monitoring progress on the 2021 Political Declaration on HIV and AIDS — Global AIDS Monitoring 2025. UNAIDS. 2024. URL: https://www.unaids.org/en/resources/documents/2024/global-aids-monitoring-guidelines [accessed 2024-12-31]
3. Neal JJ, Prybylski D, Sanchez T, Hladik W. Population size estimation methods: searching for the holy grail. JMIR Public Health Surveill 2020 Dec 3;6(4):e25076. [doi: 10.2196/25076] [Medline: 33270035]
4. Stone J, Mukandavire C, Boily MC, et al. Estimating the contribution of key populations towards HIV transmission in South Africa. J Int AIDS Soc 2021 Jan;24(1):e25650. [doi: 10.1002/jia2.25650] [Medline: 33533115]
5. Viswasam N, Lyons CE, MacAllister J, et al. The uptake of population size estimation studies for key populations in guiding HIV responses on the African continent. PLoS One 2020;15(2):e0228634. [doi: 10.1371/journal.pone.0228634] [Medline: 32101551]
6. Xu C, Jing F, Lu Y, et al. Summarizing methods for estimating population size for key populations: a global scoping review for human immunodeficiency virus research. AIDS Res Ther 2022 Feb 19;19(1):9. [doi: 10.1186/s12981-022-00434-7] [Medline: 35183203]
7. Teo AKJ, Prem K, Chen MIC, et al. Estimating the size of key populations for HIV in Singapore using the network scale-up method. Sex Transm Infect 2019 Dec;95(8):602-607. [doi: 10.1136/sextrans-2018-053747] [Medline: 31073092]
8. Sabin K, Zhao J, Garcia Calleja JM, et al. Availability and quality of size estimations of female sex workers, men who have sex with men, people who inject drugs and transgender women in low- and middle-income countries. PLOS ONE 2016;11(5):e0155150. [doi: 10.1371/journal.pone.0155150] [Medline: 27163256]
9. Edwards JK, Hileman S, Donastorg Y, et al. Estimating sizes of key populations at the national level: considerations for study design and analysis. Epidemiology 2018 Nov;29(6):795-803. [doi: 10.1097/EDE.0000000000000906] [Medline: 30119057]
10. Bozicevic I, Sharifi H, Haghdoost A, Sabry A, Hermez J. Availability of HIV surveillance data in key populations in the countries of the World Health Organization Eastern Mediterranean Region. Int J Infect Dis 2022 Aug;121:211-216. [doi: 10.1016/j.ijid.2022.05.003] [Medline: 35533830]
11. Mavragani A, Ochoa G. Google Trends in infodemiology and infoveillance: methodology framework. JMIR Public Health Surveill 2019 May 29;5(2):e13439. [doi: 10.2196/13439] [Medline: 31144671]

12. Card KG, Lachowsky NJ, Hogg RS. Using Google Trends to inform the population size estimation and spatial distribution of gay, bisexual, and other men who have sex with men: proof-of-concept study. JMIR Public Health Surveill 2021 Nov 29;7(11):e27385. [doi: 10.2196/27385] [Medline: 34618679]
13. Porn, gay porn - explore. Google Trends. URL: https://trends.google.com/trends/explore?date=2020-01-01%202020-12-31&geo=DZ&q=Porn,Gay%20Porn [accessed 2024-12-31]
14. FAQ about google trends data - trends help. Google. URL: https://support.google.com/trends/answer/4365533?hl=en [accessed 2024-12-31]
15. How many women watch porn? Fight the New Drug. URL: https://fightthenewdrug.org/how-do-men-and-womens-porn-site-searches-differ [accessed 2024-12-31]
16. Pornhub stats reveal women search for more hardcore genres than you might expect. Fight the New Drug. URL: https://fightthenewdrug.org/data-reveals-women-are-searching-hardcore-genres [accessed 2024-12-31]
17. When no one is looking, many women are watching gay porn. NBC News. URL: https://www.nbcnews.com/feature/nbc-out/when-no-one-looking-many-women-are-watching-gay-porn-n894266 [accessed 2024-12-31]
18. Solano I, Eaton NR, O'Leary KD. Pornography consumption, modality and function in a large internet sample. J Sex Res 2020 Jan 2;57(1):92-103. [doi: 10.1080/00224499.2018.1532488]
19. Træen B, Daneback K. The use of pornography and sexual behaviour among Norwegian men and women of differing sexual orientation. Sexologies 2013 Apr;22(2):e41-e48. [doi: 10.1016/j.sexol.2012.03.001]
20. PEPFAR 2022 country and regional operational plan (COP/ROP) guidance for all PEPFAR-supported countries. US Department of State.: PEPFAR; 2022. URL: https://www.state.gov/wp-content/uploads/2022/02/COP22-Guidance-Final_508-Compliant-3.pdf [accessed 2025-01-02]
21. Dangerous inequalities: world AIDS day report 2022. UNAIDS. 2022. URL: https://www.unaids.org/sites/default/files/media_asset/dangerous-inequalities_en.pdf [accessed 2025-01-02]
22. Quick start guide for Spectrum. UNAIDS. 2020. URL: https://www.unaids.org/sites/default/files/media_asset/QuickStartGuide_Spectrum_en.pdf [accessed 2025-01-02]
23. World Health Organization. Recommended population size estimates of men who have sex with men. UNAIDS. 2020. URL: https://www.unaids.org/sites/default/files/media_asset/2020-recommended-population-size-estimates-of-men-who-have-sex-with-men_en.pdf [accessed 2025-01-02]
24. The Key Populations Atlas. UNAIDS. 2022. URL: https://kpatlas.unaids.org/dashboard [accessed 2024-12-31]
25. Key Populations Atlas data sources. UNAIDS - Key Populations Atlas. 2017. URL: https://kpatlas.unaids.org/document/kp_data_sources.pdf [accessed 2025-01-02]
26. Internet world stats - usage and population statistics. ResearchGate. URL: https://www.researchgate.net/publication/258847361_Internet_World_Stats_-_Usage_and_Population_Statistics [accessed 2025-01-02]
27. World development indicators. World Bank Group. URL: https://databank.worldbank.org/source/world-development-indicators [accessed 2024-12-31]
28. Mendos LR, Botha K, Lelis RC, de la Peña EL, Savelev I, Tan D. State-Sponsored Homophobia report. ILGA World. 2020. URL: https://ilga.org/state-sponsored-homophobia-report/ [accessed 2025-01-02]
29. Most searched things on Google in 2024. SimilarWeb. URL: https://www.similarweb.com/blog/marketing/seo/top-keywords/ [accessed 2024-12-31]
30. Carneiro HA, Mylonakis E. Google trends: a web-based tool for real-time surveillance of disease outbreaks. Clin Infect Dis 2009 Nov 15;49(10):1557-1564. [doi: 10.1086/630200] [Medline: 19845471]
31. Havelka EM, Mallen CD, Shepherd TA. Using Google Trends to assess the impact of global public health days on online health information seeking behaviour in Central and South America. J Glob Health 2020 Jun;10(1):010403. [doi: 10.7189/jogh.10.010403] [Medline: 32373327]
32. Arora VS, McKee M, Stuckler D. Google Trends: opportunities and limitations in health and health policy research. Health Policy 2019 Mar;123(3):338-341. [doi: 10.1016/j.healthpol.2019.01.001] [Medline: 30660346]
33. Azzam DB, Nag N, Tran J, et al. A novel epidemiological approach to geographically mapping population dry eye disease in the United States through Google Trends. Cornea 2021 Mar 1;40(3):282-291. [doi: 10.1097/ICO.0000000000002579] [Medline: 33177410]
34. NordVPN survey reveals: users still trust free vpns. NordVPN. URL: https://nordvpn.com/blog/nordvpn-usage-survey/ [accessed 2025-01-02]

Abbreviations

GAM: Global AIDS Monitoring system
GT: Google Trend
GTPSE: Google Trend Population Size Estimate
KP: key population
LMIC: low- and middle-income country
MSM: men who have sex with men

PSE: Population size estimation
PSE: population size estimation
RSV: relative search volume
SSA: sub-Saharan Africa
UNAIDS: United Nations Programme on HIV/AIDS
VPN: virtual private network
WHO: World Health Organization

Edited by A Mavragani; submitted 20.03.24; peer-reviewed by E Tuyishime, K Card; revised version received 28.05.24; accepted 28.05.24; published 09.01.25.

Please cite as:
Malburg CM, Gutreuter S, Ruiseñor-Escudero H, Abdul-Quader A, Hladik W
Population Size Estimation of Men Who Have Sex With Men in Low- and Middle-Income Countries: Google Trends Analysis
JMIR Public Health Surveill 2025;11:e58630
URL: https://publichealth.jmir.org/2025/1/e58630
doi:10.2196/58630

Strategies to Increase Response Rate and Reduce Nonresponse Bias in Population Health Research: Analysis of a Series of Randomized Controlled Experiments during a Large COVID-19 Study

Christina J Atchison[1], PhD; Nicholas Gilby[2], BA; Galini Pantelidou[2], MSc; Sam Clemens[2], BA; Kevin Pickering[2], MSc; Marc Chadeau-Hyam[1], PhD; Deborah Ashby[1], PhD; Wendy S Barclay[3], PhD; Graham S Cooke[3], PhD; Ara Darzi[4], MD; Steven Riley[1], PhD; Christl A Donnelly[1,5], ScD; Helen Ward[1], PhD; Paul Elliott[1], PhD

[1]School of Public Health, Imperial College London, London, United Kingdom

[2]Ipsos, London, United Kingdom

[3]Department of Infectious Disease, Imperial College London, Norfolk Place, London, United Kingdom

[4]Institute of Global Health Innovation, Imperial College London, South Kensington Campus, London, United Kingdom

[5]Department of Statistics, University of Oxford, Oxford, United Kingdom

Corresponding Author:
Paul Elliott, PhD
School of Public Health, Imperial College London, London, United Kingdom

Abstract

Background: High response rates are needed in population-based studies, as nonresponse reduces effective sample size and bias affects accuracy and decreases the generalizability of the study findings.

Objective: We tested different strategies to improve response rate and reduce nonresponse bias in a national population–based COVID-19 surveillance program in England, United Kingdom.

Methods: Over 19 rounds, a random sample of individuals aged 5 years and older from the general population in England were invited by mail to complete a web-based questionnaire and return a swab for SARS-CoV-2 testing. We carried out several nested randomized controlled experiments to measure the impact on response rates of different interventions, including (1) variations in invitation and reminder letters and SMS text messages and (2) the offer of a conditional monetary incentive to return a swab, reporting absolute changes in response and relative response rate (95% CIs).

Results: Monetary incentives increased the response rate (completed swabs returned as a proportion of the number of individuals invited) across all age groups, sex at birth, and area deprivation with the biggest increase among the lowest responders, namely teenagers and young adults and those living in more deprived areas. With no monetary incentive, the response rate was 3.4% in participants aged 18 - 22 years, increasing to 8.1% with a £10 (US $12.5) incentive, 11.9% with £20 (US $25.0), and 18.2% with £30 (US $37.5) (relative response rate 2.4 [95% CI 2.0-2.9], 3.5 [95% CI 3.0-4.2], and 5.4 [95% CI 4.4-6.7], respectively). Nonmonetary strategies had a modest, if any, impact on response rate. The largest effect was observed for sending an additional swab reminder (SMS text message or email). For example, those receiving an additional SMS text message were more likely to return a completed swab compared to those receiving the standard email-SMS approach, 73.3% versus 70.2%: percentage difference 3.1% (95% CI 2.2%-4.0%).

Conclusions: Conditional monetary incentives improved response rates to a web-based survey, which required the return of a swab test, particularly for younger age groups. Used in a selective way, incentives may be an effective strategy for improving sample response and representativeness in population-based studies.

(JMIR Public Health Surveill 2025;11:e60022) doi:10.2196/60022

KEYWORDS

study recruitment; response rate; population-based research; COVID-19; SARS-CoV-2; web-based questionnaires

Introduction

In population-based studies, a high response rate from a representative sample may reduce nonparticipation bias, increase the generalizability, and improve the accuracy of study estimates [1]. However, achieving this goal is challenging, both due to the difficulty in contacting and then engaging eligible participants [2]. For example, UK Biobank, a population-based cohort study with stored biological samples from half a million participants aged 40 - 69 years in the United Kingdom, achieved an overall response rate of 5.5% [3], which was lower in men, younger people, and those living in more deprived areas [4]. The impact of nonresponse and nonrepresentativeness on the generalizability of disease prevalence and incidence rates in the UK Biobank has been widely debated [5,6].

It is important to address low or falling response rates to reduce the likelihood of systematic biases that may affect study estimates [7]. While weighting is commonly applied to correct for differential participation, it may fail to correct bias if the responders in a particular subgroup of the population are not representative of that subgroup as a whole. Furthermore, weighting to correct for observed biases worsens precision (reducing the effective sample size) [8].

Systematic reviews that have evaluated interventions to increase response rates in surveys have concluded that monetary incentives are more effective than nonmonetary incentives [9-11], although findings were inconsistent concerning web-based surveys in educational research [12]. Some studies have found incentives can increase response among under-represented sociodemographic groups, such as those with low incomes, those with low education, single parents, and minority ethnic groups, potentially reducing nonresponse bias [13], while others show mixed results [9].

Other strategies that have been shown to improve response rates in surveys have included the use of SMS text message reminders to enhance the contact method of letters and emails [14,15], using alternative motivational statements in invitation letters [16], and changing the font color of text [17]. In a United Kingdom–based study investigating the effects of augmenting the contact strategy of letters and emails with SMS text messages for a web questionnaire, the findings indicated that SMS text messages did not help to significantly increase response rates overall, although some subgroups benefited from them, such as younger panel members and those with an irregular response pattern [15].

The Real-time Assessment of Community Transmission-1 (REACT-1) study was one of the largest population surveillance studies in the world. Across 19 rounds between May 1, 2020 and March 31, 2022, it provided timely prevalence estimates of SARS-CoV-2, the virus that causes COVID-19, from random cross-sectional samples of the population in England [18,19].

Response rate varied between 11.7% and 30.5% and, like in many population surveys, varied across demographic groups [19]. For financial reasons, we could not issue more than 845,000 invitation letters by mail, so we could only achieve the minimum desired sample size adopted from round 12 (May 20 to June 7, 2021) of 100,000 by improving response [18]. The observed nonresponse biases meant REACT-1 was under-representing groups with lower vaccination rates and where COVID-19 prevalence was highest; thus, we were likely underestimating the true population prevalence despite our attempts to correct for such biases by use of weighting on known demographic variables [20,21]. Here we present results of experiments nested within the REACT-1 study to test the effectiveness of different strategies to increase response rates and participation of groups with a lower propensity to take part.

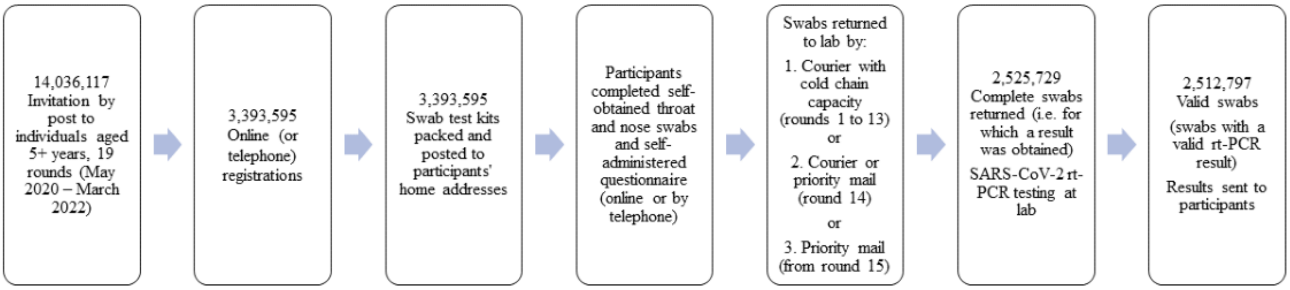
Methods

The REACT-1 Study

Methods for the study, including sample size calculations, are described in detail elsewhere [18,19]. In summary, at approximately monthly intervals, between 395,020 and 841,227 people were sent personalized invitations by mail to take part. For children (5-17 years old), the invitation was sent to or via the parents or guardians. Individuals aged 5 years and older were randomly sampled from the National Health Service (NHS) list of patients in England (with near-universal population coverage) across all 316 Lower Tier Local Authorities [18,19]. This list includes the name, address, date of birth, and sex of everyone registered with a general practitioner in England. Invitees who registered (most digitally, some by telephone) for the study received a kit by mail with instructions on how to take a throat and nose swab and send it for SARS-CoV-2 testing using reverse transcriptase polymerase chain reaction (rt-PCR). Swabs were transferred to laboratories for processing, initially being picked up by courier with cold chain capacity (rounds 1-13 and part of 14) or sent by priority mail (part of round 14 and subsequent rounds). Participants were also asked to complete a self-administered web-based or telephone questionnaire [18,19].

Over the 19 study rounds, we sent out 14,036,117 invitations, 3,393,595 registrations were made, and 2,525,729 completed swabs were returned (ie, for which a laboratory result was obtained) (Figure 1). Of these swabs, 2,512,797 (99.5% of completed swabs returned) were considered valid for analysis in REACT-1 (swabs with a valid rt-PCR result) [19]. A swab with a valid rt-PCR result was a swab for which a “cycle threshold” (Ct) value could be obtained. Therefore, not all swabs tested by the lab were considered valid. Overall, 12,932 (0.5% of completed swabs returned) were considered invalid and rejected. Reasons included inadequate sample volume, contamination during sample collection, inappropriate sample storage, or inappropriate sample transportation. All analyses in this paper are based on completed swabs returned (ie, for which a laboratory result was obtained, n=2,525,729), thus including swabs for which a Ct value could not be obtained but excluding swabs returned unused.

Figure 1. REACT-1 study process over 19 rounds of data collection: England, May 1, 2020 to March 31, 2022. Overall, across 19 rounds, we report the number of invitations sent, the number of participants registered, the number of swab test kits sent out, the number of completed swabs returned (ie, for which a laboratory result was obtained) and the number of valid swabs (swabs with a valid rt-PCR result). REACT-1: REal-time Assessment of Community Transmission-1; rt-PCR: reverse transcriptase Polymerase Chain Reaction; SARS-CoV-2: Severe Acute Respiratory Syndrome CoronaVirus 2.



All the experiments carried out to improve response rate were randomized trials, enabling an unbiased assessment of the impact of the changed survey procedure compared to a control group. Due to funding constraints, initial experiments focused on approaches which would not materially affect the survey budget, before turning to an experiment with monetary incentives.

Swab Reminder and Tailored Letter or SMS Experiments

In each round of REACT-1, those registering for a swab test were, where necessary, sent at least one reminder to complete the swab test and return it, to maximize the number of swabs returned. In round 3 (July 24 to August 11, 2020) we conducted an experiment to establish the optimal use of email and SMS text message swab return reminders, with participants randomly allocated to the experimental conditions (Table 1).

Table . Round 3 swab reminder experimental conditions, England, July 24 to August 11, 2020.

Condition	Reminder on day 4 after swab test kit received	Reminder on day 6 after swab test kit received	Reminder on day 8 after swab test kit received	Sample, n
Control group	Email (SMS if no email address)	SMS	None	11,194
Experimental group A	SMS	Email (SMS if no email address)	None	11,154
Experimental group B	Email (SMS if no email address)	SMS	Email (SMS if no email address)	96,337
Experimental group C	SMS	Email (SMS if no email address)	SMS	96,305

The tailored letters or SMS experiments are summarized in Table 2. Further details are available in Multimedia Appendix 1. The experiments tested whether it was possible to increase participation by different types of conditions (Textbox 1).

Table . Rounds 9, 11, and 12 registration invitation letter experimental conditions and rounds 10 and 12 SMS registration reminder experimental conditions, England, February 4 to June 7, 2021.

Age and letter or SMS type	Additional content for experiment (actual additional content used in bold text)	Sample, n
Round 9[a] (≥70 years)		
Standard invitation letter Adult	None	37,037
Experiment invitation letter A	"**It is still important to take part in this study if you have received a vaccination from COVID-19 or expect to be vaccinated in the near future**. Your participation will help DHSC assess the impact of the vaccines on COVID-19 infection rates." As well as a new sub-heading "**COVID-19 Testing Study: Take part to help measure COVID-19 infection rates among those aged 70 and over**."	37,037
Experiment invitation letter B	As per Experiment Letter A with additional line "**Older people are a vulnerable group, so we need your help to monitor prevalence**. It is still important to take part in this study if you have received a vaccination from COVID-19 or expect to be vaccinated in the near future."	37,036
Round 9 (5 - 12 years)		
Standard invitation letter Child (addressed to parent)	None	24,009
Experiment invitation letter C	**"We need to know how many children and young people have COVID-19, and how easily the new variant spreads amongst them."**	24,009
Experiment invitation letter D	As per Experiment Letter C with new sub-heading **"COVID-19 Testing Study: Take part to help measure how easily COVID-19 spreads among children and young people."**	24,008
Round 9 (all)		
Standard registration reminder letter	Blue text used	306,012
Experiment registration reminder letter E	Red text used	305,041
Round 11[b] (≥18 years)		
Standard invitation reminder letter	None	178,828
Experiment invitation reminder letter A	New content asking participants to take a test to help prevent the spread of COVID-19 and explaining that taking part would help the Government work out the best way to manage the pandemic. Also mentioned testing for new variants, that the study compared people who had been vaccinated with those who had not, and that taking part would help inform the vaccine strategy and help to avoid lockdowns.	178,809
Round 12[c] (all)		
Standard invitation final reminder letter	Double-sided	169,845
Shorter invitation final reminder letter	Single-sided	342,191
Round 10[d] (all)		
Standard SMS first reminder	Unchanged **"The study is closing soon, please register by 18 March if you want to take part."**	50,000

Age and letter or SMS type	Additional content for experiment (actual additional content used in bold text)	Sample, n
Experiment first SMS reminder	New SMS content “**Taking part will help inform decisions about the best time to lift restrictions**.”	430,283
Round 11 (all)		
Standard SMS second reminder	Unchanged “**The study is closing soon, please register by 3pm on 22 April if you want to take part**.”	127,028
Experiment second SMS reminder	New SMS content “**Taking part will help monitor infection rates and new variants of the virus**.”	127,028
Round 12 (all)		
Standard SMS first reminder	Unchanged “**Taking part will help inform decisions about the best time to lift restrictions**.”	321,042
Experiment first SMS reminder	New SMS Content **“Taking part will help monitor infection rates and new variants of the virus.”**	155,683
Standard SMS second reminder	Unchanged **“Taking part will help monitor infection rates and new variants of the virus.”**	272,836
Experiment second SMS reminder	New SMS Content **“Last chance to help monitor variants in your area.”**	136,026

[a]Round 9 (Feb 4-23, 2021)
[b]Round 11 (Apr 15 to May 3, 2021)
[c]Round 12 (May 20 to Jun 7, 2021)
[d]Round 10 (Mar 11-30, 2021)

Textbox 1. Types of conditions tested in the tailored letters or SMS experiments.

- Using additional content in the invitation letter, tailored for the oldest and youngest age groups.
- Using color, additional content, and varying the length of the reminder letter.
- Using additional content in the SMS reminder.

Incentives Experiment

In round 15 (October 19 to Nov 5, 2021), conditional incentives (£10 [US $12.5], £20 [US $25.0], or £30 [US $37.5] gift vouchers for returning a completed swab test) were tested in a randomized controlled trial for all age groups except 5‐ to 12-year-olds.

The process for obtaining consent in REACT-1 for children was undertaken differently based on participant age at the time of the invitation [18]. For 5‐ to 12-year-olds, the parent or guardian was contacted via letter and asked to consent on behalf of the child. Therefore, we did not include the 5‐ to 12-year-olds in the trial, as the sampled child would not be making the decision to take the swab test, and their parent would be incentivized, raising ethical and reputational concerns. For 13‐ to 17-year-olds, the parent or guardian received a letter addressed to them, asking them to pass on an enclosed invitation letter addressed to their sampled child if they agreed for their child to take part in the study. As such, children aged 13‐17 years were able to decide whether to consent to the study and take the swab test. In addition, those aged 13 to 15 years were asked at registration to confirm the name of the parent or guardian who had given them permission to take part. This was not required for those aged 16‐17 years, as in UK health research, the Health Research Authority states that young people over 16 are presumed capable of giving consent on their own behalf [22].

Participants were randomly allocated to experimental and control groups: (1) £10 (US $12.5) conditional incentive (n=10,900), (2) £20 (US $25.0) conditional incentive (n=10,900), (3) £30 (US $37.5) conditional incentive (only for 18‐ to 32-year-olds) (n=1750), and (4) control group (n=23,500). Further details of the sample size calculations are available in Multimedia Appendix 1.

The £30 (US $37.5) incentive was limited to the 18- to 32-year-olds because the response rate in REACT-1 was lowest among this age group. Also, there is evidence that incentives can be more effective among younger age groups [23]. Those in this age group were of particular interest as they were less likely to be vaccinated, had more social contact (and therefore were more likely to be at risk of infection), and had been particularly impacted by the pandemic (in terms of well-being, education, and employment) [21,24]. It was decided to test offering a larger (£30 [US $37.5]) incentive to this age group (and not the other age groups) to overcome their higher reluctance to take part and to better represent this group in the achieved sample. Based on the same rationale, we oversampled younger age groups to maximize the statistical power we had to detect an increase in the response rate due to the use of incentives among these groups.

The primary outcome was overall swab response rate, ie, the number of completed swabs returned (referred to as swabs returned forthwith) as a proportion of the number of invitations sent. For those invited, we knew age, sex at birth, and score from an area-level index of multiple deprivation, the Index of Multiple Deprivation (IMD) 2019 [25]. Participants were classified by quintiles of the deprivation score based on their residential postcode.

We used COVID-19 vaccination status (the proportion who had received at least one vaccine dose) as a proxy for attitudes to health behaviors and health care access, hypothesizing that REACT-1 responders would be more likely to be vaccinated than those who did not, indicating a responder bias. Thus, the difference in vaccination status at registration between the experimental and control groups was used as a crude indicator of how incentives might improve response rates in individuals less likely to participate in research, beyond sociodemographic characteristics. We also compared the COVID-19 vaccination status of those who returned a swab with the achieved population vaccination rate for that age group as a whole. To obtain information on dates of received COVID-19 vaccine doses, participant study data were linked to their NHS records from NHS Digital (now NHS England) on COVID-19 vaccination events [26] using their unique NHS number and other personal identifiers. This was only possible for study participants who had consented to data linkage. The source of vaccination data for the population vaccination rates was the NHS National Immunization Management System [27].

Ethical Considerations

The study was ethically approved by the South Central-Berkshire B Research Ethics Committee (IRAS ID: 283787). Participants provided informed consent when they registered for the study, and all data were handled securely in accordance with a detailed privacy notice. Collected data were deidentified; the data used in this study were anonymous and did not contain any personally identifiable information. Participants had the ability to opt out anytime during the research period. The study did not provide any specific compensation other than the monetary gift vouchers for returning a completed swab test as set out in the study's incentives experiment described above.

Statistical Analysis

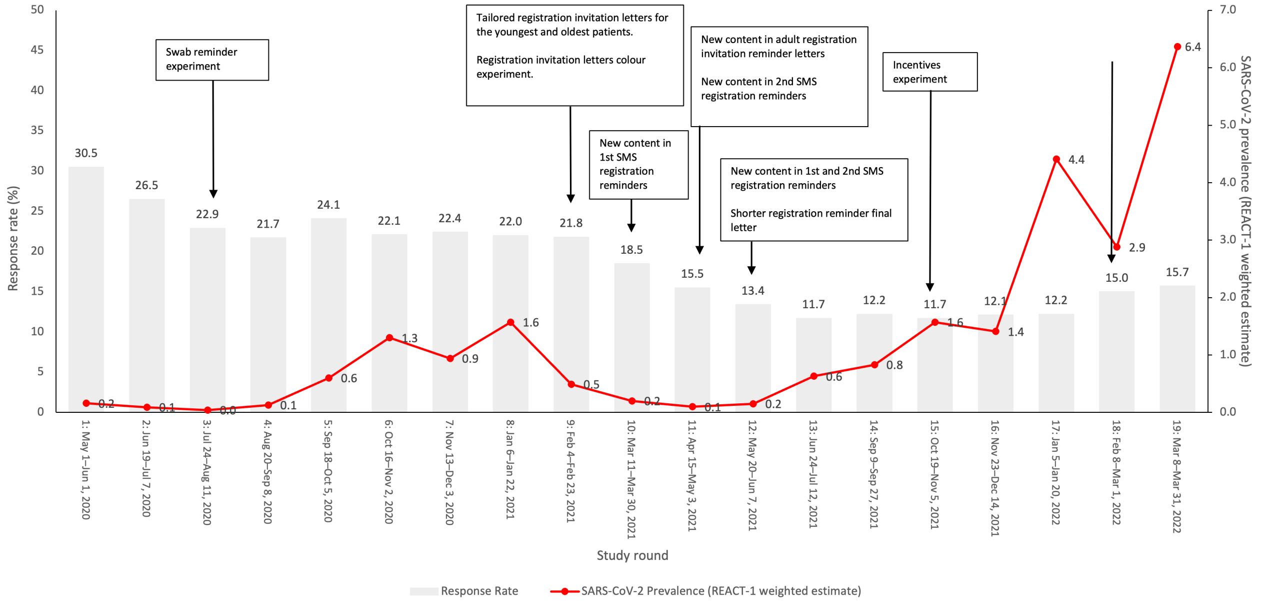
Data analysis was conducted using IBM SPSS Statistics (version 28). As the incentives experiment was skewed toward younger age groups, swab response rates for sex at birth and area deprivation (IMD) were calculated with age-standardized weighting using 2021-based population estimates for England [28]. The percentage point difference (95% CI) and independent 2-tailed *t*-tests were used to show the absolute difference in swab response rates between the experimental and control groups and were also used to show the absolute difference in vaccination rates at registration between the experimental and control groups. Using multivariable logistic regression, we tested the impact of each of the incentive conditions on swab response rate by age, sex at birth, and area deprivation (relative response rate [RRR] with 95% CI). The reference group was the no-incentive condition—eg, the response rates for females in the £10 (US $12.5), £20 (US $25.0), and £30 (US $37.5) incentives groups were compared to females in the no-incentive group (£0 [US $0.0]). We tested interaction terms for age, sex at birth, and area deprivation by incentive (incentive*age, incentive*sex at birth, and incentive*IMD), which can be interpreted as testing whether the estimated effects of incentives on swab response rates differ by each of these 3 covariates.

Results

Overview

Overall, 24.2% (3,393,595/14,036,117) of invitees registered for the study, and 74% (2,512,797/3,393,595) of those registered returned valid swabs, giving an overall response rate for the REACT-1 study (number of valid swabs/number of invitations) of 17.9% (2,512,797/14,036,117) [19]. Whilst the rate at which registered participants returned valid swabs remained relatively stable across rounds (range 67.2%-78.9%), response rates varied more widely, ranging from 11.7% in rounds 13 (98,233/841,227) (June 24 to July 12, 2021) and 15 (100,112/859,184) (October 19 to November 5, 2021) to 30.5% in round 1 (120,620/395,020) (May 1 to June 1, 2020, during the first lockdown in England) (Figure 2). The following groups were relatively underrepresented: younger people, men, ethnic minorities, and those living in the most deprived areas (comparing achieved sample profiles with population profiles) (Table S1 in Multimedia Appendix 2).

Figure 2. REACT-1 study timeline over 19 rounds of data collection showing response rates, SARS-CoV-2 prevalence (weighted), and timing of experiments to improve response. England, May 1, 2020 to March 31, 2022. REACT-1 Response Rate: number of valid swabs returned/number of invitations. We report weighted SARS-CoV-2 swab-positivity prevalence for individuals aged 5 years and older from all rounds of the REACT-1 study. REACT-1: REal-time Assessment of Community Transmission-1; SARS-CoV-2: Severe Acute Respiratory Syndrome CoronaVirus 2.



Swab Reminder and Tailored Letter or SMS Experiments

Table S2 in Multimedia Appendix 2 summarizes the results of the swab reminder and tailored letter or SMS experiments. Sending an additional reminder (email or SMS) to those who registered resulted in a small increase in response rate: those receiving a third swab reminder (experimental groups B and C) were more likely to return a completed swab compared to those receiving the standard Email-SMS approach (group B vs control: 73% vs 70.2%, percentage difference 2.8% [95% CI 1.9%-3.7%]; group C vs control 73.3% vs 70.2%, percentage difference 3.1% [95% CI 2.2%-4%]).

In round 9 (February 4-23, 2021), both experimental invitation letters A and B had a small but positive impact on response rate in participants aged ≥70 years of 0.9% (95% CI 0.2%-1.5%) and 1.2% (95% CI 0.6%-1.8%) percentage difference, respectively, compared to the standard invitation letter. For participants aged 5 - 12 years, experiment letter C generated a slightly higher response rate compared to the standard letter (16.6% vs 15.9%; percentage difference 0.7% (95% CI 0.1%-1.4%)). In round 11 (April 15 to May 3, 2021) and round 12 (May 20 to June 7, 2021), the experimental invitation reminder letters had a small positive impact on response rate compared to the standard letters: round 11 (new content), 5.6% vs 5.4%, percentage difference 0.2% (95% CI 0%-0.3%); round 12 (shorter), 2.3% vs 1.6%, percentage difference 0.8% (95% CI 0.7%-0.8%). We saw no effect on response rate for any of the other nonmonetary strategies (Table S2 in Multimedia Appendix 2).

Incentives Experiment

The conditional monetary incentives increased the response rate across all age groups but were particularly effective among the lowest responding groups, those aged 13 - 17 years and 18 - 22 years (Figure 3 and Tables S3 and S4 in Multimedia Appendix 2). Table 3 shows the RRR for each incentive level by age, sex at birth, and area deprivation. The higher the monetary value of the incentive, the higher the response rate. For example, in participants aged 18 - 22 years, the response rate in the control group was 3.4% (95% CI 2.9%-3.8%), increasing to 8.1% (95% CI 7.0%-9.2%), 11.9% (95% CI 10.6%-13.2%), and 18.2% (95% CI 15.4%-21.1%) with £10 (US $12.5), £20 (US $25.0), and £30 (US $37.5) incentives, respectively. The largest relative increase was with the £30 (US $37.5) incentive in 18 - to 22-year-olds (RRR 5.4 [95% CI 4.4-6.7]) (Table 3). All incentive conditions led to a greater increase in response rate in younger age groups. The £20 (US $25.0) incentive led to a greater increase in the more deprived areas, RRR 2.7 (95% CI 2.2-3.3) for the most deprived quintile and RRR 1.8 (95% CI 1.6-2.1) for the least deprived.

Figure 3. Swab response rates and 95% CIs for the intervention and control groups in the incentives experiment in round 15, England, October 19 to November 5, 2021. Note: Participants randomly allocated to experimental and control groups. (1) £10 (US $12.5) conditional incentive (n=10,900), (2) £20 (US $25.0) conditional incentive (n=10,900), (3) £30 (US $37.5) conditional incentive (only for 18- to 32-year-olds) (n=1750), and (4) control group (n=23,500).

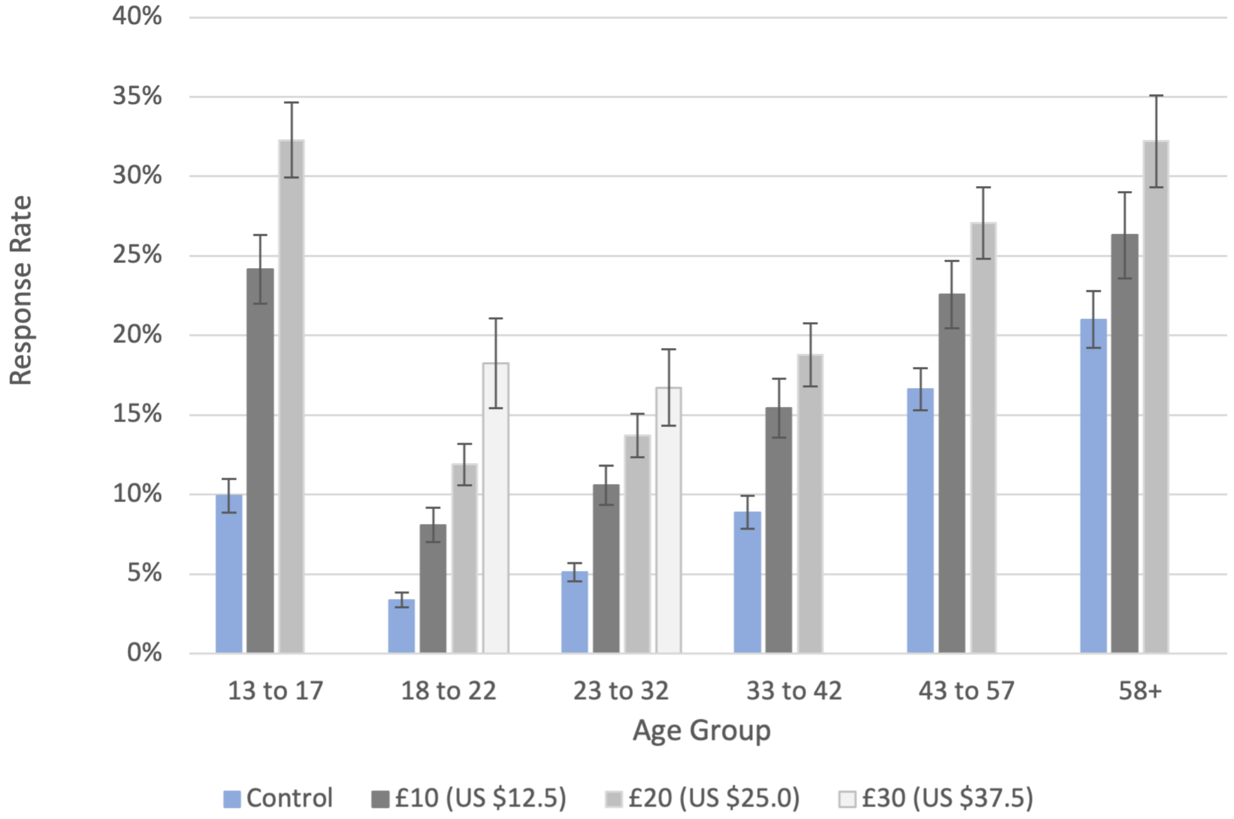


Table . Variation in relative response rates (RRR) and 95% CI for the interventions compared to the control group by age, sex at birth, and area deprivation (IMD[a]) in the incentives experiment in round 15, England, October 19 to November 5, 2021.

	£10 (US $12.5) RRR (95% CI)	£20 (US $25.0) RRR (95% CI)	£30 (US $37.5) RRR (95% CI)
Age[b] (years)			
13-17	2.4 (2.1-2.8)	3.3 (2.6-3.7)	N/A
18-22	2.4 (2.0-2.9)	3.5 (3.0-4.2)	5.4 (4.4-6.7)
23-32	2.1 (1.7-2.4)	2.7 (2.3-3.1)	3.3 (2.7-3.9)
33-42	1.7 (1.5-2.0)	2.1 (1.8-2.5)	—[c]
43-57	1.4 (1.2-1.5)	1.6 (1.5-1.8)	—
58+	1.3 (1.1-1.4)	1.5 (1.4-1.7)	—
P value for interaction between incentive and age	<.001	<.001	<.001
Sex at birth[bd]			
Male	1.4 (1.3-1.6)	1.8 (1.7-2.1)	3.7 (3.0-4.7)
Female	1.5 (1.4-1.6)	1.8 (1.6-2.0)	3.7 (3.1-4.4)
P value for interaction between incentive and sex at birth	.37	.96	.68
IMD[bd]			
1—most deprived	1.8 (1.5-2.3)	2.7 (2.2-3.3)	4.8 (3.3-7.0)
2	1.6 (1.3-1.9)	1.9 (1.6-2.3)	4.0 (2.9-5.5)
3	1.4 (1.2-1.6)	1.5 (1.3-1.8)	3.4 (2.5-4.8)
4	1.3 (1.1-1.5)	1.7 (1.5-1.9)	3.2 (2.4-4.3)
5—least deprived	1.5 (1.3-1.7)	1.8 (1.6-2.1)	3.5 (2.6-4.8)
P value for interaction between incentive and IMD	.38	.01	0.72

[a]IMD: Index of Multiple Deprivation

[b]Reference group, ie, the reference group for each row is the no incentive condition. For example, the RRR for female £10 (US $12.5), female £20 (US $25.0), and female £30 (US $37.5) is versus female £0 (US $0.0). *P* value for main effect of incentive on response rate for all row comparisons <.001.

[c]Not applicable

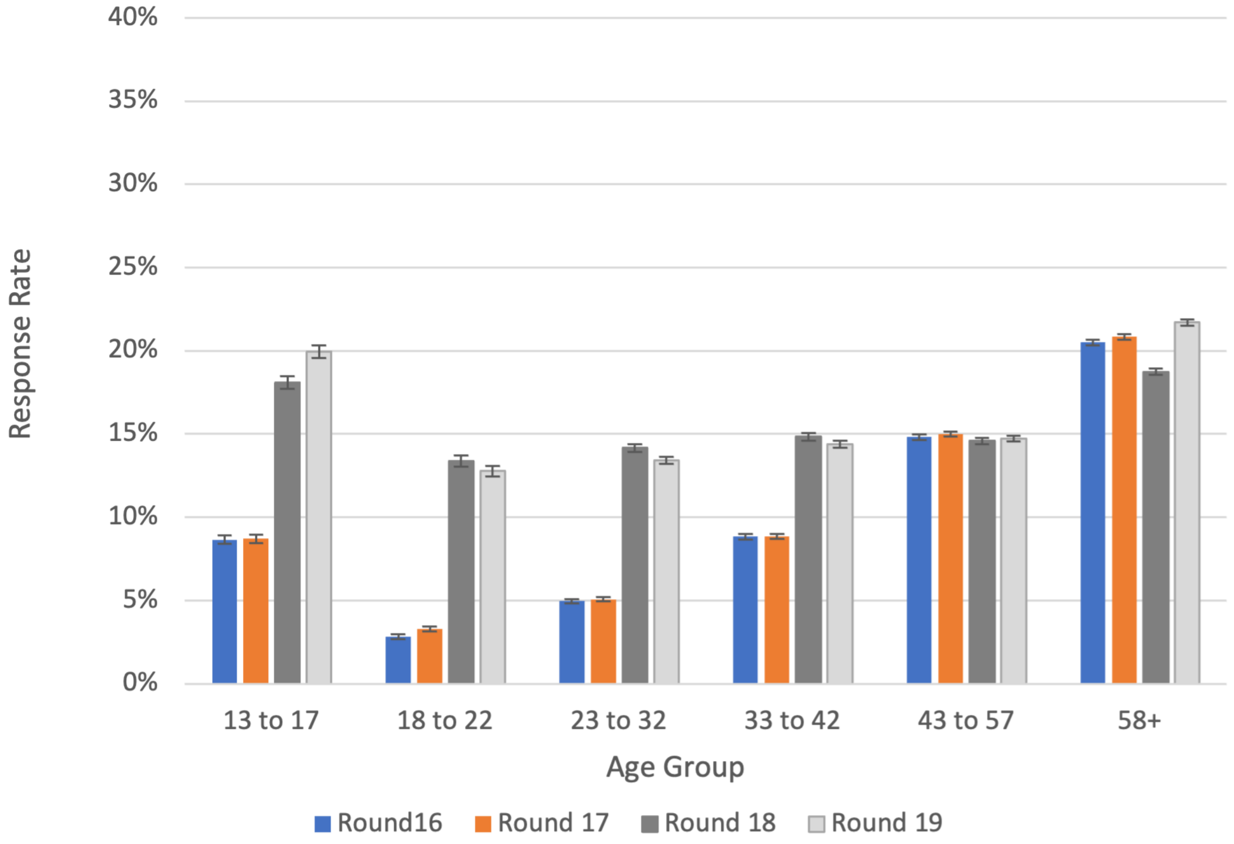
[d]Age-standardized weighting applied to calculate swab response rate with the control group totals used as the sample profiles.

Following the results of the selective use of incentives in round 15, they were introduced in rounds 18 (Feb 8-Mar 1, 2022) and 19 (Mar 8-Mar 31, 2022). For returning their completed test, those aged 13 - 17 and 35 - 44 years were offered a gift voucher worth £10 (US $12.5), while those aged 18 - 34 years were offered a voucher worth £20 (US $25.0). In these final 2 rounds, this had the effect of increasing the swab response rate in these groups and was associated with less variation in response rate by age (Figure 4), suggesting that the selective use of incentives reduced participation bias by age.

Table S5 in Multimedia Appendix 2 shows the effective sample sizes and sample efficiency for each round of REACT-1. The effective sample size measures the size of a (unweighted) simple random sample that would achieve the same precision (standard error) as the design used. The efficiency of a sample is given by the ratio of the effective sample size to the actual sample size. Rounds 18 and 19, where selective use of incentives was used, saw the fourth and second highest (respectively) effective sample sizes of any REACT-1 round, and the highest sample efficiency for any REACT-1 round.

Overall, vaccination rates were higher in REACT-1 participants than in the general population (Tables S6 and S7 in Multimedia Appendix 2). For example, by October 24, 2021, just over 3 quarters of 18 - to 22-year-olds had received at least one vaccine dose nationally (Table S6 in Multimedia Appendix 2) [29], lower than the 84.0% (95% CI 78.1%-88.6%) in the round 15 (October 19-November 5, 2021) control group for that age (Table S7 in Multimedia Appendix 2). With the incentives that proportion declined to 82.1% (95% CI 76.0%-86.8%), 73.9% (95% CI 68.6%- 78.7%), and 75.9% (95% CI 67.9%-82.5%) for £10 (US $12.5), £20 (US $25.0), and £30 (US $37.5), respectively, suggesting that the selective use of incentives reduced participation bias in relation to vaccination status as a proxy for health behaviors.

Figure 4. Swab response rates and 95% CIs for round 16 (November 23 to December 14, 2021), round 17 (January 5-20, 2022), and rounds 18 (February 8 to March 1, 2022) and 19 (March 8-31, 2022) in which incentives were used selectively, England. Note: Incentive amounts used in rounds 18 and 19: £10 (US $12.5) for 13- to 17-year-olds and 35- to 44-year-olds, £20 (US $25.0) for 18- to 34-year-olds, and no incentives for other age groups.



Discussion

Principal Findings

In this large population-based study of the prevalence of SARS-CoV-2 infection in England, we tested several measures to increase response rates and reduce nonresponder bias. We found that changes to the wording of letters, timing, and numbers of reminders made only limited differences to response rates, with a maximum increase in response rate of 3.1 percentage points for additional swab reminders sent to people who had already registered for the study. Sending an additional reminder, regardless of its form (SMS or email), increased response. This is consistent with other studies in the literature [30]. These reminder strategies may have helped slow, but did not halt, the decline in response rates over time observed during REACT-1. Nonetheless, these findings informed the swab reminder strategy and invitation letter wording in later rounds. In contrast, the offer of a financial incentive conditional on the return of a completed swab made a more substantial difference of up to 22.3 percentage points and was particularly effective in those with a lower propensity to respond: younger age groups and those living in more deprived areas. Similarly, incentives increased the return of completed swabs by unvaccinated individuals so that COVID-19 vaccination rates were more in keeping with those in the general population at the time. Thus, the selective use of incentives may reduce nonresponder bias in relation to factors of interest in population health research beyond sociodemographic characteristics.

The selective use of incentives was subsequently adopted from round 18, making the achieved sample more representative by age, with a reduction in age-based variation in response rates. Previous research suggests that ethnic minorities [31], individuals living in more deprived areas [32], those in urban areas [33], and the youngest and oldest age groups [34,35] are the least likely to respond in general population surveys. Using incentives selectively allowed us, at modest cost, to increase recruitment among such groups and hence increase the effective sample size; thus, in round 18, the effective sample size was over 10,000 greater compared to round 17, even though we received circa 7000 fewer swabs. We were able to reduce the number of invitations sent out while achieving a similar number of completed swabs returned as in earlier rounds when response rates were higher.

Using incentives selectively has been tried in UK social surveys previously and is common practice in the United States, where studies show they are cost-effective, improve response, and reduce bias [11,13,23,36]. From an ethical perspective, in the selective use of (versus universal) incentives, it was important to consider not only issues of equity but also cost and the public interest in continuing to obtain high-quality data, covering all sectors of society, to monitor the spread of a serious disease. This needs to be balanced against the possible disappointment of some participants who learn others are being offered a (larger) incentive. These considerations might apply to many population-based surveys. We accept that the argument for using incentives selectively may have been more persuasive in the context of REACT-1, a study to measure the spread of SARS-CoV-2 during the pandemic, the policy responses to which had far-reaching consequences for the way of life of every person in England.

Both unconditional and conditional financial incentives have been shown to significantly increase response rates to both postal and web-based surveys [37,38]. Although unconditional incentives appear to have the largest effect, the conditional approach is more cost-effective [37,38]. Unconditional incentives have been used in social surveys in the United Kingdom, and in experiments in how to increase response rates [39,40]. Unconditional incentives were not an option for REACT-1 due to the constraints of the survey budget.

Limitations

In terms of limitations, it was not possible to ascertain the extent to which noncontact (ie, the intended recipient did not receive the invitation letter) accounted for nonresponse. Such situational factors, for example, not informing their general practitioner of a change in address or having moved with no forwarding address (shown to be greater for young adults and lower socioeconomic groups) [32] will not be affected by the experimental conditions; therefore, our estimates of effect are likely conservative, as invitations sent out do not necessarily mean that invitations were received. In addition, the unique circumstances of carrying out such assessments of response rates during a global pandemic may not "read across" to other less pressing issues.

Conclusions

We achieved small improvements in response rates by varying the number, order, and content of invitations and reminders but much larger effects were seen through the use of monetary incentives. Lessons learnt from the REACT-1 study may help inform the design and implementation of future population-based surveys where the intent is to obtain as representative a sample as possible and to reduce nonresponse bias at reasonable cost. The results suggest selectively using incentives with younger and more deprived individuals may be justifiable to achieve these ends.

Acknowledgments

This work was supported by the Department of Health and Social Care in England. The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; and preparation, review, or approval of this article. Infrastructure support for this research was provided by the National Institute for Health (NIHR) Imperial Biomedical Research Centre (BRC). PE is director of the Medical Research Council (MRC) Centre for Environment and Health (MR/L01341X/1 and MR/S019669/1). PE acknowledges support from Health Data Research UK, the NIHR Imperial Biomedical Research Centre; NIHR Health Protection Research Units in Chemical and Radiation Threats and Hazards, and Environmental Exposures and Health; the British Heart Foundation Centre for Research Excellence at Imperial College London (RE/18/4/34215); and the UK Dementia Research Institute at Imperial College London (MC_PC_17114). HW acknowledges support from an NIHR Senior Investigator Award, the Wellcome Trust (205456/Z/16/Z), the NIHR Imperial Biomedical Research Centre, the NIHR School of Public Health Research, and the NIHR Applied Research Collaboration North West London. CAD and SR acknowledge support from MRC Centre for Global Infectious Disease Analysis (reference MR/R015600/1), jointly funded by the UK MRC and the UK Foreign, Commonwealth & Development Office (FCDO), under the MRC/FCDO Concordat agreement. CAD acknowledges support from the NIHR Health Protection Research Unit (HPRU) in Emerging and Zoonotic Infections (NIHR200907), and the NIHR funded Vaccine Efficacy Evaluation for Priority Emerging Diseases (PR-OD-1017-20007). SR acknowledges support from the NIHR HPRU in Respiratory Infections, Wellcome Trust (200861/Z/16/Z, 200187/Z/15/Z), and Centres for Disease Control and Prevention (US, U01CK0005-01-02). MCH acknowledges support from Cancer Research UK, Population Research Committee Project grant "Mechanomics" (grant 22184). MCH acknowledges support from the H2020-EXPANSE (Horizon 2020 grant 874627) and H2020-LongITools (Horizon 2020 grant 874739).We thank key collaborators on this work: K Beaver, S Finlay, J Kennedy, D Peskett, G Welch, and K Ward at Ipsos. We also thank the REACT Public Advisory Panel who provided input which changed the wording of the invitation letters and reminders and provided advice on the design and equity concerns in relation to the incentives experiment.

Data Availability

The datasets generated during and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

NG, GP, SC, SR, HW, and PE designed the study. NG conducted the analyses. KP provided statistical support. CJA, NG, HW, and PE designed the analytical strategy and helped to interpret the findings. CJA, NG, and HW conducted the literature review. CJA, NG, HW and PE drafted the manuscript. HW, GSC, CJA, AD, CAD, MCH, SR, DA, WSB, and PE provided study oversight. AD and PE obtained funding. All authors have reviewed and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary methods.
[DOCX File, 19 KB - publichealth_v11i1e60022_app1.docx]

Multimedia Appendix 2
Supplementary tables.
[DOCX File, 55 KB - publichealth_v11i1e60022_app2.docx]

References

1. Groves RM, Peytcheva E. The impact of nonresponse rates on nonresponse bias: a meta-analysis. Public Opin Q 2008;72(2):167-189. [doi: 10.1093/poq/nfn011]
2. Galea S, Tracy M. Participation rates in epidemiologic studies. Ann Epidemiol 2007 Sep;17(9):643-653. [doi: 10.1016/j.annepidem.2007.03.013] [Medline: 17553702]
3. Allen N, Sudlow C, Downey P, et al. UK Biobank: current status and what it means for epidemiology. Health Policy Technol 2012 Sep;1(3):123-126. [doi: 10.1016/j.hlpt.2012.07.003]
4. Fry A, Littlejohns TJ, Sudlow C, Doherty N, Allen NE. OP41: the representativeness of the UK biobank cohort on a range of sociodemographic, physical, lifestyle and health-related characteristics. J Epidemiol Community Health 2016 Sep;70(Suppl 1):A26. [doi: 10.1136/jech-2016-208064.41]
5. Swanson JM. The UK Biobank and selection bias.. Lancet 2012 Jul;380(9837):110. [doi: 10.1016/S0140-6736(12)61179-9]
6. Keyes KM, Westreich D. UK Biobank, big data, and the consequences of non-representativeness. Lancet 2019 Mar 30;393(10178):1297. [doi: 10.1016/S0140-6736(18)33067-8] [Medline: 30938313]
7. Andrade C. Sample size and its importance in research. Indian J Psychol Med 2020;42(1):102-103. [doi: 10.4103/IJPSYM.IJPSYM_504_19] [Medline: 31997873]
8. Gelman A, Carlin J. Poststratification and weighting adjustments. In: Groves R, Dillman D, Eltinge J, Little R, editors. Survey Nonresponse: Wiley; 2000:289-302.
9. Pforr K, Blohm M, Blom AG, et al. Are Incentive effects on response rates and nonresponse bias in large-scale, face-to-face surveys generalizable to Germany? Evidence from ten experiments. Public Opin Q 2015;79(3):740-768. [doi: 10.1093/poq/nfv014]
10. Lange C, Finger JD, Allen J, et al. Implementation of the European health interview survey (EHIS) into the German health update (GEDA). Arch Public Health 2017;75:40. [doi: 10.1186/s13690-017-0208-6] [Medline: 28936356]
11. Singer E, Hoewyk JV, Gebler N, Raghunathan TE, Mcgonagle KA. The effect of incentives on response rates in interviewer-mediated surveys. J Off Stat 1999;15(2):217-230.
12. Wu MJ, Zhao K, Fils-Aime F. Response rates of online surveys in published research: a meta-analysis. Comput Hum Behav Rep 2022 Aug;7:100206. [doi: 10.1016/j.chbr.2022.100206]
13. Nicolaas G CE, Davies B. The use of incentives to recruit and retain hard-to-get populations in longitudinal studies. NatCen Soc Res 2019 May 31.
14. Mavletova A, Couper MP. Mobile web survey design: scrolling versus paging, SMS versus e-mail invitations. J Surv Stat Methodol 2014 Dec;2(4):498-518. [doi: 10.1093/jssam/smu015]
15. Cabrera-Álvarez P, Lynn P. Text messages to facilitate the transition to web-first sequential mixed-mode designs in longitudinal surveys. J Surv Stat Methodol 2024 Jun 1;12(3):651-673. [doi: 10.1093/jssam/smae003]
16. Lynn P. Targeted appeals for participation in letters to panel survey members. PUBOPQ 2016;80(3):771-782. [doi: 10.1093/poq/nfw024]
17. Taylor KSM, Counsell CE, Harris CE, Gordon JC, Fonseca SCC, Lee AJ. In a randomized study of envelope and ink color, colored ink was found to increase the response rate to a postal questionnaire. J Clin Epidemiol 2006 Dec;59(12):1326-1330. [doi: 10.1016/j.jclinepi.2006.03.011] [Medline: 17098576]
18. Riley S, Atchison C, Ashby D, et al. REal-time Assessment of Community Transmission (REACT) of SARS-CoV-2 virus: study protocol. Wellcome Open Res 2020;5(200):200. [doi: 10.12688/wellcomeopenres.16228.2] [Medline: 33997297]
19. Elliott P, Whitaker M, Tang D, et al. Design and implementation of a National SARS-CoV-2 Monitoring Program in England: REACT-1 study. Am J Public Health 2023 May;113(5):545-554. [doi: 10.2105/AJPH.2023.307230] [Medline: 36893367]
20. Ward H, Whitaker M, Flower B, et al. Population antibody responses following COVID-19 vaccination in 212,102 individuals. Nat Commun 2022;13(1):907. [doi: 10.1038/s41467-022-28527-x]
21. Eales O, Wang H, Haw D, et al. Trends in SARS-CoV-2 infection prevalence during England's roadmap out of lockdown, January to July 2021. PLoS Comput Biol 2022 Nov;18(11):e1010724. [doi: 10.1371/journal.pcbi.1010724] [Medline: 36417468]
22. Adolescent consent. NHS Health Research Authority. 2022. URL: https://www.hra.nhs.uk/planning-and-improving-research/application-summaries/research-summaries/adolescent-consent/

#:~:text=Young%20people%20over%20the%20age%20of%2016%20are,consent%20from%20their%20parents%20provided%20they%20are%20competent [accessed 2024-12-31]
23. Wong E. Incentives in Longitudinal Studies, CLS Working Paper 2020/1: UCL Centre for Longitudinal Studies; 2020.
24. Ward H, Atchison C, Whitaker M, et al. SARS-CoV-2 antibody prevalence in England following the first peak of the pandemic. Nat Commun 2021 Feb 10;12(1):905. [doi: 10.1038/s41467-021-21237-w] [Medline: 33568663]
25. Comparisons of all-cause mortality between european countries and regions: january to june 2020. Office for National Statistics. URL: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/articles/comparisonsofallcausemortalitybetweeneuropeancountriesandregions/januarytojune2020 [accessed 2024-12-31]
26. COVID-19 vaccine data. NHS Digital. 2022. URL: https://digital.nhs.uk/about-nhs-digital/corporate-information-and-documents/directions-and-data-provision-notices/data-provision-notices-dpns/covid-19-vaccine-data [accessed 2022-12-31]
27. Tessier E, Edelstein M, Tsang C, et al. Monitoring the COVID-19 immunisation programme through a national immunisation management system - England's experience. Int J Med Inform 2023 Feb;170:104974. [doi: 10.1016/j.ijmedinf.2022.104974] [Medline: 36577202]
28. Population and household estimates, england and wales: census 2021. Office for National Statistics. 2022. URL: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/populationandhouseholdestimatesenglandandwales/census2021 [accessed 2024-12-31]
29. COVID-19 vaccinations archive. 2021. URL: https://www.england.nhs.uk/statistics/statistical-work-areas/covid-19-vaccinations/covid-19-vaccinations-archive [accessed 2024-12-31]
30. Sammut R, Griscti O, Norman IJ. Strategies to improve response rates to web surveys: a literature review. Int J Nurs Stud 2021 Nov;123:104058. [doi: 10.1016/j.ijnurstu.2021.104058] [Medline: 34454334]
31. D. Royal K. Surveying Black or African American populations: challenges and solutions in medicine and education. Educ Med J 2019 Jun;11(2):59-61. [doi: 10.21315/eimj2019.11.2.8]
32. Goodman A, Gatward R. Who are we missing? Area deprivation and survey participation. Eur J Epidemiol 2008;23(6):379-387. [doi: 10.1007/s10654-008-9248-0] [Medline: 18409005]
33. Inpatient experience survey 2018: technical report. Scottish Government. 2018 Aug 28. URL: https://www.gov.scot/publications/inpatient-experience-survey-2018-technical-report/ [accessed 2024-12-31]
34. Harrison S, Alderdice F, Henderson J, Redshaw M, Quigley MA. Trends in response rates and respondent characteristics in five National Maternity Surveys in England during 1995-2018. Arch Public Health 2020;78(46):46. [doi: 10.1186/s13690-020-00427-w] [Medline: 32509303]
35. Klein DJ, Elliott MN, Haviland AM, et al. Understanding nonresponse to the 2007 Medicare CAHPS survey. Gerontologist 2011 Dec;51(6):843-855. [doi: 10.1093/geront/gnr046] [Medline: 21700769]
36. Yu S, Alper HE, Nguyen AM, et al. The effectiveness of a monetary incentive offer on survey response rates and response completeness in a longitudinal study. BMC Med Res Methodol 2017 Apr 26;17(1):77. [doi: 10.1186/s12874-017-0353-1] [Medline: 28446131]
37. Young JM, O'Halloran A, McAulay C, et al. Unconditional and conditional incentives differentially improved general practitioners' participation in an online survey: randomized controlled trial. J Clin Epidemiol 2015 Jun;68(6):693-697. [doi: 10.1016/j.jclinepi.2014.09.013] [Medline: 25450450]
38. Abdelazeem B, Hamdallah A, Rizk MA, et al. Does usage of monetary incentive impact the involvement in surveys? A systematic review and meta-analysis of 46 randomized controlled trials. PLoS One 2023;18(1):e0279128. [doi: 10.1371/journal.pone.0279128] [Medline: 36649255]
39. Office for National Statistics. Surveys and provision of monetary voucher for participation. 2018. URL: https://www.ons.gov.uk/aboutus/transparencyandgovernance/freedomofinformationfoi/surveysandprovisionofmonetaryvoucherforparticipation [accessed 2024-12-31]
40. Office for National Statistics. Crime survey for england and wales: technical note (Y/E march 2022). 2023. URL: https://www.ons.gov.uk/file?uri=/peoplepopulationandcommunity/crimeandjustice/methodologies/crimeandjusticemethodology/csewtechnicalnoteyemarch2022.pdf [accessed 2024-12-31]

Abbreviations

Ct: cycle threshold
IMD: Index of Multiple Deprivation
NHS: National Health Service
REACT-1: Real-time Assessment of Community Transmission-1
RRR: relative response rate
rt-PCR: reverse transcriptase polymerase chain reaction

Edited by A Mavragani; submitted 29.04.24; peer-reviewed by D Falk, P Lynn; revised version received 12.11.24; accepted 13.11.24; published 09.01.25.

Please cite as:
Atchison CJ, Gilby N, Pantelidou G, Clemens S, Pickering K, Chadeau-Hyam M, Ashby D, Barclay WS, Cooke GS, Darzi A, Riley S, Donnelly CA, Ward H, Elliott P
Strategies to Increase Response Rate and Reduce Nonresponse Bias in Population Health Research: Analysis of a Series of Randomized Controlled Experiments during a Large COVID-19 Study
JMIR Public Health Surveill 2025;11:e60022
URL: https://publichealth.jmir.org/2025/1/e60022
doi:10.2196/60022

Contribution of Travelers to Plasmodium Vivax Malaria in South West Delhi, India: Cross-Sectional Survey

Deepali Savargaonkar[1], MBBS, MPH; Bina Srivastava[1], MSc; Chander Prakash Yadav[2], PhD; Mrigendra Pal Singh[3], PhD; Anup Anvikar[1,4], MBBS, MD; Amit Sharma[5], PhD; Himmat Singh[1,4*], PhD; Abhinav Sinha[1,4*], MBBS, MRES, MSc, MD, PhD

[1]ICMR-National Institute of Malaria Research, Sector 8, Dwarka, New Delhi, India
[2]ICMR-National Institute of Cancer Prevention and Research, Noida, India
[3]ICMR-National Institute for Research in Tribal Health, Jabalpur, India
[4]Academy of Scientific and Innovative Research, Ghaziabad, India
[5]International Centre for Genetic Engineering and Biotechnology, New Delhi, India
[*]these authors contributed equally

Corresponding Author:
Abhinav Sinha, MBBS, MRES, MSc, MD, PhD
ICMR-National Institute of Malaria Research, Sector 8, Dwarka, New Delhi, India

Abstract

Background: India is committed to malaria elimination by the year 2030. According to the classification of malaria endemicity, the National Capital Territory of Delhi falls under category 1, with an annual parasite incidence of <1, and was targeted for elimination by 2022. Among others, population movement across states is one of the key challenges for malaria control, as it can result in imported malaria, thus introducing local transmission in an area nearing elimination.

Objective: This descriptive study attempts to assess the contribution of such imported *Plasmodium vivax* cases to the malaria burden in South West Delhi (SWD).

Methods: A cross-sectional study was carried out at the fever clinic of the Indian Council of Medical Research-National Institute of Malaria Research in SWD from January 2017 to December 2019. Demographic and travel history data were recorded for all *P vivax* confirmed malaria cases diagnosed at the fever clinic. Vector and fever surveys along with reactive case detection were conducted in SWD and Bulandshahr district of Uttar Pradesh, 1 of the 6 geographical sources for a high number of imported malaria cases.

Results: A total of 355 *P vivax* malaria cases were reported during the study period. The proportion of imported cases was 63% (n=222). Of these, 96% (n=213) of cases were from Uttar Pradesh. The distribution of malaria cases revealed that imported cases were significantly associated with travel during the transmission season compared with that in the nontransmission season. Entomological and fever surveys and reactive case detection carried out in areas visited by imported *P vivax* malaria cases showed the presence of adults and larvae of *Anopheles* species and *P vivax* parasitemia.

Conclusions: Population movement is a key challenge for malaria elimination. Although additional *P vivax* infections and vector mosquitoes were detected at places visited by the imported malaria cases, the inability to detect the parasite in mosquitoes and the possibility of relapses associated with *P vivax* limit the significance of malaria associated with the travel. However, there remains a need to address migration malaria to prevent the introduction and re-establishment of malaria in areas with very low or 0 indigenous cases.

(JMIR Public Health Surveill 2025;11:e50058) doi:10.2196/50058

KEYWORDS
malaria; Plasmodium vivax; imported malaria; population movement; transmission; elimination; India

Introduction

Malaria is a parasitic disease transmitted by the bites of infected female *Anopheles* mosquitoes. The estimated number of malaria cases worldwide in 2021 was 247 million spread over 84 countries with 619,000 deaths. Although the World Health Organization (WHO) South East Asia region comprised ≈2% of the estimated global burden, India contributed to ≈79% of these cases, with a preponderance of *Plasmodium vivax* over *Plasmodium falciparum* [1].

Migration, both international and within-country, is a recognized social health determinant of multiple diseases across the globe, and malaria is no exception. Various factors put the migrating population at risk of contracting malaria, and these include their socioeconomic, living, working, and transit conditions [2]. The risk of malaria is also to the host communities that provide shelter to the migrants, particularly when the migration is along an epidemiological gradient from a high-burden to a low-burden or nonendemic area, putting malaria elimination efforts at risk [3-5]. Several pieces of evidence of this have been documented in the context of international migration [6-8] and within-country migration [9,10], including that in India [11].

India is committed to malaria elimination by 2030 and has formulated the National Framework for Malaria Elimination that classifies Indian states and union territories into 4 categories from 0 to 3, with category 3 being the highest-burden areas with an annual parasite incidence (API) of ≥1 per thousand persons at risk [12]. To achieve the elimination goal in the desired time frame, special focus needs to be given to the identified challenges by the National Center for Vector Borne Diseases Control (NCVBDC). Population size and migration are recognized as important challenges for malaria elimination, apart from asymptomatic parasite reservoirs, low-density infections, and parasite- and vector-resistance [13-15]. The movement of populations across and within Indian states is one of the key challenges in malaria control [12], and particularly, the migration of workers in large numbers from rural areas to cities has been reported in India [16]. Similarly, malaria cases among the mobile population contribute to a high percentage of total malaria cases in many countries [17] and have been a recognized challenge for malaria elimination [18]. Hence, it is important to carry out the screening and treatment for malaria in mobile populations for control and elimination of malaria in endemic areas, especially National Framework for Malaria Elimination category 1 areas (with API<1), and for prevention of the re-establishment of local transmission of malaria [12] in areas that have eliminated malaria (category 0 with 0 indigenous cases).

Despite being the capital of India that attracted >100,000 migrants each year since 2012 [19], no study on malaria in the migrant population in Delhi has yet been reported. Delhi falls under category 1 (API<1 per 1000 persons at risk), and its neighboring state, Uttar Pradesh (UP), is in category 2 with an API of less than 1 but with some districts having an API of 1 or more. With such a magnitude of migration, there remains a sustained threat of the introduction of *Plasmodium* infection by infected migrant populations from high-burden areas to the areas in categories 0 and 1 [16].

As Delhi and its neighboring states have *P vivax* as the predominant *Plasmodium* species causing malaria, this study was therefore carried out to assess the contribution of imported *P vivax* cases to the *P vivax* malaria burden in South West Delhi (SWD) by tracking the travel history of infected patients diagnosed at the fever clinic of Indian Council of Medical Research-National Institute of Malaria Research (ICMR-NIMR). Additionally, the study also aimed to identify mosquito breeding habitats and the presence and types of mosquito species in areas where these imported malaria cases resided/visited, and to detect additional *Plasmodium* infections/malaria cases through reactive case detection (RACD) and fever surveys in selected areas visited by these patients. These entomological and fever surveys and RACD were done to further identify any ongoing transmission in the areas visited by patients with imported malaria to support the hypothesis that these are indeed the cases likely to import malaria from areas with ongoing transmission to SWD.

Methods

Ethical Considerations

The study was approved by the Institutional Ethics Committee of ICMR-NIMR (ECR/NIMR/EC/2015/507 and ECR/NIMR/EC/2019/175). Informed consent was obtained from all human participants who were involved in the study. The participant identifier data were anonymized. No compensation was provided to the participants.

Study Sites, Samples, and Definitions

The study was initiated at the fever clinic at ICMR-NIMR, SWD, and later expanded to include the prominent catchment areas of the fever clinic of SWD and 6 villages of 1 selected district (Bulandshahr) of UP. Incoming febrile patients were screened for *Plasmodium* infection at the fever clinic of ICMR-NIMR from January 2017 to December 2019. Basic demographic data including age, gender, history of fever, and travel details (if any) were recorded using a paper-based structured questionnaire. The parasitological diagnosis at the clinic was performed by microscopy. Thick and thin blood smears were prepared, stained using the Jaswant Singh–Bhattacharji stain [20] and examined under 100× magnification. Microscopy was performed independently by 2 trained microscopists. In case of a discrepancy, a third trained microscopist examined the smears, and consensus observation by 2 trained microscopists was considered final. Those diagnosed with malaria were treated as per the national drug policy [21].

History of travel during the preceding 30 days of fever, including the places visited and duration of stay, was verbally elicited through a calendar-based recall method among all patients with malaria. Obtained travel history was used to classify the patients into 3 nonoverlapping categories: no travel history (patients neither traveled in or out of their residence in SWD), returning travelers (patients who were residing in SWD and traveled outside Delhi but came back), and incoming travelers (patients who were not residing in SWD but transiently traveled to SWD). Returning travelers were further classified based on the duration of stay outside Delhi into those returning to SWD within 7 days and those returning between 7 and 30 days. Similarly, incoming travelers were also classified into those who came to SWD within and beyond 30 days of fever.

Since the incubation period for malaria is 7 to 30 days [22-25], returning travelers with malaria who returned to SWD between 7 and 30 days of fever onset and incoming travelers who entered SWD within 30 days of fever onset were classified as imported malaria cases (acquired infection outside Delhi) for this study. The rest of the patients with malaria were considered to have indigenous infections.

Imported malaria cases were further investigated for the exact village and district of travel based on their recorded travel history. Thus, to further investigate whether the malaria cases were imported or indigenous, malaria vector (anopheline) surveys were carried out both in SWD and in the villages the imported malaria cases traveled to. These surveys were carried out by field workers adequately trained in entomology in catchment areas (Raj Nagar and Bagdola; every month from September 2018 to December 2019) of the fever clinic of ICMR-NIMR predominantly reporting malaria and also in villages (once in October 2019) of Bulandshahr district of UP state. Fever surveys and RACD were also carried out in these villages of Bulandshahr, UP. Bulandshahr district was preferred out of the 6 districts that showed significant sources of imported malaria in UP based on the burden of imported cases, logistic convenience, and operational feasibility. All 6 villages of Bulandhahr districts that had epidemiologically relevant travel connections with the imported malaria cases were surveyed.

The vector survey included the collection of mosquitoes (adults and larvae) from the houses of reported cases and their surrounding houses, species identification, and enumeration of mosquito breeding habitats. Resting adult mosquito collection was conducted in households of 6 villages of Bulandshahr district during early morning (6 AM to 8 AM) using hand aspirators. The larval collection was also done in each village from all water-bearing sites, that is, ponds, ditches, large cement tanks, drains, and seepages, in peridomestic and domestic areas of each village. In the Raj Nagar catchment locality of SWD, the houses were searched from 7 AM to 9 AM, and larval collection was also conducted simultaneously if searched houses and containers were found positive for larval presence from domestic and peridomestic water bodies and other sites including overhead tanks, large open water bodies, tires, coolers, bird pots, flowerpots, iron containers, and solid wastes in urban catchment areas of the clinic and from domestic and peridomestic containers in the houses of reported cases. The collected larvae were reared in an insectary separately up to their emergence to identify the mosquito species. Identification of species was done following the standard taxonomic key as described by Christophers [26]. Adult mosquito collection was also done using the hand catch method [27], and the collected *Anopheles* mosquitoes were screened for the presence of malaria parasites (*P falciparum, P vivax, Plasmodium malariae*, and *Plasmodium ovale*), through polymerase chain reaction (PCR) [28], in pools of mosquitoes made village-wise and species-wise. To estimate the critical density of malaria vectors, per man-hour density (PMHD) was calculated as the number of anophelines collected per hour by an insect collector using the formula:

PMHD=No.of mosquitoes collected Time spent (in hours) x no. of insect collectors

Fever surveys and RACD were carried out in October 2019 in 6 villages of the Bulandshahr (UP) with support from the local health personnel (Accredited Social Health Activists and Health Inspectors). Fever camps were organized at a central location in each village, and the local health personnel informed the villagers about the camp and motivated them to visit. Incoming febrile cases were screened for malaria by using a rapid diagnostic test (SD Bio Line Malaria Ag P.f / P.v, Standard Diagnostics, Inc, Republic of Korea) as per the manufacturer's instructions. All febrile cases were treated symptomatically, and patients with malaria were treated as per the national drug policy. RACD was done as described by the WHO [29], and blood smears were prepared from apparently healthy individuals in and around the household of the index cases. The smears were examined for the presence of malaria parasites at ICMR-NIMR, Delhi (as described previously), and the results were communicated to the concerned health personnel for further management.

Data Entry and Statistical Analysis

All the collected data were entered in a Microsoft Excel 2016 spreadsheet and presented as proportions (percentages), medians, and ranges, where appropriate. The strength of association was estimated using a chi-square test, and a P value of less than .05 was considered statistically significant.

Results

Overview

A total of 14,748 fever cases were screened for malaria by microscopy from January 2017 to December 2019. The 3-year period prevalence of malaria was 2.4% (364/14,748). Out of these 364 cases, 355 (97.5%) were *P vivax* mono-infections, 8 *P falciparum* mono-infections (2.1%), and 1 mixed infection of *P vivax* and *P falciparum*. There was male predominance among patients with fever (59%) as well as patients with *P vivax* malaria (71%). More than half of the patients with *P vivax* malaria were in the age group of 15 - 29 years (183/355, 52%) with a median age of 22 years. The parasite burden ranged from 63 to 206,187 parasites per microliter of blood.

Imported Malaria Burden

Out of the 250 *P vivax* cases with a travel history (250/355, 70%), 186 (74%) cases were returning travelers and the remaining 64 (26%) cases were incoming travelers (Figure 1). However, relevant travel history to be able to label them as imported cases was available from 63% (222/355) of patients. Out of these imported *P vivax* cases, 173 (78%) cases were returning and 49 (22%) cases were incoming travelers. Ninety-five percent (212/222) of the imported cases were from UP and 142 of them (142/212, 67%) had traveled to 1 of the 6 districts of UP viz Bareilly, Badaun, Aligarh, Hathras, Bulandshahr, and Mainpuri. The remaining 10 imported cases had traveled to Uttarakhand (3; Dehradun and Hardwar), Rajasthan (3; Bikaner, Bundi, and Sawai Madhopur), Madhya Pradesh (2; Gwalior), Haryana (1; Gurugram), and Punjab (1; Sri Muktsar Sahib), as shown in Figure 2. Out of 9 patients with *P falciparum* malaria (including 1 mixed infection), 8 had a travel history.

Figure 1. Travel history among 355 patients with *Plasmodium vivax* malaria. Based on the epidemiologically relevant travel history, cases associated with travel were categorized into "imported" (222; shown in red font) and indigenous (133; gold colored font) cases. The geographical distribution (states) of imported cases is also mentioned.

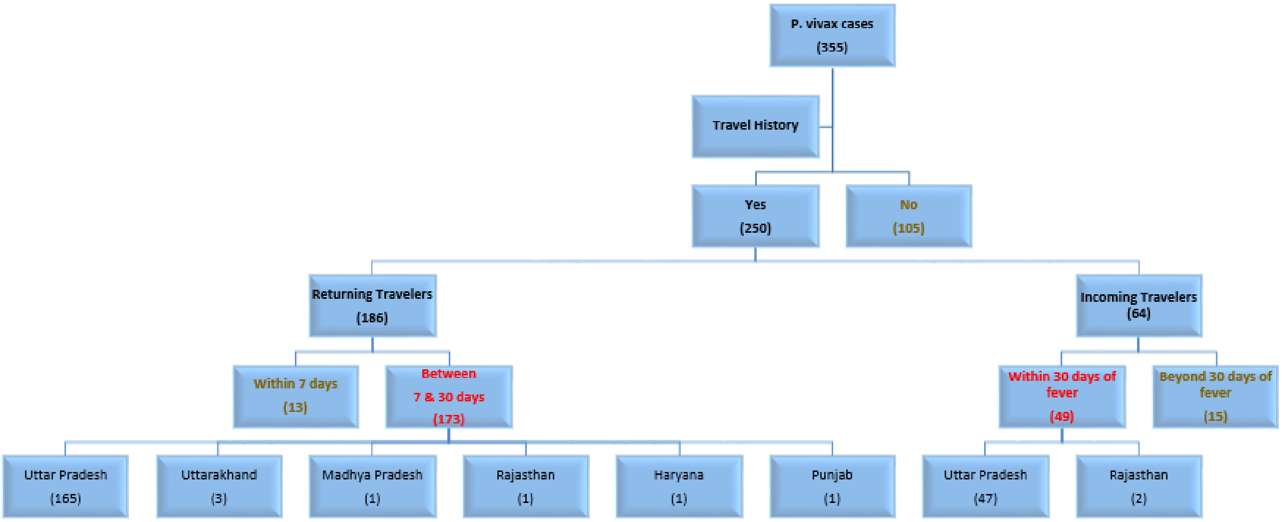
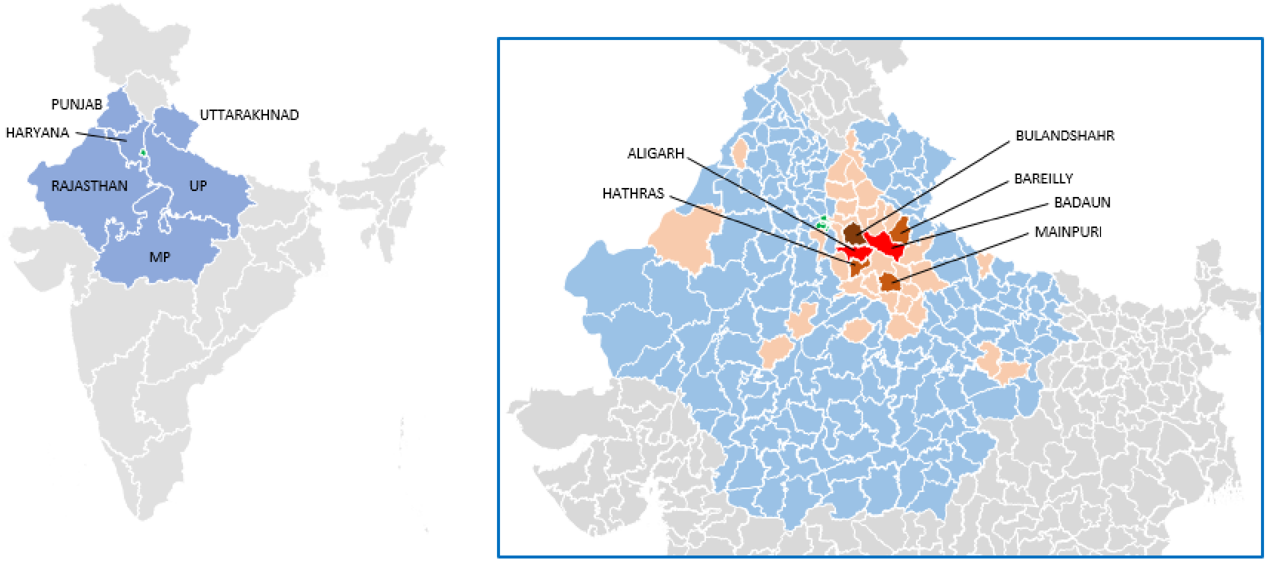


Figure 2. Areas traveled to by *Plasmodium vivax* malaria imported cases. The areas (states and union territories of India) are shown in blue (Punjab, Chandigarh, Haryana, Uttarakhand, UP, MP, and Rajasthan) whereas Delhi (study site) is shown in green. The zoomed-in image of the map in the inset shows further administrative breakdown of these 7 states and union territories (as districts) in blue. The districts within these 7 states and union territories, which are associated with the travel history of imported cases, are colored based on the number of imported cases contributed by each district: light orange (1 - 5 cases); dark orange (5 - 15 cases); darker orange (15 - 25 cases), and red (>25 cases). It is evident that UP has 3 dark orange districts: Hathras (10 cases), Mainpuri (12 cases), and Bareilly (15 cases); 1 darker orange district: Bulandshahr (25 cases); and 2 red districts: Aligarh (37 cases) and Badaun (47 cases). MP: Madhya Pradesh; UP: Uttar Pradesh.

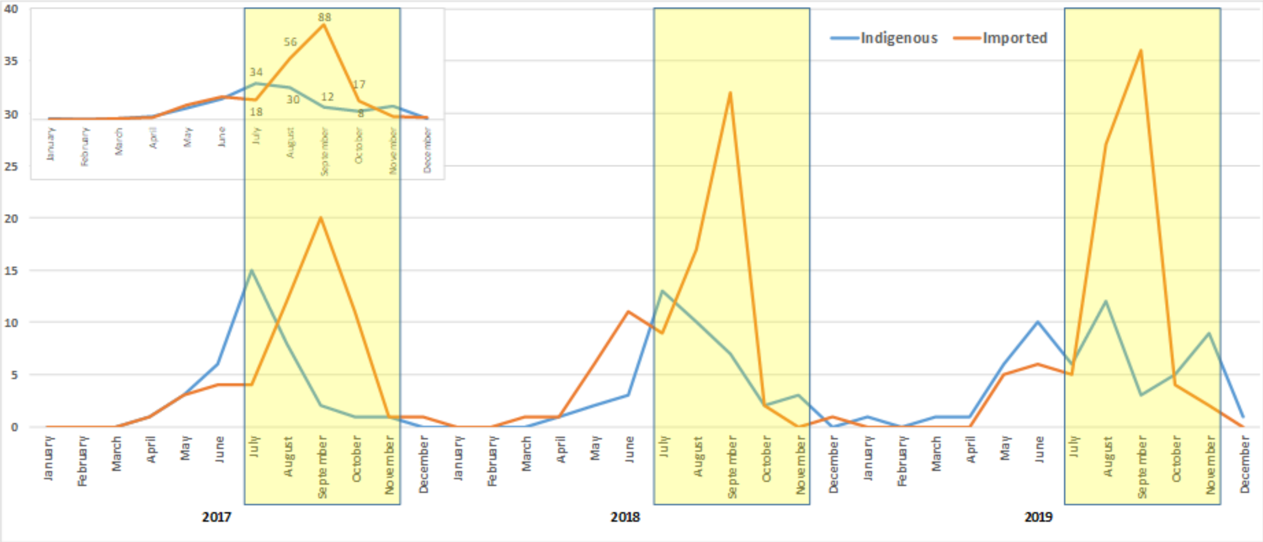


The reasons for travel in *P vivax* malaria cases included visiting their native residence in various states (mainly for returning travelers); education or employment (for incoming travelers); and visiting relatives, family, and friends during festivals (for returning and incoming travelers) as many Indian and regional festivals temporally coincide with the malaria transmission season. The minimum length of stay outside Delhi among travelers was 1 day while the maximum stay was of 111 days. Travel history was reported by the family members of the patients with malaria as well. There were 24 families with at least 2 members probably acquiring malaria after traveling.

A majority of the *P vivax* cases (279/355, 79%) were detected during the transmission season, that is, July to November. The nontransmission season (December to June) contributed to the remaining 21% (76/355) of cases.

During the transmission season (July to November; 2017 - 2019), the proportion of imported *P vivax* malaria cases diagnosed at the fever clinic was higher (65%) than that in the nontransmission season (53%; December to June), as shown in Figure 3, and the difference was statistically significant (χ^2_1=4.04; *P*=.031) at a 95% confidence level.

Figure 3. *Plasmodium vivax* malaria cases reported in Delhi (2017 - 2019). The figure shows the number of *P vivax* cases, indigenous (blue) and imported (imported), as reported each month and each year during the study period (January 2017 to December 2019) at the Indian Council of Medical Research-National Institute of Malaria Research fever clinic. The cumulative month-wise data from 2017 to 2019 are shown in the inset. The yellow semitransparent rectangles show the transmission season.



Entomological Surveys

Adult mosquito collections from the draining and catchment areas (Bagdola and Raj Nagar areas) of the fever clinic in SWD showed 4 mosquito species, with a low prevalence of *Anopheles* mosquitoes during the survey (September 2018 to December 2019). Out of 573 adult mosquitoes collected from 3395 houses, only 7 (1%) were *Anopheles stephensi* (Table 1), and all of them were found during the malaria transmission season (September). No other species of *Anopheles* were found. The majority of mosquitoes (443/573, 78%) were *Aedes aegypti* in the households.

Table . Month-wise adult mosquito collection in Bagdola and Raj Nagar, South West Delhi, between September 2018 and December 2019. Transmission season is highlighted in gold. The proportion of mosquito species identified out of the total is shown as a percentage (in parentheses).

		Houses checked, n	Mosquito species identified				
			Aedes aegypti, n (%)	*Aedes albopictus*, n (%)	*Anopheles stephensi*, n (%)	*Culex quinquefasciatus*, n (%)	Total, n
Month and year							
	September 2018[a]	316	28 (37)	0 (0)	1 (1)	46 (62)	75
	October 2018[a]	97	14 (70)	0 (0)	0 (0)	6 (30)	20
	November 2018[a]	402	63 (73)	0 (0)	0 (0)	23 (27)	86
	December 2018	220	16 (84)	1 (5)	0 (0)	2 (11)	19
	January 2019	112	1 (100)	0 (0)	0 (0)	0 (0)	1
	February 2019	70	0 (0)	0 (0)	0 (0)	0 (0)	0
	March 2019	179	2 (40)	0 (0)	0 (0)	3 (60)	5
	April 2019	60	13 (100)	0 (0)	0 (0)	0 (0)	13
	May 2019	344	10 (63)	0 (0)	0 (0)	6 (37)	16
	June 2019	367	89 (87)	0 (0)	0 (0)	13 (13)	102
	July 2019[a]	352	53 (75)	0 (0)	0 (0)	18 (25)	71
	August 2019[a]	270	61 (91)	0 (0)	6 (9)	0 (0)	67
	September 2019[a]	301	44 (90)	5 (10)	0 (0)	0 (0)	49
	October 2019[a]	211	40 (100)	0 (0)	0 (0)	0 (0)	40
	November 2019[a]	33	9 (100)	0 (0)	0 (0)	0 (0)	9
	December 2019	61	0 (0)	0 (0)	0 (0)	0 (0)	0
Total		3395	443 (78)	6 (1)	7 (1)	117 (20)	573

[a]Transmission season.

On the other hand, a total of 9 species of mosquitoes (5 anopheline, 2 *Aedes*, 1 *Culex*, and 1 *Armigeres* species) were collected from 6 villages in the Bulandshahr district (Table 2) of UP. Out of the 416 adult mosquitoes collected, 126 (30%) were *Anopheles* with *Anopheles subpictus* (21%) and *Anopheles culicifacies* (5%) dominating the anopheline burden. *Culex quinquefasciatus* was the major mosquito species found in rural Bulandshahr (235/416, 57%). Among the anophelines, *A subpictus* (PMHD 16.61) was the most abundant species followed by the main rural vector *A culicifacies* (PMHD 4.25) and *A stephensi* (PMHD 1.35).

Table . Adult mosquito collection and species identification in the surveyed 6 villages of Bulandshahr district of Uttar Pradesh during October 2019. The proportion of different mosquito species identified out of the total is shown as a percentage (in parentheses).

		Mosquito species identified									
		Aedes aegypti, n (%)	*Aedes alboipictus*, n (%)	*Anopheles stephensi*, n (%)	*Anopheles annularis*, n (%)	*Anopheles culicifacies*, n (%)	*Anopheles nigerrimus*, n (%)	*Anopheles subpictus*, n (%)	*Culex quinquefasciatus*, n (%)	*Armigeris subalbatus*, n (%)	Total, n
Villages											
	Adauli	1 (0.7)	0 (0)	0 (0)	1 (0.7)	9 (7)	4 (3)	30 (22)	86 (63)	6 (4)	137
	Lakhaoti	0 (0)	1 (2)	0 (0)	4 (6)	5 (8)	0 (0)	12 (18)	26 (40)	18 (27)	66
	Shikarpur (Kot Kalan 1)	4 (6)	2 (3)	5 (8)	0 (0)	4 (6)	0 (0)	6 (9)	45 (68)	0 (0)	66
	Kutubpur	0 (0)	0 (0)	0 (0)	1 (4)	0 (0)	0 (0)	5 (19)	20 (77)	0 (0)	26
	Mustafabad Daduwa	0 (0)	0 (0)	2 (5)	0 (0)	2 (5)	1 (2)	0 (0)	18 (41)	21 (47)	44
	Dinoul	0 (0)	0 (0)	0 (0)	0 (0)	2 (3)	0 (0)	33 (43)	40 (52)	2 (3)	77
Total		5 (1)	3 (0.7)	7 (2)	6 (1)	22 (5)	5 (1)	86 (21)	235 (57)	47 (11)	416

A total of 24 collected anopheline mosquito pools (22 mosquitoes in 6 pools of *A culicifacies*, 7 mosquitoes in 2 pools of *A stephensi*, 3 mosquitoes in 3 pools of *Anopheles annularis*, 86 mosquitoes in 11 pools of *A subpictus*, and 5 mosquitoes in 2 pools of *Anopheles nigerrimus*) were tested by PCR for the presence of malaria parasites; however, none of the pools was found positive for the presence of malaria parasites.

The vector survey to identify mosquito breeding habitats in Raj Nagar and Bagdola catchment localities revealed that out of 14,333 containers (including large containers, cemented tanks, and underground tanks) checked, *Anopheles* breeding was found only in 8 containers that included coolers, overhead tanks, cement tanks, and iron containers. There were no large water bodies in the surrounding area of the survey.

A mosquito breeding habitat survey in the 6 villages of Bulandshahr found that out of the 203 water-holding containers and water bodies, 51 (25%) had *Anopheles* breeding. Major breeding sites included drains (4/7, 57%), canals (1/2, 50%), ponds (3/6, 50%), and pits (9/20, 45%). Other sites where breeding was found included domestic and peridomestic water bodies (26/123, 21%), cemented ground tanks (cattle tanks; 5/26, 19%), and rice fields (3/19, 16%). *A culicifacies* were found mostly in canal banks and village ponds whereas *C quinquefasciatus* in sewages.

Fever Surveys

Camp-based fever surveys in the 6 villages identified 86 persons with fever, with only 1 person testing positive for *P vivax* by the rapid diagnostic tests. RACD from 22 asymptomatic persons around the households of the index cases revealed 5 additional cases (5/22, 23%) of *P vivax* by microscopy.

Discussion

Principal Findings

Between January 2017 and December 2019, 355 monoinfected *P vivax* cases were reported, and out of them, 63% (n=223) could be categorized as possible imported malaria cases based on relevant travel history, thus forming a major burden of reported malaria cases in SWD. The study also detected 5 additional *P vivax* cases through RACD done in villages visited by the imported cases and identified malaria vectors of anopheline species and their breeding habitats in such areas.

The distribution of malaria cases reported in the fever clinic at ICMR-NIMR revealed that the malaria cases were more likely to be imported than indigenous and occur in transmission season. The period July to November is considered to be the malaria transmission season in Delhi, while December to June is considered a nontransmission season [30].

Although 67% of the *P vivax* cases were imported, being associated with relevant travel history, the remaining 37% of indigenous cases could be associated with possible local transmission of *P vivax* in SWD, as suggested by the presence of anophelines in Delhi (this study) and the reported presence of malaria vectors in Delhi [31,32]. *P vivax* malaria cases during the nontransmission season or in nontravelers might also be recurrences or relapses due to the activation of hypnozoites from the liver [30].

Recent travel within the country is associated with malaria in various studies [22,23]. This study showed that the proportion of males was more than females among imported as well as indigenous malaria cases and a similar trend was seen in patients with fever as well. In similar studies, men traveling away from home in the last 30 days were reported to be strongly associated with malaria in Ethiopia [24,25].

Although Delhi shares its borders with the state of UP in the east and the state of Haryana in the remaining directions, we observed that ≈96% of the imported cases were from UP. Data highlights of the census of India in 2001 and 2011 show that Delhi receives a higher number of migrants (≈50% of the total in-migrants) from UP versus that from Haryana (≈10% of in-migrants) [33]. With >20-fold higher malaria burden in UP (than in Haryana), the findings of >95% of cases being imported from UP are explainable [12]. Reasons for migration to Delhi are cited to be due to employment, business, education, marriage, etc [33,34]. The reasons for travel reported during this study were festivals, farming, and visits to relatives. Those visiting friends and relatives in malaria-endemic areas have been reported to be at high risk of contracting malaria [35,36].

Many districts in UP contributed to the imported *P vivax* cases in SWD (Figure 2); however, 6 UP districts contributed 10 or more cases: Hathras (10 cases), Mainpuri (12 cases), Bareilly (15 cases), Bulandshahr (25 cases), Aligarh (37 cases), and Badaun (47 cases). Further investigations (vector and fever surveys) were carried out in 6 villages of Bulandshahr district only due to reasons explained earlier. Bulandshahr district of UP, located southeast of Delhi, is situated between the Ganga and Jamuna rivers, which are the major rivers in North India. The soil is very fertile where mainly sugarcane, and wheat are grown. Irrigation facilities are also well-developed and this area is canal-irrigated as well [37] which makes the area highly mosquitogenic.

During vector surveillance in 6 villages of Bulandshahr, 25% of the water bodies were positive for anopheline larval breeding, and 5 species of adult *Anopheles* mosquito were found during adult mosquito collections. Unlike Bulandshahr, where almost every village had ponds, canals, and ample water in surrounding areas providing sufficient opportunities for the breeding of anophelines, Delhi is highly urbanized and lands are not available for ponds and crop fields. In comparison to Bulandshahr, the catchment areas of the fever clinic (Raj Nagar and Bagdola localities) of SWD had a very low prevalence of *Anopheles*. Only 1 species, that is, *A stephensi* was present in these localities in contrast to Bulandshahr where 5 species of *Anopheles* were collected out of which 2 were major malaria vectors, that is, *A stephensi* and *A culicifacies*. Larval surveys suggested that urban and rural areas have different breeding habitats. In villages, natural water bodies like ponds, canals, pits, and crop fields were more prominent and harbored more breeding than the peridomestic and domesticated containers in contrast to the urban areas where natural breeding sites are limited and were confined to peridomestic and domestic containers only. Mosquito species like *A stephensi* and *A aegypti* are adapted to breed in such urban areas whereas *A culicifacies* mostly breed in outdoor natural water habitats like canal banks, village ponds, etc and *C quinquefasciatus* is found in sewage water.

Among malaria vectors, *A culicifacies* was found to be the dominating mosquito species along with an efficient malaria-transmitting vector, *A stephensi*. However, the month of the survey (October) had low vector density, which may be due to the low ambient temperature (20 - 25 °C) during the survey period. Further, the mosquitoes that were collected from the villages of Bulandshahr district did not show parasite positivity by PCR. This may be due to multiple factors, including the very short period of vector survey (30 d), a limited number of vectors collected toward the end of transmission (October), and the difference in time of mosquito collection and case reporting in the clinic, as the vector survey was carried out as a response for tracking ongoing transmission in areas previously visited by imported malaria cases.

The 6 districts of UP that contributed most to imported cases in SWD had API (2018) of 0.06 (Bulandshahr and Mainpuri), 0.1 (Hathras and Aligarh), 5.5 (Badaun), and 7.3 (Bareilly), whereas the API of Delhi during this period was 0.02 [38]. A survey was therefore, carried out in Bulandshahr wherein 1 out of 86 febrile cases (fever survey) and 5 out of 22 afebrile persons (RACD) were identified with *P vivax* infections which signifies that further studies are needed to assess the extent of asymptomatic *Plasmodium* infection and its role in transmission in such areas.

The prevalence of malaria was found lower in the camp-based fever surveys compared with the prevalence reported from the fever clinic in SWD. This may be because the camp-based fever surveys were carried out during October, which marks the end of the transmission season and therefore may have had a lower number of cases. Further, the catchment area of the camp included a village whereas the fever clinic at SWD has a much larger and densely populated catchment area.

The regions nearing malaria elimination tend to have a heterogeneous endemicity, with foci of high burden and areas with no endogenous malaria transmission. For eliminable diseases such as malaria, within-country migration is a recognized but understudied challenge in such geographically heterogeneous transmission to sustain zero-burden and prevent reintroduction and re-establishment of transmission [23,39]. Such regions often lack a robust surveillance system to deal with imported cases besides treating them, and there appears to be a lack of documented cross-reporting and targeted intervention in the foci where the infections probably originated.

This study is therefore important, as it attempted to comprehensively investigate imported malaria cases, and may be adopted and locally adapted as an implementation model in similar areas with no or few locally acquired malaria cases.

Limitations

There is an obvious limitation of this study that a limited geographical area for fever and vector surveillance was selected. Nevertheless, the study shows the presence of malaria transmission in areas where patients with malaria reporting to the fever clinic had traveled. The study was also limited by the possibility of recall bias of study participants correctly recalling the exact dates of travel for both imported and indigenous patients with malaria. The investigators, however, tried to extract the near-exact dates by relating travel to the locally relevant cultural events, festivals, and other contextual events. Misclassification bias (incorrect classification of imported malaria) could have stemmed from the possibilities of recurrences and relapses of *P vivax* infections acquired before the study period. The study did not use available molecular methods to differentiate recurrent versus new infection and therefore could not account for *P vivax* relapses. However, the possibility that only up to 40% of *P vivax* infections in the study area (that too in the nontransmission season) could be due to possible relapses [30], nonavailability of molecular methods to confidently differentiate recurring infections from new infections, and random possibilities of recurrence in both the imported and indigenous patients with malaria may have compensated for this limitation.

Last, only 2 methods for mosquito collection were used. The hand catch method using an aspirator was the only method adopted for the estimation of vector density. For larval collection, dips were taken from water bodies for assessment of breeding. No other method was adopted for mosquito collection, and this might have underestimated the frequency of vectors and their possible infection with *Plasmodium*, because the PCR results did not show any vector positivity.

Conclusions

A significant burden (63%) of *P vivax* malaria reported in SWD was found to be imported from UP. Malaria transmission possibilities (multiple breeding sites suggesting stable breeding ground of anophelines) were higher in Bulandshahr than in SWD. Indigenous cases in SWD are also a concern, as adult vectors were also found in the area. Despite the detection of additional *P vivax* cases following RACD in Bulandshahr and vector-breeding sites being identified, the conclusion that the imported cases really acquired infections from the surveyed areas in Bulandshahr is limited by the correct recall of travel or fever dates and the possibility of relapses due to *P vivax*.

Way Forward

The study reiterates that population movement is a key challenge for malaria elimination, particularly in areas with very low or 0 indigenous malaria cases, and investigations of the potential role of travelers in introducing malaria and its further spread are definitely needed. Since the epidemiology of migration malaria is contextual, appropriate tailor-made measures are needed, both at the sites where imported cases are detected and in areas where these infections might have been acquired. In addition, effective “information, education, and communication” activities to educate travelers regarding the potential risks of travel-associated malaria and its prevention should be undertaken.

Acknowledgments

The authors acknowledge the support provided by the state programme officer of Uttar Pradesh and district malaria officer of Bulandshahr. The authors are thankful to the technical staff of fever clinic and vector biology division of ICMR-NIMR for their field support. The study did not receive any specific funding from a dedicated grant and was funded by Indian Council of Medical Research-National Institute of Malaria Research (ICMR-NIMR), New Delhi, out of its intramural funds.

Disclaimer

The views expressed in the submitted article are our own and not an official position of the institution or funder. No generative artificial intelligence was used in any portion of the manuscript writing.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

DS, HS, and A Sinha conceived the idea, carried out clinical work and fieldwork, and wrote the first draft. BS carried out laboratory work. CPY and MPS carried out analysis. AA and A Sinha analyzed and interpreted the data and reviewed the manuscript. A Sharma reviewed the manuscript. HS and A Sinha contributed equally as co-corresponding authors.

Conflicts of Interest

None declared.

References

1. World malaria report 2022. : World Health Organization; 2022 URL: https://www.who.int/publications/i/item/9789240064898 [accessed 2023-06-01]
2. Chapter 2 – migration and migrants: a global overview. World Migration Report 2018. : International Organization for Migration (IOM); 2017 URL: https://onlinelibrary.wiley.com/doi/10.1002/wom3.2 [accessed 2024-04-12] [doi: 10.1002/wom3.2]
3. Lu G, Zhang D, Chen J, et al. Predicting the risk of malaria re-introduction in countries certified malaria-free: a systematic review. Malar J 2023 Jun 6;22(1):175. [doi: 10.1186/s12936-023-04604-4] [Medline: 37280626]
4. Requena-Méndez A, Cattaneo P, Bogale RT, et al. Malaria parasite prevalence among migrants: a systematic review and meta-analysis. Clin Microbiol Infect 2023 Dec;29(12):1528-1537. [doi: 10.1016/j.cmi.2023.09.010] [Medline: 37739263]
5. Cohen JM, Smith DL, Cotter C, et al. Malaria resurgence: a systematic review and assessment of its causes. Malar J 2012 Apr 24;11:122. [doi: 10.1186/1475-2875-11-122] [Medline: 22531245]

6. Pousibet-Puerto J, Lozano-Serrano AB, Soriano-Pérez MJ, et al. Migration-associated malaria from Africa in southern Spain. Parasit Vectors 2021 May 7;14(1):240. [doi: 10.1186/s13071-021-04727-0] [Medline: 33962647]
7. Rodrigues PT, Valdivia HO, de Oliveira TC, et al. Human migration and the spread of malaria parasites to the new world. Sci Rep 2018 Jan 31;8(1):1993. [doi: 10.1038/s41598-018-19554-0] [Medline: 29386521]
8. Wångdahl A, Wyss K, Saduddin D, et al. Severity of Plasmodium falciparum and non-falciparum malaria in travelers and migrants: a nationwide observational study over 2 decades in Sweden. J Infect Dis 2019 Sep 13;220(8):1335-1345. [doi: 10.1093/infdis/jiz292] [Medline: 31175365]
9. Sorichetta A, Bird TJ, Ruktanonchai NW, et al. Mapping internal connectivity through human migration in malaria endemic countries. Sci Data 2016 Aug 16;3:160066. [doi: 10.1038/sdata.2016.66] [Medline: 27529469]
10. Prothero RM. Migration and malaria risk. Health Risk Soc 2001 Mar 1;3(1):19-38. [doi: 10.1080/13698570123413]
11. Tyagi BK. Trans-border migration and malaria in desert populations. In: Desert Malaria: Springer; 2023. [doi: 10.1007/978-981-19-7693-3_21]
12. Country Office for India, World Health Organization. WHO Country Office for India. 2016. URL: https://iris.who.int/handle/10665/246096 [accessed 2023-05-31]
13. Sharma S, Verma R, Yadav B, Kumar A, Rahi M, Sharma A. What India can learn from globally successful malaria elimination programmes. BMJ Glob Health 2022 Jun;7(6):e008431. [doi: 10.1136/bmjgh-2022-008431] [Medline: 35760440]
14. Rahi M, Sharma A. Malaria control initiatives that have the potential to be gamechangers in India's quest for malaria elimination. Lancet Reg Health Southeast Asia 2022 Jul;2:100009. [doi: 10.1016/j.lansea.2022.04.005] [Medline: 37383297]
15. Khan N, Awasthi G, Das A. How can the complex epidemiology of malaria in India impact its elimination? Trends Parasitol 2023 Jun;39(6):432-444. [doi: 10.1016/j.pt.2023.03.006] [Medline: 37031071]
16. Kumar A. Some considerable issues concerning malaria elimination in India. J Vector Borne Dis 2019;56(1):25-31. [doi: 10.4103/0972-9062.257770] [Medline: 31070162]
17. Malaria in migrants and mobile populations. World Health Organization. URL: https://www.who.int/malaria/areas/high_risk_groups/migrants_mobile_populations/en [accessed 2023-05-31]
18. Karim MA, Kabir MM, Siddiqui MA, Laskar MSI, Saha A, Naher S. Epidemiology of imported malaria in Netrokona district of Bangladesh 2013-2018: analysis of surveillance data. Malar Res Treat 2019;2019:6780258. [doi: 10.1155/2019/6780258] [Medline: 31312425]
19. Migrated population in delhi from 2003 to 2022. Statista. URL: https://www.statista.com/statistics/1400863/india-migration-in-delhi/ [accessed 2023-05-31]
20. Singh J, Bhattacharji LM. Rapid staining of malarial parasites by a water soluble stain. Ind Med Gaz 1944 Mar;79(3):102-104. [Medline: 29012018]
21. Anvikar AR, Arora U, Sonal GS, et al. Antimalarial drug policy in India: past, present & future. Indian J Med Res 2014 Feb;139(2):205-215. [Medline: 24718394]
22. Poumerol G, Wilder-Smith A, editors. Vaccines for routine and selective use: rabies. In: International Travel and Health: World Health Organization; 2012. URL: https://www.who.int/publications/i/item/9789241580472 [accessed 2023-05-31]
23. Lynch CA, Bruce J, Bhasin A, Roper C, Cox J, Abeku TA. Association between recent internal travel and malaria in Ugandan highland and highland fringe areas. Trop Med Int Health 2015 Jun;20(6):773-780. [doi: 10.1111/tmi.12480] [Medline: 25689689]
24. Alemu K, Worku A, Berhane Y, Kumie A. Men traveling away from home are more likely to bring malaria into high altitude villages, northwest Ethiopia. PLoS One 2014;9(4):e95341. [doi: 10.1371/journal.pone.0095341] [Medline: 24748159]
25. Yukich JO, Taylor C, Eisele TP, et al. Travel history and malaria infection risk in a low-transmission setting in Ethiopia: a case control study. Malar J 2013 Jan 24;12:33. [doi: 10.1186/1475-2875-12-33] [Medline: 23347703]
26. Christophers SR. The Fauna of British India, Including Ceylon and Burma. Diptera, Vol. Iv, Family Culicidae, tribe Anophelini: Taylor & Francis; 1933.
27. Manual on Practical Entomology in Malaria / Prepared by the WHO Division of Malaria and Other Parasitic Diseases: World Health Organization Division of Malaria and Other Parasitic Diseases; 1995. URL: https://iris.who.int/handle/10665/42481 [accessed 2024-12-04]
28. Tassanakajon A, Boonsaeng V, Wilairat P, Panyim S. Polymerase chain reaction detection of Plasmodium falciparum in mosquitoes. Trans R Soc Trop Med Hyg 1993;87(3):273-275. [doi: 10.1016/0035-9203(93)90124-9] [Medline: 8236389]
29. Malaria surveillance, monitoring & evaluation: a reference manual. : World Health Organization; 2018 URL: https://www.who.int/publications/i/item/9789241565578 [accessed 2024-12-04]
30. Adak T, Sharma VP, Orlov VS. Studies on the Plasmodium vivax relapse pattern in Delhi, India. Am J Trop Med Hyg 1998 Jul;59(1):175-179. [doi: 10.4269/ajtmh.1998.59.175] [Medline: 9684649]
31. Savargaonkar D, Nagpal BN, Srivastava B, Anvikar AR, Valecha N. The footprints of relapsing malaria in southwest Delhi, India. J Vector Borne Dis 2015 Dec;52(4):287-292. [Medline: 26714507]
32. Singh RK, Mittal PK, Kumar G, Dhiman RC. Insecticide susceptibility status of Aedes aegypti and Anopheles stephensi larvae against temephos in Delhi, India. Int J Mosq Res 2014;1(3):69-73.

33. Census of India 2001. Data highlights: migration tables (D1, D1 (Appendix), D2 and D3 tables). The Hindu Centre For Politics and Public Policy. 2001. URL: https://www.thehinducentre.com/publications/policy-watch/article33219541.ece/binary/data_highlights_D1D2D3.pdf [accessed 2023-05-31]
34. Sarkar P. An overview of out-migration from Uttar Pradesh using census 2011. J Migr Aff 2020;2(2):58. [doi: 10.36931/jma.2020.2.2.58-66]
35. Casuccio A, D'Angelo C, Casuccio N, Di Carlo P, Immordino P. Visiting friends and relatives (VFRs) travelers and imported malaria in the Palermo district (Sicily). Ann Ist Super Sanita 2014;50(4):369-374. [doi: 10.4415/ANN_14_04_13] [Medline: 25522079]
36. Heywood AE, Zwar N, Forssman BL, et al. The contribution of travellers visiting friends and relatives to notified infectious diseases in Australia: state-based enhanced surveillance. Epidemiol Infect 2016 Dec;144(16):3554-3563. [doi: 10.1017/S0950268816001734] [Medline: 27574034]
37. Preliminary draft report. District survey report for (planning & execution of) minor mineral excavation. : Directorate of Geology & Mining; 2016 URL: https://cdn.s3waas.gov.in/s3091d584fced301b442654dd8c23b3fc9/uploads/2019/02/2019021130.pdf [accessed 2021-01-11]
38. Directorate of National Vector Borne Diseases Control Programme, 22 Sham Nath Marg, DELHI-110054. Annual report 2018. National Center for Vector Borne Diseases Control (NCVBDC). 2018. URL: https://nvbdcp.gov.in/Doc/Annual-Report-2018.pdf [accessed 2024-12-04]
39. World malaria report 2023. : World Health Organization; 2023 URL: https://www.who.int/publications/i/item/9789240086173 [accessed 2024-12-04]

Abbreviations

API: annual parasite incidence
ICMR-NIMR: Indian Council of Medical Research-National Institute of Malaria Research
NCVBDC: National Center for Vector Borne Diseases Control
PCR: polymerase chain reaction
PMHD: per man-hour density
RACD: reactive case detection
SWD: South West Delhi
SWD: South West Delhi
UP: Uttar Pradesh
WHO: World Health Organization

Edited by A Mavragani, T Sanchez; submitted 18.06.23; peer-reviewed by J Campbell, P Sahu; revised version received 16.05.24; accepted 23.05.24; published 08.01.25.

Please cite as:
Savargaonkar D, Srivastava B, Yadav CP, Singh MP, Anvikar A, Sharma A, Singh H, Sinha A
Contribution of Travelers to Plasmodium Vivax Malaria in South West Delhi, India: Cross-Sectional Survey
JMIR Public Health Surveill 2025;11:e50058
URL: https://publichealth.jmir.org/2025/1/e50058
doi:10.2196/50058

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org