

Title:

Research on Artificial Intelligence and Primary Care: A Scoping Review

Authors' names:

Jacqueline K. Kueper, Amanda L. Terry, Merrick Zwarenstein, Daniel J. Lizotte

Address and position for each author:

Department of Epidemiology & Biostatistics, Western University, London, Ontario, Canada
N6A 5C1

Jacqueline K. Kueper
PhD Candidate

Departments of Epidemiology & Biostatistics, Family Medicine, Schulich Interfaculty Program
in Public Health, Western University, London, Ontario, Canada N6A 5C1

Amanda L. Terry
Associate Professor

Departments of Epidemiology & Biostatistics, Family Medicine, Western University, London,
Ontario, Canada N6A 5C1

Merrick Zwarenstein
Professor

Departments of Epidemiology & Biostatistics, Computer Science, Schulich Interfaculty Program
in Public Health, Statistical & Actuarial Sciences, Western University, London, Ontario, Canada
N6A 5C1

Daniel J. Lizotte
Assistant Professor

Correspondence to:

J. Kueper jkueper@uwo.ca

ABSTRACT

Objective: The purpose of this study was to assess the nature and extent of the body of research on artificial intelligence (AI) and primary care.

Methods: We performed a scoping review, searching 11 published and grey literature databases with subject headings and key words pertaining to the concepts of 1) AI and 2) primary care: MEDLINE, EMBASE, Cinahl, Cochrane Library, Web of Science, Scopus, IEEE Xplore, ACM Digital Library, MathSciNet, AAI, arXiv. Screening included title and abstract and then full text stages. Final inclusion criteria: 1) research study of any design, 2) developed or used AI, 3) used primary care data and/or study conducted in a primary care setting and/or explicit mention of study applicability to primary care; exclusion criteria: 1) narrative, editorial, or textbook chapter, 2) not applicable to primary care population or settings, 3) full text inaccessible in the English Language. We extracted and summarized seven key characteristics of included studies: overall study purpose(s), author appointments, primary care functions, author intended target end user(s), target health condition(s), location of data source(s) (if any), subfield(s) of AI.

Results: Of 5,515 non-duplicate documents, 405 met our eligibility criteria. The body of literature is primarily focused on creating novel AI methods or modifying existing AI methods to support physician diagnostic or treatment recommendations, for chronic conditions, using data from higher income countries. Meaningfully more studies had at least one author with a technology, engineering, or math appointment than with a primary care appointment (57 (14%) compared to 217 (54%)). Predominant AI subfields were supervised machine learning and expert systems.

Discussion: Overall, AI research associated with primary care is at an early stage of maturity with respect to widespread implementation in practice settings. For the field to progress, more interdisciplinary research teams with end-user engagement and evaluation studies are needed.

SUMMARY BOXES

Section 1: What is already known on this topic

- Advancements in technology and the availability of health data have increased opportunities for artificial intelligence to be used for primary care purposes.
- No comprehensive review of research on artificial intelligence associated with primary care has been performed.

Section 2: What this study adds

- The body of research on artificial intelligence and primary care is driven by authors without appointments in primary care departments and is focused on developing artificial intelligence methods to support diagnostic and treatment decisions.
- There is a need for more interdisciplinary research teams and evaluation of artificial intelligence projects in ‘real world’ practice settings.

INTRODUCTION

Artificial Intelligence (AI) has gained prominence through widespread, high-profile media, academic, and industry initiatives (1–6). AI research began in the 1950s, but the recognition of its potential for adoption in health care settings is widespread and ongoing (7–9). Applications of AI to the foundational sector of health care, Primary Care (PC), are particularly compelling and promoted by advances in technology and the availability of data such as through increased use of Electronic Medical Records (EMRs) (10–12).

From its inception in the 1950s, AI was primarily concerned with processes by which computers might achieve ‘intelligence’ comparable to that of humans, and how we might recognize such intelligence (13). Turing’s (1950) seminal paper, “Computing Machinery and Intelligence,” was concerned more with the latter, but the work sparked a rich diversity of research activities (13,14). The field of AI now encompasses a wide variety of methodology, much of which falls into two broad categories: rule-centred and data-centred. Rule-centred methods came from the study of logical reasoning, and are intended to capture intelligence by explicitly writing down the rules that govern it and then deploying that intelligence to carry out different tasks (13). Data-centric methods like machine learning have focused more on learning to perform specific tasks using previously collected data rather than explicitly provided rules (13). Definitions of select AI subfields with examples are provided in Table 1S (Appendix C) and example health applications are presented below.

The first rule-centred AI application for health care, MYCIN, was not in PC. MYCIN is an expert system developed in the 1970s to diagnose blood infections using over 450 rules derived from experts, textbooks, and case reports (13,15). Although met with initial excitement, over time, limitations were identified in rule-centric methods in terms of handling the complexity of real-world decisions. At the same time, increasing amounts of health data promoted a shift towards data-centric methods in machine learning, which were thought to be better able to capture complex relationships between health-relevant features from data than could be represented using rules. Machine learning methods are now used to reveal patterns in data to answer health research questions, such as to facilitate prediction of diabetes or cancer diagnoses (16–19). Machine learning methods figure prominently in natural language processing, a sub-field of AI, which can be applied to unstructured text data to extract structured information or summarize documents (20). The use of AI to input or process EMR data into formats for future use or to output meaningful summary information could potentially remove some of the EMR-associated burden from clinicians (21–23). Other sources of health data are also processed by AI methods; for example, skin cancer diagnosis using computer vision and machine learning has advanced due to 1) skin lesion imaging data being readily available with easy digitization and image recognition technology, and 2) the suitability of using supervised machine learning for predicting the probability that a lesion is pathological (24,25). These types of techniques are purported to diagnose at least as accurately as clinicians; however, the data are mainly from settings past the point of referral from PC, where the distribution and spectrum of lesions is wider, and clinicians are restricted to visual inspection—further evidence is required to assess usefulness of these tools for PC settings (24,25). Several commentaries and reports describe these and other applications of AI to health, though not predominantly for PC (1,2,5,6,17,23,26,27).

Barbara Starfield (1998) defines PC as the “level of a health service system that provides entry into the system for all new needs and problems, provides person-focused (as opposed to disease-oriented) care over time, provides care for all but very uncommon or unusual conditions, and coordinates or integrates care provided elsewhere or by others” (10). AI may increase the range of health conditions that can be handled in PC or support care providers with information retrieval and focusing on aspects of PC like relationship building and shared decision making (9,23,26). However, there is no comprehensive review of what contribution it has made so far, and thus little guidance on how best to proceed with AI research in PC.

AI’s immediate usefulness is not guaranteed: EMRs were predicted to transform PC for the better, but led to unanticipated outcomes and barriers to adoption (11,28–30). Similarly, AI could benefit PC but also do harm. Careless or ill-informed development or use of AI may exaggerate racial, class, or gender biases if models are built with biased data or used with new populations for whom performance may be poor; liability, trust, and disrupted workflow are further concerns to be addressed (9).

Optimism surrounding the use of AI to benefit patients, care providers, and society exists, yet we have a lack of knowledge about AI for PC. Therefore, we set out to locate and summarize all research literature regarding AI and PC. The objective of this study was to assess the nature and extent of the body of research on AI for PC purposes.

METHODS

We performed a scoping review according to guidelines from Arksey (2005), Levac (2010), and Tricco (2016) whereby all literature on a topic is located through a systematic search strategy, data are extracted from each relevant study, and then findings are synthesized (31–33). We followed the PRISMA-ScR Checklist (34) (Appendix A). Our protocol was registered with the Open Science Framework on April 6, 2018 before completing our database searches (osf.io/w3n2b). There was no patient or public involvement in our study.

Search Strategy. Search strategies were developed in collaboration with a medical sciences librarian. An initial search was run in PubMed using Medical Subject Headings for AI and PC to locate relevant documents, whereby subject headings and keywords from these documents were used to develop comprehensive search strategies in conjunction with topic area knowledge and discussion amongst DJL, ALT, and JKK. Final search strategies developed for health sciences, computer science, and interdisciplinary databases included keywords and, where possible, subject headings pertaining to the concepts of 1) AI and 2) PC. Due to subject-area terminology differences, certain terms were used exclusively in health sciences or in computer science databases (e.g. “knowledge base”). Search strategies were refined in an iterative fashion to balance comprehensiveness with feasibility; all relevant documents from the initial search were re-identified using a final search strategy in MEDLINE. Appendix B contains final search strategies for the 11 published or grey literature databases that were searched with retrieved references uploaded into Covidence (35) and duplicate documents removed: MEDLINE, EMBASE, Cinahl, Cochrane Library, Web of Science, Scopus, IEEE Xplore, ACM Digital

Library, MathSciNet, AAI, arXiv. Where possible, English-language limits were set; to get a sense of the amount of literature potentially missed, searches were re-run for a subset of the databases (MEDLINE-Ovid, CINAHL, Web of Science) with language limits re-set to accept all languages except English. Fewer than 10 documents were retrieved from each search.

Study Selection. *Level 1:* Titles and abstracts were independently assessed by two reviewers (JKK, DJL) according to eligibility criteria: 1) identify the document as reporting on a research study, 2) mention or allude to the use of AI, and 3) mention a PC data source, setting, or type of personnel (e.g. general practitioner, nurse). Documents were independently rated as ‘yes’ or ‘no’ against these inclusion criteria requiring two ‘yes’ votes to be included in Level 2 screening. An initial pilot test screening of the first 25 documents was performed, during which disagreements were discussed to ensure a mutual understanding of the screening criteria and to ensure capturing of relevant literature. A second, similar meeting was held after screening an additional 100 records. Disagreements from the remainder of Level 1 screening were resolved by a third reviewer (ALT).

During this process we found a substantial amount of literature on computerized cognitive behavioural therapy. For example, computer-based programs that patients can access outside of a clinic to be guided through therapy sessions (36–40). Because the methods driving these tools were often unclear and comprehensive reviews on these systems have been performed previously, these studies were excluded (n=37).

Level 2: Two reviewers (JKK, DJL) independently reviewed the full text of all documents that made it through Level 1 screening according to the following eligibility criteria. Inclusion criteria: 1) research study of any design, 2) developed or used AI, 3) used PC data and/or study conducted in a PC setting and/or explicit mention of study applicability to PC; exclusion criteria: 1) narrative, editorial, or textbook chapter, 2) not applicable to PC population or settings, 3) full text inaccessible in the English Language. Similar to Level 1 screening, pilot testing and refinement of eligibility criteria was performed. All disagreements were resolved by discussion until consensus (JKK, DJL).

A notable challenge that arose in assessing eligibility criteria was authors’ use of terminology that overlaps with AI when the methods used are not considered AI. For example, “data mining” or “natural language processing” was used to describe methodologies like simple database queries or string matching that are not AI. We also excluded 34 studies because there was not enough information to determine whether AI was involved, even after consulting any methods references.

Data Extraction and Synthesis. An initial data extraction form was discussed amongst all reviewers (JKK, ALT, MZ, DJL) after pilot testing extraction of three randomly selected articles (41–43). A revised data extraction sheet was pilot tested by three reviewers (JKK, ALT, DJL) with five new randomly selected articles (25,44–47). Extractions were compared and discussed to ensure all relevant information could be captured in a consistent manner. Remaining documents were split alphabetically amongst three reviewers to be extracted independently (100 by ALT, 50 by DJL, 250 by JKK). The final version of the data extraction sheet collected information about publication details, study purpose(s), author appointment(s), PC function(s),

author intended target end user(s), target health condition(s), location of data source(s) (if any), utilized subfield(s) of AI, the reviewer who extracted the data, and any reviewer notes. We agreed upon definitions for each data extraction field (Table 1S, Appendix C). For fields except author appointments and additional notes, we predefined categories based on the pilot testing and on content knowledge—studies were recorded as belonging to as many categories as were relevant. An “Other” field was used to capture specifics of studies that did not fit into a predefined category and an “Unknown” category was used if not enough information was provided for category selection. Results were summarized as categorical variables for the seven data extraction fields; meaningful cross tabulations and additional summary counts across categories were performed.

RESULTS

Searches. Not including duplicates our search retrieved 5,515 documents for title and abstract screening. Of those, 727 met the eligibility criteria for full text screening and 405 met the final inclusion criteria (Figure 1; a reference list for included studies is in Appendix D). The AI and PC study with the earliest date of publication, 1986, developed a supervised machine learning method to assist clinicians with abdominal pain diagnoses (48). Characteristics of all studies are summarized below according to the seven data extraction categories, which are defined in Table 1S (Appendix C).

Study Purpose. The majority of studies (n=270, 66.7%) focused primarily on developing new or adapting existing AI methods using secondary data. The second most common study purpose (n=86, 21.2%) was analysing data using AI techniques, such as eliciting patterns from health data to facilitate research (e.g. to improve scheduling or cost management). Few studies evaluated AI in a ‘real world’ setting (n=28, 6.9%).

There were series of studies that reported on multiple stages of a project, from AI development to pilot testing; these studies included intended end users located in a PC setting (49–56). A minority of studies (n=21, 5.2%) had more than one study purpose. Figure 2 presents all combinations of these studies.

Author Appointments. We categorized author appointments into four categories: 1) Technology, Engineering, and Math (TEM) discipline, meaning an author appointed in a Department of Mathematics, Engineering, Computer Science, Informatics, and/or Statistics; 2) PC discipline, meaning an author appointed in a Department of Family Medicine, Primary Care, Community Health, and/or other analogous term; 3) Nursing discipline, and 4) Other. Authors were predominantly from TEM disciplines with 214 (52.8%) studies having at least one author with an appointment in a TEM discipline compared to 57 (14.1%) studies having at least one author with an appointment in a PC discipline. Twenty-three studies (5.7%) had a PC appointed author listed first and 27 (6.7%) had a PC appointed author listed last (senior author). These patterns did not change when unspecified or general medical appointments (i.e. any non-specialist medical appointment) were counted as PC appointments. Four studies had authors with nursing appointments. Cross tabulations between study purpose category and author appointment categories did not suggest that the types of author teams differed by study purpose. Table 1 presents a summary of the body of literature broken into PC and TEM author disciplines; Table

2S (Appendix C) contains a detailed breakdown of author appointments according to 16 categories.

Table 1: Author Appointments

Author appointment category	Number of studies n (%)
PC and TEM	27 (6.7)
PC and no TEM	30 (7.4)
TEM and no PC	187 (46.2)
Neither TEM nor PC	161 (39.8)

Note: A positive count for each row requires a study to report at least one author with an appointment in the category/categories of interest.

Primary Care Functions. Diagnostic decision support was by far the most common PC function present (n=148, 36.5%), followed by support for treatment decisions (n=56, 13.8%). The third most common function AI was used for was extracting information from data sources such as EMRs (n=49, 12.1%). Of studies that used AI for multiple functions, the most frequent combination was information extraction and description (n=21, 5.2%). Figure 3 summarizes the extent to which PC functions were targeted; Figure 1S (Appendix C) presents a more detailed breakdown.

Author Reported Target End User. The majority of studies reported physicians as a target end user of the research, either alone or in combination with other possible target end users (n=243, 60%). There appears to be no positive association between having physicians as a target end user and having at least one author with a medical appointment: the percentage of studies with at least one author with an appointment in a medical discipline of any kind was similar between studies with physician and exclusively non-physician target end-users (51.9% and 46.3%, respectively). Twenty-six (6.4 %) studies stated their research was intended for patients, 25 (6.2%) for administrative use, and 9 (2.2%) for nurses or nurse practitioners, either alone or in combination with other end users. Figure 4 shows the number of studies that included each of the target end user categories; Figure 2S (Appendix C) presents all combinations of author reported end users on a per-study basis.

Health Conditions. A large proportion of studies (n=108, 26.7%) focused on developing, using, or analyzing AI such that it would be relevant for any or for most health conditions of patients in PC settings. Of studies that targeted a particular condition, chronic physical conditions were more frequent than acute or psychiatric conditions. Target health conditions were condensed into 10 categories and are presented in Figure 5; Table 3S (Appendix C) expands the health conditions into 27 categories.

Geographical Location. Predominantly, the location of data source(s) used in a study or the intended location of the AI implementation were OECD countries. Low- and middle-income countries were poorly represented. The majority of studies used data from a single country, with the United States alone appearing to be the most common source of data for research on AI (n=79, 19.5%). Figure 6 contains a summary of location counts and per capita rates; Table 3S (Appendix C) contains a more detailed breakdown of the counts.

AI Subfield. The majority of studies (n=363, 89.6%) employed methods within a single subfield of AI and of these, supervised machine learning was the most common (n=162, 40.0%), followed by expert systems (n=90, 22.2%), and then natural language processing (n=35, 8.6%). There were no robotics papers. Expert systems had the earliest median year of publication (2007); data mining had the most recent (2015). Figure 7 presents frequencies and median year of publication for 10 subfields of AI used by studies captured in our literature review; all AI subfield combinations are presented in Figure 4S (Appendix C).

DISCUSSION

Key Findings. We located and summarized the characteristics of 405 research studies that included AI and PC. Predominant trends included 1) who: the vast majority of studies did not have any PC involvement; 2) methods: an initial focus on expert system methods shifted overtime to supervised machine learning; and 3) applications: studies most often developed AI to support diagnostic or treatment decisions, most often for chronic conditions, predominantly in higher income countries. Overall, AI and PC research is at an early stage of maturity.

The dominance of TEM-appointed authors and AI methods development research is congruent with the stage of this field. An AI-driven technology needs to be working reliably with good performance, which is accomplished through methods development research, before ‘real world’ testing, let alone widespread implementation. This is further reflected by the majority of studies having researchers as an intended end user alongside care providers, whereby further refinement seemed to be required before the AI application would be ready for use in a practice setting. For example, methods that are intended to work with many health conditions but have only been tested with a single candidate condition, or methodologically well-performing AI that needs to be incorporated into software and pilot-tested in practice settings to evaluate its impact with end-users. On the other hand, studies which focused on using AI to analyse health data may be considered distinct and are at a later stage of maturity in terms of readiness for use. These AI applications will not be used in everyday clinical practice, so while methodological performance is important, longer-term health or workflow outcomes may not need to be assessed before ‘real world’ use.

The dominant subfields of AI identified by our review mirror broader trends in AI research and make sense in terms of other characteristics of the included studies. Expert systems comprise a significant portion of the literature but are now less common (median year of publication 2007 versus 2014 for supervised machine learning), reflecting a general shift in AI research from expert systems and rule-centric AI methods to machine learning and data-centric AI methods (57). These methods are amendable to providing diagnostic and treatment recommendations as well as making predictions about future health, which supports PC activities such as primary prevention and screening. This also aligns with the focus on physicians as target eventual end users.

An underlying driver of AI research, and by extension maturation, is data availability, particularly after the shift toward data-driven machine learning methods. The United States is the single dominant country in the field, which is not surprising given its population, wealth, and

research resources and output (58–61). The high standing of the United Kingdom and Netherlands despite smaller populations may be attributable to PC data availability (62,63), facilitated by high adoption rates of EMRs (64), and strong information technology academics and industries (65,66).

Strengths and limitations. Strengths of our review include a comprehensive search strategy in peer reviewed and grey literature databases, without date restriction, with inclusive eligibility criterion, and conducted by an interdisciplinary research team. Limitations include using multiple reviewers to extract data without double coding and English language restriction. Proprietary research would not be captured by our review.

Suggestions for future research. Our next steps include further assessing the relevance and quality of the included studies and summarizing cases of exemplary research projects. We additionally recommend a review on AI for the broader primary health care system that includes care providers beyond physicians and nurses (e.g. social workers, physiotherapists).

In terms of future research studies, there is a need for more interdisciplinary research teams with PC end user engagement. Value must be placed both on methods development contributions and on the potential impact of developed AI technologies for intended end users. Knowledge and skills from TEM disciplines will contribute towards rigorous methodology and AI performance; knowledge and skills from PC disciplines will contribute towards identifying the best PC challenges to target and assessing success in terms of care delivery and longer-term health outcomes. Inclusion of nurses, patients, and administrative staff needs to increase—identifying relevant non-physician end user activities that could be augmented by AI is an outstanding research question on its own.

For future AI methods development, we expect to see a shift towards a middle-ground between rule-centric and data-centric methods because interpretable models may better support decisions and trust in the healthcare setting. For example, explainable AI is a paradigm whereby one can understand what a model is doing or why it arrives at a particular output (67–69). Interpretability of models is additionally important from an equity lens to be able to identify and then avoid AI reproducing biases in data, which is a present concern with data-driven methods (70). It is also important to remember that AI is not always a superior solution; a recent review found no benefit overall of machine learning compared to logistic regression for clinical prediction rules (71).

Finally, availability of data determines the clinical problem areas and PC functions to which AI can be applied, and thus also a limit on where it can help. Investments in data generation, quality, and access will increase future possibilities for AI to be used to strengthen PC in the corresponding region. This includes adoption and use of EMRs (64,72).

Conclusions. To our knowledge this is the first comprehensive, interdisciplinary summary of research on AI and PC. Two fundamental aims in the body of research on AI and PC emerged: 1) support for care provider decisions and 2) extracting meaningful information from PC data. Overall, AI for PC is an innovation in early stages of maturity, with few tools ready for widespread implementation. Interdisciplinary research teams including front line clinicians and

evaluation studies in PC settings will be crucial for advancement and success of AI for PC purposes.

ACKNOWLEDGEMENTS

This research was supported by funding from the Canadian Institutes of Health Research and from a TUTOR-PHC Fellowship funded by INSPIRE-PHC.

REFERENCES

1. Jiang F, Jiang Y, Zhi H, et al. Artificial intelligence in healthcare: past, present and future. *Stroke Vasc Neurol*. 2017;2(4):230–43.
2. Beam AL, Kohane IS. Big data and machine learning in health care. *JAMA*. 2018;319(13):1317.
3. Gray R. The A-Z of how artificial intelligence is changing the world. 2018 Nov [cited 2019 Mar 18]; Available from: <http://www.bbc.com/future/gallery/20181115-a-guide-to-how-artificial-intelligence-is-changing-the-world>
4. Davenport TH, Ronanki R. Artificial intelligence for the real world [Internet]. *Harvard Business Review*; 2018 Jan. Available from: <https://hbr.org/2018/01/artificial-intelligence-for-the-real-world>
5. Hitching R. Primary Care 2.0 [Internet]. M. 2019 [cited 2019 Jul 8]. Available from: <https://becominghuman.ai/the-changing-face-of-healthcare-artificial-intelligence-primary-care-an-empowered-patients-8ba638ffacd6>
6. Pratt M. Artificial intelligence in primary care [Internet]. *Medical Economics*. 2018 [cited 2019 Jul 8]. Available from: <https://www.medicaleconomics.com/business/artificial-intelligence-primary-care>
7. Gupta A, Thorpe C, Bhattacharyya O, Zwarenstein M. Promoting development and uptake of health innovations: The Nose to Tail Tool. *F1000Res*. 2016;5:361.
8. Voracek D. NASA innovation framework and center innovation fund. *Edwards Technical Symposium*; 2018; NASA Armstrong Flight Research Center.
9. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. *NEJM*. 2019;380(14):1347–58.
10. Starfield B. In: *Primary care Balancing health needs, services, and technology*. New York, NY: Oxford University Press, Inc.; 1998. p. 8–9.
11. Huang M, Gibson C, Terry A. Measuring electronic health record use in primary care: a scoping review. *Appl Clin Inform*. 2018;09(01):15–33.
12. Terry AL, Stewart M. How does Canada stack up? A bibliometric analysis of the primary healthcare electronic medical record literature. *Inform Prim Care*. 2012;20:233–40.
13. Russell S, Norvig, Peter. Chapter 1: Introduction. In: *Artificial Intelligence A Modern Approach*. 3rd ed. Pearson Education Inc.; 2010. p. 1–33.
14. Turing A. Computing machinery and intelligence. *Mind*. 1950;49:433–60.

15. Shortliffe EH, Axline SG, Buchanan BG, Green CC, Cohen SN. Computer-based consultations in clinical therapeutics: explanation and rule acquisition capabilities of the MYCIN System. *Comput Biomed Res.* 1975;8:303–20.
16. Alanazi HO, Abdullah AH, Qureshi KN. A critical review for developing accurate and dynamic predictive models using machine learning methods in medicine and health care. *J Med Syst.* 2017;41(4).
17. Yu K-H, Beam AL, Kohane IS. Artificial intelligence in healthcare. *Nat Biomed Eng.* 2018;2(10):719–31.
18. Miotto R, Wang F, Wang S, Jiang X, Dudley JT. Deep learning for healthcare: review, opportunities and challenges. *Brief Bioinform.* 2018;19(6):1236–46.
19. Miotto R, Li L, Kidd BA, Dudley JT. Deep patient: an unsupervised representation to predict the future of patients from the electronic health records. *Scientific Reports.* 2016;6.
20. Friedman C, Elhadad N. Natural language processing in health care and biomedicine. In: Shortliffe EH, Cimino JJ, editors. *Biomedical Informatics*. London: Springer London; 2014. p. 255–84.
21. Collier R. Electronic health records contributing to physician burnout. *CMAJ.* 2017;189(45):E1405–6.
22. Arndt BG, Beasley JW, Watkinson MD, et al. Tethered to the EHR: primary care physician workload assessment using EHR event log data and time-motion observations. *Ann Fam Med.* 2017;15(5):419–26.
23. Topol E. The Topol Review. Preparing the healthcare workforce to deliver the digital future. NHS; 2019.
24. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature.* 2017;542(7639):115–8.
25. Stanganelli I, Brucale A, Calori L, et al. Computer-aided diagnosis of melanocytic lesions. *Anticancer Res.* 2005;25(6C):4577–82.
26. Moore SF, Hamilton W, Llewellyn DJ. Harnessing the power of intelligent machines to enhance primary care. *Br J Gen Pract.* 2018;68(666):6–7.
27. U.S. Agency for International Development. Artificial intelligence in global health: defining a collective path moving forward [Internet]. USAID; The Rockefeller Foundation; Bill & Melinda Gates Foundation; 2019 [cited 2019 Jun 17]. Available from: <https://www.usaid.gov/cii/ai-in-global-health>
28. Greenhalgh T, Potts HWW, Wong G, Bark P, Swinglehurst D. Tensions and paradoxes in electronic patient record research: a systematic literature review using the meta-narrative method: electronic patient record research. *Milbank Q.* 2009;87(4):729–88.

29. Greenhalgh T, Hinder S, Stramer K, Bratan T, Russell J. Adoption, non-adoption, and abandonment of a personal electronic health record: case study of HealthSpace. *BMJ*. 2010;341:c5814–c5814.
30. Terry AL, Ryan BL, McKay S, et al. Towards optimal electronic medical record use: perspectives of advanced users. *Fam Pract*. 2018;35(5):607–11.
31. Tricco AC, Lillie E, Zarin W, et al. A scoping review on the conduct and reporting of scoping reviews. *BMC Med Res Methodol*. 2016;16(1).
32. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. *Int J Soc Res Methodol*. 2005;8(1):19–32.
33. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. *Implement Sci*. 2010;5(69).
34. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. *Ann Intern Med*. 2018;169(7):467.
35. Covidence [Internet]. Available from: <https://www.covidence.org/home>
36. Craske MG, Rose RD, Lang A, et al. Computer-assisted delivery of cognitive behavioral therapy for anxiety disorders in primary-care settings. *Depress Anxiety*. 2009;26(3):235–42.
37. Cucciare MA, Curran GM, Craske MG, et al. Assessing fidelity of cognitive behavioral therapy in rural VA clinics: design of a randomized implementation effectiveness (hybrid type III) trial. *Implementation Science*. 2015;11(1).
38. de Graaf LE, Gerhards SAH, Arntz A, et al. Clinical effectiveness of online computerised cognitive-behavioural therapy without support for depression in primary care: randomised trial. *Br J Psychiatry*. 2009;195(1):73–80.
39. de Graaf LE, Hollon SD, Huibers MJH. Predicting outcome in computerized cognitive behavioral therapy for depression in primary care: A randomized trial. *J Consult Clin Psychol*. 2010;78(2):184–9.
40. de Graaf LE, Gerhards SAH, Arntz A, et al. One-year follow-up results of unsupported online computerized cognitive behavioural therapy for depression in primary care: A randomized trial. *J Behav Ther Exp Psychiatry*. 2011;42(1):89–95.
41. Astilean A, Avram C, Folea S, Silvasan I, Petreus D. Fuzzy Petri nets based decision support system for ambulatory treatment of non-severe acute diseases. In: 2010 IEEE International Conference on Automation, Quality and Testing, Robotics (AQTR). Cluj-Napoca, Romania: IEEE; 2010. p. 1–6. Available from: <http://ieeexplore.ieee.org/document/5520706/>

42. Dominguez-Morales JP, Jimenez-Fernandez AF, Dominguez-Morales MJ, Jimenez-Moreno G. Deep neural networks for the recognition and classification of heart murmurs using neuromorphic auditory sensors. *IEEE Trans Biomed Circuits Sys.* 2018;12(1):24–34.
43. Zamora M, Baradad M, Amado E, et al. Characterizing chronic disease and polymedication prescription patterns from electronic health records. In: 2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA). Campus des Cordeliers, Paris, France: IEEE; 2015. p. 1–9. Available from: <http://ieeexplore.ieee.org/document/7344870/>
44. Widmer G, Horn W, Nagele B. Automatic knowledge base refinement: Learning from examples and deep knowledge in rheumatology. *Artif Intell Med.* 1993;5(3):225–43.
45. Michalowski M, Michalowski W, Wilk S, O’Sullivan D, Marc Carrier M. AFGuide system to support personalized management of atrial fibrillation. In: Joint Workshop on Health Intelligence. 2017. Available from: <https://aaai.org/ocs/index.php/WS/AAAIW17/paper/view/15215>
46. Glasspool DW, Fox J, Coulson AS, Emery J. Risk assessment in genetics: a semi-quantitative approach. *Stud Health Technol Inform.* 2001;84:459–63.
47. Flores CD, Fonseca JM, Bez MR, Respício A, Coelho H. Method for building a medical training simulator with Bayesian networks: SimDeCS. In: 2nd KES International Conference on Innovation in Medicine and Healthcare. San Sebastian, Spain: Studies in Health Technology and Informatics; 2014. p. 102–14. Available from: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-84918803850&doi=10.3233%2f978-1-61499-474-9-102&partnerID=40&md5=255c09d7e35b1245eb4691cef5613903>
48. Orient JM. Evaluation of abdominal pain: clinicians’ performance compared with three protocols. *South Med J.* 1986;79(7):793–9.
49. Sumner 2nd. W, Truszczynski M, Marek VW. Simulating patients with parallel health state networks. In: AMIA Annual Symposium Proceedings. 1998;438–42.
50. Sumner 2nd. W, Xu JZ, Roussel G, Hagen MD. Modeling relief. In: AMIA Annual Symposium Proceedings. 2007;706–10.
51. Sumner 2nd. W, Xu JZ. Modeling fatigue. In: AMIA Annual Symposium Proceedings. 2002;747–51.
52. Sumner 2nd. W, Hagen MD, Rovinelli R. The item generation methodology of an empiric simulation project. *Adv Health Sci Educ.* 1999;4:49–66.
53. Zhuang ZY, Churilov L, Sikaris K. Uncovering the patterns in pathology ordering by Australian general practitioners: a data mining perspective. In: Proceedings of the 39th Annual Hawaii International Conference on System Sciences (HICSS’06). Kauia, HI, USA: IEEE; 2006. p. 92c–92c. Available from: <http://ieeexplore.ieee.org/document/1579476/>

54. Zhuang ZY, Amarasiri R, Churilov L, Alahakoon D, Sikaris K. Exploring the clinical notes of pathology ordering by Australian general practitioners: a text mining perspective. In: 2007 40th Annual Hawaii International Conference on System Sciences (HICSS'07). Waikoloa, HI: IEEE; 2007. p. 136–136. Available from: <https://ieeexplore.ieee.org/document/4076642/>
55. Zhuang ZY, Churilov L, Burstein F, Sikaris K. Combining data mining and case-based reasoning for intelligent decision support for pathology ordering by general practitioners. *Eur J of Oper Res.* 2009;195(3):662–75.
56. Zhuang ZY, Wilkin CL, Ceglowski A. A framework for an intelligent decision support system: A case in pathology test ordering. *Decis Support Sys.* 2013;55(2):476–87.
57. Hao K. We analyzed 16,625 papers to figure out where AI is headed next. *MIT Technology Review* [Internet]. 2019 Jan 25 [cited 2019 Mar 15]; Available from: <https://www.technologyreview.com/s/612768/we-analyzed-16625-papers-to-figure-out-where-ai-is-headed-next/>
58. Scimago journal and country rank [Internet]. SJR. [cited 2019 May 31]. Available from: <https://www.scimagojr.com/countryrank.php>
59. Conte ML, Liu J, Schnell S, Omary MB. Globalization and changing trends of biomedical research output. *JCI Insight.* 2017;2(12):e95206.
60. Hajjar F, Saint-Lary O, Cadwallader J-S, et al. Development of primary care research in North America, Europe, and Australia from 1974 to 2017. *Ann Fam Med.* 2019;17(1):49–51.
61. 15 countries with highest technology in the world. 41Studio [Internet]. 2019 [cited 2019 May 31]; Available from: <https://www.41studio.com/blog/2018/15-countries-with-highest-technology-in-the-world/>
62. Sturgiss E, van Boven K. Datasets collected in general practice: an international comparison using the example of obesity. *Aust Health Rev.* 2018;42(5):563.
63. Smeets HM, Kortekaas MF, Rutten FH, et al. Routine primary care data for scientific research, quality of care programs and educational purposes: the Julius General Practitioners' Network (JGPN). *BMC Health Serv Res.* 2018;18(1).
64. WHO Global Observatory for eHealth, World Health Organization, WHO Global Observatory for eHealth, editors. *Atlas of eHealth country profiles: the use of eHealth in support of universal health coverage: based on the findings of the third global survey on eHealth, 2015.* Geneva, Switzerland: World Health Organization; 2016. 386 p.
65. Boyrikova A. Why the Netherlands is the new Silicon Valley [Internet]. *Your Sneak Preview of the Future.* 2017 [cited 2019 Jun 14]. Available from: <https://innovationorigins.com/why-netherlands-is-the-new-silicon-valley/>

66. Europe's hub for R & D innovation [Internet]. Invest in Holland. 2019 [cited 2019 Jun 14]. Available from: <https://investinholland.com/business-operations/research-and-development/>
67. Doshi-Velez F, Kim B. Towards A Rigorous Science of Interpretable Machine Learning. arXiv:170208608 [cs, stat]. 2017 Feb 27; Available from: <http://arxiv.org/abs/1702.08608>
68. Holzinger A, Biemann C, Pattichis CS, Kell DB. What do we need to build explainable AI systems for the medical domain? arXiv:171209923 [cs, stat]. 2017 Dec 28; Available from: <http://arxiv.org/abs/1712.09923>
69. Miller T. Explanation in Artificial Intelligence: Insights from the Social Sciences. arXiv:170607269 [cs]. 2017 Jun 22; Available from: <http://arxiv.org/abs/1706.07269>
70. Verghese A, Shah NH, Harrington RA. What this computer needs is a physician: humanism and artificial intelligence. *JAMA*. 2018;319(1):19.
71. Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *J Clin Epidemiol*. 2019;110:12–22.
72. Electronic health records and support for primary care teamwork [Internet]. The Commonwealth Fund; 2015 Feb [cited 2019 May 31]. Available from: <https://www.commonwealthfund.org/publications/journal-article/2015/feb/electronic-health-records-and-support-primary-care-teamwork>

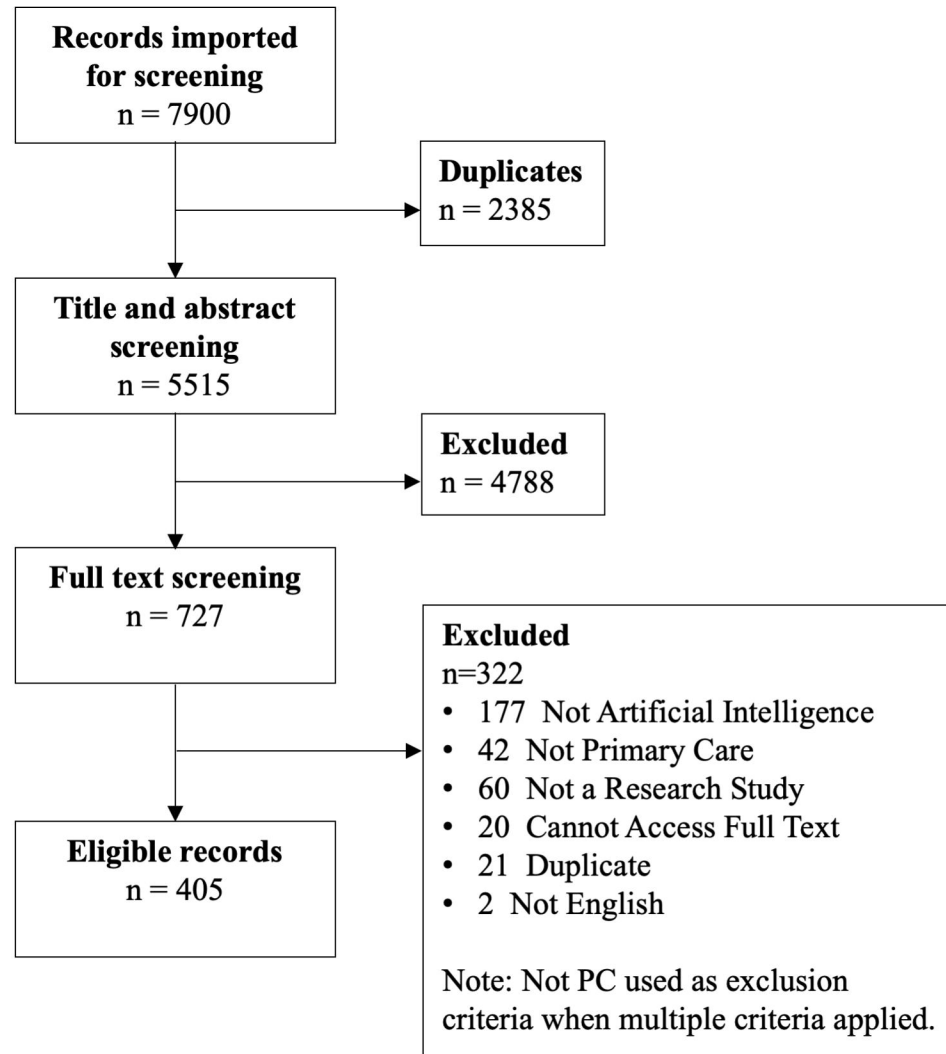


Figure 1: PRISMA flow diagram

Figure 2 Overall Study Purpose

Study Purpose

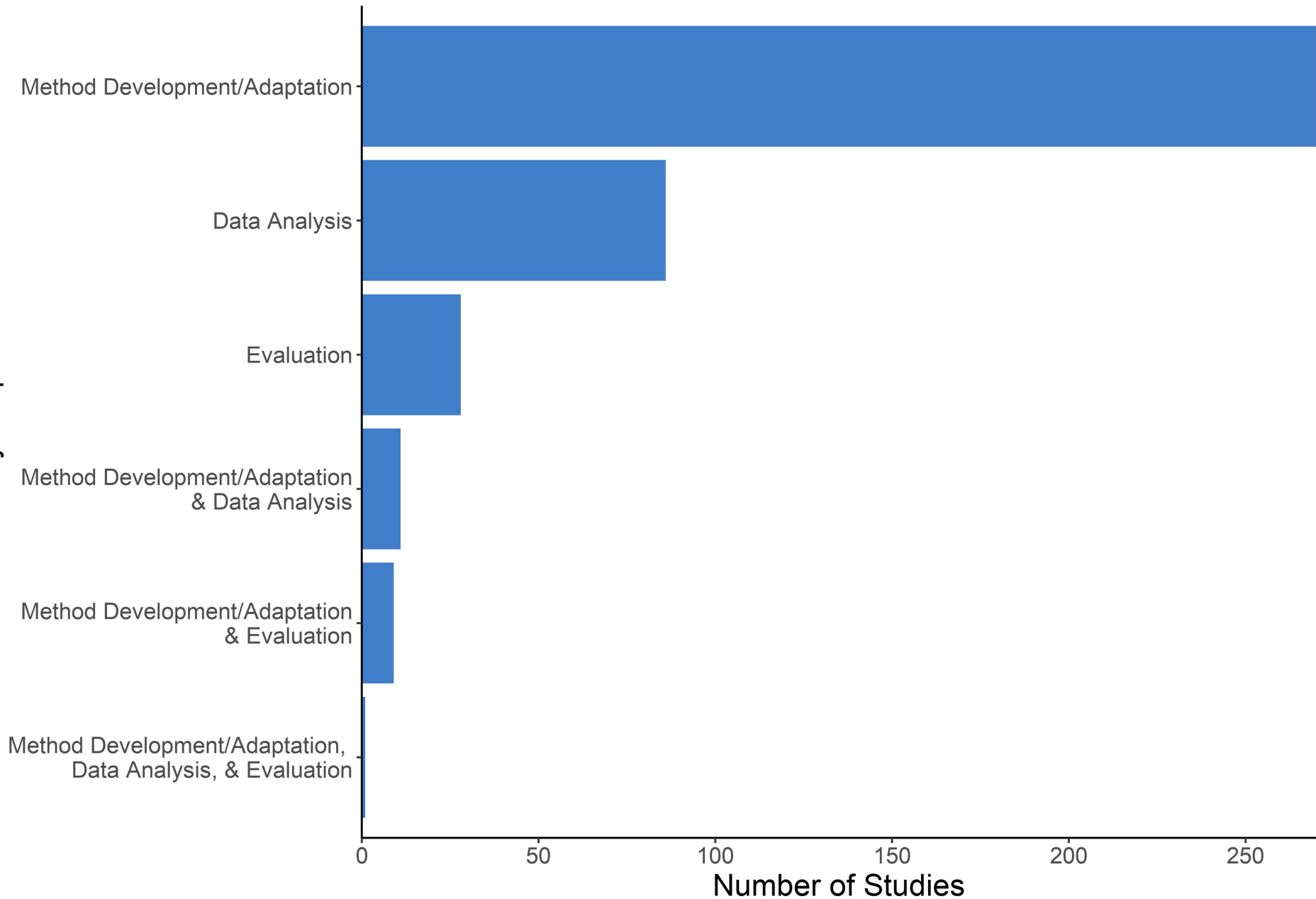


Figure 3 Primary Care Functions to be Supported with Artificial Intelligence

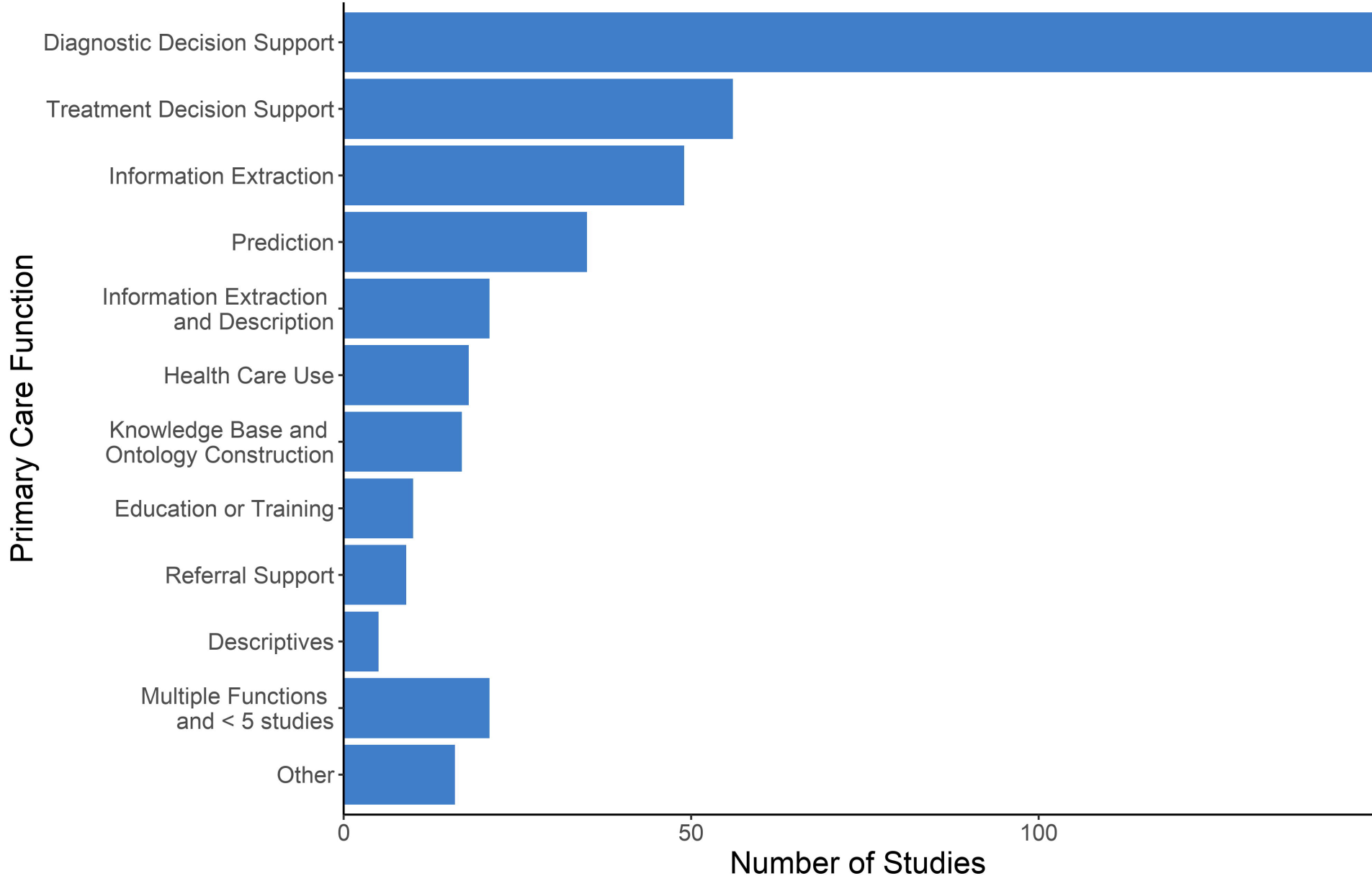
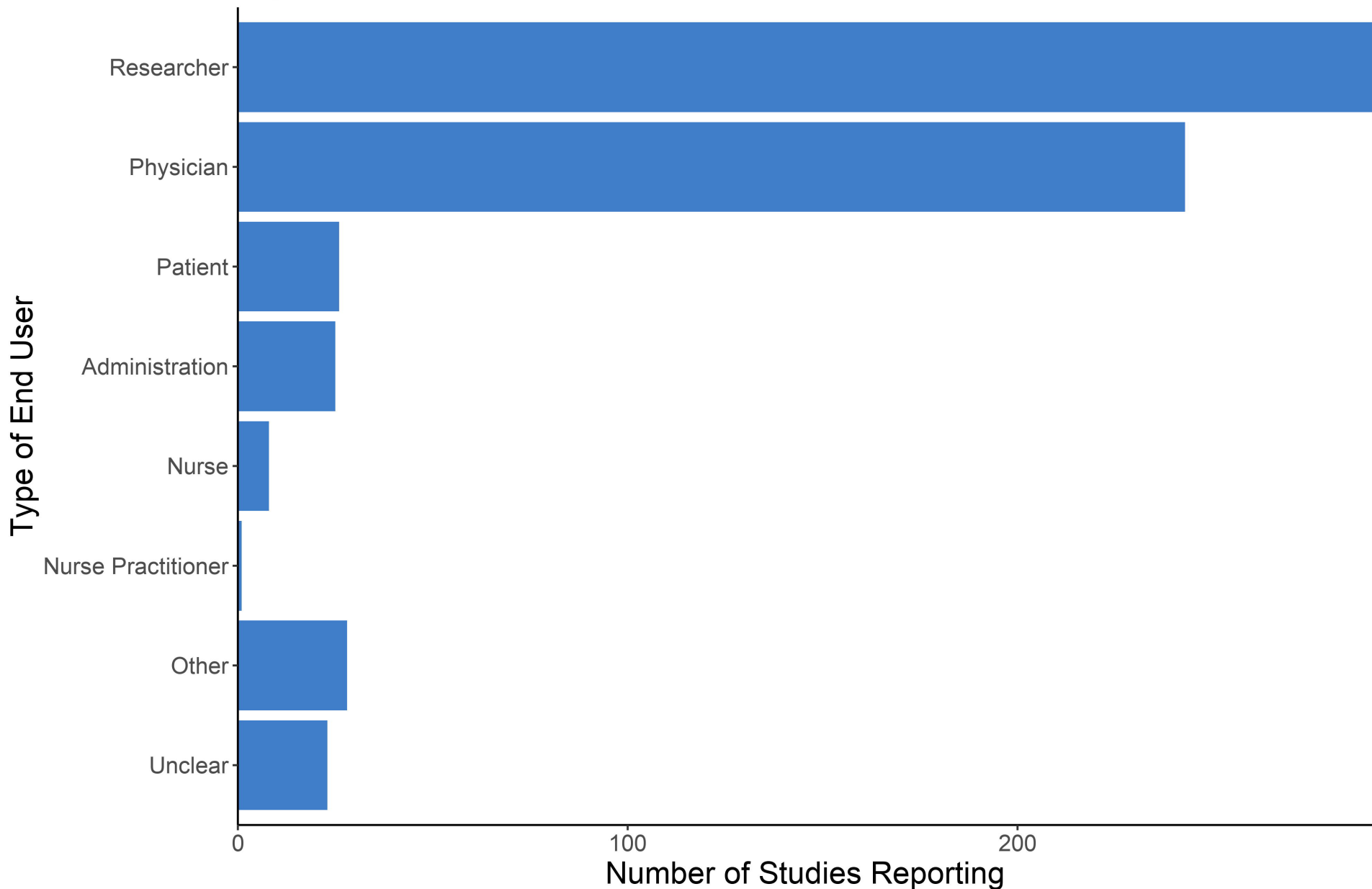


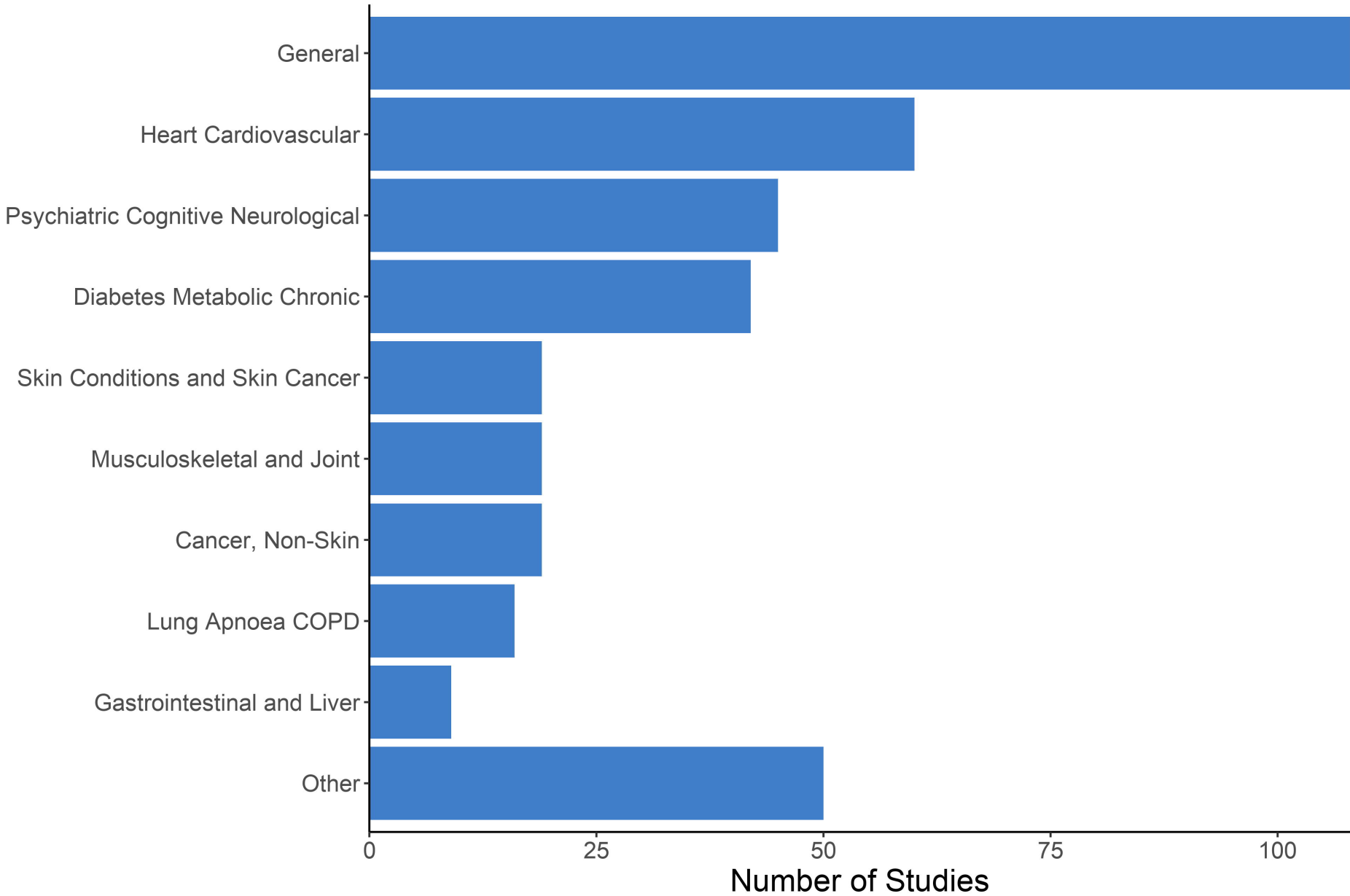
Figure 4 Author Reported End Users



Note: A single study may contribute towards the count for one or more end user categories.

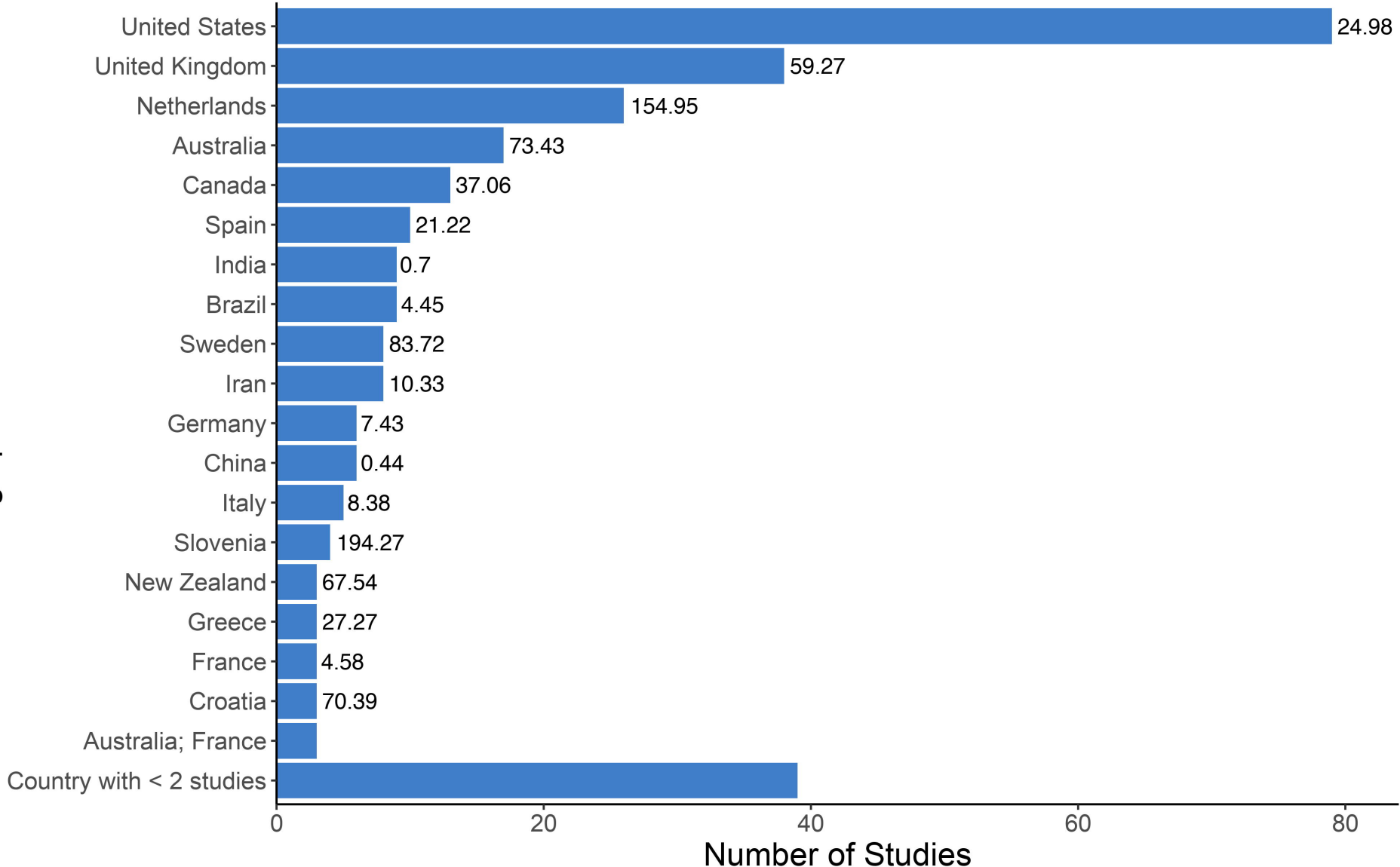
Figure 5 Health Conditions

Health Condition of Interest



Note: Only studies for which target health condition(s) could be identified are included (n = 387).

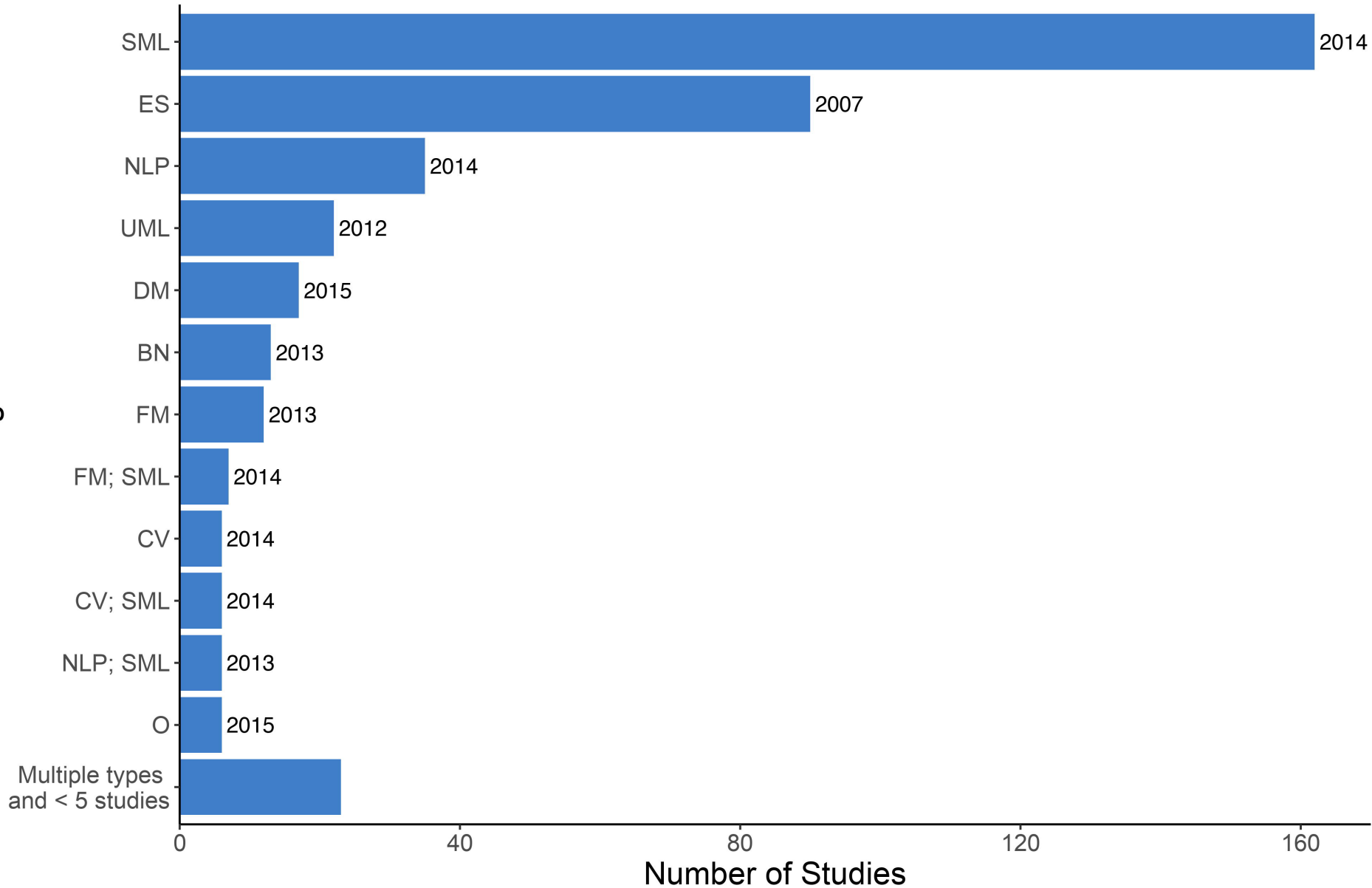
Figure 6 Location of Data Source or Intended Implementation with Per Capita Rates



Notes: Only studies with location reported are included (n=292).
Number at the end of each bar is the number of studies per 100,000,000 people, based on 2013 population estimates.

Figure 7 Subfields of Artificial Intelligence with Median Year of Publication

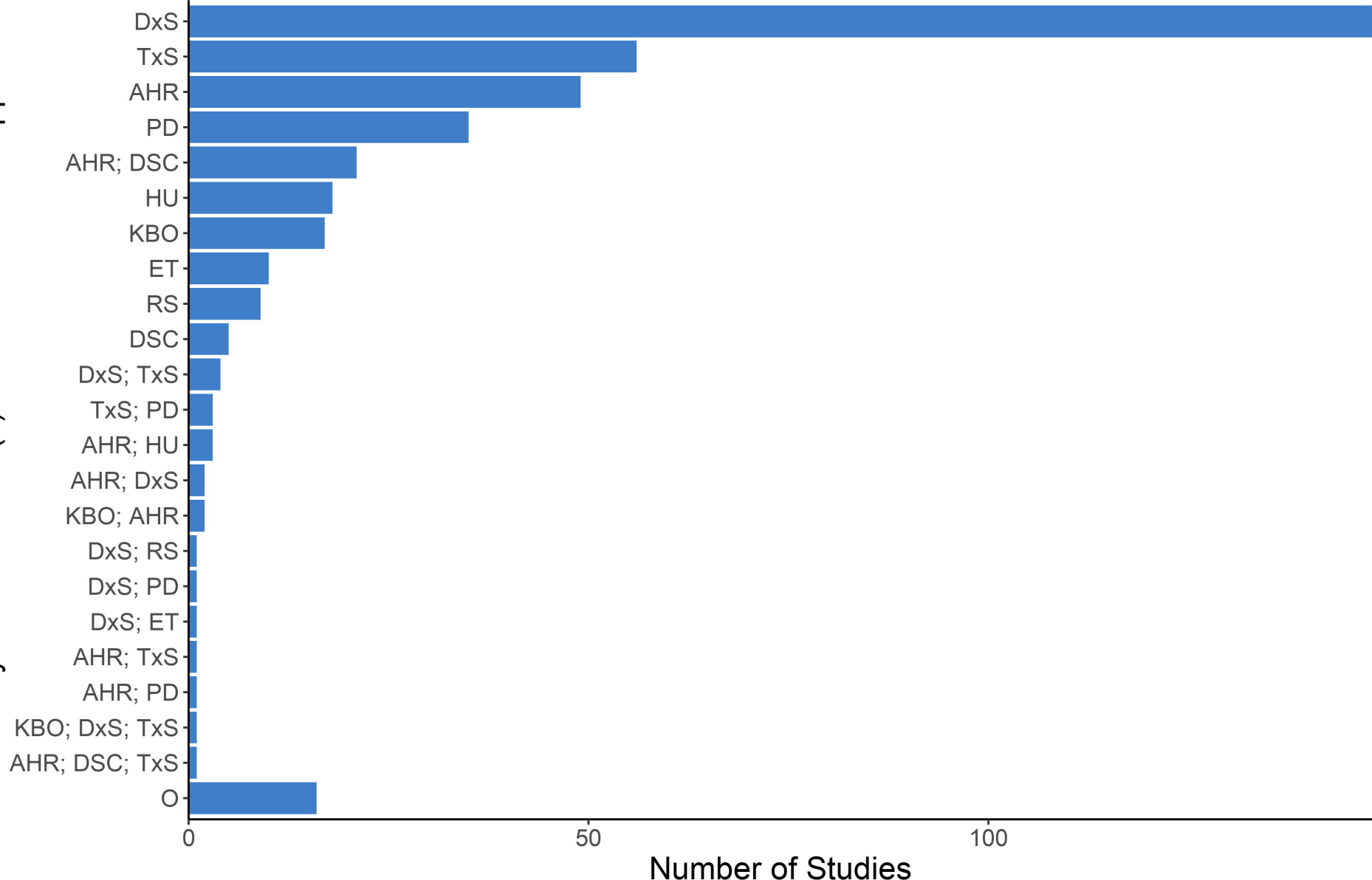
Artificial Intelligence Subfield



Legend: SML: Supervised Machine Learning, ES: Expert System, NLP: Natural Language Processing, UML: Unsupervised Machine Learning, DM: Data Mining, BN: Bayesian Network, FM: Fuzzy Models, O: Other

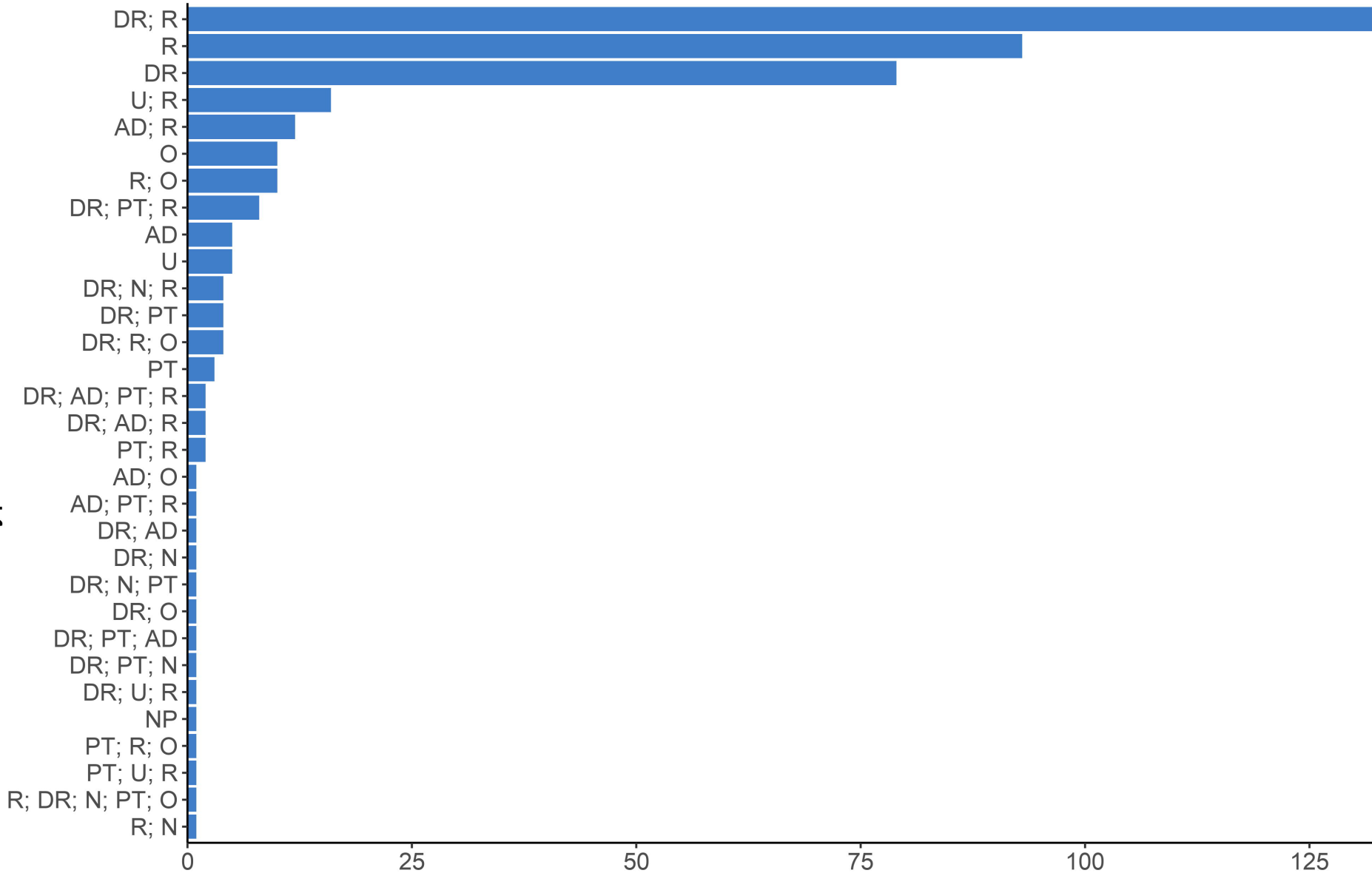
Figure 1S Detailed Breakdown of Primary Care Functions

Primary Care Function(s) AI is Intended to Support



Legend: AHR: Analyze Health Records, DSC: Descriptives, ET: Education or Training, KBO: Knowledge Base or Ontology Construction, PD: Prediction, DxS: Diagnostic Decision Support, RS: Referral Support, TxS: Treatment Decision Support, HU: Health Care Use, O: Other

Figure 2S: Detailed Breakdown of Author Reported Intended End Users



Legend: DR: Physician, R: Researcher, U: Unknown, AD: Administrator, PT: Patient, N: Nurse, NP: Nurse Practitioner, O: Other

Figure 3S Detailed Breakdown of Health Conditions

Health Condition of Interest

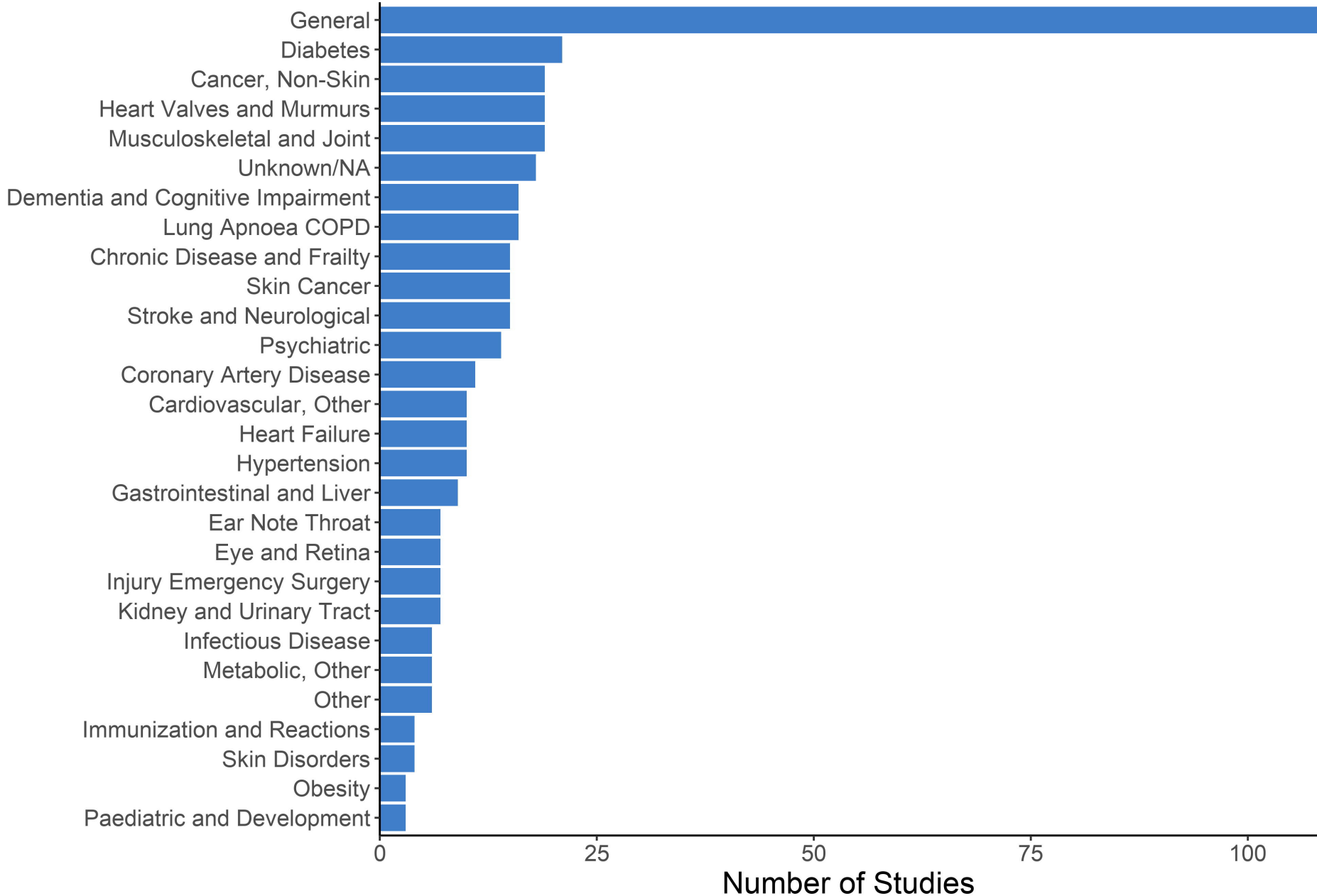
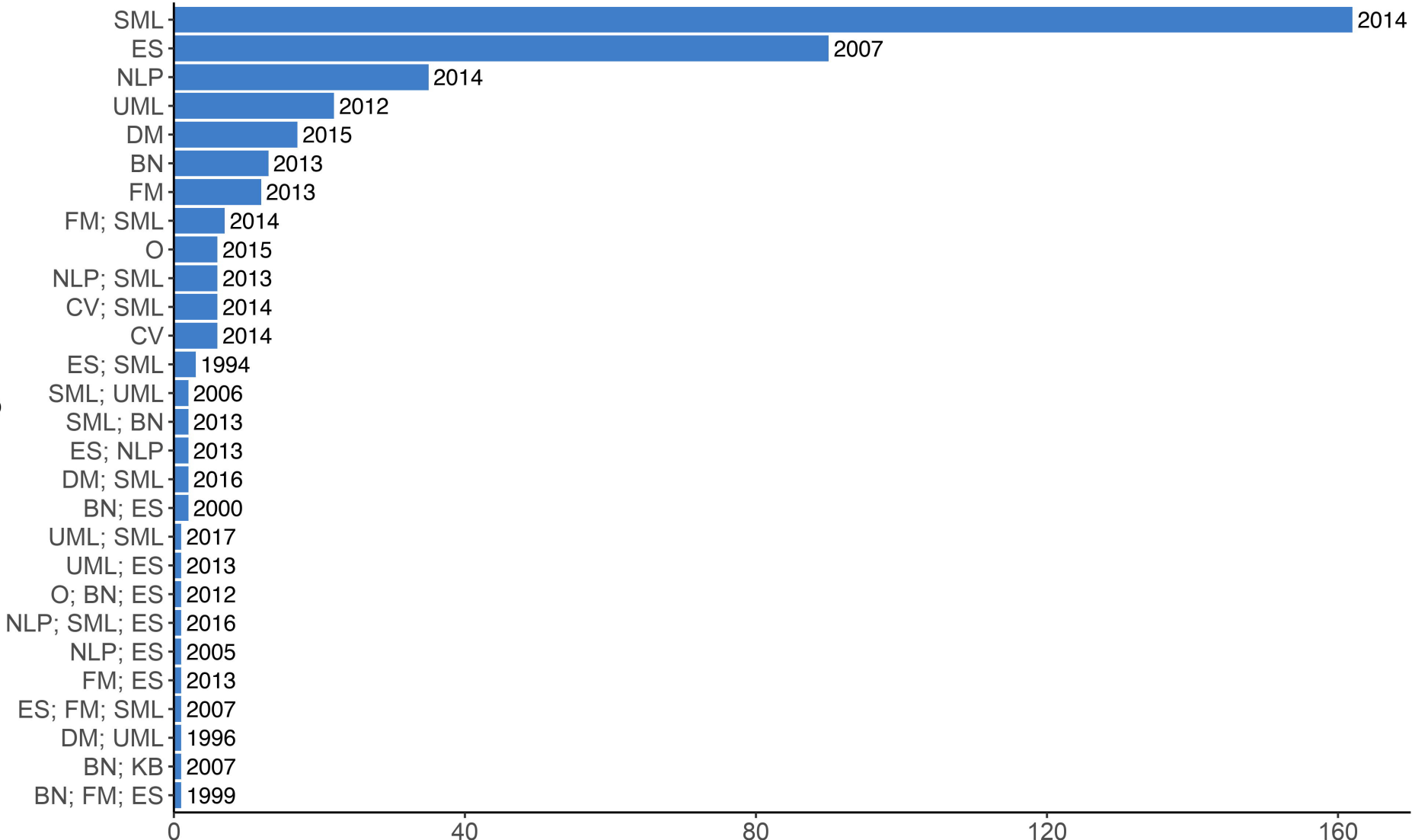


Figure 4S: Detailed Breakdown of Artificial Intelligence Subfields with Median Year of Publication

Artificial Intelligence Subfield



Legend: SML: Supervised Machine Learning, ES: Expert System, NLP: Natural Language Processing, UML: Unsupervised Machine Learning, DM: Data Mining, BN: Bayesian Network, FM: Fuzzy Models, O: Other