Demographic Patterns in Exclusion Criteria Across Phase 3 Randomised Controlled Clinical Trials for Cancer on the ClinicalTrials.gov Database

Table of Contents

Executive Summary

Background

Historically, patients have been excluded from phase 3 clinical trials on the basis of sex, age, and other pre-existing conditions. This exclusion leads to a lack of evidence for treatments in these groups. Existing literature suggests that the historical trends in exclusion have been addressed for sex in recent years (1), but continue to be a problem for both older patients (2) and those with pre-existing conditions (3).

Methods

We studied data from the ClinicalTrials.gov database to determine whether there were trends in exclusion from phase 3 trials for cancer-related treatments on the bases of sex, age, or pre-existing conditions (heart issues, HIV, pregnancy, breastfeeding, and psychiatric conditions). For sex and age, we examined the proportion of studies excluding individuals of a given sex or age across three broad cancer groups: digestive, head and neck, and respiratory cancers. To determine exclusion rates on the basis of pre-existing conditions, we determined whether studies mentioned keywords related to particular conditions in their exclusion criteria. We assessed the accuracy of this approach by manually validating a subset (10%) of our search results.

Results

Our findings were consistent with the existing literature. Within the cancer studies examined, there was negligible exclusion on the basis of sex. However, significant proportions of trials excluded elderly people, despite the growing burden of cancer among older age groups. Pregnant or breastfeeding patients, patients with heart issues, and HIV+ patients were also frequently excluded from cancer-treatment trials.

Conclusions

The age groups and pre-existing conditions we found to be most often excluded from phase three trials represent important sub-populations of the patients affected by these cancers. For elderly individuals especially, there is a large mismatch between the burden of the disease and the extent of this exclusion. Informed use of cancer-treatments by clinicians in these sub-populations requires evidence from clinical research. For some conditions, this evidence could be coming from phase 3 trials – in many cases, exclusion criteria are out-dated and should be changed (4,5). For other conditions where exclusion is more justifiable due to safety concerns (e.g., pregnancy), evidence needs to come from other sources.

Background

Underrepresentation in clinical trials refers to a mismatch between the demographics of the patients enrolled in a trial and those of the population that the treatment being tested is intended to treat. This can lead to treatments that have a poor evidence base for use in the excluded individuals, despite them being intended users of the treatment. As such, clinicians are often justifiably reluctant to give such individuals the treatment because of concerns about undiscovered harmful effects (6), which can stem from the excluded patients having different physiologies and pharmacokinetics and therefore different adverse reactions and treatment responses (7). Thus, for randomised controlled trials to be useful in a clinical context, they must be applicable to the entire patient population in question (8).

The primary aim of our study was to investigate trends in the underrepresentation of different demographic groups in the ClinicalTrials.gov registry. This underrepresentation can occur because of systemic barriers to entering studies (e.g., study location, recruitment processes, socio-economic status of participants) (9) or the exclusion criteria established for the study itself. Given that all studies in the database are required to declare their exclusion criteria, we decided to investigate patterns in exclusion criteria along the lines of sex, age, and several comorbidities.

Given that patient demographic profiles vary across diseases, we decided to focus our analysis on cancer, as it was the most widely studied disease area across phase 3 trials in the database. As cancer itself is a broad disease category, we further grouped our analysis by particular subtypes of cancer.

Exclusion by Sex

Historically, women were systematically excluded from clinical trials. In 1977, FDA guidelines recommended that women of childbearing age be excluded from the early phases of most clinical studies (10). At the time, they were primarily concerned with possible foetal harm that could be caused by experimental drugs (11). Consequently, many studies went beyond FDA guidelines and even excluded women from phase 3 trials (12). Over time, increased awareness that sex often influences drug responses led to a reversal of the FDA's recommendation (12,13). Given the historical issue of sex exclusion in clinical trials, we wanted to examine the current state of affairs in the ClinicalTrials.gov database. In other countries like Australia, evidence from 2008 suggested an absence of sex exclusion in clinical research (11). In our study, we examined whether this trend was also present in the clinicaltrials.gov data.

Exclusion by Age

Exclusion based on age often represents an attempt to exclude individuals that are considered more likely to have comorbidities (4). These might make the study unsafe for an individual, or could make it difficult for the researchers to complete the study (e.g., if a patient is likely to die before the study terminates). While exclusion based on these criteria

may be justifiable, exclusion based solely on an upper age limit is not (14). Historically, however, many studies have used this criteria (7,8). This unjustified exclusion continues to be a widespread issue in clinical trials: a systematic review of age exclusion in all randomised controlled trials (RCTs) in the BMJ, Lancet, JAMA and NEJM published from 1998 to 2015 found a minimal increase in justifiable age-based exclusions despite an increased awareness of this problem in the literature (9). In our study, we looked to see if this trend was also present in the ClinicalTrials.gov database.

Exclusion based on Pre-existing Conditions

Patients with pre-existing conditions are often excluded from clinical trials, which could be especially problematic when these are common in the patient population (11). These conditions may cause patients to react differently to treatments from those included in the study population (15). Conditions that frequently lead to automatic exclusion include cardiovascular diseases, psychiatric conditions and HIV (16). Pregnant and breastfeeding patients are often also excluded from participating in clinical trials due to a fear of liability and concerns regarding the safety of the foetus (17). To examine trends in the exclusion of patients on the basis of pre-existing conditions, we identified some of the most commonly mentioned conditions among the exclusion criteria for cancer-related clinical trials and examined their presence in studies over time.

Study Objective

In our study, we examine demographic patterns in exclusion criteria – for age, sex and other pre-existing conditions – across phase 3 RCTs in the ClinicalTrials.gov database across different subtypes of cancer. This work should help us better understand how exclusion criteria have contributed to the underrepresentation of certain patient groups. Ultimately, we intend for this information to improve the representativeness of evidence available for treatments.

Methods

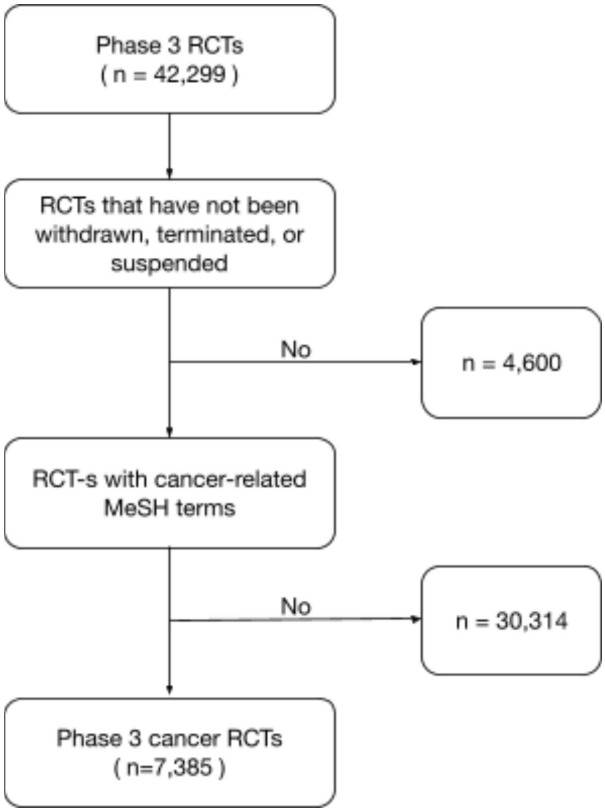
Data Source

The primary data source for this study was ClinicalTrials.gov (6). Any clinical trial with “one or more sites in the US . . . conducted under an FDA investigational new drug application or investigational device exemption . . . [or involving] a drug, biological, or device product that is manufactured in the United States or its territories and is exported for research” must be registered on this database under US federal law (19). The database was accessed through the Aggregate Analysis of ClinicalTrials.gov (AACT) (20). AACT is a relational database that downloads data from ClinicalTrials.gov on a daily basis. The database is stored on a PostgreSQL server, from which we extracted the required data fields (see data dictionary in Appendix A) for our study using a SQL query run on pgAdmin4 (version 6.19) on January 24, 2023. We downloaded the output of the query as a .csv file which was used in our analyses.

Data Preparation

We extracted 42,299 trials categorised as Phase 3, interventional RCTs from the AACT database. We then excluded any trials that were withdrawn, terminated or suspended. Additionally, we excluded any trials that were not cancer-related. Through this, we identified a total of 7,385 phase 3 cancer RCTs for inclusion in our analyses.

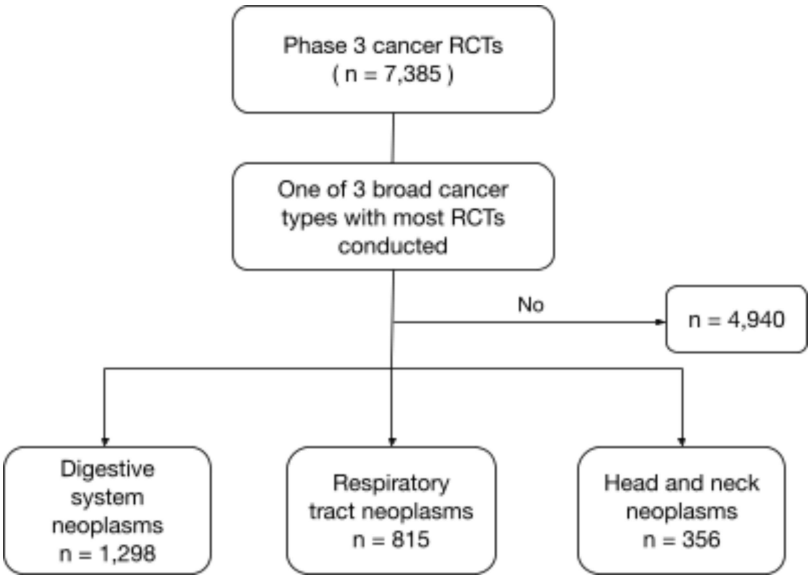
Figure 1. Flow diagram detailing how the subset of phase 3 cancer RCTs were filtered:



Selecting Cancer Groups

Medical Subject Headings (MeSH) (21) are used to categorise all studies on ClinicalTrials.gov. The highest level classification of cancer studies classifies studies by neoplasm site. We counted the number of studies investigating each neoplasm site and picked the three most common cancer types (affecting both males and females): Digestive system neoplasms" (n = 1,298), "Respiratory tract neoplasms" (n = 814) and "Head and neck neoplasms" (n = 333) as illustrated in Figure 2. Trials that did not contain exactly 1 of the 3 MeSH terms pertaining to these broad cancer categories were excluded from our study (n=4,940).

Figure 2. Flow diagram of categorisation into 3 broad cancer types:



Exclusions by Sex and Age

We compared the rates of exclusion for sex and age between these broad cancer types. We then examined trends in age exclusion over time. To do this, we grouped studies by 4-year periods (2008-2012, 2012-2016, 2016-2020 and 2020-2024) based on their start date. This left a large number of studies in each year group for each cancer type (see Table 1).

Table 1. Number of studies in each year group by cancer type

		Year Group			
		[2008, 2012)	[2012, 2016)	[2016, 2020)	[2020, 2024)
Cancer Type	Digestive System	192	228	287	225
	Head and Neck	46	69	72	77
	Respiratory	103	131	194	170

We then calculated the percentage of studies excluding a given age group within each time-period. The age groups used (0-5, 5-18, 18-40, 40-65, 65-75, and 75-101) were decided based on the age intervals that appeared most frequently together in the exclusion analysis by age. The analysis was limited to studies from 2008 onwards since this is the year when all clinical trials were mandated by US Federal law to be registered on ClinicalTrials.gov (19).

Exclusion by Medical Conditions

We began by identifying which studies had clearly delineated their inclusion and exclusion criteria within the free text of the criteria column, as marked by words such as “Inclusion Criteria:” and “Exclusion Criteria:”, respectively. This allowed us to programmatically separate the two without using semantic techniques.

N-Grams

N-grams are ordered sequences of words with lengths *n* (22). Among these studies, we observed the most frequently appearing n-grams of lengths 2 to 5 among the exclusion criteria.

Keyword Search

The most frequent n-grams among the exclusion criteria pertained to the following broad conditions: heart issues, HIV, pregnancy, or breastfeeding. Our review of past studies suggested that people with psychiatric illnesses (3) are frequently excluded by clinical trials, so we decided to also include this condition. Table 2 indicates the medical terms considered under each pre-existing condition. The keywords used for the search were based on these medical terms.

Table 2. Medical terms considered for each investigated condition:

Pre-existing Conditions	**Medical Terms**
Heart Issues	Congestive heart failure, heart failure, unstable angina, ischemic heart failure, coronary disease, hypertension, myocardial infarction, cardiovascular disease, arrhythmia, high blood pressure
HIV	Human immunodeficiency virus
Pregnancy	Pregnancy
Lactation	Breastfeeding, lactation
Mental illness	Psychiatric disorder/illness, depression, anxiety, schizophrenia

We created a keyword dictionary for these pre-existing conditions, considering modifications (such as adding punctuations or using abbreviations) as necessary to capture as many variations of each medical term as possible (see keyword search terms in Appendix B). We flagged clinical trials as excluding a given condition if it mentioned any of that condition’s keywords.

Manual Validation of Search Results

To measure the accuracy of our keyword search, we randomly sampled a set of studies for manual validation. For each condition, we randomly selected 10% of studies tagged by our search as containing that condition. We also selected 10% of studies that had not been

tagged with any condition by the keyword search. We then manually classified this entire set of studies, and calculated the accuracy of the keyword search for each condition.

Analysis of Exclusion Rates

For these conditions, we reported their overall exclusion across trials, as well as segregated by cancer type. We then evaluated how the exclusion rates for each of these conditions varied over time in each cancer group, using the same 4-year periods used for our time-trend analysis of sex-exclusion described above.

All analyses described were performed using Python version 3.8.10 on a Google Colaboratory notebook, stored in a Github repository.

Results

Exclusions by Sex

Exclusion rates by sex were below 1% across cancer categories except for digestive system studies, which excluded males in approximately 2.5% of studies. (see Table 3).

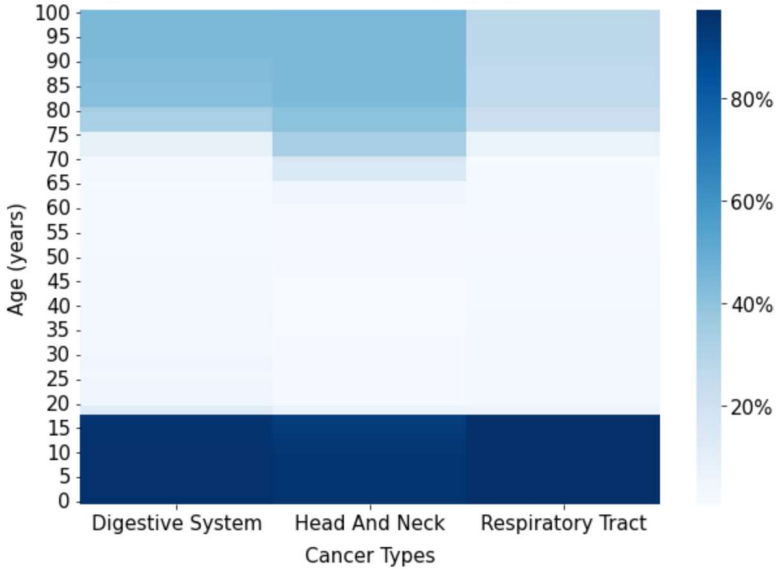
Table 3. Indicates exclusion by sex, stratified by broad cancer category:

Cancer Category	**Studies Excluding Males**	**Studies Excluding Females**
Digestive System	2.54%	0.15%
Head and Neck	0.28%	0.84%
Respiratory Tract	0.12%	0.00%
Across All Studies	1.42%	0.21%

Exclusions by Age

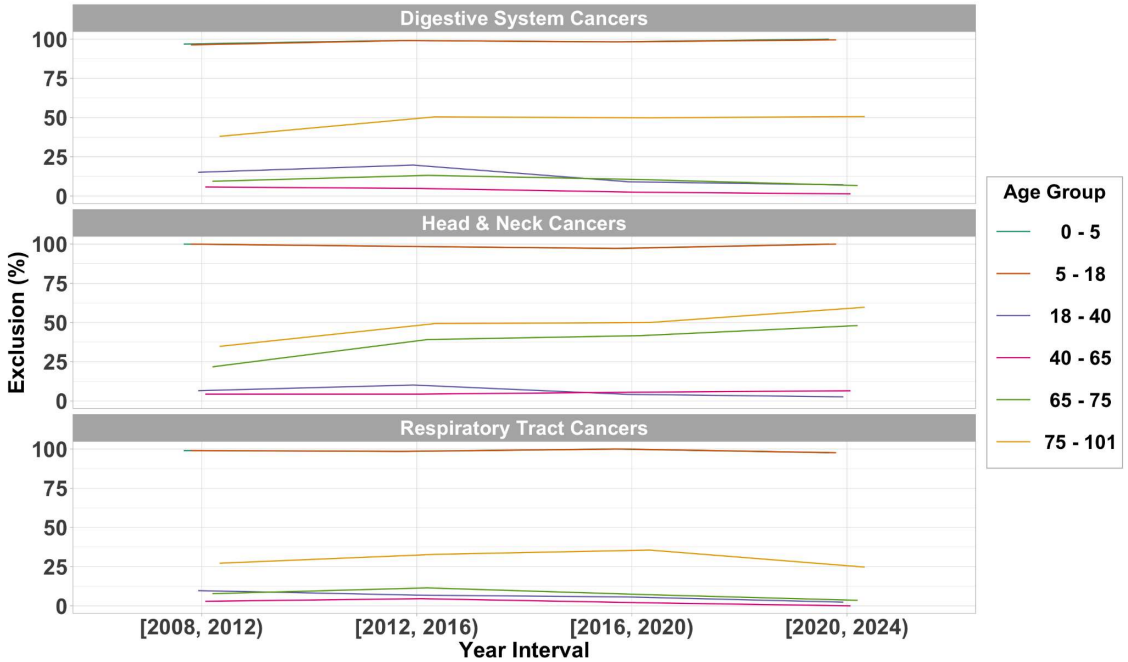
The heatmap in Figure 3, shows that almost all studies in all 3 broad cancer categories excluded individuals younger than 18-years old. Between ages 18-65 years old, exclusion rates for all categories were generally low. Above 65 years old, the rates of exclusion for all categories progressively increased with increasing age. The percentage of studies excluding elderly individuals was highest in the ‘head and neck cancer’ category and lowest in the ‘respiratory tract cancer’ category.

Figure 3. Heat-map indicating the percentage of cancer trials excluding by age and cancer type across cancer categories:



For ages 0-18, exclusion rates remained at 96%-100% over time. Ages 18-40 showed a downward trend in exclusion rates across all cancer types (going from 20%-7% in digestive system cancers, 10%-2% in head and neck, and respiratory tract cancers). There was a downward trend for ages 18-40 across all cancer types. There were minimal changes for ages 40-65 – the age group was the least excluded (below 10%) across all studies. For ages 65-101, there was a reduction in exclusion rates for digestive system and respiratory tract cancers. However, there was an increase in exclusion rates for head and neck cancers (22%-48% for ages 65-75, and 35%-60% for ages 75-101).

Figure 4. Age exclusion across cancer studies from 2008 to 2024:



Exclusion by Common Conditions

Of the studies in the sample, 2,057 (83.3%) had clearly delineated inclusion and exclusion criteria and were thus included in the keyword search:

Table 4. Studies with clearly delineated inclusion and exclusion criteria by cancer type:

Cancer Category	No. Included	No. Excluded	% Included
Digestive System	1,089	209	83.9%
Head and Neck	298	58	83.7%
Respiratory Tract	670	145	82.2%
Total	2,057	412	83.3%

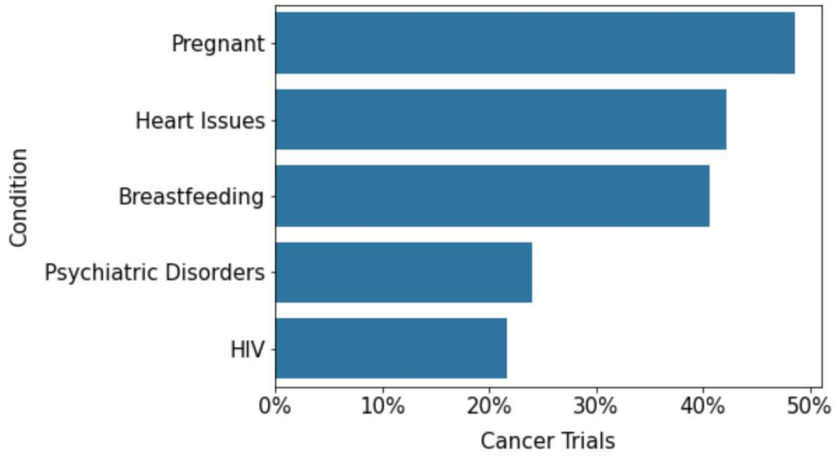
A total of 409 studies were manually validated by the team. We found the keyword search correctly tagged above 94% of the clinical trials that were randomly sampled for each category of pre-existing conditions. Table 5 details the performance of the algorithm.

Table 5. Accuracy of keyword search by condition:

Condition	No. Tagged as Excluded	No. Correctly Tagged	Accuracy	Misclassification Rate
Breastfeeding	261	389	95.1%	4.9%
HIV	156	405	99.0%	1.0%
Heart Issues	256	387	94.6%	5.3%
Pregnancy	305	408	99.8%	0.2%
Psychiatric Disorders	158	393	96.1%	3.9%

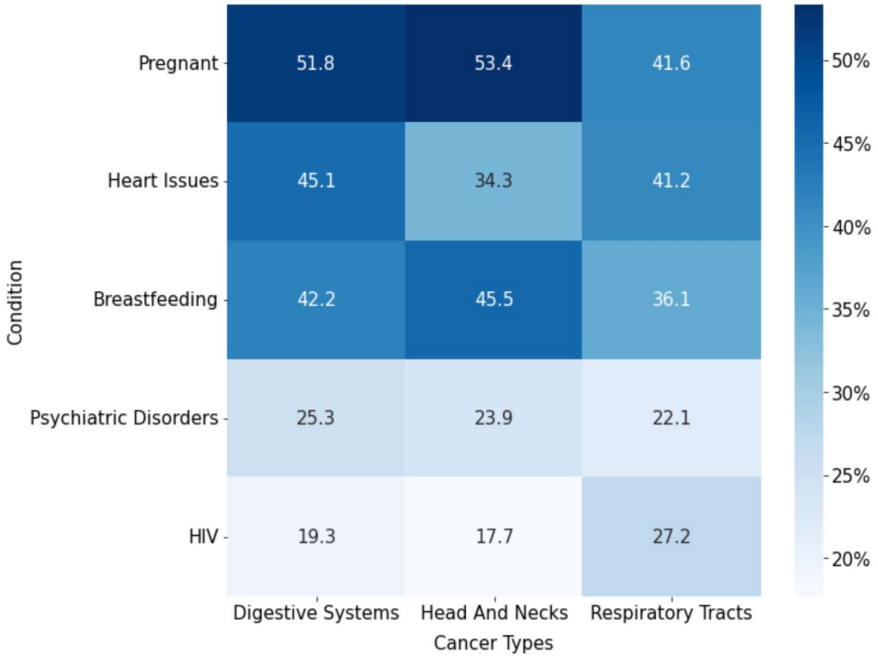
Of the five conditions that were part of the keyword search, pregnancy came up most frequently, having appeared in the exclusion criteria for almost 50% of all the clinical trials in the sample. This was followed by heart issues and breastfeeding, both of which appeared for above 40% of trials. Patients with psychiatric disorders and HIV were excluded in more than 20% of the studies.

Figure 5. Percentage of cancer trials that exclude each condition:



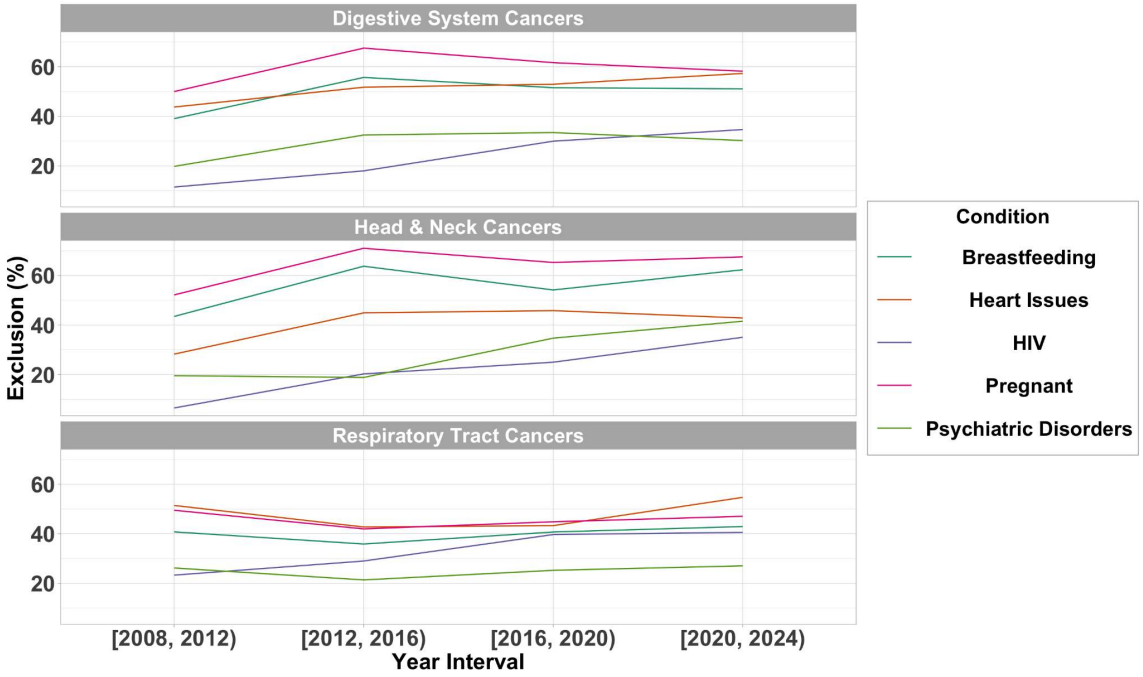
We further investigated exclusion rates stratified by cancer type. These rates did not differ by more than 10% across cancer types, as seen in Figure 6.

Figure 6. Percentage of cancer trials excluding each condition by cancer type:



The proportion of studies excluding individuals based on these conditions has consistently increased across cancer types, with every condition showing a higher rate of exclusion in the latest year-bin compared to the earliest (with the exception of heart issue exclusion in respiratory tract cancers). This trend is especially noticeable for HIV, demonstrating a clear increase in exclusion rate across every year-bin, for every cancer type.

Figure 7. Percentage of studies excluding certain conditions across three cancer categories over time:



Discussion

Data Source

The ClinicalTrials.gov database is not a comprehensive record of all clinical trials globally. Furthermore, there are variations in the quality of the information recorded across studies within the database, which limited the studies we were able to investigate. The conclusions we draw in the current study are based on those trials within the database that had explicitly stated inclusion and exclusion criteria. The patterns we identify may not extend to other studies within the database, or for all studies globally.

Exclusions by Sex

A small proportion (less than 2%) of studies excluded based on sex, with a higher exclusion rate for males (1.4%) than for females (0.2%). However, 32 of the 35 studies excluding males were investigating ovarian cancer. Our inclusion of these studies reflects a limitation of our initial MeSH term filter. We selected for digestive system, respiratory, and head and neck cancers, however some of these studies had additional MeSH terms for sex-specific cancers. If these studies are dismissed, sex-specific exclusion drops to a negligible level (0.12%). This is consistent with the findings that have previously been discussed in the literature (1).

Exclusions by Age

Our initial investigation of age exclusion showed that exclusion rates first increased dramatically at age 65 before continuing to increase with age. Our time-trend analyses

demonstrate that this is an ongoing trend, with recent studies still demonstrating high levels of exclusion of older participants. Exclusion of participants based on age alone has often been criticised as being an insufficient reason for exclusion (23,24). Instead, trials should aim to include older populations and only exclude groups for whom there are specific safety concerns. The failure of recent studies to do this is concerning, especially since incidence rates of cancer are highest in the elderly within the USA population (see Appendix C) (25).

Almost all trials across cancer types excluded children under 18 years old. This pattern may be due to the low-burden of disease of cancer in this age group (as seen in Appendix C). This could influence investigator decision-making when considering the potential risks and benefits of including children in these trials (26). In the USA, the Belmont Report (27), specifies strict limitations on research that presents more than a minimal risk to children. This report forms the basis of regulations around clinical trials in the USA, which may explain the high levels of exclusion we have seen in under 18 year-olds.

Exclusion based on Pre-existing Conditions

The five conditions we identified using n-gram frequency and reviews of related literature appeared in the exclusion criteria of 20% to 50% of cancer-related studies. Manual validation of our keyword search algorithm demonstrated a high level of accuracy (more than 95%). This suggests that the presence of a specified keyword within the exclusion criteria of a study was a good proxy for the actual exclusion of the associated condition. Thus, it is reasonable to interpret our findings in Figure 5 with a high degree of confidence.

In some cases, exclusion based on these pre-existing conditions may be justifiable. For example, it has been shown that there are often teratogenic effects of cancer treatments (28), explaining the frequent exclusion of pregnant women in clinical trials. In contrast, the exclusion of participants based on their HIV status alone, based on concerns about immunodeficiency, has been argued to be less justifiable given the modern treatment options available for HIV patients (5). Instead, eligibility criteria should focus on individuals' current antiretroviral therapy and immune status (5).

Despite their sometimes justifiable exclusion from phase 3 trials, efforts still must be made to develop an evidence base for treatments in populations with concurrent conditions. These groups represent significant sub-populations affected by the cancers we have studied and therefore it is important to understand the safety and efficacy of cancer therapies in these groups.

Limitations of the Keyword Search

We used n-grams to determine which conditions to include in our keyword search. One issue with this approach is that the most frequent n-grams tend to represent medical terms or phrases that are described consistently across different studies' eligibility criteria. Thus, n-gram summaries tend to underestimate the frequency of conditions that are often

misspelt, abbreviated, or referred to in different ways across studies. This is a potential limitation of the approach we used to select conditions for further investigation.

Having selected conditions based on n-grams, we then searched for the condition within the exclusion criteria using a keyword search. The central limitation to this approach is that it only checks for the presence of search terms without considering the context of the sentences or phrases they belong to. As such, it cannot detect negatively phrased items (e.g. “this study will not exclude patients with HIV”) or conditional exclusions (e.g. “pregnant women are excluded from this study, but may be included if conditions x, y, or z are met”).

Our keyword search would also struggle to deal with concepts like severity thresholds for exclusion. For example, in many studies excluding those with heart issues only those with the most severe cardiovascular issues are excluded. Because these thresholds are not reported in a standardised format, our keyword search would struggle to identify them. For these reasons, we suggest that any subsequent work utilising a keyword search to identify exclusion of pre-existing conditions also perform manual validation of their results.

Lastly, many studies include broad, catch-all exclusion criteria such as ‘any medical or psychiatric conditions deemed by the investigator to be unfit for the study.’ Some conditions might be frequently excluded under these criteria and thus cannot be accounted for in the analysis altogether.

Recommendations for Future Research

In our study, we sought to understand the extent and causes of underrepresentation in clinical research because this underrepresentation limits the evidence-based treatments available for often marginalised populations (e.g. elderly patients, those with pre-existing conditions). By studying the extent and causes of underrepresentation, future research could tell us when this is avoidable (e.g. with upper age limits), and how we might begin to address it. Below, we make suggestions for future research.

In our study, we focused specifically on how patterns in exclusion criteria can lead to the systematic underrepresentation of certain populations in phase 3 clinical trials for cancer treatments. Future work could build on this by examining patterns across different pre-existing conditions and disease areas.

As discussed in the background, even when certain populations are not excluded from clinical research, there may be systemic barriers (e.g. socio-economic status) to their entry into clinical trials. Existing work supports this idea – the demographics of clinical trials are often unrepresentative of certain groups, even when exclusion criteria are not present (29–31). Future research could examine the reasons which prevent under-represented groups from participating in these trials, which may ultimately help to improve the evidence-based treatments available for all populations.

References

1. Rogers WA, Ballantyne AJ. Exclusion of women from clinical research: myth or reality? Mayo Clin Proc. 2008 May;83(5):536–42.
2. Thake M, Lowry A. A systematic review of trends in the selective exclusion of older participant from randomised clinical trials. Arch Gerontol Geriatr. 2017 Sep 1;72:99–102.
3. Humphreys K, Blodgett JC, Roberts LW. The exclusion of people with psychiatric disorders from medical research. J Psychiatr Res. 2015 Nov;70:28–32.
4. McMurdo MET, Roberts H, Parker S, Wyatt N, May H, Goodman C, et al. Improving recruitment of older people to research through good practice. Age Ageing. 2011 Nov 1;40(6):659–65.
5. Uldrick TS, Ison G, Rudek MA, Noy A, Schwartz K, Bruinooge S, et al. Modernizing Clinical Trial Eligibility Criteria: Recommendations of the American Society of Clinical Oncology-Friends of Cancer Research HIV Working Group. J Clin Oncol Off J Am Soc Clin Oncol. 2017 Nov 20;35(33):3774–80.
6. Rothwell PM. External validity of randomised controlled trials: "To whom do the results of this trial apply?" The Lancet. 2005 Jan 1;365(9453):82–93.
7. Stephenson T. How children's responses to drugs differ from adults. Br J Clin Pharmacol. 2005 Jun;59(6):670–3.
8. Kennedy-Martin T, Curtis S, Faries D, Robinson S, Johnston J. A literature review on the representativeness of randomized controlled trial samples and implications for the external validity of trial results. Trials. 2015 Nov 3;16(1):495.
9. Murphy E, O'Keeffe A, O Shea N, Long E, Eustace JA, Shiely F. Patient perceptions of the challenges of recruitment to a renal randomised trial registry: a pilot questionnaire-based study. Trials. 2021 Sep 6;22(1):597.
10. Food and Drug Administration. General considerations for the clinical evaluation of drugs. Gen Consid Clin Eval Drugs. 1977;11–11.
11. Merkatz RB. Inclusion of Women in Clinical Trials: A Historical Overview of Scientific Ethical and Legal Issues. J Obstet Gynecol Neonatal Nurs. 1998 Jan 1;27(1):78–84.
12. U.S. Food and Drug Administration. Guideline for the study and evaluation of gender differences in the clinical evaluation of drugs; notice. Fed Regist. 1993 Jul 22;58(139):39406–16.
13. Merkatz RB, Temple R, Sobel S, Feiden K, Kessler DA. Women in Clinical Trials of New Drugs – A Change in Food and Drug Administration Policy. N Engl J Med. 1993 Jul 22;329(4):292–6.
14. Cherubini A, Oristrell J, Pla X, Ruggiero C, Ferretti R, Diestre G, et al. The Persistent Exclusion of Older Patients From Ongoing Clinical Trials Regarding Heart Failure. Arch Intern Med. 2011 Mar 28;171(6):550–6.
15. Bassi PU, Osakwe AI, Ogar CK, Elagbaje C, Nwankwo BB, Balogun ST, et al. Impact of comorbidity on adverse drug reaction profile in a cohort of patients treated with Artemisinin combination therapies for uncomplicated malaria in Nigeria. Pharmacol Res Perspect. 2017 Mar 24;5(2):e00302.
16. Duma N, Kothadia SM, Azam TU, Yadav S, Paludo J, Vera Aguilera J, et al. Characterization of Comorbidities Limiting the Recruitment of Patients in Early Phase Clinical Trials. The Oncologist. 2019 Jan;24(1):96–102.
17. Shields KE, Lyerly AD. Exclusion of pregnant women from industry-sponsored clinical trials. Obstet Gynecol. 2013 Nov;122(5):1077–81.
18. U.S. National Library of Medicine. ClinicalTrials.gov [Internet]. [cited 2023 Feb 2]. Available from: https://www.clinicaltrials.gov/

19. 42 CFR Part 11 -- Clinical Trials Registration and Results Information Submission [Internet]. [cited 2023 Feb 7]. Available from: https://www.ecfr.gov/current/title-42/chapter-I/subchapter-A/part-11
20. Clinical Trials Transformation Initiative. AACT Database [Internet]. [cited 2023 Feb 2]. Available from: https://aact.ctti-clinicaltrials.org/
21. Medical Subject Headings - Home Page [Internet]. U.S. National Library of Medicine; [cited 2023 Feb 8]. Available from: https://www.nlm.nih.gov/mesh/meshhome.html
22. Robertson AM, Willett P. Applications of -grams in textual information systems. J Doc. 1998 Jan 1;54(1):48–67.
23. McMurdo MET, Witham MD, Gillespie ND. Including older people in clinical research. BMJ. 2005;331(7524):1036–7.
24. Watts G. Why the exclusion of older people from clinical research must stop. BMJ [Internet]. 2012;344. Available from: https://www.bmj.com/content/344/bmj.e3445
25. International Agency for Research on Cancer. Global Cancer Observatory [Internet]. [cited 2023 Feb 2]. Available from: https://gco.iarc.fr/
26. Field MJ, Behrman RE, Children I of M (US) C on CRI. Defining, Interpreting, and Applying Concepts of Risk and Benefit in Clinical Research Involving Children [Internet]. Ethical Conduct of Clinical Research Involving Children. National Academies Press (US); 2004 [cited 2023 Feb 7]. Available from: https://www.ncbi.nlm.nih.gov/books/NBK25555/
27. The Belmont Report. Ethical principles and guidelines for the protection of human subjects of research. J Am Coll Dent. 2014 Summer;81(3):4–13.
28. Arnon J, Meirow D, Lewis-Roness H, Ornoy A. Genetic and teratogenic effects of cancer treatments on gametes and embryos. Hum Reprod Update. 2001 Aug;7(4):394–403.
29. Turner BE, Steinberg JR, Weeks BT, Rodriguez F, Cullen MR. Race/ethnicity reporting and representation in US clinical trials: A cohort study. Lancet Reg Health – Am [Internet]. 2022 Jul 1 [cited 2023 Feb 8];11. Available from: https://www.thelancet.com/journals/lanam/article/PIIS2667-193X(22)00069-2/fulltext
30. Xiao H, Vaidya R, Liu F, Chang X, Xia X, Unger JM. Sex, Racial, and Ethnic Representation in COVID-19 Clinical Trials: A Systematic Review and Meta-analysis. JAMA Intern Med. 2023 Jan 1;183(1):50–60.
31. Kong WY, Saber H, Marawar R, Basha MM. Racial and ethnic trends in antiseizure medications trial enrolment: A systematic review using ClinicalTrials.gov. Epilepsy Res. 2021 Jul 1;173:106613.

Appendices

Appendix A: Data Dictionary

This table contains the variable names, descriptions, and data types for each field in the data extract from the AACT. Descriptions were taken from the AACT data dictionary where available.

Field	Variable Name	Description	Data Type
NCT ID	nct_id	Unique identification code for each clinical study	Text
Brief Title	brief_title	Title of the clinical trial	Text
Start Date	start_date	Start date of the clinical trial	Date
Start Date Type	start_date_type	Type of start date provided, whether "Actual" or "Anticipated"	Text
Study Type	study_type	Type of clinical trial; must be "interventional" for all those in the data extract	Text
Phase	phase	Phase of clinical trial; must be "Phase 3" or "Phase 2/Phase 3" for those in the data extract	Text
Overall Status	overall_status	Current status of clinical trials that have not yet completed recruitment, whether: "Completed", "Unknown status", "Recruiting", "Terminated", "Active, not recruiting", "Not yet recruiting", "Withdrawn", "Enrolling by invitation", "Suspended"	Text
Last Known Status	last_known_status	Last known status for uncompleted studies, whether: "Recruiting", "Active, not recruiting", "Not yet recruiting", or "Enrolling by invitation"	Text
Minimum Age	minimum_age	Minimum and maximum age (respectively) required to be eligible for participation in the trial, indicated with units (ex: 5 years, 6 weeks, 7 days, etc.)	Text
Maximum Age	maximum_age		Text
Gender	gender	Included sex for the study population ("All", "Male", or "Female")	Text
Healthy Volunteers	healthy_volunteers	Whether or not a study accepts healthy volunteers: "No" or "Accepts Healthy Volunteers"	Text
Countries Removed	countries_removed	List of countries removed from the study record	Text
Countries Retained	countries_retained	List of countries in which the study has facilities/sites	
Conditions	conditions_table	Name(s) of the disease(s) or condition(s) studied in the clinical study, or the focus of the clinical	Text

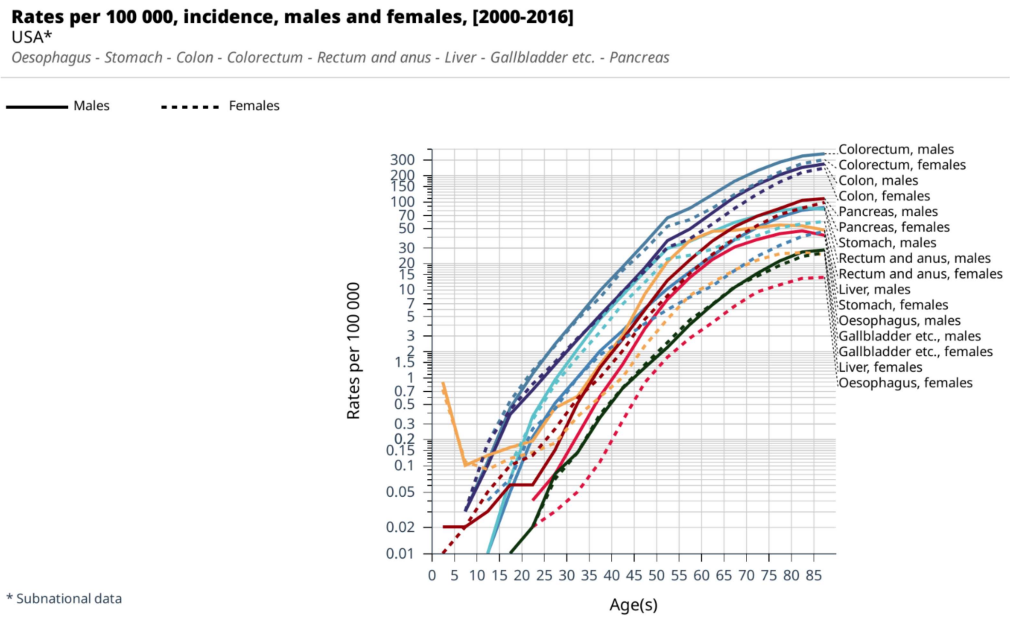
		study. Can include NLM's Medical Subject Heading (MeSH)-controlled vocabulary terms	
MeSH Terms	mesh_term	The NLM Medical Subject Heading-controlled (MeSH) terms for all conditions being studied in the clinical trial	Text
Criteria	criteria	Information about the criteria used to select participants, includes inclusion and exclusion criteria	Text

Appendix B: Keyword Search Terms

Category	Search Terms
Breastfeeding	‘breastfeed’, ‘breast feed’, ‘breast-feed’, ‘breast- feed’, ‘lactation’, ‘lactating’
Heart Issues	‘angina’, ‘arrhythmia’, ‘arrythmia’, ‘congestive heart’, ‘cardiac’, ‘cardiovasc’, ‘coronary disease’, ‘coronary artery disease’, ‘heart failure’, ‘high blood pressure’, ‘hypertension’, ‘ischemic heart’, ‘myocardial’
HIV	‘hiv-’, ‘-hiv’, ‘,hiv’, ‘hiv/aids’, ‘(hiv)’, ‘known hiv’, ‘proven hiv’, ‘hiv positiv’, ‘hiv-positiv’, ‘human immunodeficiency’, ‘or hiv’, ‘hiv infect’, ‘hiv-infect’, ‘ hiv’
Pregnancy	‘pregnan’
Psychiatric Disorders	‘psych’, ‘depressi’, ’anxiety’, ’schizophreni’, ‘mental illness’, ‘mental disorder’

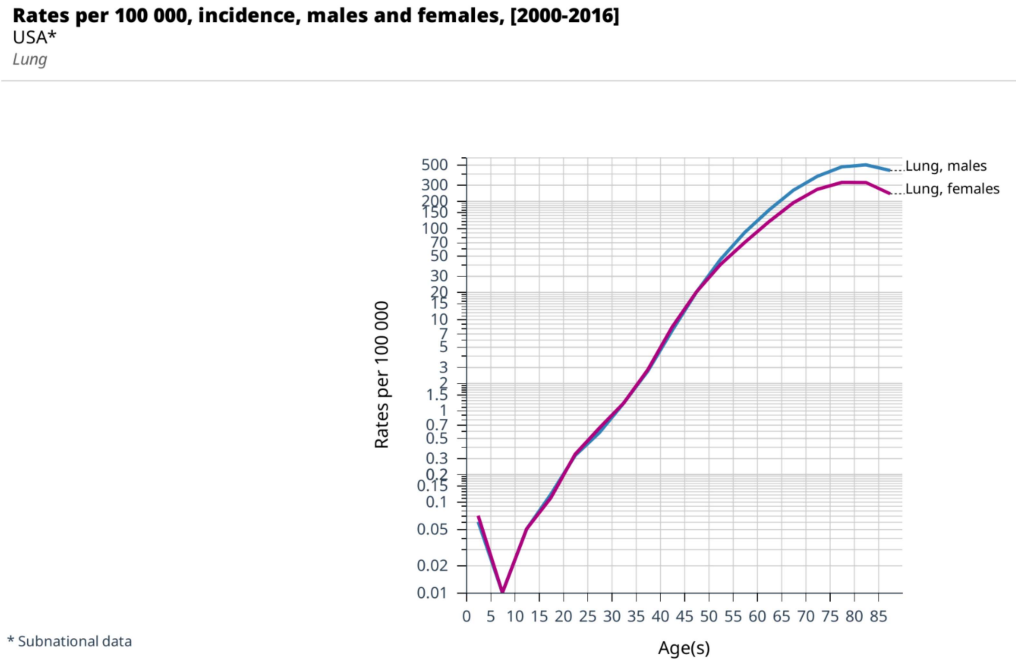
Appendix C: Cancer Age Trends

Appendix C: Figure 1. Age trends in digestive system cancers for a USA population:



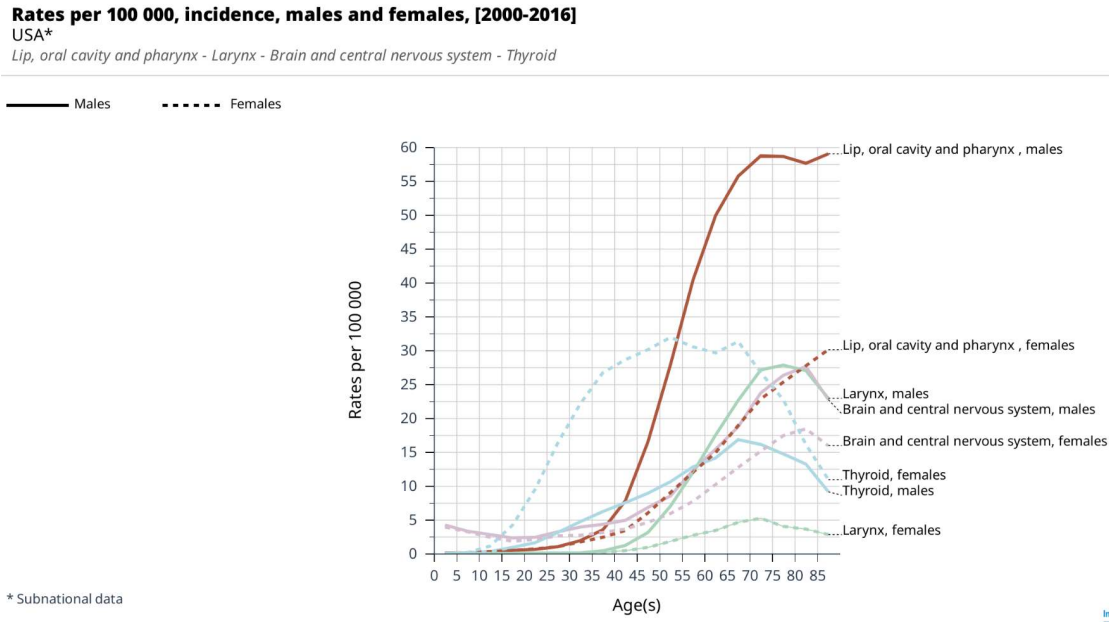
(Source: Global Cancer Observatory - International Agency for Research on Cancer)

Appendix C: Figure 2. Age trends in lung cancers for a USA population:



(Source: Global Cancer Observatory - International Agency for Research on Cancer)

Appendix C: Figure 3. Age trends in lung cancers for a USA population:



(Source: Global Cancer Observatory - International Agency for Research on Cancer)