



**HAL**  
open science

# Leveraging administrative health databases to address health challenges in farming populations: a scoping review and bibliometric analysis (1975-2024)

Pascal Petit, Nicolas Vuillerme

## ► To cite this version:

Pascal Petit, Nicolas Vuillerme. Leveraging administrative health databases to address health challenges in farming populations: a scoping review and bibliometric analysis (1975-2024). *JMIR Public Health and Surveillance*, In press. hal-04848292

**HAL Id: hal-04848292**

**<https://hal.science/hal-04848292v1>**

Submitted on 19 Dec 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

1 **Title:** Leveraging administrative health databases to address health challenges in farming populations:  
2 a scoping review and bibliometric analysis (1975-2024)

3

4 **Authors**

5 Pascal Petit<sup>1</sup>; Nicolas Vuillerme<sup>1,2</sup>

6 <sup>1</sup> Univ. Grenoble Alpes, AGEIS, 38000 Grenoble, France

7 <sup>2</sup> Institut Universitaire de France, Paris, France

8

9 **Correspondence (present address)**

10 \* Corresponding author at: Pascal Petit. Laboratoire AGEIS - Université Grenoble Alpes. Bureau 315.

11 Bâtiment Jean Roget. UFR de Médecine. Domaine de La Merci. 38706 La Tronche Cedex, France.

12 E-mail address: [pascal.petit@univ-grenoble-alpes.fr](mailto:pascal.petit@univ-grenoble-alpes.fr) (P. Petit). Telephone: + 33 4 76 63 71 04.

13

14 **ORCID**

15 PP: 0000-0001-9015-5230

16 NV: 0000-0003-3773-393X

Accepted paper (not yet proofread)

## 17 **Abstract**

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

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

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

33 *Results:* Between 1975 and April 2024, 296 publications across 118 journals from 34 countries were  
34 identified, predominantly from high-income countries. Nearly one-third of these publications were  
35 associated with well-established cohorts such as AGRICAN and AHS. The most frequently used AHDs  
36 included disease registers (158/296, 53.4%), electronic health/medical records (124/296, 41.9%),  
37 insurance claims (106/296, 35.8%), population registers (95/296, 32.1%), and hospital discharge  
38 (41/296, 13.9%). Most studies included only adult participants. Fifty studies involved over 1 million  
39 participants. Although a broad range of exposure proxies were used, most studies relied on broad  
40 proxies, which failed to capture the specifics of farming tasks. Research on the farming exposome  
41 remains underexplored, with a predominant focus on the specific external exposome, in particular  
42 pesticide exposure. Furthermore, a limited range of health events have been examined, primarily cancer,  
43 mortality, and injuries.

44 *Conclusions:* The increasing use of AHDs holds major potential to advance public health research within  
45 farming populations. However, substantial research gaps persist, particularly in low-income regions and  
46 among underrepresented farming subgroups, such as females, children, and contingent workers.  
47 Emerging issues, including exposure to PFAS, biological agents, microbiome, microplastics, and climate  
48 change, warrant further research. Major gaps also persist in understanding various health conditions,  
49 including cardiovascular, reproductive, ocular, sleep, age-related, and autoimmune diseases. Addressing  
50 these overlooked areas is essential for a more comprehensive understanding of the health risks faced by  
51 farming communities and to guide public health policies. Within this context, promoting AHD-based  
52 research, in conjunction with other digital data sources (e.g., mHealth, social health data, wearable  
53 sensors) and artificial intelligence approaches, represents a promising avenue for future exploration.

54

55 *Keywords:* farming population; digital public health; digital epidemiology; administrative health  
56 database; farming exposome; review; bibliometric analysis; digital data; data reuse; human health

Accepted paper (not yet proofread)

## 57 **Introduction**

## 58 **Background**

59 Public health research seeks to identify and understand the factors that influence population health to  
60 effectively prevent diseases and promote health and well-being for all [1,2]. A broad range of  
61 environmental determinants can impact health across the lifespan. One of the core areas of public health  
62 research, known as the exposome, investigates how cumulative environmental influences contribute to  
63 disease etiology and pathogenesis. [3-18]. The exposome, which complements genomic research, refers  
64 to the comprehensive examination of all environmental exposures experienced throughout an  
65 individual's lifetime, including physical, chemical, biological, psychosocial, and behavioral factors,  
66 from conception to death [3-18]. The exposome classically includes three overlapping domains: the  
67 general external exposome (e.g., climate, built environment), the specific external exposome (e.g.,  
68 chemical exposure, lifestyle, occupations), and the internal exposome (e.g., aging, oxidative stress,  
69 metabolism, gut microbiome) [8,14,16,17,19,20]. Understanding the exposome is crucial for enabling  
70 both population-wide and precision prevention [3,21-23]. However, fully describing the exposome is  
71 challenging due to the vast diversity and the temporal and spatial variability of environmental factors  
72 [3]. Public health research in this area requires data on both risk factors and adverse health outcomes to  
73 progress effectively [3,14,24,25].

74 The volume of data collected has grown exponentially as the world becomes increasingly reliant on  
75 technology and digitization [26,27]. Data is omnipresent in our everyday lives, leading science towards  
76 data-driven research [27,28], in particular in the health field. The digital transformation in healthcare  
77 has enabled unprecedented data availability, collection, storage, and analysis capabilities, leading to a  
78 paradigm shift in healthcare systems, with entire care pathways becoming digitized [29,30]. Health-  
79 related data now represent approximately 6% of all digital data globally, a figure that continues to rise  
80 [31]. This explosion of data has transformed research, providing new opportunities, especially in public  
81 health, to enhance disease understanding and evaluate intervention effectiveness [27,28,32-35]. The  
82 integration of digital technologies and digital data in public health has led to the emergence of “digital  
83 public health”, an evolving field focused on using digital data to achieve public health goals [33,36-41].

84 Public health research is moving from isolated data systems to more integrated, accessible, and reusable  
85 data resources [42]. Reusing data allows researchers to explore various health determinants—including  
86 environmental, occupational, behavioral, and organizational factors, fostering a holistic approach to  
87 disease prevention and health promotion strategies [14].

88 Within the digital public health framework, two main types of data are being used, namely primary and  
89 secondary data. Primary data are tailor-made, designed for a specific purpose, and often used once or  
90 repeatedly for the same goal [43-45]. Primary data are the cornerstone of traditional public health policy  
91 and decision-making. These data are derived from several types of studies [46-50], in particular  
92 observational cohorts (e.g., the Framingham Cohort study [51,52]) [46-50,53,54], case-control studies  
93 [46-50], cross-sectional surveys (e.g., the China Health and Retirement Longitudinal Study [55,56]) [46-  
94 50,53], and experimental studies [46-50]. Primary data have many advantages [46-50]. They are rich, of  
95 high quality, and designed to answer specific research questions for public health and epidemiological  
96 purposes. Primary data are usually available at the individual level and are derived from studies that  
97 control for certain biases. By contrast, they are cumbersome, time-consuming, and costly to set up and  
98 maintain [53,54,57]. The representativeness of primary data is also limited in size, geographical scope,  
99 and temporal coverage and can erode over time [46-50,53]. Primary data are not free from bias, such as  
100 selection/attrition, healthy worker, recall, or prevarication bias [53,58].

101 Unlike traditional public health, digital public health does not rely solely on primary data but takes  
102 advantage of the myriad of existing digital data that have not been generated originally for research  
103 purposes (i.e., secondary data) to overcome some limitations intrinsic to primary data and complement  
104 them [28,43,44,53,59-68]. Indeed, some data can have an additional impact when used beyond the  
105 context for which they were originally created [68,69]. Secondary data are collected for purposes other  
106 than public health or epidemiology and include contextual data (e.g., air quality, climate data)  
107 [14,24,26,29,70-74], person-generated data (e.g., social media, crowdsourcing, mHealth)  
108 [2,24,26,31,43,61,62,73,75-86], synthetic data (e.g., digital twin) [87-91], and administrative health  
109 databases (AHDs) [26,64,68,81,92-102].

110 AHD is a broad term encompassing a wide range of routinely collected data on individuals' health and  
111 sociodemographic information collected for registration, billing, record-keeping, and other

112 administrative purposes [26,64,81,93,95,98,100-102]. For this review, based on previous works  
 113 [93,95,96,103-105], AHDs included population register, claims database, disease register, electronic  
 114 health/medical record, and hospital discharge database that were collected at a local, regional, national,  
 115 or international level (Table 1) [26,61,62,93,95,96,100,103-110]. AHDs offer many advantages for  
 116 research. Such data is collected as part of routine administrative processes, reducing additional costs for  
 117 researchers. AHDs therefore offer relatively inexpensive access to a large number of individuals that can  
 118 be tracked over time for several years, guaranteeing the representativeness of the populations studied  
 119 [26,54,78,93,95,104,105,111-114]. Data recorded within AHDs is structured, coded in a standardized  
 120 way, and less affected by participation and recall biases [54,58,95,106,113,115]. AHDs enable the study  
 121 of rare events and populations underrepresented in studies using only primary data [95,111-113]. AHDs  
 122 have limitations inherent to their nature, such as the absence of some confounding factors, the limited  
 123 granularity of certain information, the data complexity, as well as confidentiality issues  
 124 [73,78,93,95,115-125].

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

	Population register	Claims database	Disease register	Electronic health/medical record	Hospital discharge database
<b>Definition/ content</b>	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	continuous and exhaustive digital collection of individual data regarding one 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 utilization with information about patients, their care, and their stay in the hospital
<b>Source</b>	local or national authorities	insurance programs or schemes, healthcare providers	healthcare institutions (e.g., hospitals)	hospitals, physicians, healthcare centers and institutions	hospitals
<b>Population</b>	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 at a geographically defined scale	all patients using the healthcare system	all patients from a hospital
<b>Purpose/ finality</b>	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
<b>Health event</b>	none	health events covered by insurance or a healthcare provider	specific health events (e.g., cancer)	health events requiring care that are reported in medical records	health events from hospital admission

127

## 128 **Rationale**

129 AHDs are increasingly employed in population-based health research due to their complementarity with  
 130 traditional sources of public health and epidemiological data (i.e., primary data)  
 131 [42,59,64,87,93,95,96,126-128]. The reuse of AHDs—referring to their application beyond their

132 original or intended purpose—holds major potential to advance public health and epidemiological  
133 research, offering insights that can guide public health decision-making [42,59,64,87,93,95,96,105,126-  
134 131]. Although several reviews have previously explored the general use of AHDs in research  
135 [42,95,107,129,132-135], others have focused on their application within specific countries [96,136],  
136 examined individual AHDs [108,137], or investigated their role in studying specific diseases and adverse  
137 health outcomes [44,93,104,138-142]. However, to our knowledge, no study has synthesized how AHDs  
138 are reused for epidemiological and public health research within a specific population group.

139 To address this gap, we conducted a comprehensive scoping review and bibliometric analysis aimed at  
140 identifying how AHDs are used to address health issues in a specific population. We selected farming  
141 populations as an illustrative example because they present unique health and disease trends [143-147].

142 Globally, approximately 27% of the workforce is engaged in occupational farming, and this group is  
143 exposed to numerous risk factors (exposome), including pesticides, biological agents, and limited access  
144 to healthcare [148]. These exposures put them at heightened risk for a wide range of adverse health  
145 outcomes [143,145,147,149]. Although agricultural safety and health have become a major public health  
146 issue in recent decades, most research on the health of farming populations has relied on traditional  
147 epidemiological and community-based studies, which often face limitations in terms of sample size,  
148 geographic scope, temporal coverage, and the range of health events examined [145,150,151].

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

156



## 157 **Methods**

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

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

### 170 **Search strategy and selection criteria**

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

182

183

184 **Textbox 1.** Search terms.

185 **Farming**

186 husbandry\* OR agriculture\* OR farming OR farm\* OR agricultural\* OR farmworker\*

187 **Administrative health databases (combined using AND)**

188 "health record" OR "health records" OR "digital record" OR "digital records" OR "health administrative register" OR "health  
189 administrative registry" OR "health register" OR "health registry" OR "medical register" OR "medical registry" OR "electronic  
190 health record" OR "electronic health records" OR "EHR" OR "EMR" OR "electronic medical record" OR "electronic medical  
191 records" OR "digital medical record" OR "digital medical records" OR "digital health record" OR "digital health records" OR  
192 "health administrative data" OR "health administrative database" OR "health administrative dataset" OR "health administrative  
193 datasets" OR "health administrative databases" OR "administrative health data" OR "administrative health database" OR  
194 "administrative health dataset" OR "administrative health datasets" OR "administrative health databases" OR "insurance data"  
195 OR "insurance database" OR "insurance databases" OR "insurance dataset" OR "insurance claim" OR "insurance claims" OR  
196 "cancer registry" OR "cancer register" OR "health insurance" OR "health surveillance program" OR "health surveillance  
197 programs" OR "Mutualite Sociale Agricole" OR "MSA" OR "health insurance system" OR "record-linkage" OR "population  
198 register" OR "population registry" OR "insurance scheme" OR "social security scheme" OR "hospital discharge" OR  
199 "administrative claim" OR "administrative claims" OR "medical claims" OR "medical claim" OR "electronic claim" OR  
200 "electronic claims" OR "mortality register" OR "mortality registry" OR "mortality record" OR "mortality records" OR "disease  
201 register" OR "disease registry" OR "illness register" OR "illness registry" OR "disorder register" OR "disorder registry"

202

203 To develop the eligibility criteria, an initial search of the literature was conducted on PubMed, with the  
204 review of the first 100 articles that used AHDs for public health and epidemiological research. In our  
205 pilot run, disease/morbidity registers were not considered as AHD because they were created for clinical  
206 and research purposes [47-50,53,76,155,156]. However, because disease registers contain some  
207 information derived from medical records, we decided to consider them as AHD for this review. The  
208 eligibility criteria are presented in Table 2. The search was restricted to original peer-reviewed records  
209 (all types were included) written in English or French but not constrained by the year of publication  
210 [93,106,157]. Publications that examined partly farming populations, with, for instance, studies  
211 reporting health risks for various sectors of activity, were included.

212

213

214

215 **Table 2.** Eligibility criteria for selection of publications

Criteria type	Inclusion Criteria	Exclusion Criteria
Article type	<ul style="list-style-type: none"> <li>- Data had to originate at least partly from AHD.</li> <li>- Study had to pertain at least partly to the farming population.</li> <li>- Study had to relate to public health or epidemiological research.</li> <li>- Original peer-reviewed publications.</li> </ul>	<ul style="list-style-type: none"> <li>- Publications not describing the use of AHD.</li> <li>- Animal or in vitro studies.</li> </ul>
Language	- Publications in English or French.	- Publications not in English or French.

216

217 The final literature search was conducted on both PubMed and Web of Science Core Collection  
 218 databases. Regarding Web of Science Core Collection, a topic search was performed. To reduce the bias  
 219 induced by daily database changes, all data collection (literature retrieval and data download) was  
 220 conducted and completed on the same day, April 15, 2024. Titles, abstracts, and full-text publications  
 221 were screened on the basis of pre-established inclusion and exclusion criteria. The inclusion criteria for  
 222 each phase of the literature search are detailed in Multimedia appendix 2. When abstracts did not contain  
 223 enough information about correspondence to inclusion or exclusion criteria, the article was considered  
 224 for full-text review. Reference lists of included publications were not searched, although they might  
 225 have also yielded new relevant studies.

226

227 **Data collection and processing**

228 Twenty-nine metadata were extracted from each publication included in the scoping review (Table 3).  
 229 The data underwent rigorous manual validation, cleaning, and harmonization following a structured  
 230 five-step process. First, duplicate items (e.g., keywords, institutions) were removed. Second, leading  
 231 and trailing white spaces were eliminated. Third, items were standardized by converting text to  
 232 lowercase, with only the first letter capitalized. In the fourth step, items were harmonized to either  
 233 singular or plural forms consistently. Finally, synonyms or terms with similar meanings (e.g., “illness”  
 234 and “disease”) were unified under a single term. For instance, “Pesticide”, “Pesticide exposure”, and  
 235 “Pesticide use” were standardized to “Pesticide”, while “Pulmonary disease copd”, “Copd”, and  
 236 “Chronic obstructive pulmonary disease” were unified as “COPD”. For cancer-related keywords, the  
 237 11<sup>th</sup> International Classification of Diseases was used to consolidate varied terms (e.g., “lung cancer”,  
 238 “lung cancer risk”, “lung and bronchus cancer”, “lung tumor”, “lung tumour”, “lung neoplasm”, “basal

239 cell carcinoma of the lung”), into standard categories (e.g., lung cancer). Quality appraisals were not  
 240 performed because they were beyond the aim of this review [106,157].

241

242 **Table 3.** List of meta-data of interest to collect from the literature search

Meta-data	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, pesticide compounds
Number of farming activities studied	10
Number of pesticide compounds studied	29
Population	adults
Sex	female
Number of participants included	[100-1000]
Country	France
Oldest data used	1991
Most recent data used	2020
Data follow-up period (in years)	4
Number of years between the most recent data used and publication year	7
Disease/Health events	Parkinson's 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
Number of citations	14

243

## 244 **Data analysis**

245 To analyze the research trends (i.e., hotspots and gaps) on the use of AHDs for public health and  
 246 epidemiological research in farming populations, a bibliometric approach was conducted [158-160].

247 This analysis examined the number of publications, countries of publications, most active journals,  
 248 institutions, authors, funding bodies, subject areas, citations of publications, and keywords of  
 249 publications (Figure 1). Seven bibliometric metrics were computed, including the h-, g-, m-, and Y-  
 250 indices, dominance factor, annual growth rate (AGR), and fractionalized frequency (Multimedia  
 251 appendix 3). The h index attempts to measure both the productivity and citation impact of the published  
 252 body of work of an entity (e.g., author, institution, journal) [161,162]. It refers to the total number of  
 253 publications by a particular entity with at least the same number of citations. The m index is calculated  
 254 by dividing the h index by the number of years of an entity's productive life (e.g., researcher) [161]. The

255 g index of an entity corresponds to the largest number  $g$  such that the top  $g$  publications have at least  $g^2$   
256 or more citations together [162]. The Y index refers to the sum of both the total number of first-authored  
257 publications and the total number of corresponding-author publications [163]. The dominance factor  
258 (DF) refers, for a particular researcher, to the proportion of multi-authored publications as specific  
259 author's rank to the total number of multi-authored publications [164]. The fractional frequency (FF)  
260 intends to reflect/measure an author's contribution. The AGR refers to the variable's change in  
261 percentage as a year-over-year statistic [165]. The most up-to-date journals' impact factors and ranks  
262 were retrieved manually using the Journal Citation Report in April 2024.

263 Spearman correlations were calculated to examine the association between the number of publications  
264 and GDP (gross domestic product), population size [166], as well as the total labor force, number of  
265 researchers in R&D (research and development) (per million people), fertilizer consumption (in both %  
266 of fertilizer production and kilograms per hectare of arable land), agricultural land ( $\text{km}^2$ ), agricultural  
267 land (% of land area), land under cereal production (hectares), permanent cropland (% of land area),  
268 cereal production (metric tons), crop production index, food production index, livestock production  
269 index, cereal yield (kg per hectare), female employment in agriculture (% of female employment), male  
270 employment in agriculture (% of male employment), employment in agriculture (% of total  
271 employment), and agriculture, forestry, and fishing, value added (% of GDP). These country  
272 characteristics were obtained from the World Bank. When available, the most recent country  
273 characteristic (e.g., GDP) was used.

274 Research trends, including hotspots and gaps, were investigated with keyword frequency, co-occurrence  
275 (counting of paired keywords), and thematic mapping analyses. Thematic mapping and keyword co-  
276 occurrence network are two complementary but distinct approaches that serve different purposes and  
277 offer different insights. In summary, thematic mapping is more focused on the strategic positioning of  
278 research themes within a field, while keyword co-occurrence networks emphasize the relationships and  
279 connections between specific keywords in the literature [158,167]. Both methods complement each  
280 other and are usually used to provide a more comprehensive understanding of research landscapes. The  
281 co-occurrence of two keywords was defined by the frequency with which they appear together in  
282 publications and was quantified using association strength (AS) or equivalent index, calculated as

283  $AS_{ij} = \frac{c_{ij}^2}{c_i c_j}$ , where  $c_{ij}$  is the number of publications in which keywords  $i$  and  $j$  co-occur, while  $c_i$  and  
284  $c_j$  are the number of publications in which each keyword appears, respectively [158,167]. AS measures  
285 how close two keywords are to each other. An AS value of 1 indicates keywords always appear together,  
286 while 0 indicates they never co-occur. These keyword co-occurrences can be visualized using a co-  
287 occurrence network graph, where a vertex/node represents a keyword, the size of the node represents  
288 the keyword frequency, and the edge represents the association between two keywords [158,167]. Based  
289 on the keyword co-occurrence network graph, a community detection procedure can be used to identify  
290 groups of words highly associated with each other [158,167]. In other words, equivalent keywords,  
291 based on AS, can be grouped together to identify research themes [158,167]. A strategic diagram or  
292 thematic map is based on Callon's centrality (x-axis) and Callon's density (y-axis) [158,167]. Callon's  
293 centrality measures the degree of interaction of a theme with other themes. It is defined as  
294 *Callon's centrality* =  $10 \times \sum AS_{kh}$ , where  $k$  is a keyword belonging to a theme and  $h$  is a keyword  
295 belonging to another theme [158,167]. Callon's centrality can be interpreted as an indicator of the  
296 importance of a particular topic within the broader/overall research landscape. Callon's density measures  
297 the internal strength of a theme. It is defined as *Callon's density* =  $100 \times \sum \frac{AS_{ij}}{w}$ , where  $i$  and  $j$  are  
298 keywords belonging to the same theme and  $w$  is the total number of keywords in a theme [158,167].  
299 Callon's density serves as a metric for assessing the progression and maturation of that topic [158,167].  
300 A strategic diagram is divided into four quadrants according to Callon's centrality and density values,  
301 which correspond to four types of topics. Hotspots or hot topics are defined by both a high density and  
302 centrality value (upper right quadrant), while basic topics are defined by a high centrality value but a  
303 low density value (lower right quadrant). Peripheral topics are defined by both low centrality and density  
304 values (lower left quadrant), while niche topics are defined by low centrality and high density values  
305 (upper left quadrant) [158,167].  
306 To focus on agricultural/farming exposome research, a bibliometric profile/picture of the "farming  
307 exposome" was constructed, which restricts the exposome concept to environmental exposures specific  
308 to farming populations [152,168]. This bibliometric farming exposome picture examined co-

309 occurrences between keywords related to potential risk factors and specific health events (e.g., cancers,  
310 reproductive disorders).

311 The bibliometric analysis was conducted and reported according to the preliminary guideline for  
312 reporting bibliometric reviews of the biomedical literature (BIBLIO) (Multimedia appendix 4) [169].

313 All analyses were performed using R software 4.3.2® (R Core Team, Vienna, Austria) for Windows  
314 10©. The bibliometric analysis was performed using the bibliometrix R package version 4.1.4 [170].

315

## 316 **Results**

### 317 **Overview**

318 After excluding 4485 irrelevant records, 296 publications were analyzed (Figure 1). The majority were  
319 articles (293/296), with a small number of reviews (2/296) and editorial materials (1/296) (Table 4).

320 Only one-third of the publications were open access (Table 4 and Multimedia appendix 5). The average  
321 publication age was 14.2 years, ranging from the oldest in 1975 [171] to the most recent in April 2024

322 [152]. Since 1975, there has been a steady increase in publications utilizing AHDs to address health  
323 issues in farming populations, with an AGR of 5.2%. Notably, almost one-third of these articles (91/296)

324 were published in the last five years, highlighting the rising interest in AHD-based public health research  
325 in this population (Multimedia appendix 6). Collectively, the publications received 9379 citations,

326 averaging 31.7 citations per publication. Multimedia appendix 7 presents the historical direct citation  
327 network. The body of work involved 1225 authors from 338 institutions, with 1882 author appearances

328 and an average of six authors per paper (Table 4). There were four (4/296, 1.4%) single-author  
329 publications. On average, each paper cited 30 references.

330 Studies were led by authors from 34 countries, predominantly high-income nations, with 24.3% of  
331 studies (72/296) involving multi-country collaborations (Multimedia Appendix 8). US-based authors

332 contributed the most publications (91/296, 30.7%), followed by France (71/296, 24.0%) and Finland  
333 (35/296, 11.8%). US authors also had the most citations (3495/9379, 37.2%), with France and Finland

334 ranking second and third, respectively.

335 The 25 most-cited publications, appearing in 17 different journals, received between 83 (83/9379, 0.9%)  
336 and 485 (485/9379, 5.2%) citations (Multimedia appendix 9) [150,172-196]. Of these, ten were  
337 published before 2000, another ten between 2000 and 2010, and five after 2010. Most of these studies  
338 focused on cancer risk (16/25, 64%), while others investigated neurodegenerative disorders (5/25, 20%),  
339 respiratory conditions (2/25, 8%), and multiple health outcomes such as sleep disorders, mental health,  
340 and musculoskeletal disorders (2/25, 8%).

341 For details on the most productive countries, prolific authors, active journals, institutions, and funding  
342 bodies, please refer to Multimedia Appendix 10.

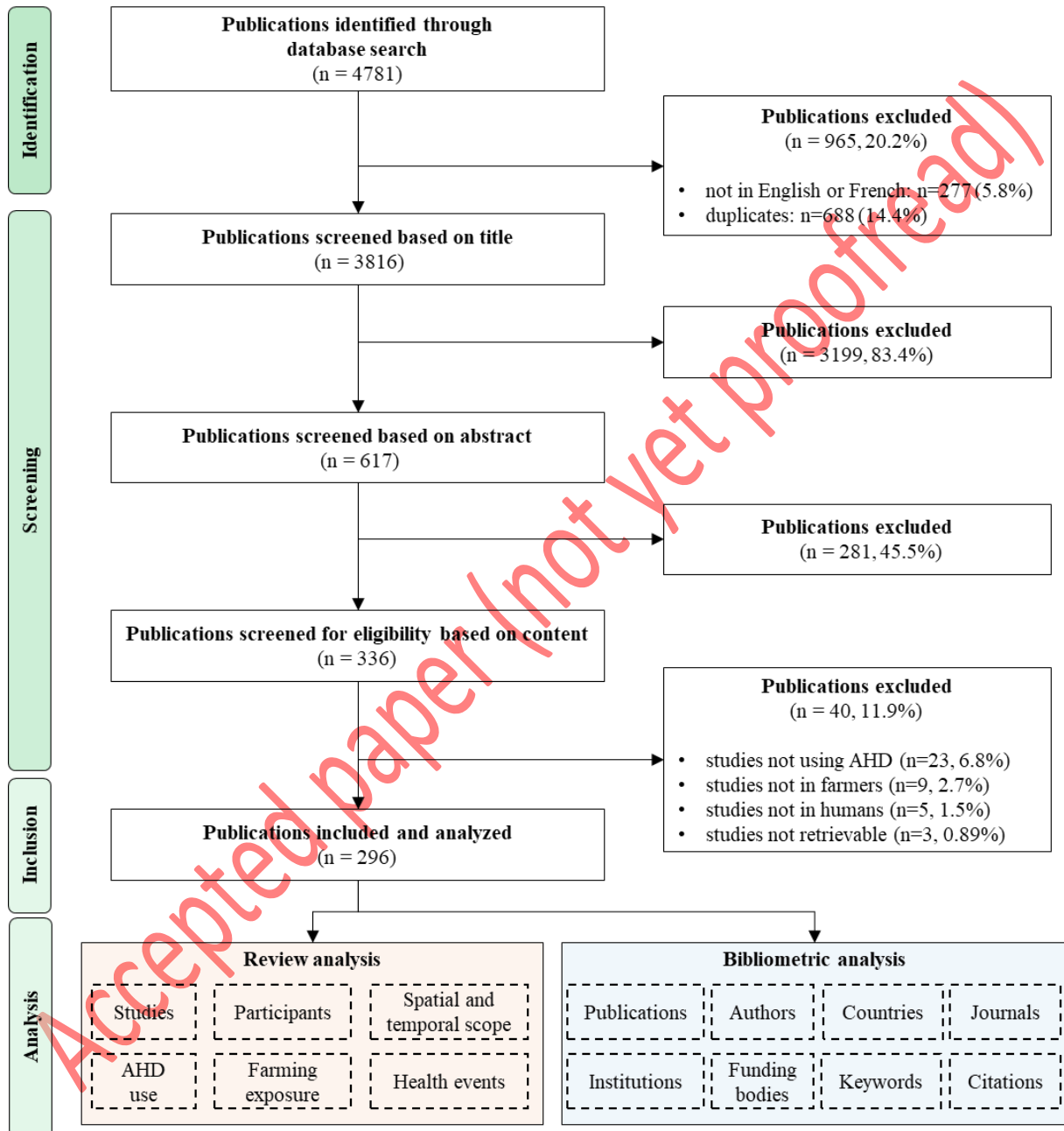
Accepted paper (not yet proofread)



343 **Figure 1.** PRISMA-ScR flow chart depicting the literature search and the evaluation process for finding  
 344 relevant records.

345 AHD : administrative health database.

346 The search, conducted on April 15, 2024, in PubMed and Web of Science, had no date restrictions.



347

348 **Table 4.** Main characteristics of the included publications

Description	Results
<b>Timespan</b>	1975-2024
<b>Publication type</b>	
total	296 (100%)
article	293 (99.0%)
review	2 (0.7%)
editorial material	1 (0.3%)
<b>Number of open access publications</b>	107 (36.1%)
<b>Document average age (years)</b>	14.2
<b>Annual growth rate (%)</b>	5.20
<b>Publication citation</b>	
Number of citations	9379
Average citations per publication	31.7
Average citations per year per publications	2.02
<b>Number of references</b>	8814
<b>Journal</b>	
Number of journals	118
Average number of publications per journal	1.86
Average number of citations per journal	79.5
<b>Author</b>	
Number of authors	1225
Number of single-authored publications	4
Author appearances	1882
Average number of co-authors per publication	6.36
Average number of publications per author	0.24
International co-authorships %	24.3
Number of author's keywords	576
<b>Author's country</b>	
Number of author's country	34
Average number of publications per country	2.86
Average number of citations per country	436.0
<b>Author's institution</b>	
Number of author's institutions	338
Average number of publications per institution	3.11
Average number of citations per institution	101.3
<b>Author's funding body</b>	
Number of funding bodies	181
Average number of publications per funding body	2.48
Average number of citations per funding body	77.7

349

350 **Study characteristics**

351 Table 5 provides an overview of the included publications. Longitudinal study designs were the most  
 352 common, including retrospective cohorts (129/296, 43.6%) and prospective cohorts (56/296, 18.9%).  
 353 Case-control studies (62/296, 20.9%), cross-sectional studies (39/296, 13.2%), and ecological studies  
 354 (17/296, 5.7%) were less common (Multimedia appendix 11). A few studies (10/296, 3.4%) utilized  
 355 multiple study designs [188,194,197-204].

356 The median follow-up period was 9.5 [IQR: 5-17 years]. On average, there was a 7-year gap [90% CI:  
 357 3-14] between the most recent data used and the year of publication, with considerable variation  
 358 depending on the publication year (Figure 2). The oldest data was from 1801 [205], and the most recent  
 359 data was from 2022 [206]. Notably, one-third of the data used were from before 2000, while nearly

360 three-quarters were from before 2015 (Multimedia appendix 6). Only ten studies used data from the last  
361 five years (since 2020), while 80 used data from the last ten years (since 2015).

362 Studies were conducted on all continents, but most participants were from Europe (249/296, 84.1%),  
363 followed by North America (85/296, 28.7%), Asia (24/296, 8.1%), Oceania (17/296, 5.7%), Africa  
364 (4/296, 1.4%), and Central and South America (4/296, 1.4%). France (70/296, 23.6%) and the US  
365 (67/296, 22.6%) were the most represented countries, followed by Finland (36/296, 12.2%), Sweden  
366 (32/296, 10.8%), Denmark (28/296, 9.5%), and Norway (26/296, 8.8%) (Figure 3 and Multimedia  
367 appendix 12). Most studies had a regional or local scope (177/296, 59.8%), in particular traditional  
368 epidemiological studies, such as AGRICAN (Agriculture and Cancer) [207] and AHS (Agricultural  
369 Health Study) [172], which used AHDs to identify potential individuals for inclusion and/or enrich their  
370 cohorts.

371 Most studies included 1000 to 10000 participants (67/296, 22.6%), followed by studies with 10000 to  
372 100000 participants (65/296, 22.0%) and 100000 to 1 million participants (53/296, 17.9%) (Table 5).  
373 Larger studies ( $\geq 1$  million participants) accounted for 16.9% (50/296). Smaller studies, with 100 to 1000  
374 participants, were less common (47/296, 15.9%), and very few (8/296, 2.7%) had fewer than 100  
375 participants. Most studies included adult participants (284/296, 95.9%). 169 studies examined males  
376 (169/296, 57.1%), 130 at females (130/296, 43.9%), and 188 at both sexes (188/296, 63.5%), but 108  
377 studies (108/296, 36.5%) did not specify the participants' sex.

378 Over half of the studies (156/296, 52.7%) aimed to explore the relationship between farming activities  
379 (e.g., dairy farming) and health events, while 131 studies (131/296, 44.3%) focused on individual  
380 characteristics like occupation, age, sex, and socioeconomic status (Multimedia Appendix 13). Among  
381 those examining individual characteristics, farming was often considered broadly and compared to other  
382 occupations (95/131, 72.5%). Conversely, in studies investigating health outcomes specifically related  
383 to farming activities, agriculture was treated as a broad category in only 27.6% (43/156) of the cases.

384 Most studies used the general population or other non-farming groups as the reference category, without  
385 differentiating farmers by job role (e.g., farm managers vs. farmworkers). Descriptive statistics and  
386 multivariable regression were the most commonly employed methods. Notably, only two studies (2/296,  
387 0.7%) incorporated artificial intelligence (AI) in their analysis [208,209].

388 There were few studies that investigated health outcomes in farmers' family members or non-farmers  
389 exposed to farming. Only three studies (3/296, 1.0%) focused on health events in farmers' partners  
390 [177,210,211], five (5/296, 1.7%) on farmers' children [179,212-215], and six (6/296, 2.0%) on non-  
391 farmers exposed to farming-related risks [209,216-221]. There were eleven studies that explored health  
392 risk in migrant workers.

393 Some publications reported findings from the same cohorts (Multimedia appendix 14). The ten most  
394 prolific cohorts included France-based AGRICAN (18/296, 6.1%) [207], the US-based AHS (17/296,  
395 5.7%) [172], NOCCA (Nordic Occupational Cancer Study) (12/296, 4.1%) from Nordic countries  
396 (Finland, Denmark, Norway, Sweden and Iceland) [189], France-based TRACTOR (Tracking and  
397 monitoring occupational risks in agriculture) (7/296, 2.4%) [222], and CNAP (Cancer in the Norwegian  
398 Agricultural Population cohort) (7/296, 2.4%) [182]. Other notable cohorts included the US-based UFW  
399 (United Farm Workers of America) (6/296, 2.0%) [223], France's BALISTIC (5/296, 1.7%) [224], the  
400 international (29 countries) consortium AGRICOH (4/296, 1.4%) [150,225], AIRBAg (4/296, 1.4%)  
401 from France [226], and the US-based NAWS (National Agricultural Workers Survey) (3/296, 1.0%)  
402 [227]. Among these top ten cohorts, only NOCCA, UFW, and TRACTOR exclusively used AHDs.

Accepted paper (not yet proofread)

403 **Table 5.** Characteristics of the 296 included studies (1975-2024).

Characteristic	n (%)
<b>Research goal</b>	
Study the association between farming and health event	156 (52.7)
Study the association between individual characteristics and health event	131 (44.3)
Other research goal	9 (3.0)
<b>Study design</b>	
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)
<b>Geographical scope</b>	
Nationwide	117 (39.5)
Regional/local	176 (59.5)
<b>Temporal scope (years, mean [IQR])</b>	
Follow-up period	9.50 [IQR: 5.00-17.0]
Gap between latest data used and publication year	7.21 [IQR: 5.00-9.00]
<b>Population</b>	
Adult	265 (89.5)
Adult and child	19 (6.4)
Child	8 (2.7)
Not reported	1 (0.3)
<b>Sex</b>	
Female	130 (43.9)
Male	169 (57.1)
Female and male	188 (63.5)
Not specified	108 (36.5)
<b>Number of participants</b>	
> 1e6	50 (16.9)
[1e5-1e6[	53 (17.9)
[1e4-1e5[	65 (22.0)
[1000-1e4[	67 (22.6)
[100-1000[	47 (15.9)
[10-100[	8 (2.7)
Not reported	3 (1.0)
<b>AHD type</b>	
Disease register	158 (53.4)
Electronic health/medical record	124 (41.9)
Insurance claim	106 (35.8)
Population register	95 (32.1)
Hospital discharge	41 (13.9)
<b>AHD use</b>	
Obtain information on socio-demographics	272 (91.9)
Obtain information on health event	269 (90.9)
Identify farmer	147 (49.7)
Identify individual	140 (47.3)
Obtain information on occupations	117 (39.5)
Exposure assessment	57 (19.3)
Obtain information on farming activity	43 (14.5)
Other usage	14 (4.7)

404 Note: AHD: administrative health database, IQR: interquartile range.

405

406 **AHD use**

407 There was a high heterogeneity in the coding systems used and the granularity of the information

408 available regarding health events (outcomes), population, and exposure determinants, depending on the

409 AHD and study considered. Regardless of the publication reviewed, AHDs and other datasets were never

410 reported as adhering to the FAIR principles (findable, accessible, interoperable, and reusable data) [228-  
411 230]. In addition, none of them could be considered as FAIR because, with a few exceptions (e.g., [222]),  
412 most AHDs were not precisely described and data availability statements were rare. Furthermore, mainly  
413 due to privacy concerns, AHDs were not available in open and free access.

414 The most commonly used AHDs were disease registers, used in over half of the studies (158/296,  
415 53.4%), followed by electronic health/medical records (124/296, 41.9%), insurance claims (106/296,  
416 35.8%), population registers (95/296, 32.1%), and hospital discharge records (41/296, 13.9%) (Table 5).  
417 Among disease registers, cancer (120/158, 75.9) and mortality registers (75/158, 47.5%) were the most  
418 frequently used (Multimedia appendices 11 and 15). Nearly one-third of the studies (91/296, 30.7%)  
419 relied on a single AHD, with disease registers being the most common (38/91, 42%), followed by  
420 insurance claims (29/91, 32%), electronic health/medical records (18/91, 20%), population registers  
421 (5/91, 5%), and hospital discharge records (1/91, 1%). Other types of digital data were used less  
422 frequently, including pesticide registration records (13/296, 4.4%), job-exposure matrices (12/296,  
423 4.1%), crop-exposure matrices (11/296, 3.7%), pesticide use records (8/296, 2.7%), climate data (7/296,  
424 2.4%), and air quality data (2/296, 0.7%). While contextual data were sometimes used (9/296, 3.0%),  
425 person-generated data, smart agriculture data, and omics were never utilized.

426 The AHDs and other digital data were primarily used to obtain socio-demographic information  
427 (272/296, 91.9%) and health event data (269/296, 90.9%). They were also used to identify farmers  
428 (147/296, 49.7%) or individuals (140/296, 47.3%), gather occupational information (117/296, 39.5%),  
429 assess exposure (57/296, 19.3%), obtain data on farming activities (43/296, 14.5%), or track climate  
430 conditions (7/296, 2.4%).

431 Nearly two-thirds of the studies (181/296, 61.1%) relied exclusively on digital data (AHDs or other),  
432 while more than a third (112/296, 37.8%) incorporated self-reported information/active data (requiring  
433 active participant involvement) as part of epidemiological cohorts. A total of 111 studies (111/296,  
434 37.5%) used participant-completed questionnaires (paper or electronic) to gather socio-demographic  
435 data and confounding factors (98/296, 33.1%), assess exposure (96/296, 32.4%), and/or collect health  
436 information (83/296, 28.0%). Some information was obtained through interviews (44/296, 14.9%) or  
437 clinical examinations (32/296, 10.8%). Biological monitoring (24/296, 8.1%) and airborne monitoring

438 (2/296, 0.7%) were sometimes employed, whereas no study reported dermal monitoring (Multimedia  
439 appendix 16).

440 Among all the AHDs used, the Mutualité Sociale Agricole (MSA) is a singularity. To the best of our  
441 knowledge, it is the only AHD specifically dedicated to the entire farming population of a country.  
442 Indeed, MSA is the French national insurance scheme that covers the entire farming workforce (5% of  
443 the overall French population) [115,128]. MSA was used in 60 studies (60/296, 20.3%). These studies  
444 were often part of cohorts with multiple publications, such as AGRICAN (18/60, 30%), TRACTOR  
445 (7/60, 12%), BALISTIC (5/60, 8%), AIRBAg (4/60, 7%), AMI (Aging Multidisciplinary Investigation)  
446 (2/60, 3%) [151], BM3R (2/60, 3%) [231], FERMA (risk factors of the rural environment and allergic  
447 and respiratory disease) (1/60, 2%) [232], and Phytoneer (1/60, 2%) [233]. Of these, TRACTOR was the  
448 only cohort using exclusively MSA data [222].

449

#### 450 **Farming exposure**

451 A variety of exposure proxies were used to assess farming-related exposure. The most common proxy  
452 was job title, which generally referred to whether the individual was a farmer (184/296, 62.2%). Other  
453 proxies included specific farming activities, such as dairy or crop farming (111/296, 37.5%), general  
454 pesticide exposure (yes/no) (62/296, 20.9%), and exposure to specific pesticide compounds (e.g.,  
455 glyphosate or paraquat) or pesticide classes (e.g., insecticides) (51/296, 17.2%) (Multimedia appendices  
456 11 and 17). The number of farming activities studied ranged from just one (e.g., [226]) to 78 [222], with  
457 an average of 8 farming activities per study. Similarly, the number of pesticide compounds assessed  
458 ranged from one (e.g., [234]) to 943 [235], with an average of 42 pesticides per study. Only one study  
459 investigated the mixture effect of exposure to multiple pesticide combinations on human health [236].  
460 Investigations into other chemical exposures were also limited, with only two papers each addressing  
461 silica exposure [237,238] and air pollution [194,217] (2/296, 0.7%). Notably, no studies examined  
462 exposure to per- and polyfluoroalkyl substances (PFAS) or micro- and nano-plastics. Research on the  
463 broader farming exposome was rare (5/296, 1.7%) and typically used farming activities as proxies (e.g.,  
464 [152]).

465 Few studies explored exposure to physical agents, with five studies (5/296, 1.7%) focusing on radiation  
466 [187,218,239-241]. No studies investigated the effects of climate change on farmers' health. Exposure  
467 to biological agents was rarely studied as well, with just three papers (3/296, 1.0%) addressing  
468 mycotoxins [241-243]. Finally, only three studies (3/296, 1.0%) examined psychological factors related  
469 to farming exposure [244-246].

470

#### 471 **Health events**

472 The most frequently studied health events were cancer (142/296, 48.0%), followed by mortality (44/296,  
473 14.9%), injuries (38/296, 12.8%), workplace accidents (32/296, 10.8%), respiratory disorders (30/296,  
474 10.1%), neurodegenerative diseases (28/296, 9.5%), and mental health issues (26/296, 8.8%)  
475 (Multimedia appendices 11 and 18). Less studied conditions included cardiovascular diseases (16/296,  
476 5.4%), autoimmune disorders (11/296, 3.7%), musculoskeletal disorders (11/296, 3.7%), reproductive  
477 disorders (3/296, 1.0%), sleep disorders (1/296, 0.3%), and frailty (1/296, 0.3%). Notably, no studies  
478 explored the farming microbiome.

479 Among cancers, lung cancer was the most commonly investigated (43/142, 30.3%), followed by prostate  
480 cancer (38/142, 26.8%), leukemia (37/142, 26.1%), colorectal cancer (35/142, 24.6%), multiple  
481 myeloma (35/142, 24.6%), non-Hodgkin lymphoma (35/142, 24.6%), bladder cancer (31/142, 21.8%),  
482 and brain cancer (31/142, 21.8%) (Multimedia appendix 19). Respiratory disorders were primarily  
483 focused on asthma (15/30, 50%) and COPD (chronic obstructive pulmonary disease) (14/30, 47%).  
484 Parkinson's disease was the most studied neurodegenerative condition (16/28, 57%), followed by  
485 multiple sclerosis (6/28, 21%). Fewer publications examined Alzheimer's disease (2/28, 7%) and  
486 amyotrophic lateral sclerosis (2/28, 7%) (Multimedia appendix 20). In the mental health field, suicide  
487 (12/26, 46%) and depression (8/26, 31%) were the most investigated issues (Multimedia appendix 21).

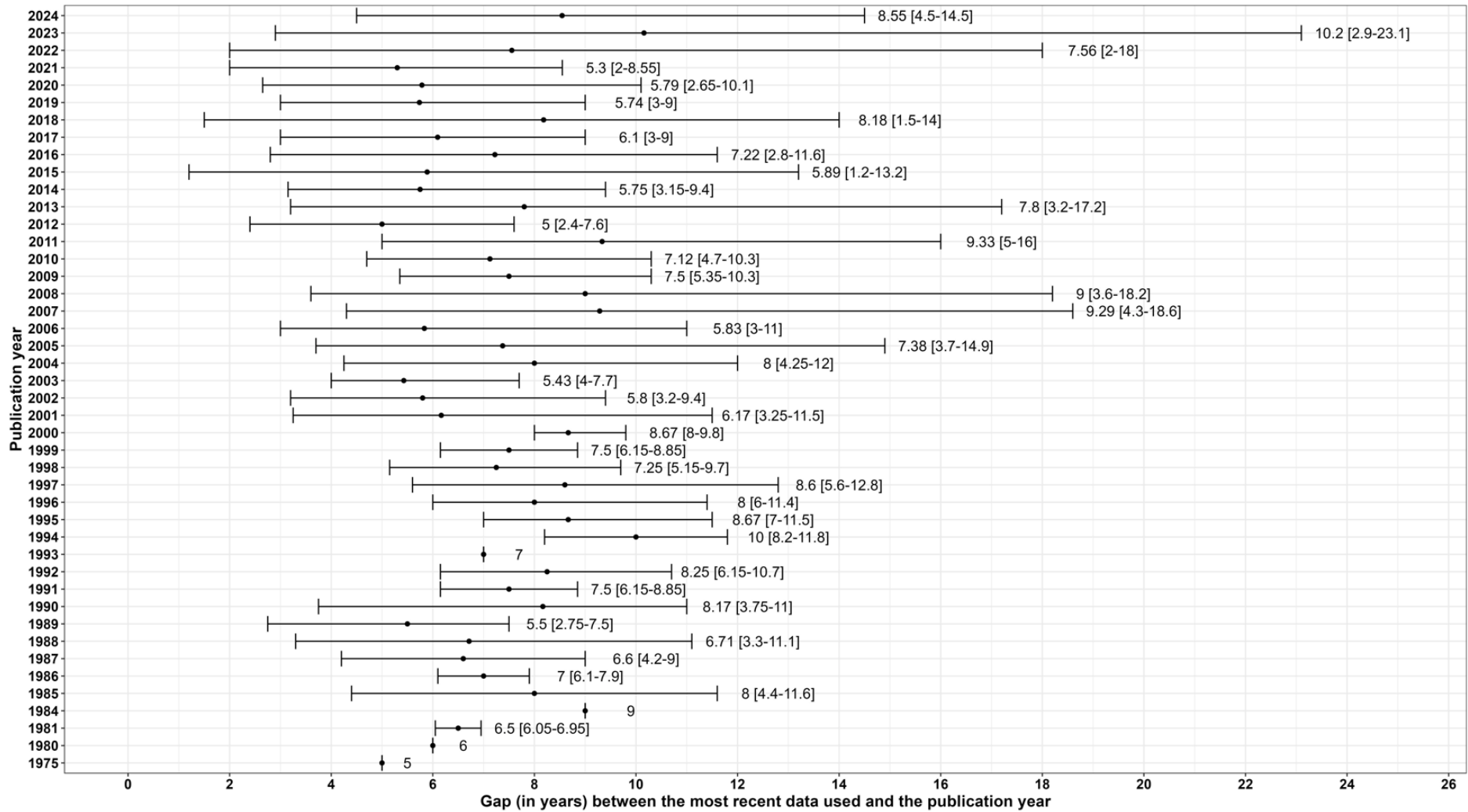


488 **Figure 2.** Number of year between the most recent data used and publication for all included articles (1975-2024).

489 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

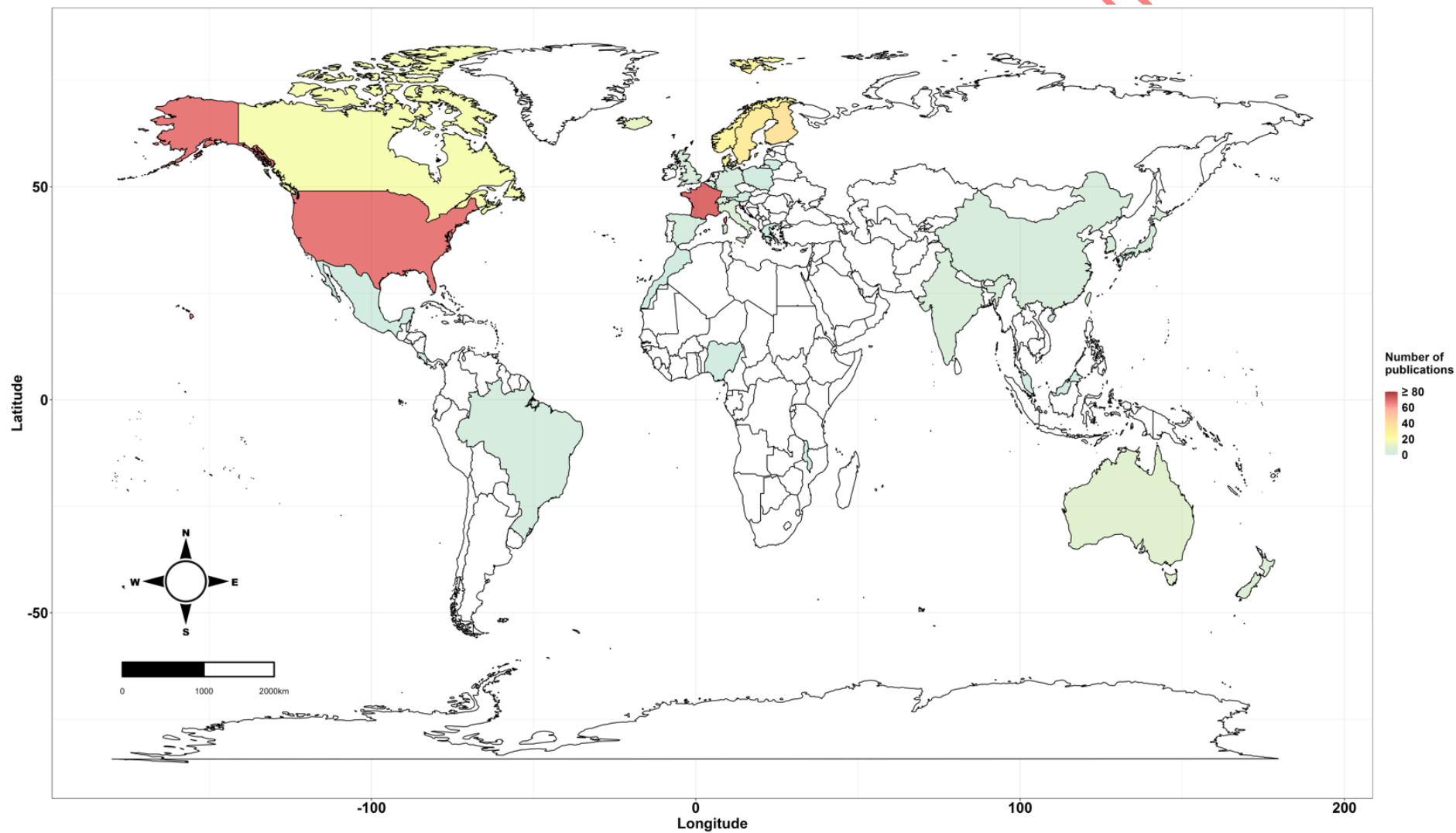
490 refer to the 90%CI of the number of years between the most recent data used and publication.

Accepted paper (not yet proofread)



ACCEPTED

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



493

## 494 **Keyword analysis**

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

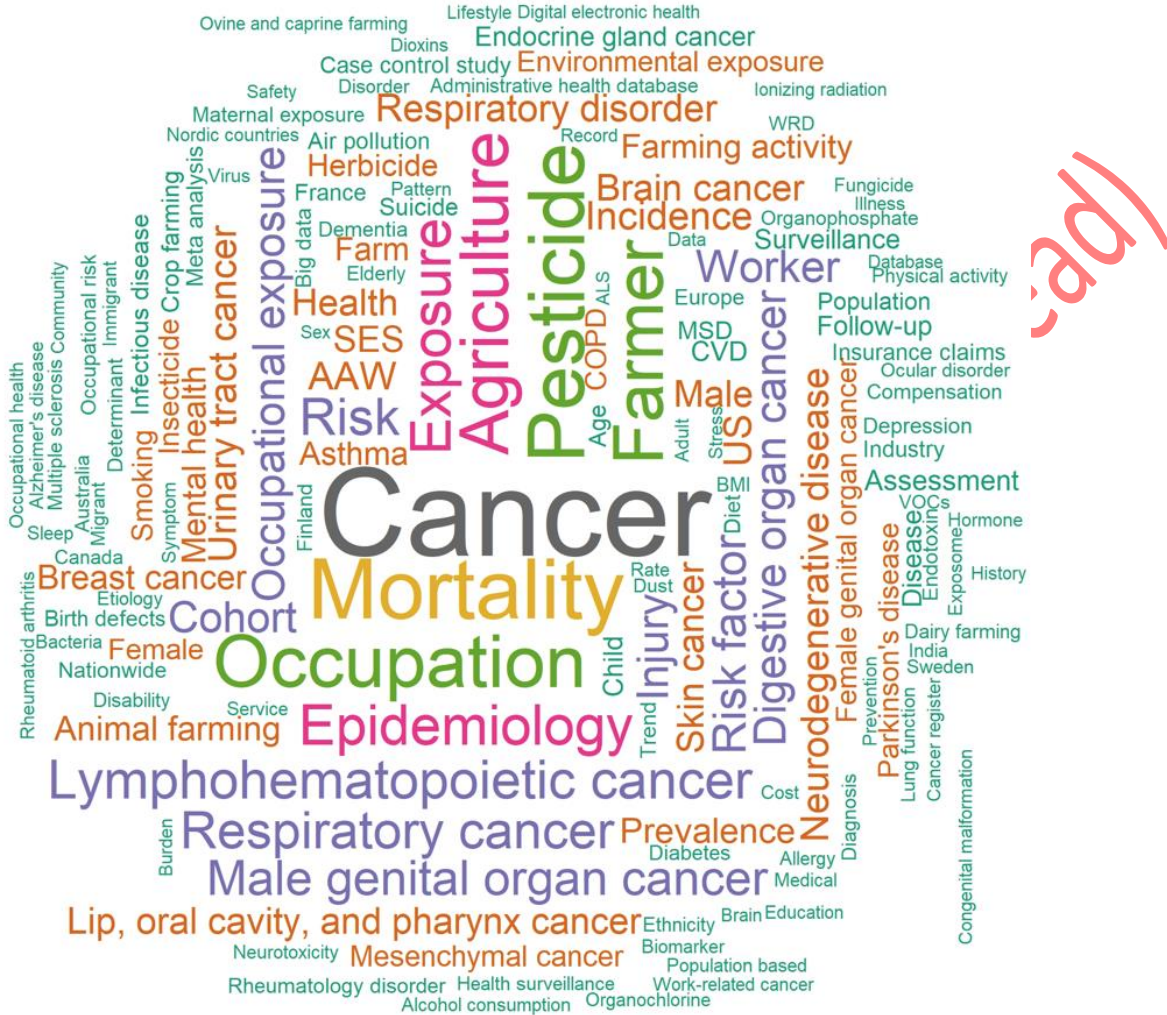
500 On average, each publication included 8.90 keywords [90% CI: 0-17], although 35 publications (35/296,  
501 11.8%) lacked any keywords, in line with the journal guidelines. Keyword analysis confirmed prior  
502 findings regarding farming exposure and health outcomes, as shown in Figure 4. It also provided deeper  
503 insights into emerging research hotspots, trends, and gaps.

504 Of the total 576 keywords, 301 (301/576, 52.3%) appeared only once, while 68 (68/576, 11.8%) were  
505 mentioned at least 10 times. More frequently used keywords included 39 that appeared at least 20 times  
506 (39/576, 6.8%) and 11 keywords that featured in at least 50 publications (11/576, 1.9%). The 50 most  
507 frequent keywords were mentioned in at least 17 publications (17/296, 5.7%), while the top 10 appeared  
508 in at least 51 publications (51/296, 17.2%) (Multimedia appendix 22). The most frequently cited  
509 keyword was “cancer” (150/296, 50.7%), followed by “mortality” (96/296, 32.4%), “pesticide” (88/296,  
510 29.7%), “occupation” (82/296, 27.7%), “farmer” (77/296, 26.0%), “agriculture” (74/296, 25.0%),  
511 “exposure” (57/296, 19.3%), and “epidemiology” (57/296, 19.3%).

512 In terms of overall citations, “cancer” (5766/9379, 61.5%), “pesticide” (3569/9379, 38.1%), and  
513 “mortality” (3097/9379, 33.0%) were the most cited. Over the past decade, the frequency of the top five  
514 keywords has drastically increased (Multimedia appendix 23). Notably, keywords like “cancer,”  
515 “mortality,” “occupation,” “pesticide,” “agriculture,” and “farmer” have been consistently present in  
516 publications spanning at least 30 years (not necessarily consecutively) (Multimedia appendix 24). In the  
517 last decade, emerging keywords such as “big data,” “administrative health database,” “dust,” and “BMI”  
518 (body mass index) have gained prominence (Multimedia appendix 25).

519 **Figure 4.** Wordcloud of the most frequent keywords among the 296 papers published between 1975-  
520 2024.

521 The larger the font size, the greater the frequency of the keywords.



522

523

### 524 ***Keyword co-occurrence***

525 A keyword co-occurrence network illustrating the frequency of keyword co-mentions in publications  
526 was constructed, thereby revealing relationships and conceptual connections (Figure 5). In this network,  
527 nodes/vertices represent keywords, with their sizes indicating frequency, while edges denote co-  
528 occurrences. The network's density and arrangement reveal topic interconnectivity, with larger vertices  
529 representing more frequently mentioned keywords. The network visualization helps identify clusters of  
530 related topics and highlights core research areas.

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

535 The top keywords for each cluster were “cancer” (red cluster), “pesticide” (purple cluster), “mortality”  
536 (green cluster), and “exposure” (blue cluster). The red cluster highlights associations between various  
537 types of cancer, reflecting the fact that studies investigating cancer risks often examine multiple types  
538 of cancer. The green cluster links “mortality” with terms like “mental health”, “injury”, and “animal  
539 farming”, explained by the association between workplace accidents, mental health issues (e.g., suicide),  
540 animal farming, and mortality. In the purple cluster, “pesticide” connects with “occupational exposure”  
541 and “farming activity”, emphasizing that pesticide exposure is primarily studied in occupational settings  
542 across different types of farming. The blue cluster connects “exposure” to terms like “neurodegenerative  
543 disease”, “respiratory disorder”, “cardiovascular disorder”, “risk factor”, “air pollution”, “age”, and  
544 “diet”, indicating the study of various risk factors in relation to several health events. These clusters  
545 highlight current research hotspots that focus on four main interconnected themes: the associations  
546 between risk factors, pesticide exposure, farming activities, and a range of diseases.

547

#### 548 *Thematic mapping - research hotspots*

549 Figure 6 presents a thematic map that illustrates current research trends. Thematic mapping visualizes  
550 the relationship between research themes or topics, enabling the identification of trends, emerging areas,  
551 and gaps in the literature. The result is a strategic diagram that shows how themes relate to each other  
552 and their relevance within a specific field. The graph is divided into four quadrants, categorizing topics  
553 based on their relevance (x-axis, Callon’s centrality) and maturity (y-axis, Callon’s density) within the  
554 broader research landscape. Each circle represents a theme or topic (i.e., a cluster of equivalent  
555 keywords), with the circle size corresponding to the frequency of the keywords associated with that  
556 theme.

557 The upper-right quadrant represents “hot topics”, which are both highly relevant and mature in the  
558 research landscape. Four key hot topics drive AHD-based public health research in farming populations.

559 These include one focused on cancer research, another on respiratory disorders, and a third  
560 encompassing neurodegenerative diseases, workplace accidents, injuries, and mental health issues. The  
561 final hot topic involves large-scale studies in France and Europe using big data and insurance claims.  
562 The lower-right quadrant contains “basic topics”, which are relevant but not yet mature in the research  
563 landscape. Only one such theme emerged: research related to pesticide exposure, mortality, and farming.  
564 In the upper-left quadrant, “niche themes” refer to mature research topics that have not yet achieved full  
565 relevance. Three niche themes were identified: the first involves studies examining aging and research  
566 conducted in Norway; the second focuses on reproductive disorders and parental exposure, a theme  
567 poised to potentially evolve into a hot topic. The final niche theme covers genetics and metabolism.  
568 Finally, the lower-left quadrant contains “peripheral topics”, which represent either emerging or  
569 declining themes with low relevance and maturity. Four peripheral topics were observed, primarily  
570 centered around research on ocular disorders, the use of electronic health/medical records, and studies  
571 conducted in India.  
572 This thematic map helps highlight both well-established and emerging areas of research, as well as gaps  
573 that may be ripe for future investigation.

Accepted paper (not yet proofread)

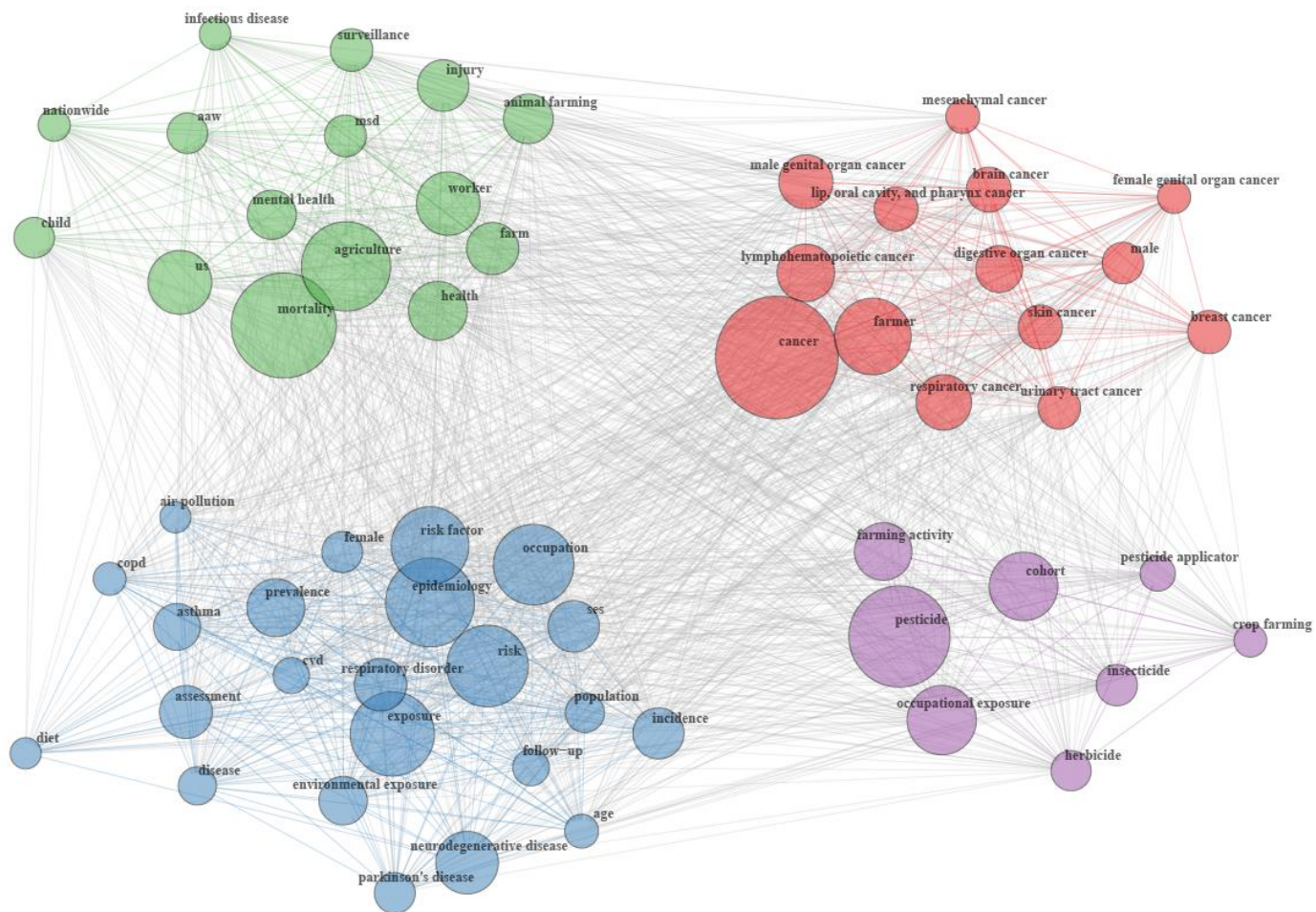
574 **Figure 5.** Keyword co-occurrence network of the 296 articles published between 1975-2024.

575 AAW: workplace accident, COPD: chronic obstructive pulmonary disease, CVD: cardiovascular disorder, MSD: musculoskeletal disorder, SES: socio-economic  
576 status, US: United States of America.

577 Each vertex/node represents a keyword, while edges represent the co-occurrence between keywords. Two keywords are connected when they co-occur in the  
578 same publication, and the size of each vertex indicates the frequency of a keyword—larger vertices represent more frequently mentioned keywords. Keywords  
579 with the same color (cluster) represent a research area.

Accepted paper (not yet proofread)



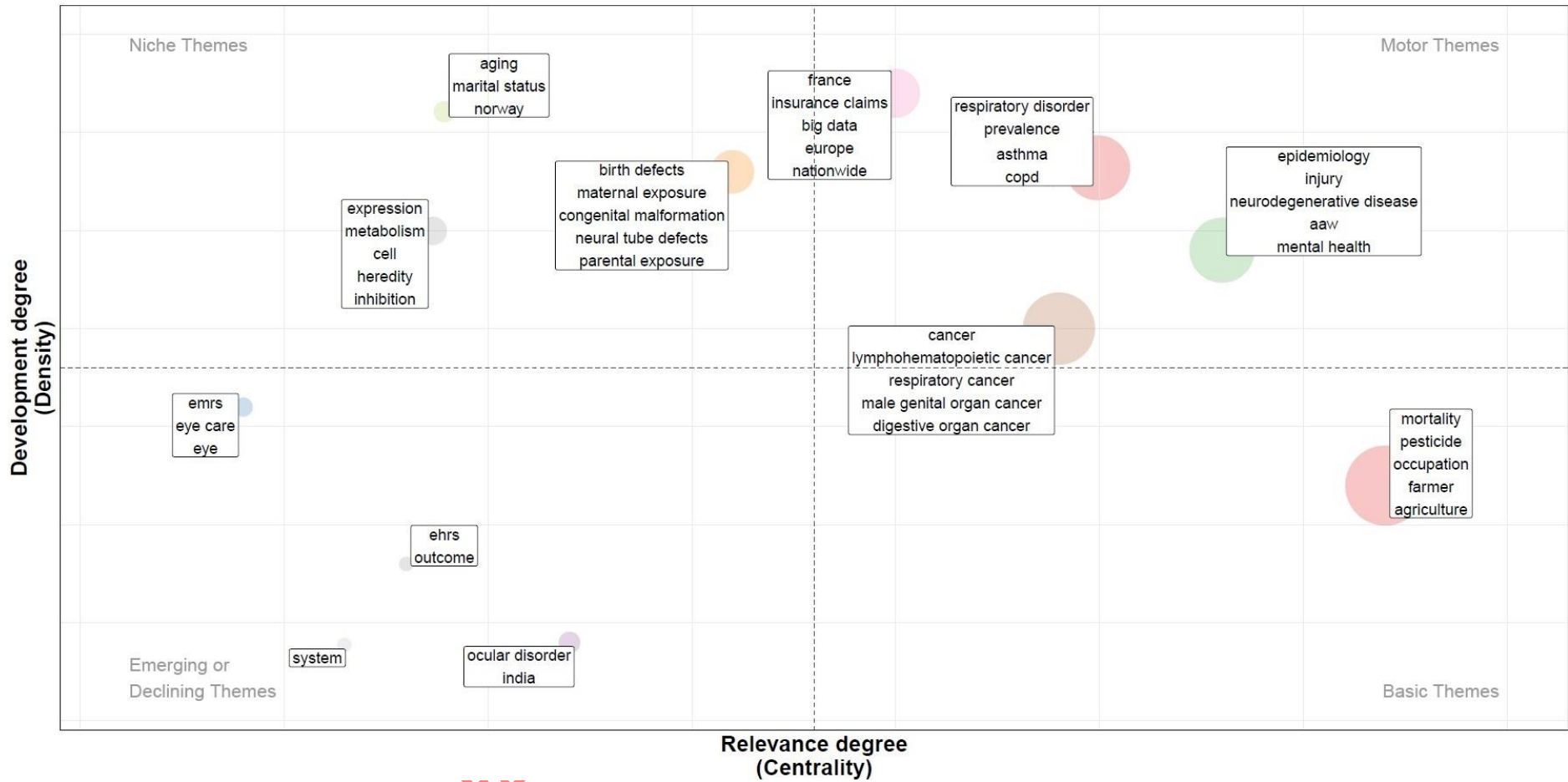


580

Accepted

581 **Figure 6.** Thematic mapping - research hotspots based on keywords from the 296 articles published between 1975-2024.  
582 AAW: workplace accident; COPD: chronic obstructive pulmonary disease, EHRs: electronic health records, EMRs: electronic medical records.  
583 The graph shows how themes relate to each other and their relevance within a specific field. This graph is divided into four quadrants, categorizing topics based  
584 on their relevance (x-axis, Callon's centrality) and maturity (y-axis, Callon's density) within the broader research landscape. Each circle represents a theme or  
585 topic (i.e., a cluster of equivalent keywords), with the circle size corresponding to the frequency of the keywords associated with that theme.

Accepted paper (not yet proofread)



586

Accepte

### 587 *Bibliometric farming exposome*

588 In order to identify research trends and gaps in the farming exposome literature, a bibliometric keyword  
589 co-occurrence analysis was conducted to explore the farming exposome by examining the co-occurrence  
590 between keywords associated with potential risk factors and specific health outcomes. This analysis was  
591 restricted to exposome-related and health event-related keywords. Out of 576 keywords, 130 were  
592 related to the exposome (130/576, 22.6%), among which 93 were related to the specific external  
593 exposome (e.g., pesticide), 19 to the general external exposome (e.g., climate), and 18 to the internal  
594 exposome (e.g., oxidative stress). There were also 70 health event-related keywords (e.g., brain cancer)  
595 (70/576, 12.2%).

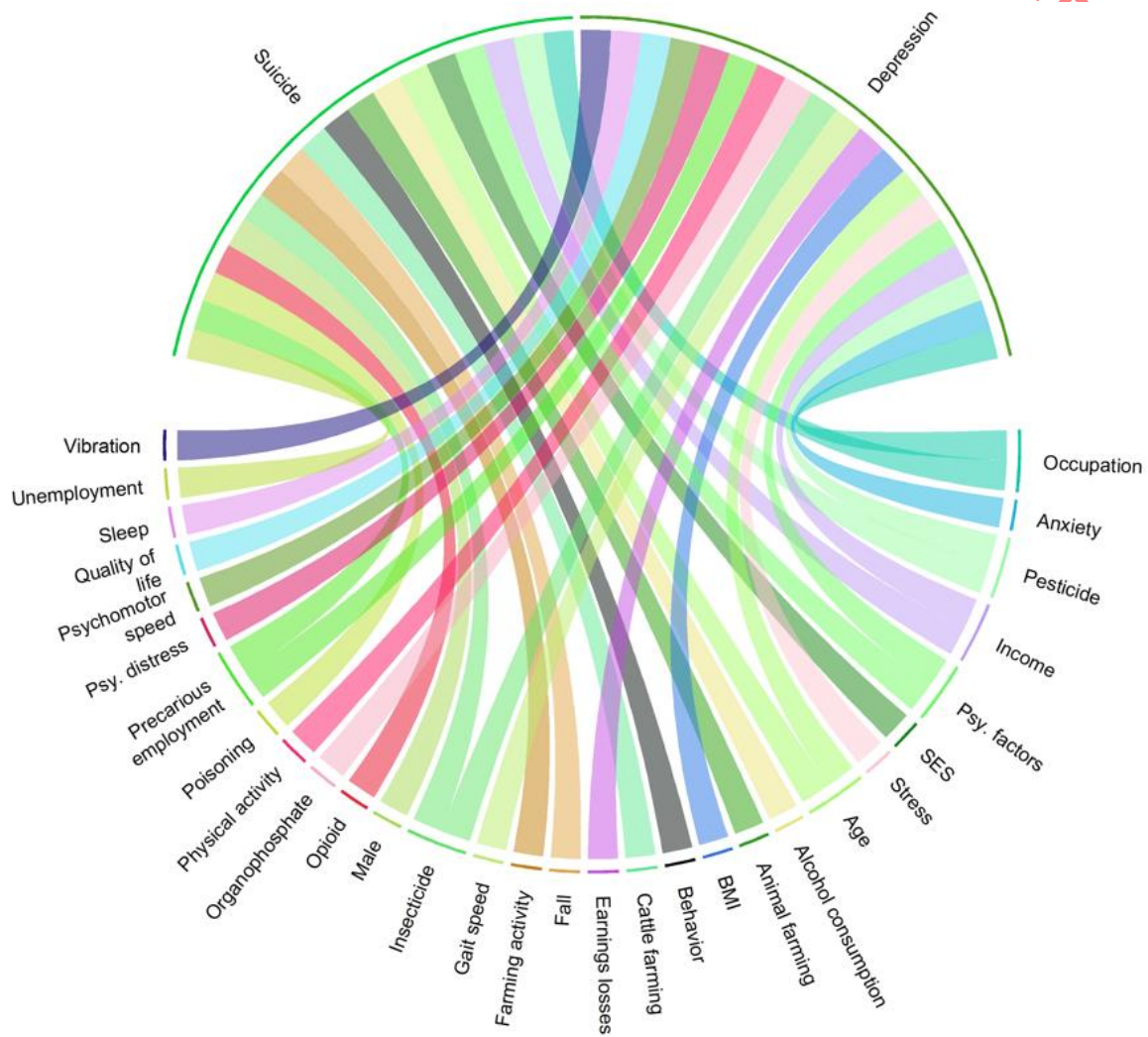
596 The results of this analysis are synthesized in Table 6 and Multimedia appendix 26, with each cell  
597 representing the percentage of occurrences of an exposome-related keyword (e.g., air pollution) in all  
598 publications mentioning a specific health event-related keyword (e.g., Alzheimer's disease). For  
599 example, a value of 33.3 indicates that an exposome-related keyword appeared in 33.3% of all  
600 publications mentioning a specified health event-related keyword. To facilitate interpretation and ease  
601 the reading of Table 6, exposome-related keywords were categorized into 19 groups (e.g., chemical  
602 agent), and health event-related keywords into 20 groups.

603 Distinct keyword exposome profiles were developed for each health event-related keyword (Multimedia  
604 appendices 27-52), as illustrated in Figure 7 for mental health disorders. Most exposome-related  
605 keywords associated with mental health disorder-related keywords pertained to the type of occupations  
606 as well as chemical, lifestyle, socio-economic, and psychological factors. Cancer-related keywords were  
607 associated mostly with keywords related to the internal (sex) and specific external exposome (chemical  
608 agents, lifestyle, and type of occupation). Autoimmune disease-related keywords co-occurred mostly  
609 with external exposome-related keywords (chemical agents, lifestyle, type of occupations, and socio-  
610 economic factors). Neurodegenerative disease-related keywords were associated mostly with keywords  
611 related to the specific external exposome (lifestyle, chemical agents, and type of occupations).  
612 Reproductive disorders co-occurred mostly with internal (sex, BMI) and specific external exposome-  
613 related keywords (chemical agents and type of occupations). Both musculoskeletal disorder and injury-  
614 related keywords were associated with keywords from all three exposome components, in particular sex,

615 type of occupations, lifestyle, biomechanical factors, chemical agents and psychological factors.  
616 Infectious disease-related keywords co-occurred with specific external exposome keywords (biological  
617 agents and type of occupations), while respiratory disorder-related keywords were associated mostly  
618 with internal (sex) and specific external exposome-related keywords (lifestyle, chemical and biological  
619 agents, and type of occupations). Cardiovascular disorder-related keywords were associated with  
620 keywords from all three exposome components, in particular the sex, type of occupations, lifestyle,  
621 chemical agents, and socio-economic and psychological factors.

Accepted paper (not yet proofread)

622 **Figure 7.** Chord diagram of keyword co-occurrence between potential risk factor and mental health  
 623 disorder keywords from the 296 articles published between 1975-2024.  
 624 BMI: body mass index, psy: psychological, SES: socio-economic status.  
 625 On the top half of the chord diagram are disease-related keywords, while exposome-related keywords  
 626 are displayed on the bottom half. Each chord or link indicates that an exposome-related keyword was  
 627 mentioned with a disease-related keyword (co-occurrence) at least once in the same publication. The  
 628 chord color differs from one exposome-related keyword to another.



629

630 **Table 6.** Co-occurrence between exposome category-related keywords and health event category-related keywords among the 296 articles published between  
 631 1975 and 2024.

Health event	Internal exposome								Specific external exposome							General external exposome			
	Age	Sex	BMI	BP	Heredity	Ethnicity	Hormone	Menopause	Metabol.	OS	Inflam.	Lifestyle	Chem.	Bio.	Biomech.	Occup.	Phys.	SE	Psy.
<b>Cardiovascular disease</b>	5.88	11.8	5.88	5.88	0	5.88	0	0	0	0	0	11.8	29.4	0	11.8	23.5	0	29.4	17.6
<b>Work-related disease</b>	0	20	0	0	0	0	0	0	0	0	0	0	20	20	20	40	0	20	0
<b>Autoimmune disease</b> (IBD, RA, vasculitis, NR)	0	0	11.1	0	0	0	0	0	0	0	0	22.2	44.4	0	0	55.6	0	22.2	0
<b>Cancer</b>	3.62	27.5	2.17	0	0.73	1.45	3.62	0.73	2.9	0.73	0	17.4	51.4	7.25	1.45	58.7	5.8	8.7	0
<b>Dental health</b>	0	0	0	0	0	0	0	0	0	0	0	50	0	0	0	0	0	50	0
<b>Ocular disorder</b>	16.7	0	0	0	0	0	0	0	0	0	0	16.7	16.7	0	0	0	0	0	16.7
<b>Frailty</b>	50	0	50	0	0	0	0	0	0	0	0	100	50	0	0	0	0	0	50
<b>Anemia</b>	100	100	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0
<b>Infectious disease</b> (malaria, Lyme disease, tuberculosis, toxoplasmosis, NR)	0	7.14	0	0	0	7.14	0	0	0	0	0	21.4	21.4	50	7.14	28.6	0	14.3	7.14
<b>Injury</b> (includ. workplace accident and disability)	4.26	8.51	2.13	0	0	2.13	0	2.13	0	0	0	12.8	14.9	2.13	14.9	23.4	2.13	4.26	6.38
<b>Chronic kidney disease</b>	0	0	0	0	0	0	0	0	0	0	0	0	100	66.7	0	0	33.3	0	0
<b>Mental health</b> (depression, suicide, NR)	9.52	4.76	4.76	0	0	0	0	0	0	0	0	28.6	23.8	0	4.76	33.3	4.76	23.8	28.6
<b>Metabolic disorder</b> (diabetes, dysthyroidism, NR)	11.1	0	11.1	11.1	0	11.1	0	0	0	0	0	33.3	44.4	0	11.1	44.4	11.1	0	22.2
<b>Mortality</b>	4.17	17.7	0	0	1.04	0	1.04	2.08	0	0	0	17.7	55.2	8.33	4.17	54.2	5.21	17.7	2.08
<b>Musculoskeletal disorder</b> (arthritis, low-back pain, NR)	7.14	14.3	7.14	0	0	0	0	0	0	0	0	14.3	21.4	7.14	14.3	42.9	7.14	0	21.4
<b>Neurodegenerative disease</b> (AD, ALS, MND, MS, PD, NR)	6.06	0	0	0	3.03	6.06	0	0	12.1	6.06	0	21.2	48.5	6.06	0	36.4	0	9.09	3.03
<b>Sensory impairment</b>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0
<b>Reproductive disorder</b> (birth defects, infertility, spontaneous abortion, NR)	0	28.6	14.3	0	0	7.14	7.14	0	7.14	0	0	7.14	35.7	7.14	0	64.3	0	7.14	0
<b>Respiratory disorder</b> (allergy, asthma, COPD, pneumonia, sarcoidosis, NR)	5.88	11.8	8.82	2.94	0	5.88	0	0	5.88	0	2.94	47.1	35.3	11.8	2.94	35.3	0	8.82	5.88
<b>Skin disorder</b> (dermatitis, NR)	0	50	0	0	0	0	0	0	0	0	0	50	50	0	0	50	0	0	0

632 Note: AD: Alzheimer’s disease, ALS: amyotrophic lateral sclerosis, Bio.: biological agent, Biomech.: biomechanical factor, BMI: body mass index, BP: blood  
 633 pressure, Chem.: chemical agent, COPD: chronic obstructive pulmonary disease, IBD: inflammatory bowel disease, includ.: including, Inflam.: inflammation,  
 634 Metabol.: metabolism, MND: motor neuron disease, MS: multiple sclerosis, NR: not reported, Occup.: occupation, OS: oxidative stress, PD: Parkinson’s disease,  
 635 Phys.: physical agent, psy: psychological factor, RA: rheumatoid arthritis, SE: socio-economic factor.

636 Each cell refers to the number of times (in %) an exposome category-related keyword (e.g., chemical agent) was mentioned among all publications in which a  
637 health event category-related keyword (e.g., cancer) appeared.

Accepted paper (not yet proofread)



## 638 Discussion

639

### 640 Main findings

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

646

### 647 Current trends in AHD use for public health research in farming populations

648 Research utilizing AHDs in farming populations has been predominantly conducted in developed  
649 countries [150,225], with the US (e.g., [172,223,227]), France (e.g., [207,222,224]), Canada (e.g.,  
650 [145,186,195]), and Scandinavian nations (e.g., [182,189])—Denmark, Finland, Norway, and  
651 Sweden—leading the field. This dominance is linked to considerable funding from these regions and  
652 international collaborations. Scandinavian countries are particularly advanced in AHD use, offering  
653 databases that are highly complete, accessible, and well-integrated into public health research. AHDs  
654 from Denmark, Sweden, Canada, and France also provide comprehensive data on a patient's digital  
655 trajectory within their respective health systems [93,98,104,108,113,115,128,247]. France stands out  
656 further, with an AHD dedicated specifically to the entire farming population (MSA). This may explain  
657 the frequent large-scale and long-term studies from these countries, some of which included over 100  
658 000 participants. However, many studies still had a regional focus, partly due to the use of AHDs by  
659 traditional epidemiological studies like AGRICAN [207] and AHS [172], which rely on limited  
660 resources [47-50,53,54,57]. These studies often used AHDs to identify farming populations for inclusion  
661 or to supplement cohort data. The international AGRICOH consortium, initiated by the US National  
662 Cancer Institute (NCI) and the International Agency for Research on Cancer (IARC), includes 11 of the  
663 29 cohorts identified in this review [150,225]. However, several cohorts in AGRICOH were not  
664 identified, potentially due to lack of publications, language barriers, or limited use of AHDs. There were

665 many publications associated with these well-established cohorts, for which many of the most prolific  
666 authors were working [172,182,189,207,222-227].

667 The most frequently used AHDs in farming population health research were disease registers followed  
668 by electronic health/medical records and insurance claims. More than two-thirds of the studies used  
669 disease registers, in particular cancer and mortality registers. This is not surprising since disease registers  
670 are created for clinical and research purposes with continuous, exhaustive, and optimal digital collection  
671 of individual data regarding one or more health events in a geographically defined population [53,105].  
672 The coding systems and the granularity of information related to health outcomes, populations, and  
673 exposure determinants varied widely across studies. Most studies used AHDs to collect socio-  
674 demographic and health event information (e.g., [222]).

675 There was no consensus on the best methods or proxies to assess farming exposure. A variety of exposure  
676 proxies and determinants were used across studies, with indirect methods being the most common. Many  
677 studies dichotomized proxy, for example, classifying individuals as farmers or non-farmers or as  
678 pesticide-exposed versus non-exposed. In nearly two-thirds of the included studies, job title (i.e., “being  
679 a farmer”) served as the primary exposure proxy. About one-third of studies took into account specific  
680 farming activities (e.g., dairy farming, crop farming) to reflect the diversity of farming practices. This  
681 approach is a valuable proxy for agricultural exposure, offering a broader representation of the farming  
682 exposome, which involves multiple stressors beyond just pesticides [147,152,188,201,248,249].  
683 Farming activity information was often derived from digitalized data, such as agricultural censuses or  
684 self-reported data from mandatory insurance enrollments [152]. Many studies combined AHDs with  
685 self-reported data (e.g., questionnaires) (e.g., [172,207]), which allowed for more comprehensive data  
686 collection but tended to restrict the scope to regional studies due to resource constraints. These studies  
687 typically yielded high-quality data, with more potential confounders considered compared to studies  
688 relying solely on AHDs. Most studies using self-reported data focused on single exposures, mainly  
689 pesticides, with only one addressing multi-exposure to pesticides [236]. Biological monitoring and  
690 airborne chemical sampling were rarely conducted, likely due to practical and financial constraints and  
691 the short half-lives of most pesticides [250]. Dermal chemical monitoring has not been reported, even  
692 though it is the main exposure route for pesticides [251]. The high number of studies investigating

693 exposure to pesticides may be explained by the fact that AHS focuses on pesticide applicators and their  
694 spouses [172] and because many pesticides have human adverse health effects such as neurotoxicity or  
695 endocrine disruption [252-256]. Beyond pesticides, farmers face exposure to other chemicals [257], such  
696 as air pollution, micro- and nanoplastics [258-263], as well as biological agents (e.g., endotoxins,  
697 zoonoses) [264-266], physical agents (e.g., UV radiation, noise, vibrations) [187,267], biomechanical  
698 factors (e.g., repetitive movements, heavy load, working posture) [198,268,269], and psychosocial  
699 stressors [270-274], which have been less studied than pesticides. Despite these multiple exposures, the  
700 broader farming exposome remains understudied.

701 In addition to AHDs, some studies integrated other secondary data, such as climate data [187,275], air  
702 quality data [194,217], job-exposure matrices (JEMs) [212,276,277], or crop-exposure matrices [278].  
703 JEMs provide exposure level estimates for various chemicals and stressors based on job categories  
704 [250,279]. Although JEMs can provide valuable exposure data, they often lack the specificity of  
705 individual-level data, making it difficult to account for task-specific risks, temporal variations, and the  
706 inclusion of specific worker subgroups like females [250,279-282]. The lack of a universal standard for  
707 JEMs further complicates their application, which may explain why many studies still rely on self-  
708 reported data for more accurate exposure assessment, despite the risk of recall bias [250,279-282].

709 The health outcomes studied were predominantly cancer (e.g., [145,150,151,171-175,177-  
710 179,182,183,185,189,190,196,207,210-212,223]), mortality (e.g., [173,186,194,195,202,205]),  
711 workplace injuries (e.g., [198,208]), respiratory disorders (e.g., asthma and COPD) (e.g.,  
712 [151,180,181,213,224,226]), neurodegenerative diseases (e.g., Parkinson's disease) (e.g.,  
713 [111,151,176,184,187,188,193,197,201,248,249,283]), and mental health issues (e.g., [151,244,273]),  
714 which represent focal points within the research field. This is not surprising because many well-  
715 established cohorts centered on cancer research, in particular AGRICAN [207], CNAP [182] and  
716 NOCCA [189]. In addition, arsenic and inorganic arsenic compounds are classified as carcinogenic to  
717 humans by IARC [252,253,284], while malathion, glyphosate, diazinon, DDT and occupational  
718 exposures in spraying and application of non-arsenical insecticides are classified as probably  
719 carcinogenic to humans (group 2A), and several other pesticides are ranked as possibly carcinogenic to  
720 humans (group 2B), such as tetrachlorvinphos and parathion. Regarding mortality, it is often cancer and

721 suicide mortalities that are investigated [186,195,202,244]. Furthermore, several pesticides are  
722 neurotoxic [252,255], but existing studies focused mainly on Parkinson's disease and multiple sclerosis,  
723 with a paucity of data on Alzheimer's disease and other neurodegenerative diseases [147,283]. In  
724 contrast, certain areas, such as cardiovascular diseases (e.g., [151,194,203,227,285,286]), autoimmune  
725 disorders (e.g., [168,219,237]), musculoskeletal disorders (e.g., [192,204,245,287,288]), reproductive  
726 disorders (e.g., [242,289,290]), sleep disorders (e.g., [191]), aging-related conditions (e.g., [151,291]),  
727 hearing impairment [267,292], and the microbiome [8,293-298], remain underexplored, despite their  
728 potential relevance to farming populations.

729

### 730 **Challenges of reusing AHD for public health research in farming populations**

731 Each AHD presents unique advantages and limitations. For example, large sample sizes and a large  
732 number of available health events are strengths, while generalizability and the absence of key  
733 confounders are challenges [64,93,95,105,115]. Access to AHDs is frequently restricted by a variety of  
734 factors, including governance and technical barriers such as language, data structure, interoperability,  
735 and coding systems. Additional challenges stem from the type of AHD (e.g., insurance claims or cancer  
736 registers), inadequate documentation (e.g., absence of a data dictionary), limited accessibility due to  
737 costs or conditions, as well as jurisdictional and legal constraints [30,62,64,81,113,115]. Identifying the  
738 optimal AHD for a given research question is also complex, especially when considering the  
739 heterogeneity in coding systems and country-specific data structures. In countries like Scandinavia,  
740 Canada, and France, individual identifiers facilitate data linkage across multiple AHDs, enhancing  
741 research opportunities [93,104,113,115,128,247]. However, many AHDs are not research-ready and  
742 require significant processing, cleaning, and understanding before they can be analyzed  
743 [93,105,113,222,299]. Another major challenge is the long lag between data access, analysis, and  
744 research publication. On average, studies used data that were seven years old at the time of publication,  
745 largely due to delays in data access, administrative approvals, and the need to prepare complex datasets  
746 for analysis [223,300]. For instance, the TRACTOR project took two years to clean and prepare its  
747 dataset for research use [222]. These delays are compounded by the time required to conduct statistical  
748 analyses and prepare manuscripts for publication, as well as review and publication times (delay from

749 submission to acceptance and from acceptance to publication). Another limitation of AHDs is the lack  
750 of detailed exposure data. AHDs rarely include exposure information due to their administrative focus,  
751 requiring researchers to supplement with additional data sources, such as JEMs or self-reported data.  
752 When exposure information is recorded in AHDs, it is often too generalized, typically only reflecting  
753 broad job classifications, such as farming, without specifying detailed activities or stressors. There are  
754 some exceptions, such as MSA data, which captures a wider range of specific farming activities (e.g.,  
755 dairy farming) [222]. However, exposure to specific stressors (e.g., chemical compounds) is largely  
756 absent from AHDs.

757 The reference populations used in farming studies vary, which precludes direct comparisons and limits  
758 the generalizability of the findings. For example, AGRICAN used the general population as a reference  
759 [207], while TRACTOR used a farming population [152,168]. Studies also differ in their focus on  
760 specific farming populations, such as the entire agricultural workforce [207], farm managers [152], or  
761 pesticide applicators [172], which may lead to distinct exposure profiles that influence health outcomes  
762 because these farming populations have different socioeconomic status, experiences, and behaviors.  
763 Hence, to avoid or lessen bias, some studies focused on specific farming populations [152,172].  
764 Moreover, the scope of farming populations included in studies is often limited, omitting subgroups like  
765 farm families, nearby residents, or consumers exposed to agricultural products, which limits the broader  
766 application of the findings. In addition, farming practices can vary significantly between countries and  
767 studies, and there is no international standardized classification for farming activities. In some cases,  
768 farming categories are derived from legal or administrative sources, as seen in the MSA data  
769 [152,168,283]. This lack of standardization limits the comparability and generalizability of findings  
770 across studies. In addition, the generalizability of the findings to other countries when using AHDs may  
771 be limited because of the differences in healthcare systems [93].

772 There are several well-known limitations of AHDs that complicate the investigation of health outcomes  
773 [301]. Health events captured in AHDs are typically limited to those requiring medical attention, which  
774 may not reflect the true incidence of diseases. Additionally, the level of detail varies across diseases,  
775 even within the same AHD [92]. Although diagnostic accuracy is generally higher in disease registers,  
776 these are often geographically limited and cover only a subset of health conditions. For example, in

777 France, cancer registers only cover 23 out of 96 administrative regions [155]. In addition, certain  
778 conditions, such as depression, are not covered by any registers. Identifying health outcomes in AHDs  
779 often requires complex algorithms that combine data from multiple sources, such as drug  
780 reimbursements, disease declarations, or medical procedures [104,105,128,152,302-304]. Additionally,  
781 inconsistencies in case definitions and algorithms across studies and countries hinder the ability to  
782 compare and pool risk estimates [104,302,304]. Some AHDs also lack critical clinical information, such  
783 as laboratory results and genetic data [115,128,305], and the recorded diagnosis or treatment date may  
784 not correspond to the actual onset of the disease. Furthermore, diagnosis codes are not always indicative  
785 of a confirmed diagnosis.

786

### 787 **Emerging opportunities and research needs**

788 While AHDs are well-utilized in certain countries, there are underexplored opportunities in regions like  
789 the UK, where AHDs exist but are underutilized for research [113]. For low- and middle-income  
790 countries, the development and access to AHDs remain limited, and international support is needed to  
791 expand this research infrastructure. As already reported [306], here is also a notable lack of sex-specific  
792 data, even though occupational exposures and health outcomes can vary significantly between sexes due  
793 to genetic, physiological, psychological, and behavioral factors [307-311]. Future research should  
794 address these disparities to provide a more comprehensive understanding of health risks in farming  
795 communities. Although there are inherent delays in using AHDs due to the time required for data  
796 generation, consolidation, and access, we advocate for the continued publication of studies, even those  
797 using older data. Historical data remain vital for better understanding long-term health trends,  
798 particularly for diseases like cancer, where tumor initiation can span decades [312,313]. Editors should  
799 encourage the publication of studies using older datasets, especially when addressing long-term health  
800 outcomes (e.g., cancer, neurodegenerative diseases) or when recent data are not available [312].

801 None of the analyzed AHDs fully adhere to the FAIR principles, possibly because most were developed  
802 prior to the establishment of these principles in 2016 [228-230]. Moreover, the assessment of AHD FAIR  
803 compliance relied solely on information presented in publications, which may not provide a  
804 comprehensive evaluation. Nevertheless, there is a critical need to advocate for the integration of FAIR

805 principles within AHDs to enhance public health research [228-230]. Currently, the landscape is  
806 favorable for data reuse, particularly with initiatives like the forthcoming European Health Data Space  
807 [314-316]. Data reuse extends beyond mere access; it encompasses data discovery, a fundamental aspect  
808 of FAIR that involves recognizing the existence of databases [228-230]. To facilitate this, the creation  
809 of data catalogs is essential [228]. Numerous data repositories, such as Re3Data [317], Zenodo [318],  
810 CANUE [319], Figshare [320], “Epidémiologie – France” [321], data.gouv [322], Dataverse [320], or  
811 Data Europa [320], already exist. In addition, specialized multidisciplinary open-access and peer-  
812 reviewed journals like *Scientific Data* and *Data in Brief* publish datasets [318]. A dataset search can also  
813 be conducted using Google [323]. However, the documentation and access conditions for datasets can  
814 highly vary across inventories, complicating the selection process for researchers. The absence of  
815 indicators or scores for data reusability further hampers efforts to identify the most suitable datasets for  
816 specific research questions [45,69]. To our knowledge, no comprehensive catalog of AHDs currently  
817 exists to date. An online inventory of AHDs, modeled after existing resources like  
818 OccupationalCohorts.net [324], OccupationalExposureTools.net [325], and TEDI (Toxicological and  
819 Exposure Database Inventory) [10], could greatly enhance research endeavors. The motivation for  
820 analyzing AHDs often stems from the data they contain. Consequently, as data availability increases,  
821 researchers will be better positioned to formulate research questions and engage in a parallel process of  
822 “datagraphy” or “datagraphic search” akin to traditional bibliographic research. The objective of  
823 “datagraphy” would be to determine which datasets are best suited for addressing specific research  
824 questions, highlighting the need for accessible catalogs to support this goal.

825 There is also an opportunity to integrate other secondary data, such as person-generated data (e.g.,  
826 mobile health, social media, digital footprints, wearable sensors), contextual data (e.g., climate and air  
827 quality), and smart agriculture data [2,83,84,326-328]. These datasets, largely untapped in farming  
828 population research, could provide new insights into health outcomes and environmental exposures  
829 [101].

830 Nationwide studies using big data were a hotspot. AI, such as machine learning (ML), is particularly  
831 useful for analyzing big data and holds substantial promise for future research [329-331], particularly  
832 for identifying predictors of health outcomes in farming populations [332]. To date, AI has been

833 underutilized, with only two studies using it, one for identifying occupational injuries in agriculture  
834 [208] and one reviewing the development of chronic kidney disease risk prediction models [209].  
835 Incorporating AI, along with cohort enrichment and interdisciplinary expert interpretation, could open  
836 new avenues for research.

837 Many studies continue to examine agriculture as a broad category, highlighting the need for more  
838 detailed investigations into specific farming activities and tasks [147,152,188,201,248,249]. Our  
839 analysis reveals major research gaps in understanding environmental and occupational exposures among  
840 farming populations, particularly with regard to emerging concerns like PFAS, biological agents,  
841 micro/nanoplastics, and the impact of climate change. Climate change is a critical issue for agriculture,  
842 as it may drive shifts in pests, diseases, and farming practices [274,333-339]. Parental exposure appears  
843 to be a theme that will soon become a hot topic. Further research is also needed to explore the farming  
844 exposome, particularly focusing on the “mixture effects” of multiple simultaneous exposures [340,341].  
845 Omics data, which has not been utilized in farming population studies to date, represents a promising  
846 avenue for future research since genetics and metabolism were found to be a niche theme. Omics data  
847 refers to the large-scale datasets generated from various omics technologies that analyze biological  
848 molecules (e.g., genomics, transcriptomics, proteomics, metabolomics), which provide comprehensive  
849 insights into different biological layers and processes [11,342,343].

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

858 The prominence of topics such as cancer, neurodegenerative diseases, mortality, injuries, and mental  
859 health issues underscores the need for targeted prevention strategies. The thematic map analysis  
860 indicates that reproductive disorders (e.g., birth defects, endometriosis, infertility) are on the verge of



861 becoming a central research focus. Emerging and understudied health conditions, including ocular  
862 disorders, autoimmune diseases (e.g., inflammatory bowel disease, rheumatoid arthritis), sleep disorders  
863 (e.g., sleep apnea), cardiovascular diseases, and musculoskeletal conditions (e.g., low back pain),  
864 warrant increased attention and further research. Aging-related health issues such as frailty also  
865 represent promising avenues for future research, particularly given the growing aging population and  
866 associated healthcare challenges [24].

867

## 868 **Limitations**

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

887 While our keyword analysis helped map the farming exposome in AHD-based public health research,  
888 this profile is incomplete and potentially biased. Because our review focused on AHD-based studies, we

889 likely missed relevant epidemiological studies not using AHDs, leaving gaps in our understanding of  
890 the full farming exposome across public health. Additionally, the variability in keyword reporting  
891 practices across journals introduced bias into our keyword analysis. Some journals limit the number of  
892 keywords, and the lack of standardized keyword ontologies adds further variability. To mitigate this bias,  
893 we manually harmonized the keywords (e.g., use of one unique term for a given entity). While this  
894 approach is time-consuming, it allows for a more accurate analysis. For instance, if this approach was  
895 not performed, the same entity could be designated by various terms that would have been considered  
896 separate entities/terms, potentially resulting in underestimates (e.g., in the number of publications).  
897 Despite these challenges, the findings from our scoping review were consistent with the keyword  
898 analysis.

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

905

## 906 **Conclusions**

907 Technological advancements have greatly increased the volume of research data available, positioning  
908 AHDs as critical resources for population-based public health studies [41]. Our review underscores the  
909 broad public health implications of AHDs, providing actionable insights for researchers, physicians, and  
910 policymakers (Textbox 2). Addressing the identified research gaps is crucial to comprehensively  
911 understanding health risks in farming populations.

912

913

914

915

916

917

918 **Textbox 2.** Take-home messages.

919 **Farming population**

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

923

924 **AHD use**

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

930

931 **Farming exposure**

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

937

938 **Health outcomes**

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

944

945 The insights derived from AHDs can inform meaningful recommendations for policymakers and guide  
946 future research directions, ultimately aiding health services and health policy development. Our findings  
947 underscore the necessity of comprehensive, interdisciplinary approaches to better understand and  
948 mitigate the health risks encountered by farming populations. Such efforts will improve data

949 comparability and research quality while also supporting the formulation of targeted prevention  
950 strategies. This, in turn, will enhance health outcomes for farming populations and promote the  
951 sustainability of agriculture in an increasingly dynamic environment. The findings from this review offer  
952 insights that are not only relevant to farming populations but also potentially generalizable to other  
953 populations.

Accepted paper (not yet proofread)

954 **Acknowledgments**

955 None.

956

957 **Funding**

958 This work was partly by MIAI@Grenoble Alpes [ANR-19-P3IA-0003, 2019] and the French National  
959 Research Agency, within the framework of the “Investissements d’avenir” program [ANR-10-AIRT-05,  
960 2012; ANR-15-IDEX-02, 2016]. The funding sources had no role in the study design and conduct of the  
961 study; in the collection, management, analysis, and interpretation of data; in the writing of the report; in  
962 the preparation, review, or approval of the article; or in the decision to submit the paper for publication.  
963 The authors were not precluded from accessing data in the study, and all authors took final responsibility  
964 for the decision to submit for publication.

965

966 **Conflicts of Interest**

967 None declared.

968

969 **Abbreviations**

970 90%CI: 90% confidence interval

971 AAW: workplace accident

972 AGEIS: Autonomy, Gerontology, E-health, Imaging & Society (Autonomie, Gérontologie, E-santé,  
973 Imagerie et Société in French)

974 AGR: annual growth rate

975 AGRICAN: Agriculture and Cancer

976 AHD: administrative health database

977 AHS: Agricultural Health Study

978 AI: artificial intelligence

979 AMI: Aging Multidisciplinary Investigation

980 ANR: French National Research Agency (Agence Nationale de la Recherche in French)

- 981 BIBLIO: preliminary guideline for reporting bibliometric reviews of the biomedical literature
- 982 BMI: body mass index
- 983 COPD: chronic obstructive pulmonary disease
- 984 CNAP: Cancer in the Norwegian Agricultural population cohort
- 985 CVD: cardiovascular disorder
- 986 DDT: dichlorodiphenyltrichloroethane
- 987 EHR: electronic health record
- 988 EMR: electronic medical record
- 989 FAIR: findable, accessible, interoperable and reusable
- 990 FERMA: risk factors of the rural environment and the allergic and respiratory disease
- 991 FF: fractional frequency
- 992 GDP: gross domestic product
- 993 IARC: International Agency for Research on Cancer
- 994 IBD: inflammatory bowel disease
- 995 ICMJE: International Committee of Medical Journal Editors
- 996 IQR: interquartile range
- 997 JEM: job-exposure matrix
- 998 mHealth: mobile health
- 999 MIAI: Multidisciplinary Institute in Artificial Intelligence
- 1000 ML: machine learning
- 1001 MSA: National Health Insurance Fund for Agricultural Workers and Farmers (Mutualité Sociale  
1002 Agricole in French)
- 1003 NAWS: National Agricultural Workers Survey
- 1004 NCI: National Cancer Institute
- 1005 NOCCA: Nordic Occupational Cancer Study
- 1006 ORCID: Open Researcher and Contributor ID
- 1007 PFAS: per- and polyfluoroalkyl substances

1008 PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for  
1009 Scoping Reviews and evidence maps  
1010 R&D: research and development  
1011 TEDI: toxicological and exposure database inventory  
1012 TRACTOR: Tracking and monitoring occupational risks in agriculture  
1013 UFW: United Farm Workers of America  
1014 UGA: Université Grenoble Alpes  
1015 UK: United Kingdom  
1016 US: United States  
1017 UV: ultraviolet

1018

#### 1019 **Author contributions**

1020 PP: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data  
1021 Curation, Writing - Original Draft, Writing - Review & Editing, Visualization, Project administration,  
1022 Funding acquisition. NV: Conceptualization, Writing - Review & Editing, Resources, Funding  
1023 acquisition.

1024 All authors meet all four criteria for authorship in the ICMJE recommendations. All authors affirm that  
1025 the manuscript is an honest, accurate, and transparent account of the study being reported; that no  
1026 important aspects of the study have been omitted; and that any potential discrepancies from the study as  
1027 originally planned (and, if relevant, registered) have been explained.

1028

#### 1029 **Data Availability Statement**

1030 The datasets generated during and/or analyzed during the current study are available in the Multimedia  
1031 Appendices and from the corresponding author on reasonable request.

1032

1033 **Multimedia Appendices**

1034 **Multimedia appendix 1:** Preferred Reporting Items for Systematic reviews and Meta-Analyses  
1035 extension for Scoping Reviews (PRISMA-ScR) Checklist.

1036 **Multimedia appendix 2:** Criteria used for the publication selection.

1037 **Multimedia appendix 3:** Bibliometric index metrics.

1038 **Multimedia appendix 4:** The BIBLIO checklist for reporting the bibliometric reviews of the biomedical  
1039 literature.

1040 **Multimedia appendix 5:** Publication type.

1041 **Multimedia appendix 6:** Scientific productivity in terms of publications per period of time.

1042 **Multimedia appendix 7:** Historical direct citation network.

1043 **Multimedia appendix 8.** Most prolific authors' country.

1044 **Multimedia appendix 9.** Top 25 of the most cited publications.

1045 **Multimedia appendix 10.** Other bibliometric analysis information.

1046 **Multimedia appendix 11.** Publications included and analyzed.

1047 **Multimedia appendix 12.** Number of publications per country.

1048 **Multimedia appendix 13.** Main goal addressed by study.

1049 **Multimedia appendix 14.** Number of publications per cohort.

1050 **Multimedia appendix 15.** Type of digital data used.

1051 **Multimedia appendix 16.** Active data used.

1052 **Multimedia appendix 17.** Most frequent farming exposure proxy.

1053 **Multimedia appendix 18.** Most frequent health events studied.

1054 **Multimedia appendix 19.** Most frequent cancer studied.

1055 **Multimedia appendix 20.** Most frequent neurodegenerative disease studied.

1056 **Multimedia appendix 21.** Most frequent mental health issue studied.

1057 **Multimedia appendix 22.** Top 50 of the most frequent keywords.

1058 **Multimedia appendix 23.** Top 5 of the most frequent keywords – growth trends.



1059 **Multimedia appendix 24.** Top ten of the most popular keywords by total number of years of  
1060 appearance.

1061 **Multimedia appendix 25.** Top 10 of the most frequent keywords by time period.

1062 **Multimedia appendix 26.** Co-occurrence between exposome-related keywords and health event-related  
1063 keywords.

1064 **Multimedia appendix 27.** Keyword co-occurrence network between potential risk factor and mental  
1065 health disorder keywords.

1066 **Multimedia appendix 28.** Chord diagram of keyword co-occurrence between potential risk/protective  
1067 factor and neurodegenerative disease keywords.

1068 **Multimedia appendix 29.** Keyword co-occurrence network between potential risk/protective factor and  
1069 neurodegenerative disease keywords.

1070 **Multimedia appendix 30.** Chord diagram of keyword co-occurrence between potential risk/protective  
1071 factor and autoimmune disease keywords.

1072 **Multimedia appendix 31.** Keyword co-occurrence network between potential risk/protective factor and  
1073 autoimmune disease keywords.

1074 **Multimedia appendix 32.** Chord diagram of keyword co-occurrence between potential risk/protective  
1075 factor and breast and genital organ cancer keywords.

1076 **Multimedia appendix 33.** Keyword co-occurrence network between potential risk/protective factor and  
1077 breast and genital organ cancer keywords.

1078 **Multimedia appendix 34.** Chord diagram of keyword co-occurrence between potential risk/protective  
1079 factor and digestive organ, lip, oral cavity and pharynx cancer keywords.

1080 **Multimedia appendix 35.** Keyword co-occurrence network between potential risk/protective factor and  
1081 digestive organ, lip, oral cavity and pharynx cancer keywords.

1082 **Multimedia appendix 36.** Chord diagram of keyword co-occurrence between potential risk/protective  
1083 factor and lymphohematopoietic and mesenchymal cancer keywords.

1084 **Multimedia appendix 37.** Keyword co-occurrence network between potential risk/protective factor and  
1085 lymphohematopoietic and mesenchymal cancer keywords.

1086 **Multimedia appendix 38.** Chord diagram of keyword co-occurrence between potential risk/protective  
1087 factor and brain, skin, ocular and endocrine gland cancer keywords.

1088 **Multimedia appendix 39.** Keyword co-occurrence network between potential risk/protective factor and  
1089 brain, skin, ocular and endocrine gland cancer keywords.

1090 **Multimedia appendix 40.** Chord diagram of keyword co-occurrence between potential risk/protective  
1091 factor and respiratory and urinary tract cancer keywords.

1092 **Multimedia appendix 41.** Keyword co-occurrence network between potential risk/protective factor and  
1093 respiratory and urinary tract cancer keywords.

1094 **Multimedia appendix 42.** Chord diagram of keyword co-occurrence between potential risk/protective  
1095 factor and infectious disease keywords.

1096 **Multimedia appendix 43.** Keyword co-occurrence network between potential risk/protective factor and  
1097 infectious disease keywords.

1098 **Multimedia appendix 44.** Chord diagram of keyword co-occurrence between potential risk/protective  
1099 factor and respiratory disorder keywords.

1100 **Multimedia appendix 45.** Keyword co-occurrence network between potential risk/protective factor and  
1101 respiratory disorder keywords.

1102 **Multimedia appendix 46.** Chord diagram of keyword co-occurrence between potential risk/protective  
1103 factor and reproductive disorder keywords.

1104 **Multimedia appendix 47.** Keyword co-occurrence network between potential risk/protective factor and  
1105 reproductive disorder keywords.

1106 **Multimedia appendix 48.** Chord diagram of keyword co-occurrence between potential risk/protective  
1107 factor and injury and work-related disease keywords.

1108 **Multimedia appendix 49.** Keyword co-occurrence network between potential risk/protective factor and  
1109 injury and work-related disease keywords.

1110 **Multimedia appendix 50.** Chord diagram of keyword co-occurrence between potential risk/protective  
1111 factor and other health event keywords.

1112 **Multimedia appendix 51.** Keyword co-occurrence network between potential risk/protective factor and  
1113 other health event keywords.

1114 **Multimedia appendix 52.** Keyword co-occurrence network between potential risk/protective factor and  
1115 mortality keywords.

Accepted paper (not yet proofread)

1116 **References**

- 1117 1. Azari R, Borisch B. What is public health? a scoping review. Arch Public Health.  
1118 2023;81(1):86. doi:10.1186/s13690-023-01091-6.
- 1119 2. Velmovitsky PE, Kirolos M, Alencar P, Leatherdale S, Cowan D, Morita PP. Leveraging  
1120 mHealth Technologies for Public Health. JMIR Public Health Surveill. 2024;10:e49719.  
1121 doi:10.2196/49719.
- 1122 3. Académie Nationale de Médecine. Rapport 24-02. L'exposome, une contribution majeure pour  
1123 la prévention. Bull Acad Natl Med. 2024;208:523-35. doi:10.1016/j.banm.2024.03.004.
- 1124 4. Barouki R, Audouze K, Becker C, et al. The Exposome and Toxicology: A Win-Win  
1125 Collaboration. Toxicol Sci. 2022;186(1):1-11. doi:10.1093/toxsci/kfab149.
- 1126 5. Bennett B, Workman T, Smith MN, Griffith WC, Thompson B, Faustman EM. Characterizing  
1127 the Neurodevelopmental Pesticide Exposome in a Children's Agricultural Cohort. Int J Environ  
1128 Res Public Health. 2020;17(5):1479. doi:10.3390/ijerph17051479.
- 1129 6. Goodrich JA, Wang H, Jia Q, et al. Integrating Multi-Omics with environmental data for  
1130 precision health: A novel analytic framework and case study on prenatal mercury induced  
1131 childhood fatty liver disease. Environ Int. 2024;190:108930. doi:10.1016/j.envint.2024.108930.
- 1132 7. Haddad N, Andrianou XD, Makris KC. A Scoping Review on the Characteristics of Human  
1133 Exposome Studies. Curr Pollut Rep. 2019;5(4):378-393. doi:10.1007/s40726-019-00130-7.
- 1134 8. Merra G, Gualtieri P, La Placa G, et al. The Relationship between Exposome and Microbiome.  
1135 Microorganisms. 2024;12(7):1386. doi:10.3390/microorganisms12071386.
- 1136 9. Pandics T, Major D, Fazekas-Pongor V, et al. Exposome and unhealthy aging: environmental  
1137 drivers from air pollution to occupational exposures. Geroscience. 2023;45(6):3381-3408.  
1138 doi:10.1007/s11357-023-00913-3.
- 1139 10. Petit P. Toxicological and Exposure Database Inventory: A review. Int J Hyg Environ Health.  
1140 2022;246:114055. doi:10.1016/j.ijheh.2022.114055.
- 1141 11. Price EJ, Vitale CM, Miller GW, et al. Merging the exposome into an integrated framework for  
1142 "omics" sciences. iScience. 2022;25(3):103976. doi:10.1016/j.isci.2022.103976.

- 1143 12. Santos S, Maitre L, Warembourg C, et al. Applying the exposome concept in birth cohort  
1144 research: a review of statistical approaches. *Eur J Epidemiol.* 2020;35(3):193-204.  
1145 doi:10.1007/s10654-020-00625-4.
- 1146 13. Siroux V, Agier L, Slama R. The exposome concept: a challenge and a potential driver for  
1147 environmental health research. *Eur Respir Rev.* 2016;25(140):124-129.  
1148 doi:10.1183/16000617.0034-2016.
- 1149 14. Tagliaferro S, Maio S, Pirona F, et al. Assessing external exposome by implementing an  
1150 Environmental Data Management System using Open Data. *Sci Rep.* 2024;14(1):17142.  
1151 doi:10.1038/s41598-024-62924-0.
- 1152 15. Vineis P, Barouki R. The exposome as the science of social-to-biological transitions. *Environ*  
1153 *Int.* 2022;165:107312. doi:10.1016/j.envint.2022.107312.
- 1154 16. Wei X, Huang Z, Jiang L, et al. Charting the landscape of the environmental exposome. *Imeta.*  
1155 2022;1(4):e50. doi:10.1002/imt2.50.
- 1156 17. Wild CP. The exposome: from concept to utility. *Int J Epidemiol.* 2012;41(1):24-32.  
1157 doi:10.1093/ije/dyr236.
- 1158 18. Zhang P, Carlsten C, Chaleckis R, et al. Defining the Scope of Exposome Studies and Research  
1159 Needs from a Multidisciplinary Perspective. *Environ Sci Technol Lett.* 2021;8(10):839-852.  
1160 doi:10.1021/acs.estlett.1c00648.
- 1161 19. Ibanez A, Melloni L, Świeboda P, et al. Neuroecological links of the exposome and One  
1162 Health. *Neuron.* 2024;112(12):1905-1910. doi:10.1016/j.neuron.2024.04.016.
- 1163 20. Liang KH, Colombijn JMT, Verhaar MC, Ghannoum M, Timmermans EJ, Vernooij RWM. The  
1164 general external exposome and the development or progression of chronic kidney disease: A  
1165 systematic review and meta-analyses. *Environ Pollut.* 2024;358:124509.  
1166 doi:10.1016/j.envpol.2024.124509.
- 1167 21. Canfell OJ, Littlewood R, Burton-Jones A, Sullivan C. Digital health and precision prevention:  
1168 shifting from disease-centred care to consumer-centred health. *Aust Health Rev.*  
1169 2022;46(3):279-283. doi:10.1071/AH21063.

- 1170 22. Mess F, Blaschke S, Schick TS, Friedrich J. Precision prevention in worksite health-A scoping  
1171 review on research trends and gaps. *PLoS One*. 2024;19(6):e0304951.  
1172 doi:10.1371/journal.pone.0304951.
- 1173 23. Wang Y, Zhu M, Ma H, Shen H. Polygenic risk scores: the future of cancer risk prediction,  
1174 screening, and precision prevention. *Med Rev (2021)*. 2022;1(2):129-149. doi:10.1515/mr-  
1175 2021-0025.
- 1176 24. Kalia V, Belsky DW, Baccarelli AA, Miller GW. An exposomic framework to uncover  
1177 environmental drivers of aging. *Exposome*. 2022;2(1):osac002.  
1178 doi:10.1093/exposome/osac002.
- 1179 25. Chung MK, House JS, Akhtari FS, et al. Decoding the exposome: data science methodologies  
1180 and implications in exposome-wide association studies (ExWASs). *Exposome*.  
1181 2024;4(1):osae001. doi:10.1093/exposome/osae001.
- 1182 26. Andreu-Perez J, Poon CC, Merrifield RD, Wong ST, Yang GZ. Big data for health. *IEEE J*  
1183 *Biomed Health Inform*. 2015;19(4):1193-1208. doi:10.1109/JBHI.2015.2450362.
- 1184 27. Li C, Huang M. Environmental Sustainability in the Age of Big Data: Opportunities and  
1185 Challenges for Business and Industry. *Environ Sci Pollut Res Int*. 2023;30(56):119001-119015.  
1186 doi:10.1007/s11356-023-30301-5.
- 1187 28. Pastorino R, De Vito C, Migliara G, et al. Benefits and challenges of Big Data in healthcare: an  
1188 overview of the European initiatives. *Eur J Public Health*. 2019;29(Supplement\_3):23-27.  
1189 doi:10.1093/eurpub/ckz168.
- 1190 29. Flahault A, Geissbuhler A, Guessous I, et al. Precision global health in the digital age. *Swiss*  
1191 *Med Wkly*. 2017;147:w14423. doi:10.4414/smw.2017.14423.
- 1192 30. Manaouil C, Chamot S, Petit P. Le médecin confronté à l'IA (Intelligence artificielle) : Éthique  
1193 et responsabilité. *Médecine & Droit*. 2024. doi:10.1016/j.meddro.2024.02.001.
- 1194 31. Samuel G, Lucassen AM. The environmental sustainability of data-driven health research: A  
1195 scoping review. *Digit Health*. 2022;8:20552076221111297. doi:10.1177/20552076221111297.
- 1196 32. Adam NR, Wieder R, Ghosh D. Data science, learning, and applications to biomedical and  
1197 health sciences. *Ann N Y Acad Sci*. 2017;1387(1):5-11. doi:10.1111/nyas.13309.

- 1198 33. Iyamu I, Xu AXT, Gómez-Ramírez O, et al. Defining Digital Public Health and the Role of  
1199 Digitization, Digitalization, and Digital Transformation: Scoping Review. *JMIR Public Health*  
1200 *Surveill.* 2021;7(11):e30399. doi:10.2196/30399.
- 1201 34. Mooney SJ, Pejaver V. Big Data in Public Health: Terminology, Machine Learning, and  
1202 Privacy. *Annu Rev Public Health.* 2018;39:95-112. doi:10.1146/annurev-publhealth-040617-  
1203 014208.
- 1204 35. Wesson P, Hswen Y, Valdes G, Stojanovski K, Handley MA. Risks and Opportunities to Ensure  
1205 Equity in the Application of Big Data Research in Public Health. *Annu Rev Public Health.*  
1206 2022;43:59-78. doi:10.1146/annurev-publhealth-051920-110928.
- 1207 36. Iyamu I, Gómez-Ramírez O, Xu AX, et al. Challenges in the development of digital public  
1208 health interventions and mapped solutions: Findings from a scoping review. *Digit Health.*  
1209 2022;8:20552076221102255. doi:10.1177/20552076221102255.
- 1210 37. Iyamu I, McKee G, Haag D, Gilbert M. Defining the role of digital public health in the evolving  
1211 digital health landscape: policy and practice implications in Canada. Définition du rôle de la  
1212 santé publique numérique dans le paysage changeant de la santé numérique : répercussions sur  
1213 les politiques et les pratiques au Canada. *Health Promot Chronic Dis Prev Can.* 2024;44(2):66-  
1214 69. doi:10.24095/hpcdp.44.2.04.
- 1215 38. Khan S, Kandukuri DK, Subramaniyan EU, MohanaSundaram A. Harnessing the untapped  
1216 potential of digital twin technology in digital public health interventions. *Explor Digit Health*  
1217 *Technol.* 2023;1:11-16. doi:10.37349/edht.2023.00003.
- 1218 39. Maaß L, Pan CC, Freye M. Mapping Digital Public Health Interventions Among Existing  
1219 Digital Technologies and Internet-Based Interventions to Maintain and Improve Population  
1220 Health in Practice: Protocol for a Scoping Review. *JMIR Res Protoc.* 2022;11(3):e33404.  
1221 doi:10.2196/33404.
- 1222 40. Odone A, Buttigieg S, Ricciardi W, Azzopardi-Muscat N, Staines A. Public health digitalization  
1223 in Europe. *Eur J Public Health.* 2019;29(Supplement\_3):28-35. doi:10.1093/eurpub/ckz161.

- 1224 41. Wienert J, Jahnel T, Maaß L. What are Digital Public Health Interventions? First Steps Toward  
1225 a Definition and an Intervention Classification Framework. *J Med Internet Res.*  
1226 2022;24(6):e31921. doi:10.2196/31921.
- 1227 42. Burgun A, Bernal-Delgado E, Kuchinke W, et al. Health Data for Public Health: Towards New  
1228 Ways of Combining Data Sources to Support Research Efforts in Europe. *Yearb Med Inform.*  
1229 2017;26(1):235-240. doi:10.15265/IY-2017-034.
- 1230 43. Canfell OJ, Davidson K, Sullivan C, Eakin E, Burton-Jones A. Data sources for precision public  
1231 health of obesity: a scoping review, evidence map and use case in Queensland, Australia. *BMC*  
1232 *Public Health.* 2022;22(1):584. doi:10.1186/s12889-022-12939-x.
- 1233 44. Canfell OJ, Davidson K, Woods L, et al. Precision Public Health for Non-communicable  
1234 Diseases: An Emerging Strategic Roadmap and Multinational Use Cases. *Front Public Health.*  
1235 2022;10:854525. doi:10.3389/fpubh.2022.854525.
- 1236 45. Sielemann K, Hafner A, Pucker B. The reuse of public datasets in the life sciences: potential  
1237 risks and rewards. *PeerJ.* 2020;8:e9954. doi:10.7717/peerj.9954.
- 1238 46. Deal JA, Betz J, Lin FR, Reed NS. Interpreting Results from Epidemiologic Studies. *Semin*  
1239 *Hear.* 2021;42(1):3-9. doi:10.1055/s-0041-1725996.
- 1240 47. DiPietro NA. Methods in epidemiology: observational study designs. *Pharmacotherapy.*  
1241 2010;30(10):973-984. doi:10.1592/phco.30.10.973.
- 1242 48. dos Santos Silva I. *Cancer Epidemiology: Principles and Methods.* Lyon : CIRC ; 1999. ISBN  
1243 : 978-92-832-0405-3.
- 1244 49. Munnangi S, Boktor SW. *Epidemiology Of Study Design.* In: StatPearls. Treasure Island (FL):  
1245 StatPearls Publishing; April 24, 2023.
- 1246 50. Röhrig B, du Prel JB, Wachtlin D, Blettner M. Types of study in medical research: part 3 of a  
1247 series on evaluation of scientific publications. *Dtsch Arztebl Int.* 2009;106(15):262-268.  
1248 doi:10.3238/arztebl.2009.0262.
- 1249 51. Himali JJ, Baril AA, Cavuoto MG, et al. Association Between Slow-Wave Sleep Loss and  
1250 Incident Dementia. *JAMA Neurol.* 2023;80(12):1326-1333.  
1251 doi:10.1001/jamaneurol.2023.3889.



- 1252 52. Mahmood SS, Levy D, Vasan RS, Wang TJ. The Framingham Heart Study and the  
1253 epidemiology of cardiovascular disease: a historical perspective. *Lancet*. 2014;383(9921):999-  
1254 1008. doi:10.1016/S0140-6736(13)61752-3.
- 1255 53. Salines G, De Launay C. Les cohortes : intérêt, rôle et position de l'InVS. Saint-Maurice (Fra)  
1256 : Institut de veille sanitaire, janvier 2010, 20 p. ISBN: 978-2-11-099073-0.
- 1257 54. Zins M, Goldberg M. L'apport des bases de données d'origine administrative aux cohortes  
1258 épidémiologiques : l'exemple de la cohorte Constances. *Statistique et société*. 2014;2(2):43-48.
- 1259 55. Chen W, Wang X, Chen J, et al. Household air pollution, adherence to a healthy lifestyle, and  
1260 risk of cardiometabolic multimorbidity: Results from the China health and retirement  
1261 longitudinal study. *Sci Total Environ*. 2023;855:158896. doi:10.1016/j.scitotenv.2022.158896.
- 1262 56. Zhao Y, Hu Y, Smith JP, Strauss J, Yang G. Cohort profile: the China Health and Retirement  
1263 Longitudinal Study (CHARLS). *Int J Epidemiol*. 2014;43(1):61-68. doi:10.1093/ije/dys203.
- 1264 57. Kesse-Guyot E, Assmann K, Andreeva V, et al. Lessons Learned From Methodological  
1265 Validation Research in E-Epidemiology. *JMIR Public Health Surveill*. 2016;2(2):e160.  
1266 doi:10.2196/publichealth.5880.
- 1267 58. Goldberg M, Quantin C, Guéguen A, Zins M. Bases de données médico-administratives et  
1268 épidémiologie : intérêts et limites. *Courrier des statistiques*. 2008;124:59-70.
- 1269 59. Babik KR, Shockey TM, Moore LL, Wurzelbacher SJ. Standardizing industrial hygiene data  
1270 collection forms used by workers' compensation insurers. *J Occup Environ Hyg*.  
1271 2018;15(9):676-685. doi:10.1080/15459624.2018.1490022.
- 1272 60. Bosward R, Braunack-Mayer A, Frost E, Carter S. Mapping precision public health definitions,  
1273 terminology and applications: a scoping review protocol. *BMJ Open*. 2022;12(2):e058069.  
1274 doi:10.1136/bmjopen-2021-058069.
- 1275 61. Canfell OJ, Kodiyattu Z, Eakin E, et al. Real-world data for precision public health of  
1276 noncommunicable diseases: a scoping review. *BMC Public Health*. 2022;22(1):2166.  
1277 doi:10.1186/s12889-022-14452-7.

- 1278 62. Chao K, Sarker MNI, Ali I, Firdaus RBR, Azman A, Shaed MM. Big data-driven public health  
1279 policy making: Potential for the healthcare industry. *Heliyon*. 2023;9(9):e19681.  
1280 doi:10.1016/j.heliyon.2023.e19681.
- 1281 63. Dolley S. Big Data's Role in Precision Public Health. *Front Public Health*. 2018;6:68.  
1282 doi:10.3389/fpubh.2018.00068.
- 1283 64. Doyle CM, Lix LM, Hemmelgar BR, Michael Paterson J, Renoux C. Data variability across  
1284 Canadian administrative health databases: Differences in content, coding, and completeness.  
1285 *Pharmacoepidemiol Drug Saf*. 2020;29:68-77. doi:10.1002/pds.4889.
- 1286 65. Gallini A, Moisan F, Maura G, et al. Identification des maladies neurodégénératives dans les  
1287 bases de données médicoadministratives en France : revue systématique de la littérature  
1288 [Identification of neurodegenerative diseases in administrative databases in France: A  
1289 systematic review of the literature]. *Rev Epidemiol Sante Publique*. 2017;65 Suppl 4:S183-  
1290 S197. doi:10.1016/j.respe.2017.01.115.
- 1291 66. Khoury MJ, Iademarco MF, Riley WT. Precision Public Health for the Era of Precision  
1292 Medicine. *Am J Prev Med*. 2016;50(3):398-401. doi:10.1016/j.amepre.2015.08.031.
- 1293 67. McCullough LE, Maliniak ML, Amin AB, et al. Epidemiology beyond its limits. *Sci Adv*.  
1294 2022;8(23):eabn3328. doi:10.1126/sciadv.abn3328.
- 1295 68. Näher AF, Vorisek CN, Klopfenstein SAI, et al. Secondary data for global health  
1296 digitalisation. *Lancet Digit Health*. 2023;5(2):e93-e101. doi:10.1016/S2589-7500(22)00195-9.
- 1297 69. Koesten L, Vougiouklis P, Simperl E, Groth P. Dataset Reuse: Toward Translating Principles  
1298 to Practice. *Patterns (N Y)*. 2020;1(8):100136. doi:10.1016/j.patter.2020.100136.
- 1299 70. Hu H, Galea S, Rosella L, Henry D. Big Data and Population Health: Focusing on the Health  
1300 Impacts of the Social, Physical, and Economic Environment. *Epidemiology*. 2017;28(6):759-  
1301 762. doi:10.1097/EDE.0000000000000711.
- 1302 71. Shaw RJ, Harron KL, Pescarini JM, et al. Biases arising from linked administrative data for  
1303 epidemiological research: a conceptual framework from registration to analyses. *Eur J*  
1304 *Epidemiol*. 2022;37(12):1215-1224. doi:10.1007/s10654-022-00934-w.

- 1305 72. Subramanian M, Wojtuszczyzn A, Favre L, et al. Precision medicine in the era of artificial  
1306 intelligence: implications in chronic disease management. *J Transl Med.* 2020;18(1):472.  
1307 doi:10.1186/s12967-020-02658-5.
- 1308 73. Velmovitsky PE, Bevilacqua T, Alencar P, Cowan D, Morita PP. Convergence of Precision  
1309 Medicine and Public Health Into Precision Public Health: Toward a Big Data Perspective. *Front*  
1310 *Public Health.* 2021;9:561873. doi:10.3389/fpubh.2021.561873.
- 1311 74. Zhang H, Hu H, Diller M, et al. Semantic standards of external exposome data. *Environ Res.*  
1312 2021;197:111185. doi:10.1016/j.envres.2021.111185.
- 1313 75. Daniels K, Bonnechère B. Harnessing digital health interventions to bridge the gap in prevention  
1314 for older adults. *Front Public Health.* 2024;11:1281923. doi:10.3389/fpubh.2023.1281923.
- 1315 76. Liu F, Panagiotakos D. Real-world data: a brief review of the methods, applications, challenges  
1316 and opportunities. *BMC Med Res Methodol.* 2022;22(1):287. doi:10.1186/s12874-022-01768-  
1317 6.
- 1318 77. Maaß L, Freye M, Pan CC, Dassow HH, Niess J, Jähnel T. The Definitions of Health Apps and  
1319 Medical Apps From the Perspective of Public Health and Law: Qualitative Analysis of an  
1320 Interdisciplinary Literature Overview. *JMIR Mhealth Uhealth.* 2022;10(10):e37980.  
1321 doi:10.2196/37980.
- 1322 78. Mooney SJ, Garber MD. Sampling and Sampling Frames in Big Data Epidemiology. *Curr*  
1323 *Epidemiol Rep.* 2019;6(1):14-22. doi:10.1007/s40471-019-0179-y.
- 1324 79. Pearson AL, Tribby C, Brown CD, Yang JA, Pfeiffer K, Jankowska MM. Systematic review of  
1325 best practices for GPS data usage, processing, and linkage in health, exposure science and  
1326 environmental context research. *BMJ Open.* 2024;14(2):e077036. doi:10.1136/bmjopen-2023-  
1327 077036.
- 1328 80. Prospero M, Min JS, Bian J, Modave F. Big data hurdles in precision medicine and precision  
1329 public health. *BMC Med Inform Decis Mak.* 2018;18(1):139. doi:10.1186/s12911-018-0719-2.
- 1330 81. Safran C. Update on Data Reuse in Health Care. *Yearb Med Inform.* 2017;26(1):24-27.  
1331 doi:10.15265/IY-2017-013.

- 1332 82. Severo M, Feredj A, Romele A. Soft data and public policy: can social media offer alternatives  
1333 to official statistics in urban policymaking? *Pol Internet*. 2016;8:354–372.  
1334 doi:10.1002/poi3.127.
- 1335 83. Shakeri Hossein Abad Z, Butler GP, Thompson W, Lee J. Physical Activity, Sedentary  
1336 Behavior, and Sleep on Twitter: Multicountry and Fully Labeled Public Data Set for Digital  
1337 Public Health Surveillance Research. *JMIR Public Health Surveill*. 2022;8(2):e32355.  
1338 doi:10.2196/32355.
- 1339 84. Singh T, Roberts K, Cohen T, et al. Social Media as a Research Tool (SMaaRT) for Risky  
1340 Behavior Analytics: Methodological Review. *JMIR Public Health Surveill*. 2020;6(4):e21660.  
1341 doi:10.2196/21660.
- 1342 85. Zins M, Cuggia M, Goldberg M. Les données de santé en France. Abondantes mais complexes.  
1343 *médecine/sciences*. 2021;37(2):179-84. doi:10.1051/medsci/2021001.
- 1344 86. Motahari-Nezhad H, Fgaier M, Mahdi Abid M, Péntek M, Gulácsi L, Zrubka Z. Digital  
1345 Biomarker-Based Studies: Scoping Review of Systematic Reviews. *JMIR Mhealth Uhealth*.  
1346 2022;10(10):e35722. doi:10.2196/35722.
- 1347 87. Ayilara OF, Platt RW, Dahl M, et al. Generating synthetic data from administrative health  
1348 records for drug safety and effectiveness studies. *Int J Popul Data Sci*. 2023;8(1):2176.  
1349 doi:10.23889/ijpds.v8i1.2176.
- 1350 88. Fadel M, Petot J, Gourraud PA, Descatha A. Flexibility of a large blindly synthesized avatar  
1351 database for occupational research: Example from the CONSTANCES cohort for stroke and  
1352 knee pain. *PLoS One*. 2024;19(7):e0308063. doi:10.1371/journal.pone.0308063.
- 1353 89. Giuffrè M, Shung DL. Harnessing the power of synthetic data in healthcare: innovation,  
1354 application, and privacy. *NPJ Digit Med*. 2023;6(1):186. doi:10.1038/s41746-023-00927-3.
- 1355 90. Kamel Boulos MN, Zhang P. Digital Twins: From Personalised Medicine to Precision Public  
1356 Health. *J Pers Med*. 2021;11(8):745. doi:10.3390/jpm11080745.
- 1357 91. Vallevik VB, Babic A, Marshall SE, et al. Can I trust my fake data - A comprehensive quality  
1358 assessment framework for synthetic tabular data in healthcare. *Int J Med Inform*.  
1359 2024;185:105413. doi:10.1016/j.ijmedinf.2024.105413.

- 1360 92. Chamot S, Manaouil C, Fantoni S et al. Système national des données de santé (SNDS) :  
1361 perspectives en santé au travail et environnementale. Arch Mal Prof Environ.  
1362 2023;84(6):101875. doi:10.1016/j.admp.2023.101875.
- 1363 93. Cohen S, Gilutz H, Marelli AJ, et al. Administrative health databases for addressing emerging  
1364 issues in adults with CHD: a systematic review. Cardiol Young. 2018;28(6):844-853.  
1365 doi:10.1017/S1047951118000446.
- 1366 94. Fraser HSF, Mugisha M, Bacher I, et al. Factors Influencing Data Quality in Electronic Health  
1367 Record Systems in 50 Health Facilities in Rwanda and the Role of Clinical Alerts: Cross-  
1368 Sectional Observational Study. JMIR Public Health Surveill. 2024;10:e49127.  
1369 doi:10.2196/49127.
- 1370 95. Gavriellov-Yusim N, Friger M. Use of administrative medical databases in population-based  
1371 research. J Epidemiol Community Health. 2014;68(3):283-287. doi:10.1136/jech-2013-202744.
- 1372 96. Hinds A, Lix LM, Smith M, Quan H, Sanmartin C. Quality of administrative health databases  
1373 in Canada: A scoping review. Can J Public Health. 2016;107(1):e56-e61.  
1374 doi:10.17269/cjph.107.5244.
- 1375 97. Howell CR, Zhang L, Clay OJ, et al. Social Determinants of Health Phenotypes and  
1376 Cardiometabolic Condition Prevalence Among Patients in a Large Academic Health System:  
1377 Latent Class Analysis. JMIR Public Health Surveill. 2024;10:e53371. doi:10.2196/53371.
- 1378 98. Jones KH, Heys S, Tingay KS, Jackson P, Dibben C. The Good, the Bad, the Clunky: Improving  
1379 the Use of Administrative Data for Research. Int J Popul Data Sci. 2018;4(1):587.  
1380 doi:10.23889/ijpds.v4i1.587.
- 1381 99. Namulanda G, Qualters J, Vaidyanathan A, et al. Electronic health record case studies to  
1382 advance environmental public health tracking. J Biomed Inform. 2018;79:98-104.  
1383 doi:10.1016/j.jbi.2018.02.012.
- 1384 100. Nicholls SG, Quach P, von Elm E, et al. The REporting of Studies Conducted Using  
1385 Observational Routinely-Collected Health Data (RECORD) Statement: Methods for Arriving  
1386 at Consensus and Developing Reporting Guidelines. PLoS One. 2015;10(5):e0125620.  
1387 doi:10.1371/journal.pone.0125620.

- 1388 101. Silverwood RJ, Rajah N, Calderwood L, De Stavola BL, Harron K, Ploubidis GB.  
1389 Examining the quality and population representativeness of linked survey and administrative  
1390 data: guidance and illustration using linked 1958 National Child Development Study and  
1391 Hospital Episode Statistics data. *Int J Popul Data Sci.* 2024;9(1):2137.  
1392 doi:10.23889/ijpds.v9i1.2137.
- 1393 102. Zuiderwijk A, Türk BO, Brazier F. Identifying the most important facilitators of open  
1394 research data sharing and reuse in Epidemiology: A mixed-methods study. *PLoS One.*  
1395 2024;19(2):e0297969. doi:10.1371/journal.pone.0297969.
- 1396 103. Hobensack M, Song J, Scharp D, Bowles KH, Topaz M. Machine learning applied to  
1397 electronic health record data in home healthcare: A scoping review. *Int J Med Inform.*  
1398 2023;170:104978. doi:10.1016/j.ijmedinf.2022.104978.
- 1399 104. Tassi MF, le Meur N, Stéfic K, Grammatico-Guillon L. Performance of French medico-  
1400 administrative databases in epidemiology of infectious diseases: a scoping review. *Front Public*  
1401 *Health.* 2023;11:1161550. doi:10.3389/fpubh.2023.1161550.
- 1402 105. Yoshida S, Takeuchi M, Tanaka-Mizuno S, et al. Clinical epidemiology and  
1403 pharmacoepidemiology studies with real-world databases. *Proc Jpn Acad Ser B Phys Biol Sci.*  
1404 2022;98(10):517-528. doi:10.2183/pjab.98.026.
- 1405 106. Howren A, Aviña-Zubieta JA, Puyat JH, Esdaile JM, Da Costa D, De Vera MA.  
1406 Defining Depression and Anxiety in Individuals With Rheumatic Diseases Using  
1407 Administrative Health Databases: A Systematic Review. *Arthritis Care Res (Hoboken).*  
1408 2020;72(2):243-255. doi:10.1002/acr.24048.
- 1409 107. Kim MK, Roupael C, McMichael J, Welch N, Dasarathy S. Challenges in and  
1410 Opportunities for Electronic Health Record-Based Data Analysis and Interpretation. *Gut Liver.*  
1411 2024;18(2):201-208. doi:10.5009/gnl230272.
- 1412 108. Lee S, Xu Y, D Apos Souza AG, et al. Unlocking the Potential of Electronic Health  
1413 Records for Health Research. *Int J Popul Data Sci.* 2020;5(1):1123.  
1414 doi:10.23889/ijpds.v5i1.1123.

- 1415 109. Pletscher F, Mändli Lerch K, Glinz D. Willingness to share anonymised routinely  
1416 collected clinical health data in Switzerland: a cross-sectional survey. *Swiss Med Wkly.*  
1417 2022;152:w30182. doi:10.4414/smw.2022.w30182.
- 1418 110. Yuan H, Ali MS, Brouwer ES, et al. Real-World Evidence: What It Is and What It Can  
1419 Tell Us According to the International Society for Pharmacoepidemiology (ISPE) Comparative  
1420 Effectiveness Research (CER) Special Interest Group (SIG). *Clin Pharmacol Ther.*  
1421 2018;104(2):239-241. doi:10.1002/cpt.1086.
- 1422 111. Andrew A, Zhou J, Gui J, et al. Pesticides applied to crops and amyotrophic lateral  
1423 sclerosis risk in the U.S. *Neurotoxicology.* 2021;87:128-135. doi:10.1016/j.neuro.2021.09.004.
- 1424 112. Chazard E, Ficheur G, Caron A, et al. Secondary Use of Healthcare Structured Data:  
1425 The Challenge of Domain-Knowledge Based Extraction of Features. *Stud Health Technol*  
1426 *Inform.* 2018;255:15-19. PMID: 30306898.
- 1427 113. Grath-Lone LM, Jay MA, Blackburn R, et al. What makes administrative data  
1428 "research-ready"? A systematic review and thematic analysis of published literature. *Int J Popul*  
1429 *Data Sci.* 2022;7(1):1718. doi:10.23889/ijpds.v6i1.1718.
- 1430 114. Hashimoto RE, Brodt ED, Skelly AC, Dettori JR. Administrative database studies:  
1431 goldmine or goose chase?. *Evid Based Spine Care J.* 2014;5(2):74-76. doi:10.1055/s-0034-  
1432 1390027.
- 1433 115. Bezin J, Duong M, Lassalle R, et al. The national healthcare system claims databases in  
1434 France, SNIIRAM and EGB: Powerful tools for pharmacoepidemiology. *Pharmacoepidemiol*  
1435 *Drug Saf.* 2017;26(8):954-962. doi:10.1002/pds.4233.
- 1436 116. Al-Sahab B, Leviton A, Loddenkemper T, Paneth N, Zhang B. Biases in Electronic  
1437 Health Records Data for Generating Real-World Evidence: An Overview. *J Healthc Inform Res.*  
1438 2023;8(1):121-139. doi:10.1007/s41666-023-00153-2.
- 1439 117. Bastarache L, Brown JS, Cimino JJ, et al. Developing real-world evidence from real-  
1440 world data: Transforming raw data into analytical datasets. *Learn Health Syst.*  
1441 2021;6(1):e10293. doi:10.1002/lrh2.10293.

- 1442 118. Fang J, Kapral MK, Richards J, Robertson A, Stamplecoski M, Silver FL. The Registry  
1443 of Canadian Stroke Network : an evolving methodology. *Acta Neurol Taiwan*. 2011;20(2):77-  
1444 84. PMID: 21739386.
- 1445 119. Fillinger S, de la Garza L, Peltzer A, Kohlbacher O, Nahnsen S. Challenges of big data  
1446 integration in the life sciences. *Anal Bioanal Chem*. 2019;411(26):6791-6800.  
1447 doi:10.1007/s00216-019-02074-9.
- 1448 120. Fonteneau L, Le Meur N, Cohen-Akenine A, et al. Apport des bases médico-  
1449 administratives en épidémiologie et santé publique des maladies infectieuses [The use of  
1450 administrative health databases in infectious disease epidemiology and public health]. *Rev*  
1451 *Epidemiol Sante Publique*. 2017;65 Suppl 4:S174-S182. doi:10.1016/j.respe.2017.03.131.
- 1452 121. Silenou BC, Avalos M, Helmer C et al. Health administrative data enrichment using  
1453 cohort information: Comparative evaluation of methods by simulation and application to real  
1454 data. *PLoS One* 2019;14(1): e0211118. doi:10.1371/journal.pone.0211118.
- 1455 122. Wang SV, Schneeweiss S. Data Checks Before Registering Study Protocols for Health  
1456 Care Database Analyses. *JAMA*. 2024;331(17):1445-1446. doi:10.1001/jama.2024.2988.
- 1457 123. Yamana H, Yamasaki K, Jo T, Yatsushashi H, Yasunaga H. A descriptive analysis of  
1458 acute viral hepatitis using a database with electronic medical records and claims data. *Ann Clin*  
1459 *Epidemiol*. 2023;5(4):107-112. doi:10.37737/ace.23014.
- 1460 124. Gianfrancesco MA, Goldstein ND. A narrative review on the validity of electronic  
1461 health record-based research in epidemiology. *BMC Med Res Methodol*. 2021;21(1):234.  
1462 doi:10.1186/s12874-021-01416-5.
- 1463 125. Denisov NS, Kamenskikh EM, Fedorova OS. Trends in Population-Based Studies:  
1464 Molecular and Digital Epidemiology (Review). *Sovrem Tekhnologii Med*. 2022;14(4):60-70.  
1465 doi:10.17691/stm2022.14.4.07.
- 1466 126. Hamm NC, Hamad AF, Wall-Wieler E, Roos LL, Plana-Ripoll O, Lix LM.  
1467 Multigenerational health research using population-based linked databases: an international  
1468 review. *Int J Popul Data Sci*. 2021;6(1):1686. doi:10.23889/ijpds.v6i1.1686.



- 1469 127. Poulain R, Tarek MFB, Beheshti R. Improving Fairness in AI Models on Electronic  
1470 Health Records: The Case for Federated Learning Methods. *FACCT 23 (2023)*. 2023;2023:1599-  
1471 1608. doi:10.1145/3593013.3594102.
- 1472 128. Tuppin P, Rudant J, Constantinou P, et al. Value of a national administrative database  
1473 to guide public decisions: From the système national d'information interrégimes de l'Assurance  
1474 Maladie (SNIIRAM) to the système national des données de santé (SNDS) in France. *Rev  
1475 Epidemiol Sante Publique*. 2017;65 Suppl 4:S149-S167. doi:10.1016/j.respe.2017.05.004.
- 1476 129. Casey JA, Schwartz BS, Stewart WF, Adler NE. Using Electronic Health Records for  
1477 Population Health Research: A Review of Methods and Applications. *Annu Rev Public Health*.  
1478 2016;37:61-81. doi:10.1146/annurev-publhealth-032315-021353.
- 1479 130. Chan JM, Carroll MW, Smyth M, et al. Comparing Health Administrative and Clinical  
1480 Registry Data: Trends in Incidence and Prevalence of Pediatric Inflammatory Bowel Disease in  
1481 British Columbia. *Clin Epidemiol*. 2021;13:81-90. doi:10.2147/CLEP.S292546.
- 1482 131. Chowdhury TT, Hemmelgarn B. Evidence-based decision-making 6: Utilization of  
1483 administrative databases for health services research. *Methods Mol Biol*. 2015;1281:469-484.  
1484 doi:10.1007/978-1-4939-2428-8\_28.
- 1485 132. Chishtie J, Sapiro N, Wiebe N, et al. Use of Epic Electronic Health Record System for  
1486 Health Care Research: Scoping Review. *J Med Internet Res*. 2023;25:e51003.  
1487 doi:10.2196/51003.
- 1488 133. Jutte DP, Roos LL, Brownell MD. Administrative record linkage as a tool for public  
1489 health research. *Annu Rev Public Health*. 2011;32:91-108. doi:10.1146/annurev-publhealth-  
1490 031210-100700.
- 1491 134. Mazzali C, Duca P. Use of administrative data in healthcare research. *Intern Emerg  
1492 Med*. 2015;10(4):517-524. doi:10.1007/s11739-015-1213-9.
- 1493 135. McPheeters ML, Sathe NA, Jerome RN, Carnahan RM. Methods for systematic reviews  
1494 of administrative database studies capturing health outcomes of interest. *Vaccine*. 2013;31  
1495 Suppl 10:K2-K6. doi:10.1016/j.vaccine.2013.06.048.

- 1496 136. Sato S, Yasunaga H. A Review of Studies Using Japanese Nationwide Administrative  
1497 Claims Databases. *Ann Clin Epidemiol.* 2023;5(2):58-64. doi:10.37737/ace.23008.
- 1498 137. Mues KE, Liede A, Liu J, et al. Use of the Medicare database in epidemiologic and  
1499 health services research: a valuable source of real-world evidence on the older and disabled  
1500 populations in the US. *Clin Epidemiol.* 2017;9:267-277. doi:10.2147/CLEP.S105613.
- 1501 138. Bernstein CN. Large Registry Epidemiology in IBD. *Inflamm Bowel Dis.*  
1502 2017;23(11):1941-1949. doi:10.1097/MIB.0000000000001279.
- 1503 139. Bloom PP, Tapper EB. The Use of Administrative Data to Investigate the Population  
1504 Burden of Hepatic Encephalopathy. *J Clin Med.* 2020;9(11):3620. doi:10.3390/jcm9113620.
- 1505 140. Hudson M, Tascilar K, Suissa S. Comparative effectiveness research with  
1506 administrative health data in rheumatoid arthritis. *Nat Rev Rheumatol.* 2016;12(6):358-366.  
1507 doi:10.1038/nrrheum.2016.34.
- 1508 141. Mitchell E, Waring T, Ahern E, O'Donovan D, O'Reilly D, Bradley DT. Predictors and  
1509 consequences of homelessness in whole-population observational studies that used  
1510 administrative data: a systematic review. *BMC Public Health.* 2023;23(1):1610.  
1511 doi:10.1186/s12889-023-16503-z.
- 1512 142. Ulrich EH, So G, Zappitelli M, Chanchlani R. A Review on the Application and  
1513 Limitations of Administrative Health Care Data for the Study of Acute Kidney Injury  
1514 Epidemiology and Outcomes in Children. *Front Pediatr.* 2021;9:742888.  
1515 doi:10.3389/fped.2021.742888.
- 1516 143. Curl CL, Spivak M, Phinney R, Montrose L. Synthetic Pesticides and Health in  
1517 Vulnerable Populations: Agricultural Workers. *Curr Environ Health Rep.* 2020;7(1):13-29.  
1518 doi:10.1007/s40572-020-00266-5.
- 1519 144. Demos K, Sazakli E, Jelastopulu E, Charokopos N, Ellul J, Leotsinidis M. Does farming  
1520 have an effect on health status? A comparison study in west Greece. *Int J Environ Res Public*  
1521 *Health.* 2013;10(3):776-792. doi:10.3390/ijerph10030776.
- 1522 145. Kachuri L, Harris MA, MacLeod JS, Tjepkema M, Peters PA, Demers PA. Cancer risks  
1523 in a population-based study of 70,570 agricultural workers: results from the Canadian census

- 1524 health and Environment cohort (CanCHEC). *BMC Cancer*. 2017;17(1):343.  
1525 doi:10.1186/s12885-017-3346-x.
- 1526 146. Ramos AK, Girdžiūtė L, Starič J, Rautianinen RH. Identifying "Vulnerable Agricultural  
1527 Populations" at Risk for Occupational Injuries and Illnesses: A European Perspective. *J*  
1528 *Agromedicine*. 2021;26(3):340-345. doi:10.1080/1059924X.2020.1771498.
- 1529 147. Sturm ET, Castro C, Mendez-Colmenares, et al. Risk Factors for Brain Health in  
1530 Agricultural Work: A Systematic Review. *Int J Environ Res Public Health*. 2022;19(6):3373.  
1531 doi:10.3390/ijerph19063373.
- 1532 148. Guindo Y, Parent ME, Richard H, Luce D, Barul C. Expert-based assessment of  
1533 chemical and physical exposures, and organizational factors, in past agricultural jobs. *Environ*  
1534 *Res*. 2024. doi:10.1016/j.envres.2024.120238.
- 1535 149. Komarek AM, De Pinto A, Smith VH. A review of types of risks in agriculture: what  
1536 we know and what we need to know. *Agric Syst*. 2020;178:102738.  
1537 doi:10.1016/j.agsy.2019.102738.
- 1538 150. Leon ME, Schinasi LH, Lebailly P, et al. Pesticide use and risk of non-Hodgkin  
1539 lymphoid malignancies in agricultural cohorts from France, Norway and the USA: a pooled  
1540 analysis from the AGRICOH consortium. *Int J Epidemiol*. 2019;48(5):1519-1535.  
1541 doi:10.1093/ije/dyz017.
- 1542 151. Pérès K, Matharan F, Allard M, Amieva H, Baldi I, Barberger-Gateau P et al. Health  
1543 and aging in elderly farmers: the AMI cohort. *BMC Public Health*. 2012;12:558.  
1544 doi:10.1186/1471-2458-12-558.
- 1545 152. Petit P, Chamot S, Al-Salameh A, Cancé C, Desaillood R, Bonnetterre V. Farming  
1546 activity and risk of treated thyroid disorders: Insights from the TRACTOR project, a nationwide  
1547 cohort study. *Environ Res*. 2024;249:118458. doi:10.1016/j.envres.2024.118458.
- 1548 153. Tricco AC, Lillie E, Zarin W, et al. PRISMA Extension for Scoping Reviews (PRISMA-  
1549 ScR): Checklist and Explanation. *Ann Intern Med*. 2018;169(7):467-473. doi:10.7326/M18-  
1550 0850.

- 1551 154. Joanna Briggs Institute Reviewers' manual 2015: methodology for JBI scoping reviews.  
1552 The Joanna Briggs Institute. 2015. URL: [https://reben.com.br/revista/wp-](https://reben.com.br/revista/wp-content/uploads/2020/10/Scoping.pdf)  
1553 [content/uploads/2020/10/Scoping.pdf](https://reben.com.br/revista/wp-content/uploads/2020/10/Scoping.pdf) [accessed 2024-04-03].
- 1554 155. Beydon M, Pinto S, De Rycke Y, et al. Risk of cancer for patients with rheumatoid  
1555 arthritis versus general population: a national claims database cohort study. *Lancet Reg Health*  
1556 *Eur.* 2023;35:100768. doi:10.1016/j.lanepe.2023.100768.
- 1557 156. Moysan M. Des archives à la base de données : valeur(s) des données en épidémiologie.  
1558 *Histoire, médecine et santé.* 2022;22 :151-164. doi:10.4000/hms.6274.
- 1559 157. Syed R, Eden R, Makasi T, et al. Digital Health Data Quality Issues: Systematic Review.  
1560 *J Med Internet Res.* 2023;25:e42615. doi:10.2196/42615.
- 1561 158. Cobo MJ, Lopez-Herrera AG, Herrera-Viedma E, Herrera F. An approach for detecting,  
1562 quantifying, and visualizing the evolution of a research field: A practical application to the  
1563 Fuzzy Sets Theory field. *J Inform.* 2011;5(1):146-166. doi:10.1016/j.joi.2010.10.002.
- 1564 159. Donthu N, Kumar S, Mukherjee D, Pandey N, Lim WM. How to conduct a bibliometric  
1565 analysis: An overview and guidelines. *J Bus Res.* 2021;133:285-296.  
1566 doi:10.1016/j.jbusres.2021.04.070.
- 1567 160. Moral-Muñoz JA, Herrera-Viedma E, Santisteban-Espejo A, Cobo MJ. Software tools  
1568 for conducting bibliometric analysis in science: An up-to-date review. *Profesional De La*  
1569 *Información.* 2020;29(1). doi:10.3145/epi.2020.ene.03.
- 1570 161. Bienert JR, Oliveira RC, Andrade PB, Caramori CA. Bibliometric indexes, databases  
1571 and impact factors in cardiology. *Rev Bras Cir Cardiovasc* 2015;30(2):254–259.  
1572 doi:10.5935/1678-9741.20150019.
- 1573 162. Bihari A, Tripathi S, Deepak A. A review on h-index and its alternative indices. *J Inf*  
1574 *Sci.* 2023;49(3):624–665. doi:10.1177/0165551521101447.
- 1575 163. Shao Y, Chien TW, Jang FL. The use of radar plots with the Yk-index to identify which  
1576 authors contributed the most to the journal of Medicine in 2020 and 2021: A bibliometric  
1577 analysis. *Medicine (Baltimore).* 2022;101(45):e31033.  
1578 doi:10.1097%2FMD.00000000000031033.

- 1579 164. Kumar S, Kumar S. Trends of collaborative research in journals of oilseeds research  
1580 (India), 1993-2004. *Indian J Agric Library Inf Services*. 2008;24:80–90.
- 1581 165. Wu T, Duan Y, Zhang T, Tian W, Liu H, Deng Y. Research Trends in the Application  
1582 of Artificial Intelligence in Oncology: A Bibliometric and Network Visualization Study. *Front*  
1583 *Biosci (Landmark Ed)*. 2022;27(9):254 doi:10.31083/j.fbl2709254.
- 1584 166. Wang B, Zhao P. Worldwide research productivity in the field of back pain: A  
1585 bibliometric analysis. *Medicine*. 2018;97(40):e11566. doi:10.1097/MD.00000000000011566.
- 1586 167. Aria M, Cuccurullo C, D’Aniello L, Misuraca M, Spano M. Thematic Analysis as a  
1587 New Culturomic Tool: The Social Media Coverage on COVID-19 Pandemic in Italy.  
1588 *Sustainability*. 2022;14(6):3643. doi:10.3390/su14063643.
- 1589 168. Petit P, Leroyer A, Chamot S, Fumery M, Bonneterre V. Farming Activities and Risk  
1590 of Inflammatory Bowel Disease: A French Nationwide Population-based Cohort Study. *J*  
1591 *Crohns Colitis*. 2024;18(9):1415-1429. doi:10.1093/ecco-jcc/jjae050.
- 1592 169. Montazeri A, Mohammadi S, M Hesari P, Ghaemi M, Riazi H, Sheikhi-Mobarakeh Z.  
1593 Preliminary guideline for reporting bibliometric reviews of the biomedical literature (BIBLIO):  
1594 a minimum requirements. *Syst Rev*. 2023;12(1):239. doi:10.1186/s13643-023-02410-2.
- 1595 170. Aria M, Cuccurullo C. bibliometrix: An R-tool for comprehensive science mapping  
1596 analysis. *J Informetrics*. 2017;11(4):959-975. doi:10.1016/j.joi.2017.08.007.
- 1597 171. Spitzer WO, Hill GB, Chambers LW, Helliwell BE, Murphy HB. The occupation of  
1598 fishing as a risk factor in cancer of the lip. *N Engl J Med*. 1975;293(9):419-424.  
1599 doi:10.1056/NEJM197508282930903.
- 1600 172. Alavanja MC, Samanic C, Dosemeci M, et al. Use of agricultural pesticides and prostate  
1601 cancer risk in the Agricultural Health Study cohort. *Am J Epidemiol*. 2003;157(9):800-814.  
1602 doi:10.1093/aje/kwg040.
- 1603 173. Alavanja MC, Dosemeci M, Samanic C, et al. Pesticides and lung cancer risk in the  
1604 agricultural health study cohort. *Am J Epidemiol*. 2004;160(9):876-885.  
1605 doi:10.1093/aje/kwh290.

- 1606 174. Andersen A, Barlow L, Engeland A, Kjaerheim K, Lynge E, Pukkala E. Work-related  
1607 cancer in the Nordic countries. *Scand J Work Environ Health*. 1999;25 Suppl 2:1-116. PMID:  
1608 10507118.
- 1609 175. Andreotti G, Koutros S, Hofmann JN, et al. Glyphosate Use and Cancer Incidence in  
1610 the Agricultural Health Study. *J Natl Cancer Inst*. 2018;110(5):509-516.  
1611 doi:10.1093/jnci/djx233.
- 1612 176. Elbaz A, Clavel J, Rathouz PJ, et al. Professional exposure to pesticides and Parkinson  
1613 disease. *Ann Neurol*. 2009;66(4):494-504. doi:10.1002/ana.21717.
- 1614 177. Engel LS, Hill DA, Hoppin JA, et al. Pesticide use and breast cancer risk among farmers'  
1615 wives in the agricultural health study. *Am J Epidemiol*. 2005;161(2):121-135.  
1616 doi:10.1093/aje/kwi022.
- 1617 178. Eriksson M, Karlsson M. Occupational and other environmental factors and multiple  
1618 myeloma: a population based case-control study. *Br J Ind Med*. 1992;49(2):95-103.  
1619 doi:10.1136/oem.49.2.95.
- 1620 179. Hemminki K, Saloniemi I, Salonen T, Partanen T, Vainio H. Childhood cancer and  
1621 parental occupation in Finland. *J Epidemiol Community Health*. 1981;35(1):11-15.  
1622 doi:10.1136/jech.35.1.11.
- 1623 180. Karjalainen A, Kurppa K, Martikainen R, Klaukka T, Karjalainen J. Work is related to  
1624 a substantial portion of adult-onset asthma incidence in the Finnish population. *Am J Respir*  
1625 *Crit Care Med*. 2001;164(4):565-568. doi:10.1164/ajrccm.164.4.2012146.
- 1626 181. Karjalainen A, Martikainen R, Klaukka T, Saarinen K, Uitti J. Risk of asthma among  
1627 Finnish patients with occupational rhinitis. *Chest*. 2003;123(1):283-288.  
1628 doi:10.1378/chest.123.1.283.
- 1629 182. Kristensen P, Andersen A, Irgens LM, Laake P, Bye AS. Incidence and risk factors of  
1630 cancer among men and women in Norwegian agriculture. *Scand J Work Environ Health*.  
1631 1996;22(1):14-26. doi:10.5271/sjweh.104.

- 1632 183. Lee WJ, Sandler DP, Blair A, Samanic C, Cross AJ, Alavanja MC. Pesticide use and  
1633 colorectal cancer risk in the Agricultural Health Study. *Int J Cancer*. 2007;121(2):339-346.  
1634 doi:10.1002/ijc.22635.
- 1635 184. Levecque C, Elbaz A, Clavel J, et al. Association between Parkinson's disease and  
1636 polymorphisms in the nNOS and iNOS genes in a community-based case-control study. *Hum*  
1637 *Mol Genet*. 2003;12(1):79-86. doi:10.1093/hmg/ddg009.
- 1638 185. Mills PK, Yang R. Prostate cancer risk in California farm workers. *J Occup Environ*  
1639 *Med*. 2003;45(3):249-258. doi:10.1097/01.jom.0000058339.05741.0c.
- 1640 186. Morrison H, Savitz D, Semenciw R, et al. Farming and prostate cancer mortality. *Am J*  
1641 *Epidemiol*. 1993;137(3):270-280. doi:10.1093/oxfordjournals.aje.a116674.
- 1642 187. Orton SM, Wald L, Confavreux C, et al. Association of UV radiation with multiple  
1643 sclerosis prevalence and sex ratio in France. *Neurology*. 2011;76(5):425-431.  
1644 doi:10.1212/WNL.0b013e31820a0a9f.
- 1645 188. Pouchieu C, Piel C, Carles C, et al. Pesticide use in agriculture and Parkinson's disease  
1646 in the AGRICAN cohort study. *Int J Epidemiol*. 2018;47(1):299-310. doi:10.1093/ije/dyx225.
- 1647 189. Pukkala E, Martinsen JI, Lynge E, et al. Occupation and cancer - follow-up of 15 million  
1648 people in five Nordic countries. *Acta Oncol*. 2009;48(5):646-790.  
1649 doi:10.1080/02841860902913546.
- 1650 190. Reif J, Pearce N, Fraser J. Cancer risks in New Zealand farmers. *Int J Epidemiol*.  
1651 1989;18(4):768-774. doi:10.1093/ije/18.4.768.
- 1652 191. Stokes L, Stark A, Marshall E, Narang A. Neurotoxicity among pesticide applicators  
1653 exposed to organophosphates. *Occup Environ Med*. 1995;52(10):648-653.  
1654 doi:10.1136/oem.52.10.648.
- 1655 192. Vingård E, Alfredsson L, Goldie I, Hogstedt C. Occupation and osteoarthritis of the  
1656 hip and knee: a register-based cohort study. *Int J Epidemiol*. 1991;20(4):1025-1031.  
1657 doi:10.1093/ije/20.4.1025.

- 1658 193. Vukusic S, Van Bockstael V, Gosselin S, Confavreux C. Regional variations in the  
1659 prevalence of multiple sclerosis in French farmers. *J Neurol Neurosurg Psychiatry*.  
1660 2007;78(7):707-709. doi:10.1136/jnnp.2006.101196.
- 1661 194. Weichenthal S, Villeneuve PJ, Burnett RT, et al. Long-term exposure to fine particulate  
1662 matter: association with nonaccidental and cardiovascular mortality in the agricultural health  
1663 study cohort. *Environ Health Perspect*. 2014;122(6):609-615. doi:10.1289/ehp.1307277.
- 1664 195. Wigle DT, Semenciw RM, Wilkins K, et al. Mortality study of Canadian male farm  
1665 operators: non-Hodgkin's lymphoma mortality and agricultural practices in Saskatchewan. *J*  
1666 *Natl Cancer Inst*. 1990;82(7):575-582. doi:10.1093/jnci/82.7.575.
- 1667 196. Wiklund K, Dich J, Holm LE, Eklund G. Risk of cancer in pesticide applicators in  
1668 Swedish agriculture. *Br J Ind Med*. 1989;46(11):809-814. doi:10.1136/oem.46.11.809.
- 1669 197. Horwitz H, Ahlgren B, Nærum E. Effect of occupation on risk of developing MS: an  
1670 insurance cohort study. *BMJ Open*. 2013;3(6):e002894. doi:10.1136/bmjopen-2013-002894.
- 1671 198. Karttunen JP, Rautiainen RH, Quendler E. Gender Division of Farm Work and  
1672 Occupational Injuries. *J Agric Saf Health*. 2019;25(3):117-127. doi:10.13031/jash.13177.
- 1673 199. Keller JE, Howe HL. Risk factors for lung cancer among nonsmoking Illinois residents.  
1674 *Environ Res*. 1993;60(1):1-11. doi:10.1006/enrs.1993.1001.
- 1675 200. Marant Micallef C, Shield KD, Vignat J, et al. Cancers in France in 2015 attributable to  
1676 occupational exposures. *Int J Hyg Environ Health*. 2019;222(1):22-29.  
1677 doi:10.1016/j.ijheh.2018.07.015.
- 1678 201. Moisan F, Spinosi J, Dupupet JL, et al. The relation between type of farming and  
1679 prevalence of Parkinson's disease among agricultural workers in five French districts. *Mov*  
1680 *Disord*. 2011;26(2):271-279. doi:10.1002/mds.23370.
- 1681 202. Toch-Marquardt M, Menvielle G, Eikemo TA, et al. Occupational class inequalities in  
1682 all-cause and cause-specific mortality among middle-aged men in 14 European populations  
1683 during the early 2000s. *PLoS One*. 2014;9(9):e108072. doi:10.1371/journal.pone.0108072.
- 1684 203. Tüchsen F, Andersen O, Costa G, Filakti H, Marmot MG. Occupation and ischemic  
1685 heart disease in the European Community: a comparative study of occupations at potential high



- 1686 risk. Am J Ind Med. 1996;30(4):407-414. doi:10.1002/(SICI)1097-  
1687 0274(199610)30:4<407::AID-AJIM5>3.0.CO;2-S.
- 1688 204. Wang CC, Lin GL, Lin YJ, Chen WL, Wu WT. Occupational health surveillance and  
1689 detection of emerging occupational diseases among Taiwan farmers, through analysis of  
1690 national-based farmers' and medico-administrative databases. Am J Ind Med. 2023;66(1):85-  
1691 93. doi:10.1002/ajim.23443.
- 1692 205. Langholz PL, Sommerseth HL. Old-age mortality and social class in northern Norway  
1693 in the first half of the twentieth century, Scand Econ Hist Rev. 2023.  
1694 doi:10.1080/03585522.2023.2255599.
- 1695 206. Das AV, Bagga B, Joseph J. Clinical Profile and Demographic Distribution of  
1696 Acanthamoeba Keratitis: An Electronic Medical Record-Driven Data Analytics from an Eye  
1697 Care Network in India. Ocul Immunol Inflamm. 2023. doi:10.1080/09273948.2023.2199331.
- 1698 207. Lemarchand C, Tual S, Levêque-Morlais N, et al. Cancer incidence in the AGRICAN  
1699 cohort study (2005-2011). Cancer Epidemiol. 2017;49:175-185.  
1700 doi:10.1016/j.canep.2017.06.003.
- 1701 208. Scott E, Hirabayashi L, Levenstein A, Krupa N, Jenkins P. The development of a  
1702 machine learning algorithm to identify occupational injuries in agriculture using pre-hospital  
1703 care reports. Health Inf Sci Syst. 2021;9(1):31. doi:10.1007/s13755-021-00161-9.
- 1704 209. Wilke RA, Qamar M, Lupu RA, Gu S, Zhao J. Chronic Kidney Disease in Agricultural  
1705 Communities. Am J Med. 2019;132(10):e727-e732. doi:10.1016/j.amjmed.2019.03.036.
- 1706 210. Engel LS, Werder E, Satagopan J, et al. Insecticide Use and Breast Cancer Risk among  
1707 Farmers' Wives in the Agricultural Health Study. Environ Health Perspect.  
1708 2017;125(9):097002. doi:10.1289/EHP1295.
- 1709 211. Louis LM, Lerro CC, Friesen MC, et al. A prospective study of cancer risk among  
1710 Agricultural Health Study farm spouses associated with personal use of organochlorine  
1711 insecticides. Environ Health. 2017;16(1):95. doi:10.1186/s12940-017-0298-1.
- 1712 212. Elbaek Pedersen J, Hansen J. Risk of breast cancer in daughters of agricultural workers  
1713 in Denmark. Environ Res. 2024;240(Pt 1):117374. doi:10.1016/j.envres.2023.117374.

- 1714 213. Kasten-Arias C, Hodes T, Marino M, et al. Healthcare utilization for asthma  
1715 exacerbation among children of migrant and seasonal farmworkers. *Prev Med Rep.*  
1716 2024;38:102598. doi:10.1016/j.pmedr.2024.102598.
- 1717 214. Mengist Dessie A, Fenta Feleke S, Getaye Workie S, Getinet Abebe T, Mossu Chanie  
1718 Y, Kassa Yalew A. Prevalence of Skin Disease and Its Associated Factors Among Primary  
1719 Schoolchildren: A Cross-Sectional Study from a Northern Ethiopian Town. *Clin Cosmet*  
1720 *Investig Dermatol.* 2022;15:791-801. doi:10.2147/CCID.S361051.
- 1721 215. Mills PK, Zahm SH. Organophosphate pesticide residues in urine of farmworkers and  
1722 their children in Fresno County, California. *Am J Ind Med.* 2001;40(5):571-577.  
1723 doi:10.1002/ajim.10007.
- 1724 216. Andrew NE, Beare R, Ravipati T, et al. Developing a linked electronic health record  
1725 derived data platform to support research into healthy ageing. *Int J Popul Data Sci.*  
1726 2023;8(1):2129. doi:10.23889/ijpds.v8i1.2129.
- 1727 217. Contiero P, Borgini A, Bertoldi M, et al. An Epidemiological Study to Investigate Links  
1728 between Atmospheric Pollution from Farming and SARS-CoV-2 Mortality. *Int J Environ Res*  
1729 *Public Health.* 2022;19(8):4637. doi:10.3390/ijerph19084637.
- 1730 218. Lotterman A, Baliatsas C, de Rooij MMT, et al. Increased risk of pneumonia amongst  
1731 residents living near goat farms in different livestock-dense regions in the Netherlands. *PLoS*  
1732 *One.* 2023;18(7):e0286972. doi:10.1371/journal.pone.0286972.
- 1733 219. Nerich V, Monnet E, Weill A, et al. Fine-scale geographic variations of inflammatory  
1734 bowel disease in France: correlation with socioeconomic and house equipment variables.  
1735 *Inflamm Bowel Dis.* 2010;16(5):813-821. doi:10.1002/ibd.21122.
- 1736 220. Nilles JD, Lim D, Boyer MP, et al. The occurrence of bone and joint cancers and their  
1737 association with rural living and radon exposure in Iowa. *Environ Geochem Health.*  
1738 2023;45(3):925-940. doi:10.1007/s10653-022-01261-5.
- 1739 221. Yzermans CJ, Moleman YP, Spreeuwenberg P, et al. Risk of pneumonia in the vicinity  
1740 of goat farms: a comparative assessment of temporal variation based on longitudinal health data.  
1741 *Pneumonia (Nathan).* 2023;15(1):13. doi:10.1186/s41479-023-00115-7.

- 1742 222. Petit P, Bosson-Rieutort D, Maugard C, et al. The TRACTOR Project: TRACKing and  
1743 MoniToring Occupational Risks in Agriculture Using French Insurance Health Data (MSA).  
1744 Ann Work Expo Health. 2022;66(3):402-411. doi:10.1093/annweh/wxab083.
- 1745 223. Mills PK, Kwong S. Cancer incidence in the United Farmworkers of America (UFW),  
1746 1987-1997. Am J Ind Med. 2001;40(5):596-603. doi:10.1002/ajim.1125.
- 1747 224. Guillien A, Laurent L, Soumagne T, et al. Anxiety and depression among dairy farmers:  
1748 the impact of COPD. Int J Chron Obstruct Pulmon Dis. 2017;13:1-9.  
1749 doi:10.2147/COPD.S143883.
- 1750 225. AGRICOH: a consortium of agricultural cohort studies. AGRICOH. 2024. URL:  
1751 <https://agricoh.iarc.fr/> [accessed 2024-05-17].
- 1752 226. Jouneau S, Pinault M, Gouyet T, et al. Étude AIRBAG : résultats préliminaires après un  
1753 an de dépistage de la BPCO agricole [AIRBAG study: preliminary results after one year of  
1754 screening for COPD in dairy farmers]. Rev Mal Respir. 2015;32(3):240-248.  
1755 doi:10.1016/j.rmr.2014.03.014.
- 1756 227. Soto S, Yoder AM, Nuño T, Aceves B, Sepulveda R, Rosales CB. Health conditions  
1757 among farmworkers in the Southwest: An analysis of the National Agricultural Workers Survey.  
1758 Front Public Health. 2022;10:962085. doi:10.3389/fpubh.2022.962085.
- 1759 228. Contaxis N, Clark J, Dellureficio A, et al. Ten simple rules for improving research data  
1760 discovery. PLoS Comput Biol. 2022;18(2):e1009768. doi:10.1371/journal.pcbi.1009768.
- 1761 229. Vesteghem C, Brøndum RF, Sønderkær M, et al. Implementing the FAIR Data  
1762 Principles in precision oncology: review of supporting initiatives. Brief Bioinform.  
1763 2020;21(3):936-945. doi:10.1093/bib/bbz044.
- 1764 230. Wilkinson MD, Dumontier M, Aalbersberg IJ, et al. The FAIR Guiding Principles for  
1765 scientific data management and stewardship. Sci Data. 2016;3:160018.  
1766 doi:10.1038/sdata.2016.18.
- 1767 231. Roux P, Guillien A, Soumagne T, et al. Smoking habits in French farmers: a cross-  
1768 sectional study. BMC Public Health. 2017;17(1):166. doi:10.1186/s12889-017-4030-4.

- 1769 232. Audi C, Baiz N, Maesano CN, et al. Serum cytokine levels related to exposure to volatile  
1770 organic compounds and PM2.5 in dwellings and workplaces in French farmers - a mechanism  
1771 to explain nonsmoking COPD. *Int J Chron Obstruct Pulmon Dis.* 2017;12:1363-1374.  
1772 doi:10.2147/COPD.S117866.
- 1773 233. Baldi I, Gruber A, Rondeau V, Lebailly P, Brochard P, Fabrigoule C. Neurobehavioral  
1774 effects of long-term exposure to pesticides: results from the 4-year follow-up of the  
1775 PHYTONER study. *Occup Environ Med.* 2011;68(2):108-115. doi:10.1136/oem.2009.047811.
- 1776 234. Christensen CH, Platz EA, Andreotti G, et al. Coumaphos exposure and incident cancer  
1777 among male participants in the Agricultural Health Study (AHS). *Environ Health Perspect.*  
1778 2010;118(1):92-96. doi:10.1289/ehp.0800446.
- 1779 235. Achard P, Maugard C, Cancé C, et al. Medico-administrative data combined with  
1780 agricultural practices data to retrospectively estimate pesticide use by agricultural workers. *J*  
1781 *Expo Sci Environ Epidemiol.* 2020;30(4):743-755. doi:10.1038/s41370-019-0166-x.
- 1782 236. Hohenadel K, Harris SA, McLaughlin JR, et al. Exposure to multiple pesticides and risk  
1783 of non-Hodgkin lymphoma in men from six Canadian provinces. *Int J Environ Res Public*  
1784 *Health.* 2011;8(6):2320-2330. doi:10.3390/ijerph8062320.
- 1785 237. Chung EYM, Risi D, Holt JL, et al. Retrospective study on the epidemiology of  
1786 antineutrophil cytoplasmic autoantibodies-associated vasculitis in two Australian health  
1787 districts. *Intern Med J.* 2022;52(4):605-613. doi:10.1111/imj.15098.
- 1788 238. Li X, Sundquist J, Sundquist K. Socioeconomic and occupational risk factors for  
1789 rheumatoid arthritis: a nationwide study based on hospitalizations in Sweden. *J Rheumatol.*  
1790 2008;35(6):986-991. PMID: 18464310.
- 1791 239. Apalla Z, Lallas A, Sotiriou E, et al. Farmers develop more aggressive histologic  
1792 subtypes of basal cell carcinoma. Experience from a Tertiary Hospital in Northern Greece. *J Eur*  
1793 *Acad Dermatol Venereol.* 2016;30 Suppl 3:17-20. doi:10.1111/jdv.13605.
- 1794 240. Lope V, Pollán M, Gustavsson P, et al. Occupation and thyroid cancer risk in Sweden.  
1795 *J Occup Environ Med.* 2005;47(9):948-957. doi:10.1097/01.jom.0000169564.21523.5d.

- 1796 241. Nordby KC, Andersen A, Kristensen P. Incidence of lip cancer in the male Norwegian  
1797 agricultural population. *Cancer Causes Control*. 2004;15(6):619-626.  
1798 doi:10.1023/B:CACO.0000036169.90864.e2.
- 1799 242. Kristensen P, Andersen A, Irgens LM. Hormone-dependent cancer and adverse  
1800 reproductive outcomes in farmers' families--effects of climatic conditions favoring fungal  
1801 growth in grain. *Scand J Work Environ Health*. 2000;26(4):331-337. doi:10.5271/sjweh.550.
- 1802 243. Døssing M, Petersen KT, Vyberg M, Olsen JH. Liver cancer among employees in  
1803 Denmark. *Am J Ind Med*. 1997;32(3):248-254. doi:10.1002/(sici)1097-  
1804 0274(199709)32:3<248::aid-ajim10>3.0.co;2-v.
- 1805 244. Klingelschmidt J, Chastang JF, Khireddine-Medouni I, Chérié-Challine L, Niedhammer  
1806 I. Occupational factors associated with suicide among French employees from the special  
1807 agricultural social security scheme (MSA) working between 2007 and 2013. *Rev Epidemiol*  
1808 *Sante Publique*. 2020;68(1):1-8. doi:10.1016/j.respe.2019.09.008.
- 1809 245. Manninen P, Heliövaara M, Riihimäki H, Mäkelä P. Does psychological distress predict  
1810 disability?. *Int J Epidemiol*. 1997;26(5):1063-1070. doi:10.1093/ije/26.5.1063.
- 1811 246. Zöller B, Li X, Sundquist J, Sundquist K. Socioeconomic and occupational risk factors  
1812 for venous thromboembolism in Sweden: a nationwide epidemiological study. *Thromb Res*.  
1813 2012;129(5):577-582. doi:10.1016/j.thromres.2011.07.050.
- 1814 247. Lucyk K, Lu M, Sajobi T, Quan H. Administrative health data in Canada: lessons from  
1815 history. *BMC Med Inform Decis Mak*. 2015;15:69. doi:10.1186/s12911-015-0196-9.
- 1816 248. Kab S, Spinosi J, Chaperon L, et al. Agricultural activities and the incidence of  
1817 Parkinson's disease in the general French population. *Eur J Epidemiol*. 2017;32(3):203-216.  
1818 doi:10.1007/s10654-017-0229-z.
- 1819 249. Perrin L, Spinosi J, Chaperon L, Kab S, Moisan F, Ebaz A. Pesticides expenditures by  
1820 farming type and incidence of Parkinson disease in farmers: A French nationwide study.  
1821 *Environ Res*. 2021;197:111161. doi:10.1016/j.envres.2021.111161.

- 1822 250. Carles C, Bouvier G, Lebailly P, Baldi I. Use of job-exposure matrices to estimate  
1823 occupational exposure to pesticides: A review. *J Expo Sci Environ Epidemiol*. 2017;27(2):125-  
1824 140. doi:10.1038/jes.2016.25.
- 1825 251. Tudi M, Li H, Li H, et al. Exposure Routes and Health Risks Associated with Pesticide  
1826 Application. *Toxics*. 2022;10(6):335. doi:10.3390/toxics10060335.
- 1827 252. Ahmad MF, Ahmad FA, Alsayegh AA, et al. Pesticides impacts on human health and  
1828 the environment with their mechanisms of action and possible countermeasures. *Heliyon*.  
1829 2024;10(7):e29128. doi:10.1016/j.heliyon.2024.e29128.
- 1830 253. Cavalier H, Trasande L, Porta M. Exposures to pesticides and risk of cancer: Evaluation  
1831 of recent epidemiological evidence in humans and paths forward. *Int J Cancer*.  
1832 2023;152(5):879-912. doi:10.1002/ijc.34300.
- 1833 254. Cresto N, Forner-Piquer I, Baig A, et al. Pesticides at brain borders: Impact on the  
1834 blood-brain barrier, neuroinflammation, and neurological risk trajectories. *Chemosphere*.  
1835 2023;324:138251. doi:10.1016/j.chemosphere.2023.138251.
- 1836 255. Richardson JR, Fitsanakis V, Westerink RHS, Kanthasamy AG. Neurotoxicity of  
1837 pesticides. *Acta Neuropathol*. 2019;138(3):343-362. doi:10.1007/s00401-019-02033-9.
- 1838 256. Sun J, Fang R, Wang H, et al. A review of environmental metabolism disrupting  
1839 chemicals and effect biomarkers associating disease risks: Where exposomics meets  
1840 metabolomics. *Environ Int*. 2022;158:106941. doi:10.1016/j.envint.2021.106941.
- 1841 257. Blair A, Freeman LB. Epidemiologic studies in agricultural populations: observations  
1842 and future directions. *J Agromedicine*. 2009;14(2):125-131. doi:10.1080/10599240902779436.
- 1843 258. Carvalho JGR, Augusto HC, Ferraz R, Delerue-Matos C, Fernandes VC.  
1844 Micro(nano)plastic and Related Chemicals: Emerging Contaminants in Environment, Food and  
1845 Health Impacts. *Toxics*. 2024;12(10):762. doi:10.3390/toxics12100762.
- 1846 259. Clarke K, Manrique A, Sabo-Attwood T, Coker ES. A Narrative Review of  
1847 Occupational Air Pollution and Respiratory Health in Farmworkers. *Int J Environ Res Public*  
1848 *Health*. 2021;18(8):4097. doi:10.3390/ijerph18084097.

- 1849 260. Domingo NGG, Balasubramanian S, Thakrar SK, et al. Air quality-related health  
1850 damages of food. *Proc Natl Acad Sci U S A*. 2021;118(20):e2013637118.  
1851 doi:10.1073/pnas.2013637118.
- 1852 261. Hao Y, Sun H, Zeng X, et al. Smallholder vegetable farming produces more soil  
1853 microplastics pollution than large-scale farming. *Environ Pollut*. 2023;317:120805.  
1854 doi:10.1016/j.envpol.2022.120805.
- 1855 262. Ng EL, Huerta Lwanga E, Eldridge SM, et al. An overview of microplastic and  
1856 nanoplastic pollution in agroecosystems. *Sci Total Environ*. 2018;627:1377-1388.  
1857 doi:10.1016/j.scitotenv.2018.01.341.
- 1858 263. van Kersen W, Bossers A, de Steenhuijsen P, et al. Air pollution from  
1859 livestock farms and the oropharyngeal microbiome of COPD patients and controls. *Environ Int*.  
1860 2022;169:107497. doi:10.1016/j.envint.2022.107497.
- 1861 264. Ganter M. Zoonotic risks from small ruminants. *Vet Microbiol*. 2015;181(1-2):53-65.  
1862 doi:10.1016/j.vetmic.2015.07.015.
- 1863 265. Golomazou E, Mamedova S, Eslahi AV, Karanis P. Cryptosporidium and agriculture:  
1864 A review. *Sci Total Environ*. 2024;916:170057. doi:10.1016/j.scitotenv.2024.170057.
- 1865 266. McDaniel CJ, Cardwell DM, Moeller RB Jr, Gray GC. Humans and cattle: a review of  
1866 bovine zoonoses. *Vector Borne Zoonotic Dis*. 2014;14(1):1-19. doi:10.1089/vbz.2012.1164.
- 1867 267. Arora K, Xu L, Bhagianadh D. Dementia and Cognitive Decline in Older Adulthood:  
1868 Are Agricultural Workers at Greater Risk? *J Gerontol B Psychol Sci Soc Sci*. 2021;76(8):1629-  
1869 1643. doi:10.1093/geronb/gbab005.
- 1870 268. Douphrate DI, Rosecrance JC, Reynolds SJ, Stallones L, Gilkey DP. Tractor-related  
1871 injuries: an analysis of workers' compensation data. *J Agromedicine*. 2009;14(2):198-205.  
1872 doi:10.1080/10599240902773215.
- 1873 269. Dos Anjos Magri C, Garófallo Garcia R, Binotto E, et al. Occupational risk factors in  
1874 health of broiler-farm workers: A systematic review. *Arch Environ Occup Health*.  
1875 2021;76(8):482-493. doi:10.1080/19338244.2020.1832036.

- 1876 270. Barbosa Junior M, Sokulski CC, Salvador R, Pinheiro E, de Francisco AC, Trojan F.  
1877 What kills the agricultural worker? A systematic review on suicide. *Rural Remote Health*.  
1878 2021;21(3):6067. doi:10.22605/RRH6067.
- 1879 271. Garrett-Wright D, Malin C, Jones MS. Mental Health in Farming Communities. *J*  
1880 *Psychosoc Nurs Ment Health Serv*. 2023;61(10):39-43. doi:10.3928/02793695-20230424-02.
- 1881 272. Kolstrup CL, Kallioniemi M, Lundqvist P, Kymäläinen HR, Stallones L, Brumby S.  
1882 International perspectives on psychosocial working conditions, mental health, and stress of  
1883 dairy farm operators. *J Agromedicine*. 2013;18(3):244-255.  
1884 doi:10.1080/1059924X.2013.796903.
- 1885 273. Petit P, Gandon G, Dubuc M, Vuillerme N, Bonneterre V. Agricultural activities and  
1886 risk of treatment for depressive disorders among the entire French agricultural workforce: the  
1887 TRACTOR project, a nationwide retrospective cohort study. *Lancet Reg Health Eur*.  
1888 2023;31:100674. doi:10.1016/j.lanpe.2023.100674.
- 1889 274. Yazd SD, Wheeler SA, Zuo A. Key Risk Factors Affecting Farmers' Mental Health: A  
1890 Systematic Review. *Int J Environ Res Public Health*. 2019;16(23):4849.  
1891 doi:10.3390/ijerph16234849.
- 1892 275. Nordby KC, Andersen A, Irgens LM, Kristensen P. Indicators of mancozeb exposure in  
1893 relation to thyroid cancer and neural tube defects in farmers' families. *Scand J Work Environ*  
1894 *Health*. 2005;31(2):89-96. doi:10.5271/sjweh.855.
- 1895 276. Kauppinen T, Uuksulainen S, Saalo A, Mäkinen I, Pukkala E. Use of the Finnish  
1896 Information System on Occupational Exposure (FINJEM) in epidemiologic, surveillance, and  
1897 other applications. *Ann Occup Hyg*. 2014;58(3):380-396. doi:10.1093/annhyg/met074.
- 1898 277. Shim YK, Mlynarek SP, van Wijngaarden E. Parental exposure to pesticides and  
1899 childhood brain cancer: U.S. Atlantic coast childhood brain cancer study. *Environ Health*  
1900 *Perspect*. 2009;117(6):1002-1006. doi:10.1289/ehp.0800209.
- 1901 278. Baldi I, Carles C, Blanc-Lapierre A, et al. A French crop-exposure matrix for use in  
1902 epidemiological studies on pesticides: PESTIMAT. *J Expo Sci Environ Epidemiol*.  
1903 2017;27(1):56-63. doi:10.1038/jes.2015.72.



- 1904 279. Ultramare C, Mediouni Z, Shoman Y, Hopf NB, Graczyk H, Berthet A. Determinants  
1905 of Pesticide Exposure in Occupational Studies: A Meta-Analysis. *Toxics*. 2023;11(7):623.  
1906 doi:10.3390/toxics11070623.
- 1907 280. Burstyn I, Lavoue J, Van Tongeren M. Aggregation of exposure level and probability  
1908 into a single metric in job-exposure matrices creates bias. *Ann Occup Hyg*. 2012;56(9):1038-  
1909 1050. doi:10.1093/annhyg/mes031.
- 1910 281. Fadel M, Evanoff BA, Andersen JH, et al. Not just a research method: If used with  
1911 caution, can job-exposure matrices be a useful tool in the practice of occupational medicine and  
1912 public health? *Scand J Work Environ Health*. 2020;46(5):552-553. doi:10.5271/sjweh.3900.
- 1913 282. Peters S, Vienneau D, Sampri A, et al. Occupational Exposure Assessment Tools in  
1914 Europe: A Comprehensive Inventory Overview. *Ann Work Expo Health*. 2022;66(5):671-686.  
1915 doi:10.1093/annweh/wxab110.
- 1916 283. Petit P, Gondard E, Gandon G, Moreaud O, Sauvée M, Bonnetterre V. Agricultural  
1917 activities and risk of Alzheimer's disease: the TRACTOR project, a nationwide retrospective  
1918 cohort study. *Eur J Epidemiol*. 2024;39(3):271-287. doi:10.1007/s10654-023-01079-0.
- 1919 284. International Agency for Research on Cancer: IARC monographs on the identification  
1920 of carcinogenic hazards to humans. List of Classifications. Agents classified by the IARC  
1921 Monographs, Volumes 1-136. IARC. 2024. URL: [https://monographs.iarc.who.int/list-of-](https://monographs.iarc.who.int/list-of-classifications)  
1922 [classifications](https://monographs.iarc.who.int/list-of-classifications) [accessed 2024-10-30].
- 1923 285. Desai B, Sahni S, Jordan H, et al. Risk of Cardiovascular Disease in Male Farmers Over  
1924 the Age of 45: A Review of Literature. *Cureus*. 2022;14(5):e24642. doi:10.7759/cureus.24642.
- 1925 286. Kim H, Jung W, Jung S, et al. Is Farming a Risk Occupation for Cardio-cerebrovascular  
1926 Diseases? A Scoping Review on Cardio-cerebrovascular Disease Risk in Farmers. *J Prev Med*  
1927 *Public Health*. 2024. doi:10.3961/jpmp.24.302.
- 1928 287. Hulkkonen S, Shiri R, Auvinen J, Miettunen J, Karppinen J, Ryhänen J. Risk factors of  
1929 hospitalization for carpal tunnel syndrome among the general working population. *Scand J*  
1930 *Work Environ Health*. 2020;46(1):43-49. doi:10.5271/sjweh.3835.

- 1931 288. Kim J, Youn K, Park J. Risk Factors for Musculoskeletal Disorders in Korean Farmers:  
1932 Survey on Occupational Diseases in 2020 and 2022. *Healthcare (Basel)*. 2024;12(20):2026.  
1933 doi:10.3390/healthcare12202026.
- 1934 289. Hemminki K, Niemi ML, Saloniemi I, Vainio H, Hemminki E. Spontaneous abortions  
1935 by occupation and social class in Finland. *Int J Epidemiol*. 1980;9(2):149-153.  
1936 doi:10.1093/ije/9.2.149.
- 1937 290. Strand K, Wergeland E, Bjerkedal T. Fertility patterns according to occupational  
1938 grouping in Norway, 1989. *Scand J Soc Med*. 1996;24(1):50-54.  
1939 doi:10.1177/140349489602400108.
- 1940 291. Liu Y, Huang L, Hu F, Zhang X. Investigating Frailty, Polypharmacy, Malnutrition,  
1941 Chronic Conditions, and Quality of Life in Older Adults: Large Population-Based Study. *JMIR*  
1942 *Public Health Surveill*. 2024;10:e50617. doi:10.2196/50617.
- 1943 292. Lie A, Skogstad M, Johannessen HA, et al. Occupational noise exposure and hearing: a  
1944 systematic review. *Int Arch Occup Environ Health*. 2016;89(3):351-372. doi:10.1007/s00420-  
1945 015-1083-5.
- 1946 293. Kraemer JG, Aebi S, Oppliger A, Hilty M. The Indoor-Air Microbiota of Pig Farms  
1947 Drives the Composition of the Pig Farmers' Nasal Microbiota in a Season-Dependent and Farm-  
1948 Specific Manner. *Appl Environ Microbiol*. 2019;85(9):e03038-18. doi:10.1128/AEM.03038-  
1949 18.
- 1950 294. Mucci N, Tommasi E, Chiarelli A, et al. WORKbiota: A Systematic Review about the  
1951 Effects of Occupational Exposure on Microbiota and Workers' Health. *Int J Environ Res Public*  
1952 *Health*. 2022;19(3):1043. doi:10.3390/ijerph19031043.
- 1953 295. Shukla SK, Ye Z, Sandberg S, Reyes I, Fritsche TR, Keifer M. The nasal microbiota of  
1954 dairy farmers is more complex than oral microbiota, reflects occupational exposure, and  
1955 provides competition for staphylococci. *PLoS One*. 2017;12(8):e0183898.  
1956 doi:10.1371/journal.pone.0183898.

- 1957 296. Sun J, Liao XP, D'Souza AW, et al. Environmental remodeling of human gut microbiota  
1958 and antibiotic resistome in livestock farms. *Nat Commun.* 2020;11(1):1427.  
1959 doi:10.1038/s41467-020-15222-y.
- 1960 297. Vilela C, Araújo B, Soares-Guedes C, et al. From the Gut to the Brain: Is Microbiota a  
1961 New Paradigm in Parkinson's Disease Treatment?. *Cells.* 2024;13(9):770.  
1962 doi:10.3390/cells13090770.
- 1963 298. Gois MFB, Fernández-Pato A, Huss A, et al. Impact of occupational pesticide exposure  
1964 on the human gut microbiome. *Front Microbiol.* 2023;14:1223120.  
1965 doi:10.3389/fmicb.2023.1223120.
- 1966 299. Christen P, Schnell R. Thirty-three myths and misconceptions about population data:  
1967 from data capture and processing to linkage. *Int J Popul Data Sci.* 2023;8(1):2115.  
1968 doi:10.23889/ijpds.v8i1.2115.
- 1969 300. Dattani N, Hardelid P, Davey J, Gilbert R. Accessing electronic administrative health  
1970 data for research takes time. *Arch Dis Child.* 2013;98(5):391-392. doi:10.1136/archdischild-  
1971 2013-303730.
- 1972 301. Verheij RA, Curcin V, Delaney BC, McGilchrist MM. Possible Sources of Bias in  
1973 Primary Care Electronic Health Record Data Use and Reuse. *J Med Internet Res.*  
1974 2018;20(5):e185. doi:10.2196/jmir.9134.
- 1975 302. Bousquet PJ, Caillet P, Coeuret-Pellicier M, et al. Recherche d'algorithmes  
1976 d'identification des cancers dans les bases médico-administratives : premiers résultats des  
1977 travaux du groupe REDSIAM Tumeurs sur les cancers du sein, du côlon-rectum et du poumon  
1978 [Using cancer case identification algorithms in medico-administrative databases: Literature  
1979 review and first results from the REDSIAM Tumors group based on breast, colon, and lung  
1980 cancer]. *Rev Epidemiol Sante Publique.* 2017;65 Suppl 4:S236-S242.  
1981 doi:10.1016/j.respe.2017.04.057.
- 1982 303. Goldberg M, Carton M, Doussin A, et al. Le réseau REDSIAM (Réseau données  
1983 Sniiram) - Spécial REDSIAM [The REDSIAM network]. *Rev Epidemiol Sante Publique.*  
1984 2017;65 Suppl 4:S144-S148. doi:10.1016/j.respe.2017.06.001.

- 1985 304. Lipscombe LL, Hwee J, Webster L, Shah BR, Booth GL, Tu K. Identifying diabetes  
1986 cases from administrative data: a population-based validation study. *BMC Health Serv Res.*  
1987 2018;18(1):316. doi:10.1186/s12913-018-3148-0.
- 1988 305. Caputo M, Pecere A, Sarro A, et al. Incidence and prevalence of hyperthyroidism: a  
1989 population-based study in the Piedmont Region, Italy. *Endocrine.* 2020;69(1):107-112.  
1990 doi:10.1007/s12020-020-02222-7.
- 1991 306. Habib RR, Hojeij S, Elzein K. Gender in occupational health research of farmworkers:  
1992 a systematic review. *Am J Ind Med.* 2014;57(12):1344-1367. doi:10.1002/ajim.22375.
- 1993 307. Biswas A, Harbin S, Irvin E, et al. Sex and Gender Differences in Occupational Hazard  
1994 Exposures: a Scoping Review of the Recent Literature. *Curr Environ Health Rep.*  
1995 2021;8(4):267-280. doi:10.1007/s40572-021-00330-8.
- 1996 308. Biswas A, Tiong M, Irvin E, et al. Gender and sex differences in occupation-specific  
1997 infectious diseases: a systematic review. *Occup Environ Med.* 2024;81(8):425-432.  
1998 doi:10.1136/oemed-2024-109451.
- 1999 309. Comfort N, Re DB. Sex-Specific Neurotoxic Effects of Organophosphate Pesticides  
2000 Across the Life Course. *Curr Environ Health Rep.* 2017;4(4):392-404. doi:10.1007/s40572-017-  
2001 0171-y.
- 2002 310. Wang W, Jin J, He R, Gong H. Gender differences in pesticide use knowledge, risk  
2003 awareness and practices in Chinese farmers. *Sci Total Environ.* 2017;590-591:22-28.  
2004 doi:10.1016/j.scitotenv.2017.03.053.
- 2005 311. Xu L, Xu X, Wu X, Kuang H, Xu C. Sex-Dependent Environmental Health Risk  
2006 Analysis of Flupyradifurone. *Environ Sci Technol.* 2022;56(3):1841-1853.  
2007 doi:10.1021/acs.est.1c07726.
- 2008 312. Ketchen DJ Jr, Roccapiore AY, Connelly BL. Using Old Data: When Is It Appropriate?  
2009 *J Manag.* 2023;49(8):2541-2548. doi:10.1177/01492063231177785.
- 2010 313. Vallot C. Epigenomic tumor evolution under the spotlight: the promises of single-cell  
2011 approaches. *C R Biol.* 2022;345(1):11-16. doi:10.5802/crbio1.75.

- 2012 314. Horgan D, Hajduch M, Vrana M, et al. European Health Data Space-An Opportunity  
2013 Now to Grasp the Future of Data-Driven Healthcare. *Healthcare (Basel)*. 2022;10(9):1629.  
2014 doi:10.3390/healthcare10091629.
- 2015 315. Martani A, Geneviève LD, Wangmo T, et al. Sensing the (digital) pulse. Future steps  
2016 for improving the secondary use of data for research in Switzerland. *Digit Health*.  
2017 2023;9:20552076231169826. doi:10.1177/20552076231169826.
- 2018 316. Raab R, Küderle A, Zakreuskaya A, et al. Federated electronic health records for the  
2019 European Health Data Space. *Lancet Digit Health*. 2023;5(11):e840-e847. doi:10.1016/S2589-  
2020 7500(23)00156-5.
- 2021 317. Pampel H, Weisweiler NL, Strecker D, et al. re3data - Indexing the Global Research  
2022 Data Repository Landscape Since 2012. *Sci Data*. 2023;10(1):571. doi:10.1038/s41597-023-  
2023 02462-y.
- 2024 318. Suhr B, Dungal J, Stocker A. Search, reuse and sharing of research data in materials  
2025 science and engineering-A qualitative interview study. *PLoS One*. 2020;15(9):e0239216.  
2026 doi:10.1371/journal.pone.0239216.
- 2027 319. Noaen M, Doiron D, Syer J, Brook J. Advancing Population Health Through Open  
2028 Environmental Data Platforms. *Curr Top Behav Neurosci*. 2024. doi:10.1007/7854\_2024\_512.
- 2029 320. Dul J, van Raaij E, Caputo A. Advancing scientific inquiry through data reuse:  
2030 Necessary condition analysis with archival data. *Strategic Change*. 2024;33:35-40.  
2031 doi:10.1002/jsc.2562.
- 2032 321. Portail Epidémiologie – France | Health Databases. Catalogue des bases de données  
2033 individuelles en santé en France. INSERM. 2020. URL: <https://epidemiologie-france.aviesan.fr/>  
2034 [accessed 2024-10-30].
- 2035 322. data.gouv: Plateforme ouverte des données publiques françaises. data.gouv. 2024. URL:  
2036 <https://www.data.gouv.fr/fr/> [accessed 2024-10-30].
- 2037 323. Dataset Search. Google. 2024. URL: <https://datasetsearch.research.google.com/>  
2038 [accessed 2024-10-30].

- 2039 324. Kogevinas M, Schläunssen V, Mehlum IS, Turner MC. The OMEGA-NET International  
2040 Inventory of Occupational Cohorts. *Ann Work Expo Health*. 2020;64(6):565-568.  
2041 doi:10.1093/annweh/wxaa039.
- 2042 325. Peters S. Although a valuable method in occupational epidemiology, job-exposure -  
2043 matrices are no magic fix. *Scand J Work Environ Health*. 2020;46(3):231-234.  
2044 doi:10.5271/sjweh.3894.
- 2045 326. Nittas V, Puhan MA, von Wyl V. Toward a Working Definition of eCohort Studies in  
2046 Health Research: Narrative Literature Review. *JMIR Public Health Surveill*. 2021;7(1):e24588.  
2047 doi:10.2196/24588.
- 2048 327. Singh AV, Bansod G, Mahajan M, et al. Digital Transformation in Toxicology:  
2049 Improving Communication and Efficiency in Risk Assessment. *ACS Omega*.  
2050 2023;8(24):21377-21390. doi:10.1021/acsomega.3c00596.
- 2051 328. Soussi A, Zero E, Sacile R, Trincherro D, Fossa M. Smart Sensors and Smart Data for  
2052 Precision Agriculture: A Review. *Sensors (Basel)*. 2024;24(8):2647. doi:10.3390/s24082647.
- 2053 329. Bukhtiyarova O, Abderrazak A, Chiu Y, Sparano S, Simard M, Sirois C. Major areas of  
2054 interest of artificial intelligence research applied to health care administrative data: a scoping  
2055 review. *Front Pharmacol*. 2022;13:944516. doi:10.3389/fphar.2022.944516.
- 2056 330. Caruana A, Bandara M, Musial K, Catchpoole D, Kennedy PJ. Machine learning for  
2057 administrative health records: A systematic review of techniques and applications. *Artif Intell*  
2058 *Med*. 2023;144:102642. doi:10.1016/j.artmed.2023.102642.
- 2059 331. Leal Neto O, Von Wyl V. Digital Transformation of Public Health for  
2060 Noncommunicable Diseases: Narrative Viewpoint of Challenges and Opportunities. *JMIR*  
2061 *Public Health Surveill*. 2024;10:e49575. doi:10.2196/49575.
- 2062 332. Wang J, Kharrat FGZ, Gariépy G, et al. Predicting the Population Risk of Suicide Using  
2063 Routinely Collected Health Administrative Data in Quebec, Canada: Model-Based Synthetic  
2064 Estimation Study. *JMIR Public Health Surveill*. 2024;10:e52773. doi:10.2196/52773.

- 2065 333. Ben-Ari T, Boé J, Ciais P, Lecerf R, Van der Velde M, Makowski D. Causes and  
2066 implications of the unforeseen 2016 extreme yield loss in the breadbasket of France. *Nat*  
2067 *Commun.* 2018;9(1):1627. doi:10.1038/s41467-018-04087-x.
- 2068 334. Bhawra J, Elsahli N, Patel J. Applying Digital Technology to Understand Human  
2069 Experiences of Climate Change Impacts on Food Security and Mental Health: Scoping Review.  
2070 *JMIR Public Health Surveill.* 2024;10:e54064. doi:10.2196/54064.
- 2071 335. Comi M, Becot F, Bendixsen C. Automation, Climate Change, and the Future of Farm  
2072 Work: Cross-Disciplinary Lessons for Studying Dynamic Changes in Agricultural Health and  
2073 Safety. *Int J Environ Res Public Health.* 2023;20(6):4778. doi:10.3390/ijerph20064778.
- 2074 336. Ebi KL, Vanos J, Baldwin JW, et al. Extreme Weather and Climate Change: Population  
2075 Health and Health System Implications. *Annu Rev Public Health.* 2021;42:293-315.  
2076 doi:10.1146/annurev-publhealth-012420-105026.
- 2077 337. El Khayat M, Halwani DA, Hneiny L, Alameddine I, Haidar MA, Habib RR. Impacts  
2078 of Climate Change and Heat Stress on Farmworkers' Health: A Scoping Review. *Front Public*  
2079 *Health.* 2022;10:782811. doi:10.3389/fpubh.2022.782811.
- 2080 338. Shahzad A, Ullah S, Dar AA, et al. Nexus on climate change: agriculture and possible  
2081 solution to cope future climate change stresses. *Environ Sci Pollut Res Int.* 2021;28(12):14211-  
2082 14232. doi:10.1007/s11356-021-12649-8.
- 2083 339. Gibb K, Beckman S, Vergara XP, Heinzerling A, Harrison R. Extreme Heat and  
2084 Occupational Health Risks. *Annu Rev Public Health.* 2024;45(1):315-335.  
2085 doi:10.1146/annurev-publhealth-060222-034715.
- 2086 340. Lefèvre-Arbogast S, Chaker J, Mercier F, et al. Assessing the contribution of the  
2087 chemical exposome to neurodegenerative disease. *Nat Neurosci.* 2024;27(5):812-821.  
2088 doi:10.1038/s41593-024-01627-1.
- 2089 341. Shan L, Heusinkveld HJ, Paul KC, et al. Towards improved screening of toxins for  
2090 Parkinson's risk. *NPJ Parkinsons Dis.* 2023;9(1):169. doi:10.1038/s41531-023-00615-9.

- 2091 342. Abdelzaher H, Tawfik SM, Nour A, et al. Climate change, human health, and the  
2092 exposome: Utilizing OMIC technologies to navigate an era of uncertainty. *Front Public Health*.  
2093 2022;10:973000. doi:10.3389/fpubh.2022.973000.
- 2094 343. DeBord DG, Carreón T, Lentz TJ, Middendorf PJ, Hoover MD, Schulte PA. Use of the  
2095 "Exposome" in the Practice of Epidemiology: A Primer on -Omic Technologies. *Am J*  
2096 *Epidemiol*. 2016;184(4):302-14. doi:10.1093/aje/kwv325.
- 2097 344. Waffenschmidt S, Knelangen M, Sieben W, Bühn S, Pieper D. Single screening versus  
2098 conventional double screening for study selection in systematic reviews: a methodological  
2099 systematic review. *BMC Med Res Methodol*. 2019;19(1):132. doi:10.1186/s12874-019-0782-  
2100 0.
- 2101 345. Waffenschmidt S, Sieben W, Jakubeit T, et al. Increasing the efficiency of study  
2102 selection for systematic reviews using prioritization tools and a single-screening approach. *Syst*  
2103 *Rev*. 2023;12(1):161. doi:10.1186/s13643-023-02334-x.

Accepted paper (not yet proofread)