manuscript\_v1

Nicole White, Rex Parsons, David Borg, Adrian Barnett

08/07/2022

# Introduction

Accurate patient risk information is critical for medical decision-making. Patient risk summaries support decision-making by estimating the chances of developing a disease or health condition [ref], or experiencing a future event based on an individual’s current health status [ref]. For instance, early knowledge of an individual’s increased risk of heart disease may prompt preventative measures to delay or reduce the likelihood of onset [ref]. Conversely, identifying patients at low risk of cancer recurrence may help prioritize high-cost interventions to improve prognosis for vulnerable patients [ref].

Routinely collected data, including demographics, medical history, and clinical test results, are often used to infer patient risk. As advances in health technology add to the volume and variety of data available, advances in quantitative methods are also needed to transform these data into reliable and interpretable outputs. Clinical prediction models offer a general framework for risk estimation based on multiple data sources. When successfully validated and implemented, clinical prediction models have the potential to progress the shift from evidence-based to personalized medicine [ref].

======= Despite their value for personalized medicine, clinical prediction models are a growing source of research waste. Systematic reviews of diagnostic and prognostic prediction models have consistently found that most models are poorly reported and unsuitable for use in practice [refs]. The Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD) statement was introduced in 2015 to help researchers better report the development, validation and updating of clinical prediction models [ref]. Related guidelines to assess risk of bias have also been developed [ref]. Whilst the TRIPOD statement has led to modest improvements in conduct and reporting [refs], poor-quality models are still being published [ref]. At the start of the COVID-19 pandemic, a statistician-led systematic review found 20 diagnostic and 10 prognostic models for predicting COVID-19 outcomes [refs]. All models were assessed as being at high risk of bias due to low sample sizes, inappropriate choice of predictors, and inadequate model assessment. By July 2022, the number of published prognostic models had increased to 606, with 381 newly developed models and 225 validation studies on existing models. Updated results showed that 90% of models were unsuitable for clinical use. These findings show that best practices for clinical prediction modelling are often ignored, and errors made during the design stage are major contributors to poor research quality [ref].

Existing reviews of clinical prediction models are likely affected by publication bias, as studies that report favourable results are more likely to be submitted and published [ref]. Publication bias severely hinders progress in health and medical research, as the failure to publish “negative” findings inevitably leads to duplicated efforts by other researchers to address the same research question(s). If publication bias exists in the clinical prediction modelling literature, then many more models will have been planned but never completed or published. Gathering data on planned studies gives an opportunity to examine non-publication and current areas of research focus, including studies aiming to predict the same outcome(s). Sources of bias in proposed models may also be identified early and remedied, to support improvements in research quality.

Study registries are a useful way to evaluate current research before it is published [ref]. Clinicaltrials.gov is an online study registry launched in 2000 to improve access to information on planned, ongoing and completed clinical studies. Since then, over 420,000 studies have been registered from 221 countries (<https://www.clinicaltrials.gov/>; last viewed: 8 July 2022). Previous research has analysed data collected by clinicaltrials.gov to assess reporting completeness and outcomes, including non-publication [refs].

In this paper, we analysed registered studies that included the development and/or validation of a clinical prediction model as part of their planned research. Our analyses had three objectives. First, we summarised trends in diagnostic and prognostic prediction models registered for all diseases and health conditions. Second, we reviewed reporting of planned versus actual sample sizes, as a common source of bias in published clinical prediction models. Third, we estimated the time to publication for all registered studies, by linking clinicaltrials.gov study identifiers to publication records in PubMed [ref].

# Data and Methods

We systematically screened clinicaltrials.gov to identify observational studies that planned to develop a new clinical prediction model or validate an existing model, for any disease or health condition. We focused on observational studies only, given the common use of existing datasets in published studies [ref - placeholder txt; could be worded better]. Studies registered on clinicaltrials.gov since 2000 were downloaded for analysis (data downloaded on 3 March 2022). Study records followed a standardised template and were downloaded in XML format. Data were extracted from compulsory fields, which covered the disease or condition of interest, planned commencement and end dates, study design, outcomes measures and planned sample size (Table 1). The disease or condition of interest was inferred from study indexing to Medical Subject Heading (MeSH) terms, based on terms included in the XML file. Code written to process the XML data is available on GitHub [link here].

A clinical prediction model was defined as any multivariable model designed to estimate the individual-level risk of a diagnostic or prognostic outcome [ref]. Risk estimates could be summarised as a clinical score, probability, or another model-based estimate such as mean survival time. No restrictions were placed on model data sources or choice of predictor variables, provided variables were defined at the individual level. Predictor variables could therefore take the form of patient-level characteristics (e.g., demographics, clinical parameters) or processed features (e.g., extracted from medical images). Studies could propose a clinical prediction model as the primary study objective, or as an objective within a larger study. Studies that planned to identify or validate independent risk factors without mention of developing a subsequent prediction model were excluded.

Relevant studies were identified over two stages. Study records were initially screened using targeted keywords (Table 1). Search strategies used by published reviews of clinical prediction models were used to inform the keyword list [refs]. Keywords could appear in any study field [rex - can you pls confirm the final search used]. Matching studies were then manually screened for eligibility using the Rayyan application (ref). All studies were independently reviewed by at least two study authors. Decision conflicts were resolved by discussion among three authors (NW, RP, DB).

For the final set of included studies, additional data were extracted on study updates over time. Study updates captured changes to anticipated and actual sample sizes, overall study status and details of resulting publications.

* rex: can you pls add a sentence or two on the PubMed matching tool.

# Data and Methods

Adrian: This paragraph is a repeat from the introduction

Clinicaltrials.gov is an online database designed to improve access to information about planned, ongoing and completed clinical studies. The database was launched in 2000 by the National Institute of Health’s National Library of Medicine, in response to United States legislation mandating the registration of funded clinical trials. Since then, over 420,000 studies have been registered, comprising both interventional and observational studies from 221 countries (<https://www.clinicaltrials.gov/>; last viewed: 8 July 2022).

Study registration is completed by a nominated study investigator using a standardised template. Compulsory fields cover the disease or condition of interest, planned commencement and end dates, study type and study design details including interventions (if any), outcomes measures and planned sample size. Additional information on participant recruitment, funding and regulatory oversight is also required. A statistical analysis plan is not required, but can be included as part of an optional detailed summary. Study records can be updated at any time until project completion, including providing details of resulting publications.

For this analysis, we identified and analysed outcomes of registered studies that planned to develop or validate a clinical prediction model (data downloaded on 3 March 2022). A clinical prediction model was defined as any multivariable model designed to estimate the individual-level risk of being diagnosed with a single disease or health condition (diagnostic model) or experiencing a future health-related outcome following diagnosis (prognostic model) (Hemingway et al. (2013)). Future outcomes considered by prognostic models included both pre-defined clinical endpoints, such as treatment response and mortality, and measures of an individual’s health state following diagnosis, such as health-related quality of life.

Adrian: excluded trials because we wanted to focus on prediction models although some trials may include prediction models as a secondary aim

All records classified as observational study designs were included. All other study types were excluded [need to provide rationale here for excluding RCTS]. Dependent variables could be defined as a binary, continuous or time-to-event outcome. No restrictions were placed on model data sources or choice of independent variables, provided variables were defined at the individual level and at least two variables were planned for inclusion in the proposed model. This meant that independent variables could take the form of patient-level characteristics (e.g., demographics, clinical parameters) or processed features (e.g., extracted from images).

Eligible studies proposed a clinical prediction model as the primary study objective or as an objective within a larger study. Studies that planned to identify or validate independent risk factors associated with outcomes without mention of developing a prediction model were excluded.

All study records posted on clinicaltrials.gov until 3 March 2022 were downloaded in XML for analysis (n = xxx). Relevant studies were identified over two stages. In the first stage, we scanned study record fields for keywords reflecting approaches to clinical prediction modelling and their application to diagnosis and prognosis (Table here - search terms + fields searched). Matching studies were then manually screened for eligibility using the web application Rayyan (Ouzzani et al. (2016)). All studies were independently reviewed by at least two study authors.

## Statistical methods

We did not use a formal sample size calculation. Rather, the sample size was determined pragmatically, by the total number of eligible studies. The analysis comprised several steps.

First, trends in diagnostic and prognostic prediction models were summarized descriptively. This included… word cloud of MeSH terms…cumulative number of…the top 10 ranked disease or health condition, according to whether models were being developed, validated, or both.

To address the second aim, we… planned versus actual sample sizes

Third, the time to publication for all registered studies was estimated using a Kaplan–Meier plot to show the publication of results after study completion.

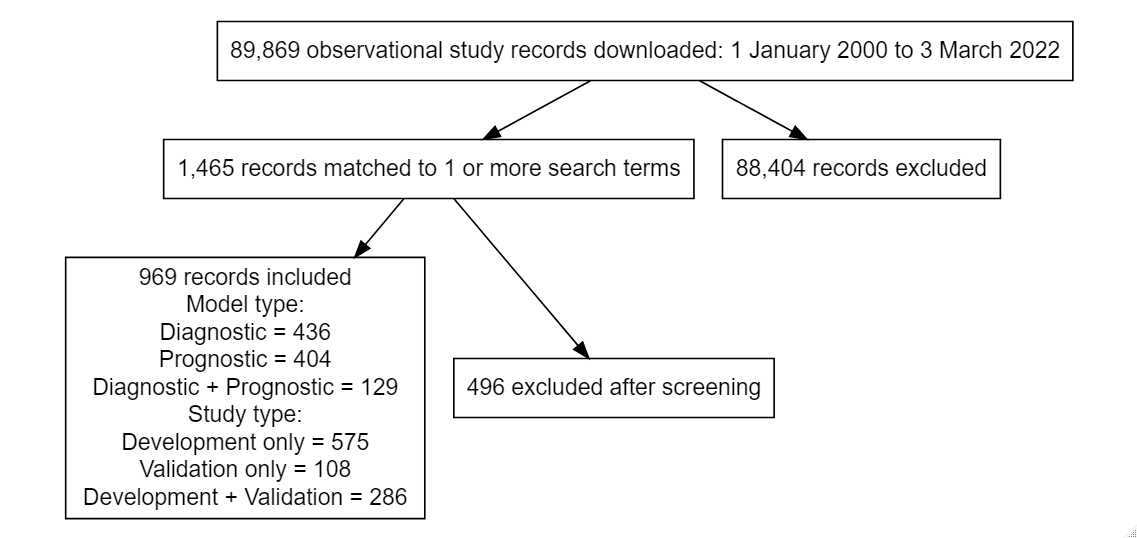
Data management and analysis were performed in R [ref]. The data are codes are available at <https://www.doi.org/xxx>. We report our results using the STROBE checklist for cohort studies [ref].

* Table 1 info: Screening criteria
* Search terms were:“machine learning”,“artificial intelligence”,“deep learning”, “prediction model”,“predictive model”,“prediction score”,“predictive score”, “warning score”,“risk score”,“risk prediction”, “prognostic model”,“diagnostic model” =======
* As table: Search terms were: “machine learning”, “artificial intelligence”, “deep learning”, “prediction model”, “predictive model”, “prediction score”, “predictive score”, “warning score”, “risk score”, “risk prediction”, “prognostic model”, “diagnostic model”

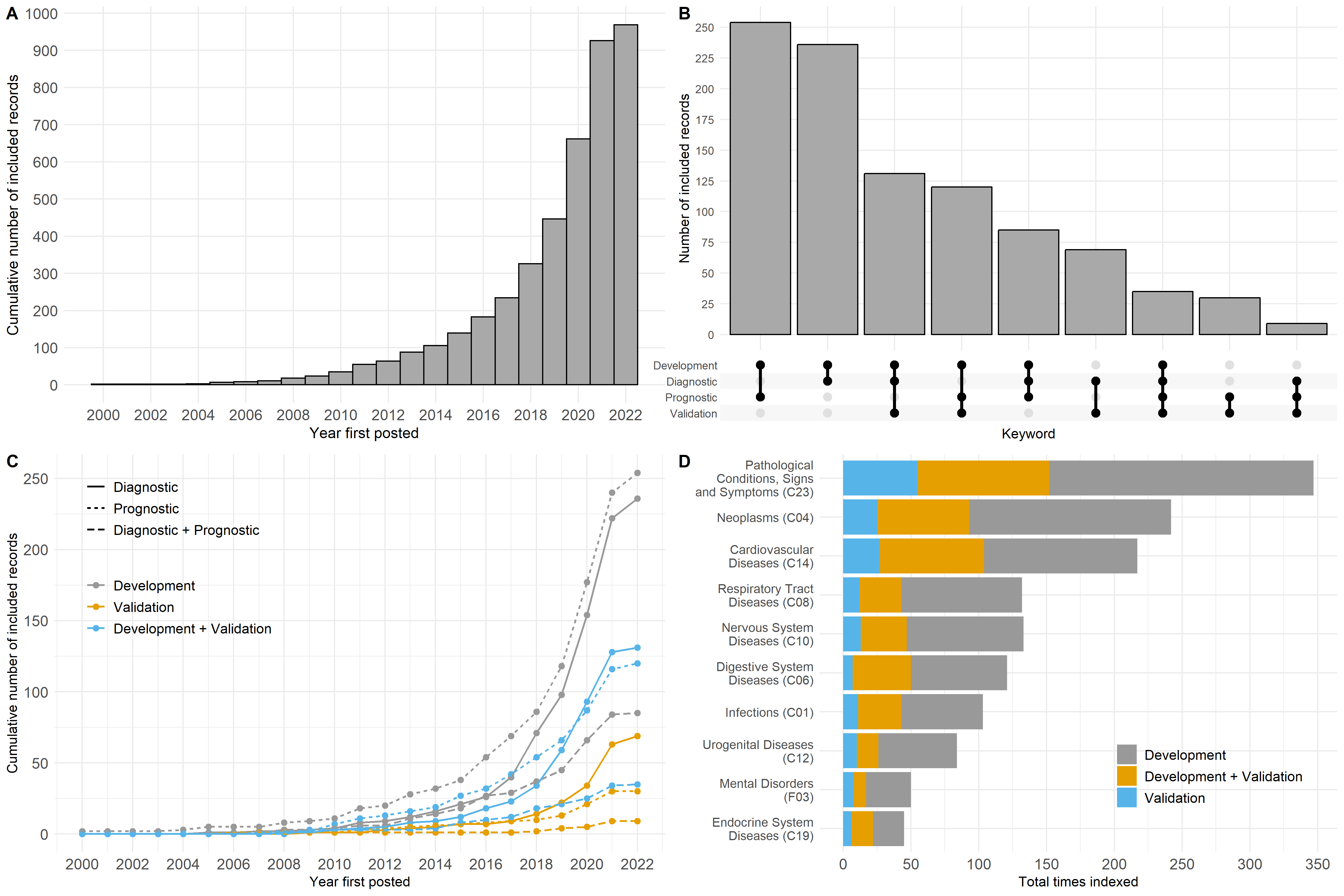
# Results

The first stage of our search returned 1,465 records, from an initial sample size of 89,869 records (Figure 1). The most frequent keyword matches were [to do - data on RDSS…], which changed/did not change over time (Figure 2A).

After screening, 969 studies were included. Most studies planned model development only (n = 575/969; 59%) or model development with mention of internal and/or external model validation (n = 286/969; 30%). Approximately one in ten studies considered the validation of existing prediction models, of which xx were clinical scores or early warning scores [more detail here]. Study types were evenly distributed across diagnostic and prognostic outcomes (Figure 2B). The total number of model development studies increased over time. Large increases in study numbers in later years were attributed to COVID-19 models (check). Or correlated with increases in the mention of deep learning/machine learning/ai.

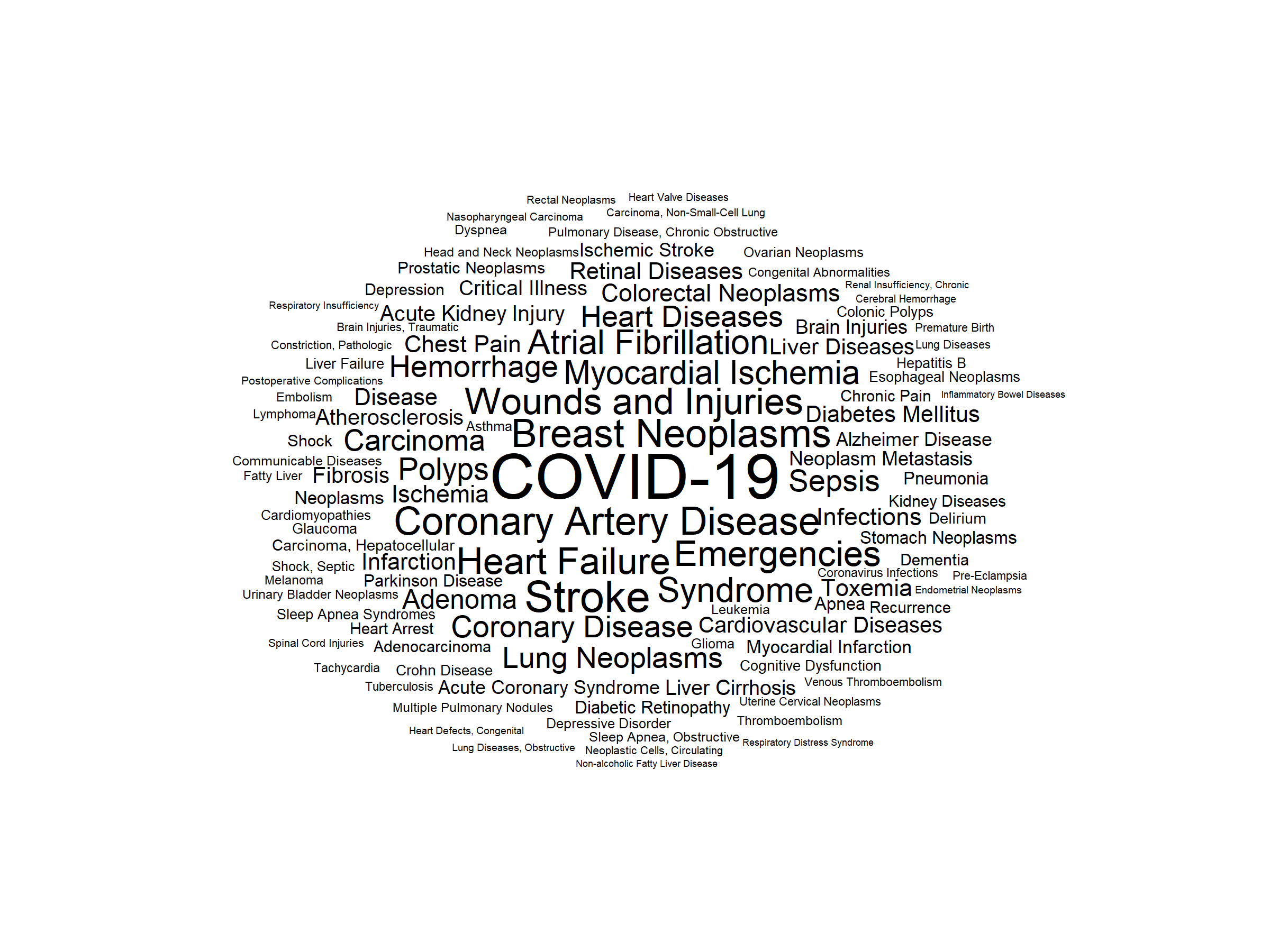


Search strategy



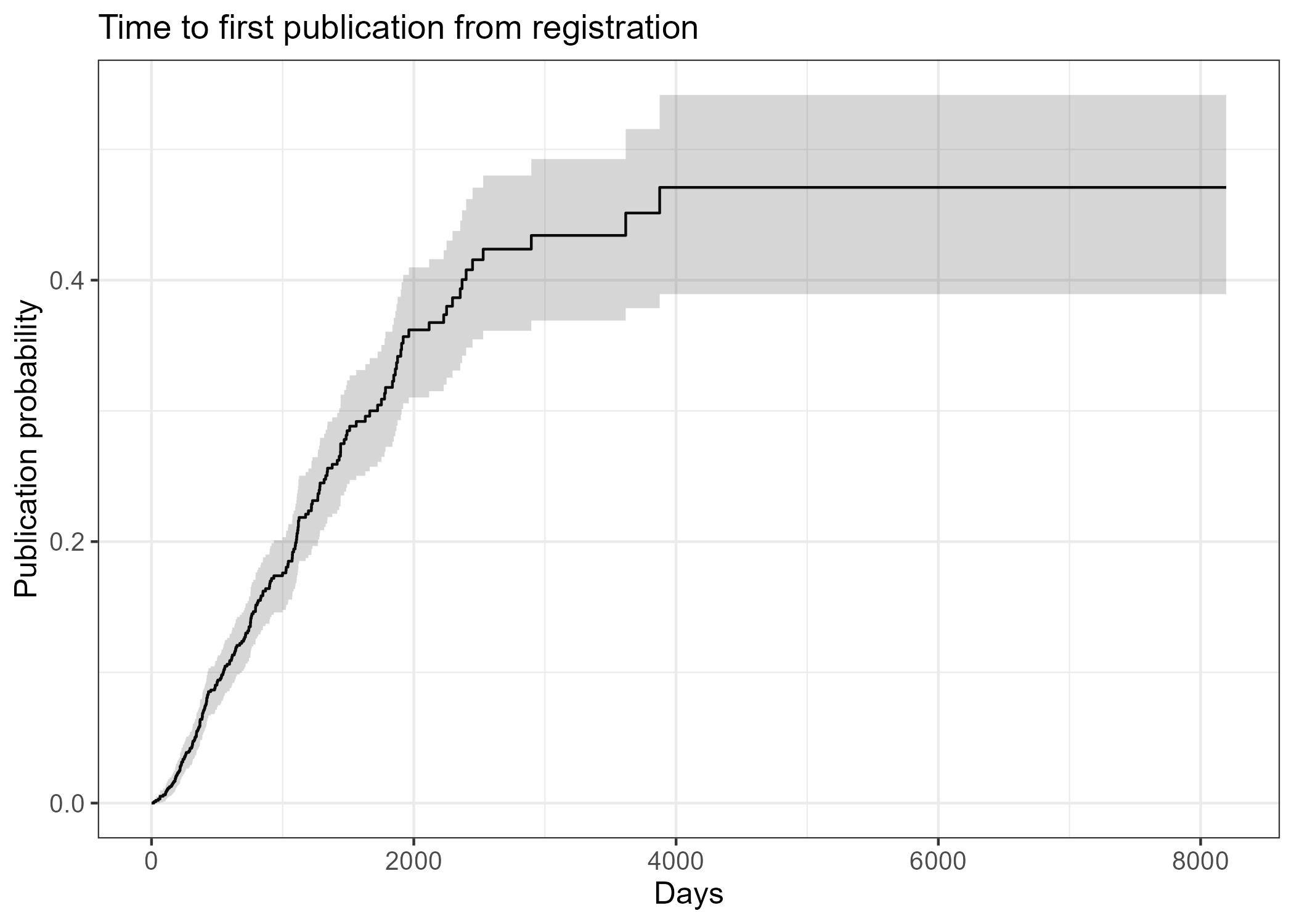
Included study summary

| keyword | n |
| --- | --- |
| prediction model | 248 |
| machine learning | 178 |
| artificial intelligence | 166 |
| predictive model | 151 |
| deep learning | 117 |
| risk score | 110 |
| risk prediction | 89 |
| prognostic model | 64 |
| diagnostic model | 35 |
| predictive score | 30 |
| prediction score | 24 |
| warning score | 4 |



Word cloud of indexed MeSH terms

Change x-axis to years. When does the clock start? Is it time since study completion.



CIF: time to first publication (check caption)

The median follow-up time after trial completion was xx months (25th to 75th centiles xx to xx months). The total follow-up time was xx,xxx years.

## Analysis

Adrian: Not sure if we need the random sample as the non-publication rate will be bad in its own right. We could say it’s relatively better or worse than other designs (e.g., trials), but it would still be a huge problem.

* Study status; all identified studies, possibly compared with random sample of observational studies that don’t return match to search terms?
* Sample size history for final sample
* Frequency of targeted keywords (e.g, (.\*)validation, TRIPOD)
* Links to publications using python tool (Rex to lead)

Need to specify what publications are included. For example, does it include preprints?

Report missing data for sample size.

# Discussion

* Comment on enormous non-publication rate and the consequences for practice. Similar non-publication rate to randomised trials (<https://pubmed.ncbi.nlm.nih.gov/22214755/>)
* need to scope existing published models for a given outcome before developing another one! [refs - riley]
* Before embarking on a prediction modelling study, it is important to ensure a sufficient sample size is available (<https://www.bmj.com/content/368/bmj.m441>) and that proposed predictors/outcome data collection is of sufficient quality. Just as per any planned statistical analysis. Poor choice will likely lead to poor performance down this line (garbage in = garbage out). ref: <https://www.bmj.com/content/338/bmj.b604> (royston2009)

# References

Hemingway, Harry, Peter Croft, Pablo Perel, Jill A Hayden, Keith Abrams, Adam Timmis, Andrew Briggs, et al. 2013. “Prognosis Research Strategy (PROGRESS) 1: A Framework for Researching Clinical Outcomes.” *Bmj* 346.

Ouzzani, Mourad, Hossam Hammady, Zbys Fedorowicz, and Ahmed Elmagarmid. 2016. “Rayyan—a Web and Mobile App for Systematic Reviews.” *Systematic Reviews* 5 (1): 210. <https://doi.org/10.1186/s13643-016-0384-4>.