University of Thessaly

School of Health Sciences

Faculty of Medicine

Methodology of Biomedical Research, Biostatistics and Clinical Bioinformatics

Master Thesis

A meta-analysis support software in Python

Ένα πρόγραμμα υποβοήθησης μετα-αναλύσεων σε γλώσσα προγραμματισμού Python

Three-member advisory committee

Professor Elias Zintzaras

Axel Kowald (supervisor)

Chrisoula Doxani

Zormpas Petros

Larissa, 2022

Table of Contents

Table of Figures

Abbreviations

- RCT – Randomized Control Trial
- NCBI – National Centre for Biotechnology Information
- TI – Title
- JT – Journal Title
- PMID – PubMed Intentifier
- PMC – PubMed Central
- AB – Abstract
- AU – Author
- DP – Date of Publication
- PT – Publication Type

Abstract

Background:

Programming can help us solve problems, automate tasks and process large sets of data. Python is an object-oriented programming language, powerful, fast and easy for the non-programmer to understand and use it. It offers a plethora of built-in tools and third-party modules that are designed by other developers and can be used by anyone to accomplish a specific and complicated task.

Objectives:

In this thesis, I aim to write a program using the python programming language in order to facilitate the process of data accumulating. These data can ultimately be used in order to perform a meta-analysis or a systematic review in any topic the user prefers.

Methods:

Python programming language offers a vast set of ready to use tools, one of those is BioPython. The aforementioned module is specifically designed to aid in bioinformatics as it offers an easy way to link the data in the National Centre for Biotechnology Information (NCBI) libraries with the Python programming language.

Results:

A python program which asks the user for an input and outputs an excel table with all the corresponding randomized controlled trials found in PubMed, sorted by citations number, also providing information on the total participants and other information provided by PubMed.

Conclusion:

Programming can be of valuable help when time-consuming data gathering can be automated. Ultimately this can improve results, accelerate timelines and augment productivity.

Keywords: python, data accumulation, meta-analysis, bioinformatics

Περίληψη

Εισαγωγή:

Ο προγραμματισμός μπορεί να μας βοηθήσει να λύσουμε προβλήματα, να αυτοματοποιήσουμε εργασίες και να επεξεργαστούμε μεγάλα πακέτα δεδομένων. Η Python είναι μια αντικειμενοστραφής γλώσσα προγραμματισμού, ισχυρή, γρήγορη και εύκολη για έναν αρχάριο να την κατανοήσει και να την χρησιμοποιήσει. Προσφέρει μια πληθώρα ενσωματωμένων εργαλείων και προγραμματικές ενότητες σχεδιασμένες απο ανεξάρτητους προγραμματιστές οι οποίες μπορούν να χρησιμοποιηθούν απο τον καθένα με στόχο την διεκπεραίωση εξειδικευμένων και πολύπλοκων διεργασιών.

Στόχοι:

Σε αυτη τη διπλωματική εργασία, στόχος είναι η συγγραφή ενός προγράμματος χρησιμοποιώντας τη γλώσσα προγραμματισμού python με σκοπό τη διευκόλυνση της διαδικασίας συσσώρευσης δεδομένων. Τα δεδομένα αυτά θα μπορούν να χρησιμοποιηθούν σε μετα-αναλύσεις ή συστηματικές ανασκοπήσεις σε οποιοδήποτε θέμα επιθυμεί ο χρήστης.

Μέθοδοι:

Η γλώσσα προγραμματισμού python προσφέρει μια ευρία βιβλιοθήκη έτοιμων εργαλείων, ένα από τα οποία είναι το BioPython. Το προαναφερθέν εργαλείο είναι ειδικά σχεδιασμένο για την υποβοήθηση της βιοπληροφορικής καθώς προσφέρει έναν εύκολο τρόπο γεφύρωσης των δεδομένων στις βιβλιοθήκες του εθνικού κέντρου πληροφοριών βιοτεχνολογίας των ΗΠΑ (NCBI) με τη γλώσσα προγραμματισμού python.

Αποτελέσματα:

Ένα πρόγραμμα python το οποίο δέχεται πληροφορίες από τον χρήστη και εξάγει έναν πίνακα excel με όλες τις αντίστοιχες τυχαιοποιημένες μελέτες μαρτύρων που αντλούνται απο τη βάση δεδομένων PubMed, καταγμένες με βάση τον αριθμό των παραπομπών προσφέροντας επίσης δεδομένα για το σύνολο των συμμετεχόντων και άλλες πληροφορίες που είναι διαθέσιμες απο την βάση PubMed.

Συμπέρασμα:

Ο προγραμματισμός μπορεί να αποτελέσει πολύτιμη βοήθεια όταν χρονοβόρες διαδικασίες συλλογής δεδομένων μπορούν να αυτοματοποιηθούν. Εν κατακλείδι, κατ' αυτόν τον τρόπο μπορούν να βελτιωθούν τα αποτελέσματα, να επιταχυνθούν χρονοδιαγράμματα και να αυξηθεί η παραγωγικότητα.

Λέξεις κλειδιά: γλώσσα προγραμματισμού python, συσσώρευση δεδομένων, μετα-ανάλυση, βιοπληροφορική

Introduction

Python is a programming language which emphasizes code readability and simplicity. It can be used in an object-oriented, functional or procedural way deeming the language suitable to help programmers write clear, logical code for small and larger-scale projects. Created by Guido van Rossum and released in 1991 Python's name originates from Monty Python's flying circus which aired on BBC at the time. Python is mainly used for web development, mathematics and system scripting. Namely, some of its advantages are: the simple syntax which resembles the English language, the fact that it doesn't require compilation and the elasticity in variable definition in contrast with C and Java respectively. It would be fair to state that a non-programmer could read a simple python program and understand its purpose and results. The simplicity of the syntax makes Python an excellent language for beginners to learn programming.

But one could argue that these are not the major advantages of Python. Indeed, one of the uttermost essential things, that comprise the Python language, is its large standard library and the limitless open source third party modules available by the Python Package Index. One of its greatest strengths, as it provides tools for many different tasks. These tools are called modules and can be used for a plethora of things including but not limited to: creating graphical user interfaces, connecting to and interacting with internet databases, generating pseudorandom numbers, manipulating regular expressions, performing data analysis, processing images, performing web scraping and complex scientific computing. Python's package index, the official python third-party software repository, is continuously updated and maintained by a large and active community of independent programmers. Bottomline, probably there is a python module for anything one can think of. In the case that there isn't, one can program one such module and make it available to the public contributing to the ever-growing open-source Python's package index.

Meta-analysis, a term first described by Gene V Glass, is the procedure of analyzing analyses. The term is used until today to describe the statistical analysis of a set of analysis results from several studies in order to group results and draw general conclusions [1]. Frequently, but not axiomatically, meta-analyses analyze the results of randomized control trials. Conclusions from such a statistical analysis are considered to be more precise, than individual study results which take part in the pooled analysis [2]. Synoptically, meta-analyses offer a quantitative summarization of many, if not all, the available research data regarding a specific scientific question. It is safe to say that meta-analyses depend on large sets of data, if not all the data available ideally, in order to produce the best outcomes. The quantity of the data renders the meta-analysis

better and more trustworthy and in fact if the researcher gathers all the available data regarding a specific problem and conducts his meta-analysis there should be no need for another one, until more scientific data are produced. Data accumulation is a time-consuming task which poses many risks. First of all, there is the risk of omitting a number of published studies. As this number grows, the probability of inaccurate outcomes grows with it. Secondly, after ignoring some research trials, one could also fall victim to cherry-picking. Cherry-picking, a selection bias, can be explained in short as picking the studies that produce the outcome of interest, while ignoring those who produce contradicting conclusions. Ideally a researcher must base his meta-analysis on the whole of existing studies published or not.

Methods

In order to bridge Python's functionality and simplicity with data accumulation for meta-analysis conduction, third party modules are necessary. The pillar of this endeavor is the Biopython module. Quoting the Biopython official webpage:

"Biopython is a set of freely available tools for biological computation written in Python by an international team of developers."

Once introduced into our code, the Biopython module lets us communicate and gather data from a number of online bioinformatics destinations such as National Centre for Biotechnology Information (NCBI) BLAST, Entrez, PubMed and ExPASy's, Swiss-Prot etc. [3]

Biopython is a means of communication for the python interpreter with the NCBI's databases. This wouldn't be feasible without Entrez and Entrez Programming Utilities(E-utilities) framework. Entrez is a database system for molecular biology databases produced by the NCBI publicly available via the Internet. The E-utilities are a set of programs which implement an interface of communication with the Entrez system. At the time 38 databases are available, offering a variety of data in the field of bioinformatics such as nucleotides, protein sequences, gene records, three-dimensional structures and the biomedical literature. [4]

In the present thesis, the primary purpose is to create a python tool to search the biomedical literature and provide key details about the results. Those details are the number of citations for articles listed in the PubMed Central database, the total number

of participants in the trials in case the study refers to a randomized control trial (RCT) and basic information available to us by PubMed.

The user is asked to input a search term and a date in years. The program searches the PubMed database for studies corresponding with the search term from the date of input and onwards up to the present day. It is critical to state that the search itself utilizes the capabilities of the PubMed search engine allowing us to specify specific search fields for each word and use logical operators AND, OR to further specialize our search query. The output produced is an excel file with the following as columns: "*Citations, Total Subjects, Title (TI), Journal Title (JT), PubMed Identifier (PMID), PubMed Central Identifier (PMC), Abstract (AB), Date Published (DP), Author (AU), Publication Type (PT)*". These columns may or may not be populated depending on the study characteristics and the data provided to PubMed from the authors. To make the tool accessible to the coding-naïve public a graphical user interface was designed and implemented using the python tkinter module.

Onwards, I will analyze the programmatic characteristics of the program.

Results

Opening the program, you are greeted with the graphical user interface seen below:

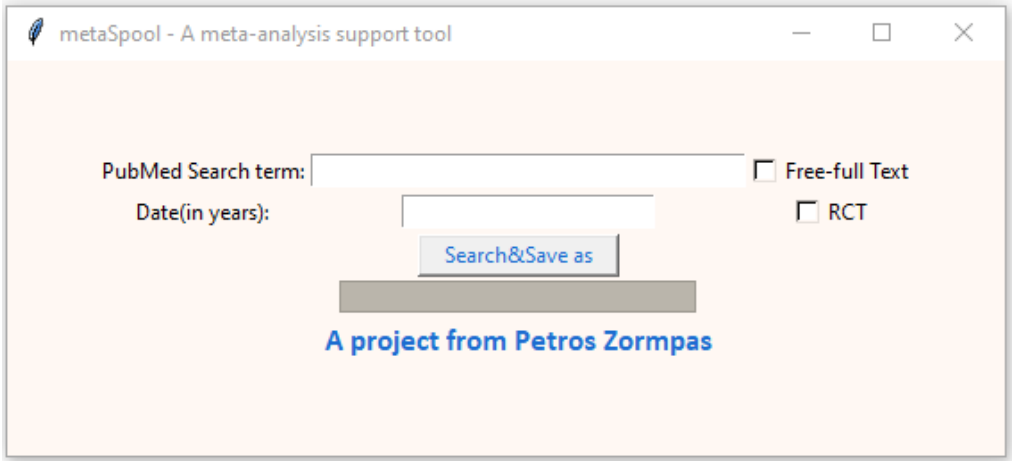
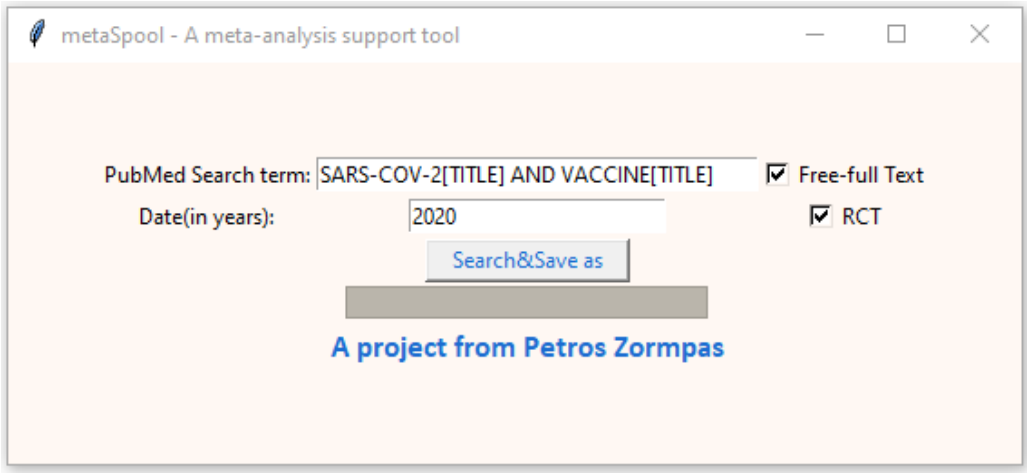


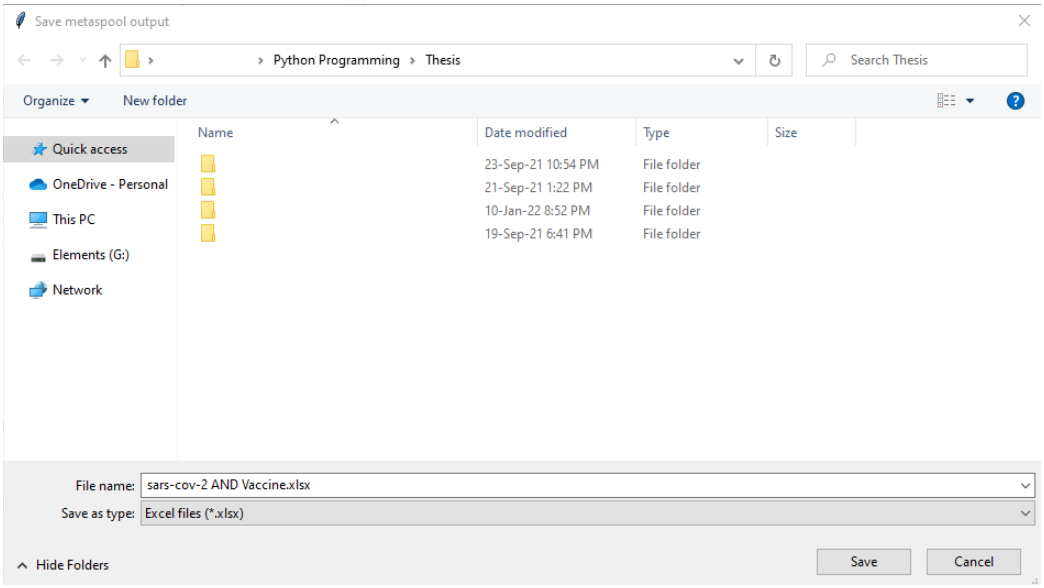
Figure 1, Graphical user interface

There are two fields the user must fill, the PubMed Search term and Date. The check-boxes seen on the right might or might not be checked depending on the user's criteria. If marked, the Free-full Text check box limits our search results to studies whose full text is freely available without the need of a special license or payment. Likewise, the RCT check box, if marked, limits our search results to studies labelled as Randomized Control Trials by PubMed (these contain "Randomized Control Trial" in their PubMed

publication type). After filling the mandatory fields, the user clicks the “Search&Save as” button initiating the download sequence.

Let’s see this with an example:





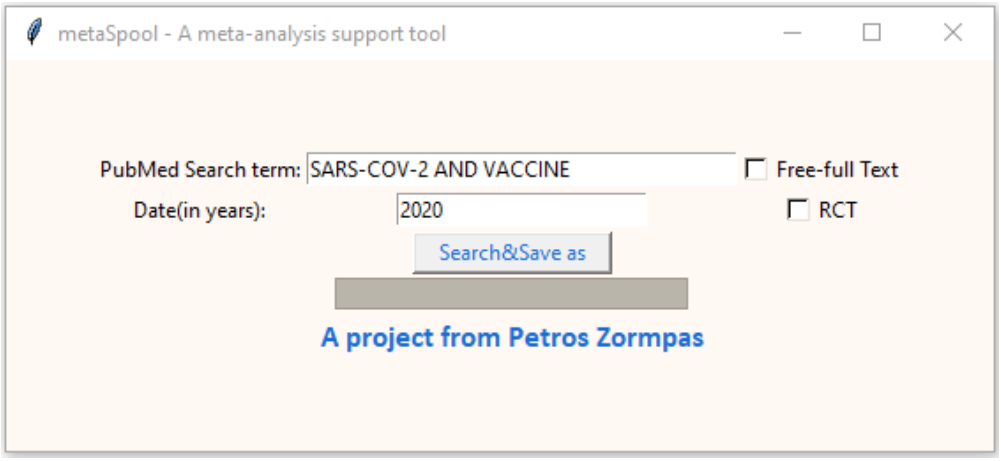
After the search and download is completed, we are introduced with this save file dialog. It offers to save file in the folder where we started the program. The file we save is an excel file with xlsx extension.

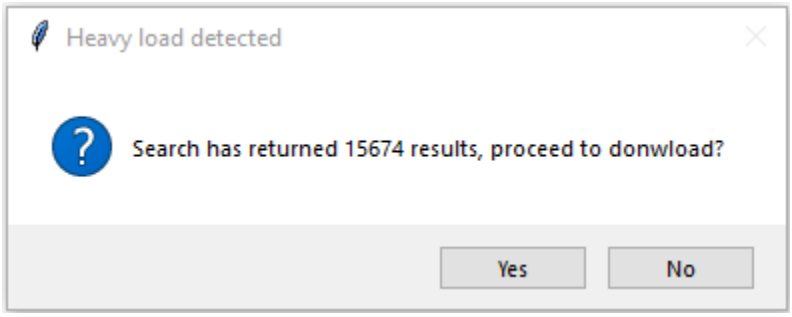
	CITATIONS	Total Subjects	TI	JT	PMID	PMC	AB	DP	AU	PT
2	1989	30420	Efficacy and Safety of the mRNA-1273 SARS-CoV-2 Vaccine.	The New England journal of medicine	33378609	PMC77872	BACKGROUND: Vaccines are ne	2021 Feb 4	['Baden LF	['Clinical Trial, Phase III'
3	1285	23848	Safety and efficacy of the ChAdOx1 nCoV-19 vaccine (AZD1222) against SARS-CoV-2: an interim a	Lancet (London, England)	33306989	PMC77234	BACKGROUND: A safe and effic	2021 Jan 9	['Voysey M	['Journal Article', 'Rando
4	838	1077	Safety and immunogenicity of the ChAdOx1 nCoV-19 vaccine against SARS-CoV-2: a preliminary	Lancet (London, England)	32702298	PMC74454	BACKGROUND: The pandemic c	2020 Aug 1	['Folegatti	['Clinical Trial, Phase I',
5	400	131	Phase 1-2 Trial of a SARS-CoV-2 Recombinant Spike Protein Nanoparticle Vaccine.	The New England journal of medicine	32877576	PMC74942	BACKGROUND: NVX-CoV2373 is	2020 Dec 1	['Keech C'	['Clinical Trial, Phase I',
6	343	743	Safety, tolerability, and immunogenicity of an inactivated SARS-CoV-2 vaccine in healthy adults	The Lancet. Infectious diseases	33217362	PMC78324	BACKGROUND: With the unpre	2021 Feb	['Zhang Y',	['Clinical Trial, Phase I',
7	265	448	Safety and immunogenicity of an inactivated SARS-CoV-2 vaccine, BBIBP-CorV: a randomised, do	The Lancet. Infectious diseases	33069281	PMC75613	BACKGROUND: The ongoing CO	2021 Jan	['Xia S', 'Zh	['Clinical Trial, Phase I',
8	242	320	Effect of an Inactivated Vaccine Against SARS-CoV-2 on Safety and Immunogenicity Outcomes: I	JAMA	32789505	PMC74268	Importance: A vaccine against c	2020 Sep 8	['Xia S', 'Di	['Clinical Trial, Phase I',
9	191	8534	Efficacy of ChAdOx1 nCoV-19 (AZD1222) vaccine against SARS-CoV-2 variant of concern 202012/0	Lancet (London, England)	33798499	PMC80096	BACKGROUND: A new variant o	2021 Apr 1	['Emary KR	['Clinical Trial, Phase II',
10	115	421	Safety, tolerability, and immunogenicity of an inactivated SARS-CoV-2 vaccine (CoronaVac) in he	The Lancet. Infectious diseases	33548194	PMC79066	BACKGROUND: A vaccine again	2021 Jun	['Wu Z', 'H	['Journal Article', 'Rando
11	99	827	Safety and immunogenicity of an inactivated SARS-CoV-2 vaccine, BBV152: a double-blind, rand	The Lancet. Infectious diseases	33485468	PMC78258	BACKGROUND: To mitigate the	2021 May	['Ella R', 'V	['Clinical Trial, Phase I',
12	77	921	Safety and immunogenicity of an inactivated SARS-CoV-2 vaccine, BBV152: interim results from	The Lancet. Infectious diseases	33705727	PMC82217	BACKGROUND: BBV152 is a whc	2021 Jul	['Ella R', 'R	['Clinical Trial, Phase I',
13	67	11303	Efficacy and safety of an inactivated whole-virion SARS-CoV-2 vaccine (CoronaVac): interim resu	Lancet (London, England)	34246358	PMC82663	BACKGROUND: CoronaVac, an i	2021 Jul 1	['Tanriove	['Clinical Trial, Phase III'
14	38	600	A preliminary report of a randomized controlled phase 2 trial of the safety and immunogenicity	Vaccine	33707061	PMC78717	BACKGROUND: Vaccines are ur	2021 May	['Chu L', 'N	['Clinical Trial, Phase II',
15	32	0	Immunogenicity of Ad26.COV2.S vaccine against SARS-CoV-2 variants in humans.	Nature	34107529	PMC8357(	The Ad26.COV2.S vaccine(1-3)	2021 Aug	['Alter G',	['Clinical Trial, Phase I',
16	29	550	Safety, tolerability, and immunogenicity of an inactivated SARS-CoV-2 vaccine (CoronaVac) in he	The Lancet. Infectious diseases	34197764	PMC82384	BACKGROUND: A vaccine again	2021 Dec	['Han B', 'S	['Clinical Trial, Phase I',
17	22		SARS-CoV-2 Neutralization with BNT162b2 Vaccine Dose 3.	The New England journal of medicine	34525276	PMC8461567		2021 Oct 2	['Falsey AI	['Clinical Trial, Phase I',
18	17	441	Safety and immunogenicity of SARS-CoV-2 recombinant protein vaccine formulations in healthy	The Lancet. Infectious diseases	33887209	PMC80552	BACKGROUND: CoV2 preS dTM	2021 Sep	['Goepfert	['Clinical Trial, Phase I',
19	17	20	Safety and immunogenicity of SARS-CoV-2 variant mRNA vaccine boosters in healthy adults: an	Nature medicine	34526698	PMC86047	The emergence of SARS-CoV-2	2021 Nov	['Choi A', '	['Clinical Trial, Phase II',
20	16	30415	Efficacy of the mRNA-1273 SARS-CoV-2 Vaccine at Completion of Blinded Phase.	The New England journal of medicine	34551225	PMC84828	BACKGROUND: At interim analy	2021 Nov 4	['El Sahly H	['Clinical Trial, Phase III'
21	16	3732	Evaluation of mRNA-1273 SARS-CoV-2 Vaccine in Adolescents.	The New England journal of medicine	34379915	PMC83855	BACKGROUND: The incidence c	2021 Dec 9	['Ali K', 'Be	['Clinical Trial, Phase II',
22	14	192	The safety and immunogenicity of an inactivated SARS-CoV-2 vaccine in Chinese adults aged 18-	Vaccine	33875266	PMC80405	BACKGROUND: This study exam	2021 May	['Pu J', 'Yu	['Clinical Trial, Phase I',
23	13	314	Safety and immunogenicity of an MF59-adjuvanted spike glycoprotein-clamp vaccine for SARS-C	The Lancet. Infectious diseases	33887208	PMC80552	BACKGROUND: Given the scale	2021 Oct	['Chappell	['Clinical Trial, Phase I',
24	13	25798	Efficacy, safety, and lot-to-lot immunogenicity of an inactivated SARS-CoV-2 vaccine (BBV152): i	Lancet (London, England)	34774196	PMC85848	BACKGROUND: We report the c	2021 Dec 1	['Ella R', 'R	['Clinical Trial, Phase III'
25	11	245	Safety and immunogenicity of an mRNA-lipid nanoparticle vaccine candidate against SARS-CoV-	Wiener klinische Wochenschrift	34378087	PMC83545	BACKGROUND: We used the RN	2021 Sep	['Kremsne	['Clinical Trial, Phase I',
26	9	104	Safety and immunogenicity of the ChAdOx1 nCoV-19 (AZD1222) vaccine against SARS-CoV-2 in p	The lancet. HIV	34416193	PMC83725	BACKGROUND: People living w	2021 Sep	['Madhi SA	['Clinical Trial, Phase I',
27	8	1288	Different dose regimens of a SARS-CoV-2 recombinant spike protein vaccine (NVX-CoV2373) in v	PLoS medicine	34597298	PMC84861	BACKGROUND: NVX-CoV2373 is	2021 Oct	['Formica	['Clinical Trial, Phase II',
28	4	1620	A phase III, observer-blind, randomized, placebo-controlled study of the efficacy, safety, and im	Vaccine	34620531	PMC84612	BACKGROUND: The WHO decla	2021 Oct 2	['Fadlyana	['Clinical Trial, Phase III'
29	3	153	Efficacy of ChAdOx1 nCoV-19 (AZD1222) vaccine against SARS-CoV-2 lineages circulating in Brazil	Nature communications	34615860	PMC84945	Several COVID-19 vaccines hav	2021 Oct 6	['Clemens	['Clinical Trial, Phase III'
30	1	2263	Immune Profile and Clinical Outcome of Breakthrough Cases After Vaccination With an Inactivat	Frontiers in immunology	34659237	PMC8511(	Constant efforts to prevent infi	2021	['Duarte LI	['Clinical Trial, Phase III'
31	0	0	Recognition of Variants of Concern by Antibodies and T Cells Induced by a SARS-CoV-2 Inactivate	Frontiers in immunology	34858404	PMC86307	Background: Severe acute resp	2021	['Melo-Go	['Clinical Trial, Phase III'
32	0	809	Safety and immunogenicity of inactivated SARS-CoV-2 vaccine in high-risk occupational populat	Infectious diseases of poverty	34933684	PMC8692(	BACKGROUND: Severe acute re	2021 Dec 2	['Feng Y', '	['Journal Article', 'Rando
33	0	4173	Safety and immunogenicity of CpG 1018 and aluminium hydroxide-adjuvanted SARS-CoV-2 S-2P	The Lancet. Respiratory medicine	34655522	PMC85141	BACKGROUND: MVC-COV1901,	2021 Dec	['Hsieh SM	['Clinical Trial, Phase II',

The excel consists of the columns: *Citations, Total Subjects, Title (TI), Journal Title (JT), PubMed Identifier (PMID), PubMed Central Identifier (PMC), Abstract (AB), Date Published (DP), Author (AU), Publication Type (PT).* Citations are portrayed as numbers, although the program harvests the actual citing articles. Total Subjects refer to the total number of patients enrolled in the study if it is a randomized control trial or a trial enrolling patients. Journal Title (JT) is the title of the scientific journal where the publication is made. PubMed identifier (PMID) refers to a unique number for every PubMed article. PubMed Central identifier (PMC) is the similar to PMID but for the PMC library. Abstract (AB) refers to the abstract of the articles as seen on the PubMed library. Date published (DP) is the date of publication on PubMed. Author (AU) is the full list of authors for the article. Last, Publication Type (PT) refers to the type of the study and may be composed of more than one types for example: Clinical Trial, Randomized Controlled Trial, Journal Article etc.

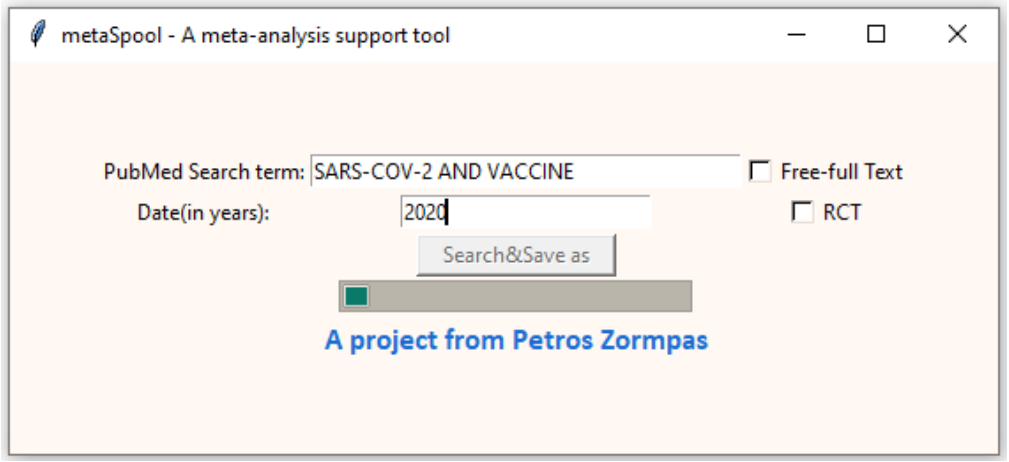
The above search consists of 32 found papers found on PubMed. The usage of the functions of the PubMed search engine is very critical as seen here. Using the [TITLE] field, the “AND” logical operator and checking both the “Free-full Text” and “RCT” check boxes narrows our search results to a bare minimum.

Let’s see what happens if we omit these features.





The program detects a massive volume of papers corresponding with our search query, warning us before initiating the download sequence. Of course, the user could continue with the download. The image below shows the progress bar filing while the download is ongoing.



Here is a flow chart, a schematic description of what this software accomplishes.

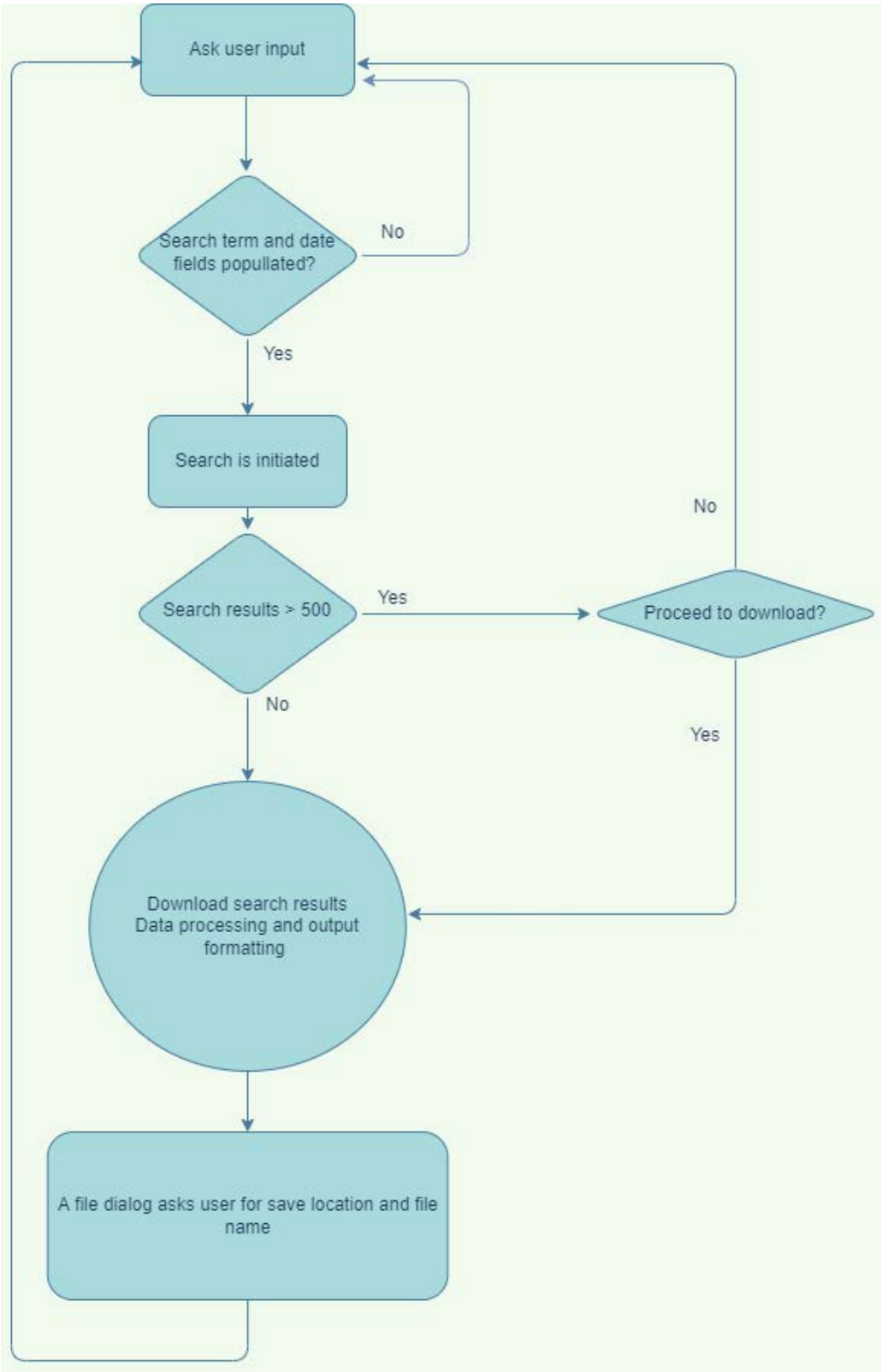


Figure 2, Flowchart

Python Code

Behind the scenes let's take a look at the code making all these possible.

```
import re
import pandas as pd
import os
import threading
from Bio import Entrez, Medline
from tkinter import *
from tkinter import messagebox, ttk
from tkinter.filedialog import asksaveasfile
```

Figure 3, Imported modules

We begin by importing the modules we need.

The first module "re" stands for regex which stands for Regular Expressions. This module enables us to search the main body of the papers, the abstract, in order to find out how many participants are in each randomized control trial. Next, we are importing the pandas module, a way of transforming our data from python lists and dictionaries into data frames which are ultimately exported in excel tables. The os (operating system) module enables python to communicate with the operating system to save files in our file system. The threading module introduces the usage of threads, a way of asynchronously running more than one processes at the same time. The most essential module, as discussed earlier on, the Bio (Biopython) module is then imported and the submodules we need Entrez, Medline are specifically stated there. The tkinter imports are necessary for the graphical user interface.

Out of these modules, Biopython and pandas are third party modules and must be installed separately through the python package index using a package manager. The other modules: re, os, threading, tkinter are a part of python's standard library which come built-in with python. Therefore, in order for the program to run, besides python the user must go on and install Biopython and pandas.

```
Entrez.email = 'pmedcod@gmail.com'
Entrez.tool = 'metaspool, a meta-analysis support tool test developed by Petros Zormpas'
Entrez.api_key =
```

Figure 4, Entrez variables

In order to use the Biopython module we must state our identity and purposes. The "email" and "tool" variables are necessary while the "api_key" variable is optional,

accelerating our search and download speed. The aforementioned key can be obtained through the official PubMed website by creating an account.

```
def pubmed_search(search_term, search_date, rct, freetext):
    """
        pubmed_search
    Returns the search handler information, number of records, WebEnv parameters etc.
    """

    is_rct = ' AND randomized controlled trial[Publication Type]'
    is_freetext = ' AND free full text[Filter]'
    date = (2022 - int(search_date)) * 365
    if rct:
        search_term += is_rct
    if freetext:
        search_term += is_freetext
    search_results = Entrez.read(
        Entrez.esearch(
            db='pubmed', term=search_term, sort='relevance',
            retmax=200, reldate=date,
            datetype='pdat', usehistory='y'
        )
    )
    return search_results
```

Figure 5, Pubmed_search function

The pubmed_search function takes as arguments: search_term, search_date, rct and freetext status. Using the search_term and search_date the user specifies the function returns a handler object, the search results containing all the information required to download and process the results. If the variables rct, freetext are set to true the string variables is_rct, is_freetext are appended to the search_term.

```python
def pubmed_fetch(search_results, batch_size=100):
    """
    Fetches results in batch_size = 100 (def)
    Saves the records in a list
    """

    count = int(search_results['Count'])
    full_records = []
    progress_bar['value']=0
    for start in range(0, count, batch_size):
        end = min(count, start + batch_size)
        print(f'Downloading record {start + 1} to {end}')
        handle = Entrez.efetch(
            db='pubmed', rettype='medline', retmode='text', retstart=start, retmax=batch_size,
            webenv=search_results['WebEnv'], query_key=search_results['QueryKey']
        )
        records = Medline.parse(handle)
        full_records += [record for record in records]
        if count > end:
            progress_bar['value'] += (batch_size/count)*100
        else:
            progress_bar['value'] = 100
        root.update_idletasks()
    return full_records
```

Figure 6, Pubmed_fetch function

The pubmed_fetch function takes as argument the handler object returned by the pubmed_search function. It then downloads the results into a readable format in batches of 100 in order to minimize the server load. Here, a progress bar is used to improve user experience, filling as the records are downloaded. The downloaded records are then stored in a human readable python list variable. This will be referred to as records from now on.

```python
def pubmed_save(records):
    """
    Stores the fetched records data in a pandas DataFrame
    Cleans up the 'unwanted' columns and exports as csv
    """

    df = pd.DataFrame.from_records(records)
    data_cols = ['TI','JT', 'PMID', 'PMC', 'TA', 'AB', 'DP', 'AU', 'PT']

    df_clean = pd.DataFrame(df, columns=data_cols)

    return df_clean
```

Figure 7, Pubmed_save function

The output of the pubmed_fetch function is then passed to the pubmed_save function which utilizes the pandas module to produce a DataFrame, which will be exported as an

excel table archive later on. The most important columns are specified here (also referenced above) in order to keep the output clean and amplify readability.

```python
def ranks(records):
    """
    An approach to find the n of the study using ReGex
    This function has a greedy approach, it will return the biggest n matched  """

    ranks_dict = {}

    for record in records:
        try:
            abstract = record['AB']
            pmid = record['PMID']
        except KeyError:
            continue
        matches = re.findall(
            r'n\s?=\s?\d+|\d+[ ,.]?\d+ individuals|\d+[ ,.]?\d+ subjects|\d+[ ,.]?\d+ healthy|\d+[ ,.]?\d+' \
            r' participants|\d+[ ,.]?\d+ patients|\d+[ ,.]?\d+ cases|\d+[ ,.]?\d+ adults|\d+[ ,.]?\d+' \
            r' controls|enrolling \d+[ ,.]?\d+|\d+[ ,.]?\d+ people|\d+[ ,.]?\d+ randomized|\d+[ ,.]?\d+ analyzed' \
            r' enrolled \d+[ ,.]?\d+|\d+[ ,.]?\d+ volunteers|\d+[ ,.]?\d+ fully vaccinated',
            abstract.lower())
        clean_matches = [int(re.sub(r'[^\d]','',i)) for i in matches]
        clean_matches.sort(reverse=True)

        try:
            ranks_dict[pmid] = clean_matches[0]
        except IndexError:
            ranks_dict[pmid] = 0

    ranks_list = sorted(ranks_dict.items(), reverse=True, key=lambda x: x[1])
    ranks_df = pd.DataFrame(ranks_list, columns=['PMID','Total Subjects'])

    return ranks_df
```

Figure 8, Ranks function

The ranks function takes as argument the records list. Python offers a simplistic approach for searching text strings, the find method. This method can find one or more characters (a substring) and return the position of the first occurrence in the string. While this can be of use for very basic text searching, the complexity of the task in hand surpasses this functions usage. Regular expressions, a mini programming language specifically for text searching, inside python are the only way to yield the outputs we desire. Going in-depth and explaining the actual regex code is far beyond the scope of this thesis. Regular expressions are being utilized here to find the number of participants in the study. The regular expression used can identify any set of numbers which are present after specific characters/words called patterns. The patterns used here are: n=*, n =*, n =*, * individuals, * subjects, * healthy, * participants, * patients, * cases, * adults, * controls, enrolling *, * people, * randomized, * analyzed, enrolled *, * volunteers, * fully vaccinated. The asterisk indicates a number in the form of consecutive digits or digits separated with a single space, comma or dot. For example, the regular expression can match the following: “5,867 patients”, “20 489 cases” and “n=3.895”. After matching a set of numbers these numbers are sorted and the highest number is kept and linked to the record. All the records are being searched one by one

and at the end a possible patient number is assigned to each record. If none of the patterns match the study gets assigned a zero total of patients. The end result is a DataFrame consisting of two columns: “PMID”, “Total Subjects”.

```python
def pubmed_link(records):
    """
    pubmed_link
    Using a list of PMIDs it get's each PMID's citations which are given in !!!PMC format!!!
    The function pubmed_link, it can only find citations inside PubMedCentral library
    Creates a dictionary, keys=PMIDs and values=number of citations
    Returns a pandas dataframe used to store values in csv or excel

    """

    citations_dict = {}
    pmids_list = [i['PMID'] for i in records]
    citations_results = Entrez.read(
        Entrez.elink(
            dbfrom='pubmed', dbto='pmc', id=pmids_list, linkname='pubmed_pmc_refs'
        )
    )
    for i in citations_results:
        pmid = i['IdList'][0]
        try:
            citations = [k['Id'] for k in i['LinkSetDb'][0]['Link']]
            citations_dict[pmid] = len(citations)
        except IndexError:
            citations_dict[pmid] = 0

    # sort by descending value(citation count)
    citations_list = sorted(citations_dict.items(), reverse=True, key=lambda x: x[1])
    citations_df = pd.DataFrame(citations_list, columns=['PMID', 'CITATIONS'])

    return citations_df
```

Figure 9, Pubmed_link function

The pubmed_link function takes as argument the records list. For each study in the list the function retrieves its citations in the PubMed Central Library. Naturally, articles that cite an article in our search results may or may not be published in the PubMed Central Library. If not, the citation won’t be “visible” to our program because of limitations of the Biopython module. Because of these limitations, we must specify that the actual citation number will be equal or greater to the retrieved amount. But relative quantitative relations between studies are not expected to differ.

```python
def download_sequence(search_results):
    records = pubmed_fetch(search_results)
    records_df = pubmed_save(records)
    ranks_df = ranks(records)
    citations_df = pubmed_link(records)
    merged_df = records_df.merge(citations_df, how='left', on='PMID').merge(ranks_df, how='left', on='PMID')
    output_df = merged_df.sort_values(by=['CITATIONS'], ascending=False).set_index('CITATIONS')
    data_cols = ['Total Subjects','TI','JT', 'PMID', 'PMC', 'AB', 'DP', 'AU', 'PT']
    final_output = pd.DataFrame(output_df,columns=data_cols)
    try:
        with asksaveasfile(filetypes=[('Excel files','*.xlsx')],defaultextension='.xlsx',
        initialdir=os.getcwd(),title='Save metaspool output') as file:
            final_output.to_excel(file.name)
    except AttributeError:
        print('Save cancelled')
```

Figure 10, download_sequence function

The download_sequence function initializes the download and processing of the search results based on the user input. Step by step it calls the functions defined before and assigns their outputs into variables. Using the pandas module these variables are merged into a single DataFrame sorted by citation count in a descending manner. Studies with higher citation count are higher in the hierarchy. The file is saved in excel format (*.xlsx) in a folder based on user choice.

```python
def final_output(term, date, rct, freetext):
    search_results = pubmed_search(term, date, rct, freetext)
    if int(search_results['Count'])<500 or messagebox.askyesno('Heavy load detected',
    f'Search has returned {search_results["Count"]} results. Proceed to donwload?',
    default='no',icon='question'):
        search_button['state']='disabled'
        download_sequence(search_results)
        search_button['state']='normal'
    else:
        messagebox.showwarning('Warning','Download cancelled')
```

Figure 11, final_output function

The final_output function calls the initial pubmed_search function passing the arguments specified by the user in the graphical environment application. An important feature of this function is the fact that it asks for permission to start the download when the search results exceed 500 to reduce lengthy downloads.

This is the code necessary for the graphical user environment. Further elaboration on the features outscores the purpose of this thesis.

```
"""GUI impementation"""
BEIGE = "#FFF8F3"
GREEN = "#097969"
BLUE = "#1C6DD0"
root = Tk()
root.title("metaSpool - A meta-analysis support tool")
root.config(padx=50, pady=50, bg=BEIGE)
# Labels
search_label = Label(text="PubMed Search term:", bg=BEIGE, highlightthickness=0)
search_label.grid(row=1, column=0)
date_label = Label(text="Date(in years):", bg=BEIGE, highlightthickness=0)
date_label.grid(row=2, column=0)
# Entries
search_term_entry = Entry(width=40, highlightthickness=0)
search_term_entry.grid(row=1, column=1, columnspan=2)
search_term_entry.focus()
date_entry = Entry(width=23, highlightthickness=0)
date_entry.grid(row=2, column=1, columnspan=2)

# Buttons
search_button = Button(root, text= "Search&Save as",command=lambda:threading.Thread(target=final_output, args=(search_term_entry.get(),date_entry.get(), rct.get(), freetext.get())).start
(), activebackground=GREEN, fg=BLUE,width=15, highlightthickness=0)
search_button.grid(row=3,column=1,pady=1)

pztext = Label(root,text='A project from Petros Zormpas',bg=BEIGE, foreground=BLUE, highlightthickness=0, font='Calibri 13 bold')
pztext.grid(row=5,column=1,pady=1)

# Check boxes
freetext = BooleanVar()
fft_btn = Checkbutton(root,text='Free-full Text',bg=BEIGE, highlightthickness=0, variable=freetext)
fft_btn.grid(row=1,column=3)

rct = BooleanVar()
rct_btn = Checkbutton(root,text='RCT',bg=BEIGE, highlightthickness=0, variable=rct)
rct_btn.grid(row=2,column=3)

#Style
s = ttk.Style()
s.theme_use('clam')
s.configure("blue.Horizontal.TProgressbar",throughcolor=BLUE, background=GREEN)

# Progress Bar
progress_bar = ttk.Progressbar(root, style="blue.Horizontal.TProgressbar", orient='horizontal', mode='determinate',length=200)
progress_bar.grid(row=4,column=1,pady=1)

root.mainloop()
```

Figure 12, Graphical User Interface code

Examples

Below are some examples of usage and a preview of their outputs:

1) Search Term: “Diabetic Nephropathy”[Title] AND “genetic association studies”[All Fields]
   Search Date: 2010
   76 results

CITATIONS	Total Subjects	TI	JT	PMID	PMC	AB	DP	AU
81	0	Genetic associations in diabetic nephropathy: a meta-analysis.	Diabetolo	21127830	PMC3034(	AIMS/HYP	2011 Mar	['Mooyaart AL', 'Valk EJ', 'van Es LA', 'Bruijn JA', 'de Heer E', 'Freedman
21	1767	Association study of genetic variants of 17 diabetes-related genes/loci and cardiovascular risk and diab	Journal of	23298195		BACKGRO	2013 Jun	['Chen G', 'Xu Y', 'Lin Y', 'Lai X', 'Yao J', 'Huang B', 'Chen Z', 'Huang H', 'Fu
20	0	Meta-analysis of diabetic nephropathy associated genetic variants in inflammation and angiogenesis in	BMC med	25280384	PMC44118	BACKGRO	2014 Oct 4	['Nazir N', 'Siddiqui K', 'Al-Qasim S', 'Al-Naqeb D']
18	12472	Association between genetic polymorphism of the angiotensin-converting enzyme and diabetic nephr	Journal of	21810896		INTRODU(	2012 Mar	['Wang F', 'Fang Q', 'Yu N', 'Zhao D', 'Zhang Y', 'Wang J', 'Wang Q', 'Zhou
18	1599	An updated meta-analysis of methylenetetrahydrofolate reductase gene 677C/T polymorphism with di	Diabetes r	22056717		Studies in	2012 Jan	['Niu W', 'Qi Y']
16	455	Examination of association with candidate genes for diabetic nephropathy in a Mexican American popu	Clinical jo	20299368	PMC28792	BACKGRO	2010 Jun	['Kim S', 'Abboud HE', 'Pahl MV', 'Tayek J', 'Snyder S', 'Tamkin J', 'Alcor
16	118	Matrix Gla protein T-138C polymorphism is associated with carotid intima media thickness and predicts	Journal of	28734846		AIMS: We	2017 Oct	['Roumeliotis S', 'Roumeliotis A', 'Panagoutsos S', 'Giannakopoulou E'
15	0	Association of ELMO1 gene polymorphisms with diabetic nephropathy in Chinese population.	Journal of	22842811		BACKGRO	2013 May	['Wu HY', 'Wang Y', 'Chen M', 'Zhang X', 'Wang D', 'Pan Y', 'Li L', 'Liu D', '
13	0	A systematic review and meta-analysis of genetic association studies for the role of inflammation and t	Clinical ki	28616206	PMC5466(	Backgrour	2017 Jun	['Tziastoudi M', 'Stefanidis I', 'Hadjigeorgiou GM', 'Stravodimos K', 'Zin
12	0	Association of interleukin-10 polymorphisms with cytokines in type 2 diabetic nephropathy.	Diabetes t	20809684		OBJECTIVI	2010 Oct	['Kung WJ', 'Lin CC', 'Liu SH', 'Chaung HC']
12	501	Allelic variations in superoxide dismutase-1 (SOD1) gene are associated with increased risk of diabetic	Molecular	21963083		BACKGRO	2011 Dec	['Mohammedi K', 'Maimaitiming S', 'Emery N', 'Bellili-Munoz N', 'Rous
11	1066	SIRT1 rs10823108 and FOXO1 rs17446614 responsible for genetic susceptibility to diabetic nephropathy.	Scientific	28860538	PMC5579(	SIRT1 and	2017 Aug 3	['Zhao Y', 'Wei J', 'Hou X', 'Liu H', 'Guo F', 'Zhou Y', 'Zhang Y', 'Qu Y', 'Gu .
11	464	Genetic variations in key inflammatory cytokines exacerbates the risk of diabetic nephropathy by influ	Gene	29605608		BACKGRO	2018 Jun 3	['Hameed I', 'Masoodi SR', 'Malik PA', 'Mir SA', 'Ghazanfar K', 'Ganai BA
11	1418	IGF2BP2 and IGF2 genetic effects in diabetes and diabetic nephropathy.	Journal of	22770937		OBJECTIVI	2012 Sep-(	['Gu T', 'Horova E', 'Mollsten A', 'Seman NA', 'Falhammar H', 'Prazny M
11	0	Mechanisms of diabetic nephropathy--old buddies and newcomers part 1.	Experime	20658438		Diabetic n	2010 Oct	['Nawroth PP', 'Isermann B']
11	756	Increased levels of circulating (TNF-alpha) is associated with (-308G/A) promoter polymorphism of TNF	Internatio	29042282		The crucia	2018 Feb	['Umapathy D', 'Krishnamoorthy E', 'Mariappanadar V', 'Viswanathan V
11	0	Oxidative stress and its association with TNF-alpha-308 G/C and IL-1alpha-889 C/T gene polymorphisms	Gene	25732517		Diabetic n	2015 May	['Dabhi B', 'Mistry KN']
10	4905	Association between two genetic polymorphisms of the renin-angiotensin-aldosterone system and dia	Molecular	21607620		The widel	2012 Feb	['Ding W', 'Wang F', 'Fang Q', 'Zhang M', 'Chen J', 'Gu Y']
10	868	Association of 2184AG Polymorphism in the RAGE Gene with Diabetic Nephropathy in Chinese Patients	Journal of	26770981	PMC46818	OBJECTIVI	2015	['Cai W', 'Li J', 'Xu JX', 'Liu Y', 'Zhang W', 'Xiao JR', 'Zhu LY', 'Liu JY']
10	0	Association of polymorphisms in the MyD88, IRAK4 and TRAF6 genes and susceptibility to type 2 diabet	Molecular	27062898		Type 2 dia	2016 Jul 5	['Guo C', 'Zhang L', 'Nie L', 'Zhang N', 'Xiao D', 'Ye X', 'Ou M', 'Liu Y', 'Zha
9	312	Rs12976445 Polymorphism is Associated with Risk of Diabetic Nephropathy Through Modulating Expres	Medical sc	26563755	PMC46481	BACKGRO	2015 Nov	['Li C', 'Lei T']
9	101	eNOS 4a/b polymorphism and its interaction with eNOS G894T variants in type 2 diabetes mellitus: moc	Disease m	23594559	PMC38103	To investi	2013	['Rahimi Z', 'Rahimi Z', 'Shahvaisi-Zadeh F', 'Sadeghei S', 'Vessal M', 'Ya
8	485	Association of NOS2 and NOS3 gene polymorphisms with susceptibility to type 2 diabetes mellitus and	IUBMB life	27192959		Inducible	2016 Jul	['Chen F', 'Li YM', 'Yang LQ', 'Zhong CG', 'Zhuang ZX']
8	0	Association of EPHX2 R287Q Polymorphism with Diabetic Nephropathy in Chinese Type 2 Diabetic Patie	Journal of	29629376	PMC58321	The aim o	2018	['Ma L', 'Yan M', 'Kong X', 'Jiang Y', 'Zhao T', 'Zhao H', 'Liu Q', 'Zhang H', '
8	100	Association between end-stage diabetic nephropathy and MTHFR (C677T and A1298C) gene polymorphi	Nephrolo	29227003		AIM: Meth	2019 Feb	['Ramanathan G', 'Harichandana B', 'Kannan S', 'Elumalai R', 'Sfd P']
8	140	Interaction of MTHFR 1298C with ACE D allele augments the risk of diabetic nephropathy in Western Ira	DNA and c	21942443		The aim o	2012 Apr	['Rahimi Z', 'Hasanvand A', 'Felehgari V']
8	3246	Association of paraoxonase gene polymorphisms with diabetic nephropathy and retinopathy.	Molecular	24100645		Emerging	2013 Dec	['Wang J', 'Yang MM', 'Rong SS', 'Ng TK', 'Li YB', 'Liu XM']
8	749	Haplotype association analysis of genes within the WNT signalling pathways in diabetic nephropathy.	BMC neph	23777469	PMC37015	BACKGRO	2013 Jun 1	['Kavanagh DH', 'Savage DA', 'Patterson CC', 'McKnight AJ', 'Crean JK', '
8	0	Is C677T polymorphism in methylenetetrahydrofolate reductase gene a risk factor for diabetic nephrop	Archives c	22209973		BACKGRO	2012 Jan	['Cui WP', 'Du B', 'Jia Y', 'Zhou WH', 'Liu SM', 'Cui YC', 'Ma FZ', 'Luo P', 'M
7	100	Association of tumor necrosis factor (TNF) promoter polymorphisms with plasma TNF-alpha levels and	Journal of	25704106		AIM: The	2015 Apr	['Gupta S', 'Mehndiratta M', 'Kalra S', 'Kalra OP', 'Shukla R', 'Gambhir JK
7	0	Genetic associations in diabetic nephropathy.	Clinical an	24129556		Diabetic n	2014 Apr	['Mooyaart AL']
7	3568	An ACACB variant implicated in diabetic nephropathy associates with body mass index and gene expres	PloS one	23460794	PMC3584(	Acetyl coe	2013	['Ma L', 'Murea M', 'Snipes JA', 'Marinelarena A', 'Kruger J', 'Hicks PJ', 'L
7	0	The genetic map of diabetic nephropathy: evidence from a systematic review and meta-analysis of gen	Clinical ki	33123356	PMC7577	Despite th	2020 Oct	['Tziastoudi M', 'Stefanidis I', 'Zintzaras E']
6	0	Association between genetic polymorphisms of ACE & eNOS and diabetic nephropathy.	Molecular	25227524		Diabetic n	2015 Jan	['Huo P', 'Zhang D', 'Guan X', 'Mei Y', 'Zheng H', 'Feng X']
6	564	Association of CTG repeat polymorphism in carnosine dipeptidase 1 (CNDP1) gene with diabetic nephrc	The Indian	27834323	PMC5116(	BACKGRO	2016 Jul	['Yadav AK', 'Sinha N', 'Kumar V', 'Bhansali A', 'Dutta P', 'Jha V']

2) Search Term: “Covid-19”[Title] AND RCT, Free-full Text
Search Date: 2020
565 results

CITATIONS	Total Subjects	TI	JT	PMID	PMC	AB	DP	AU
3325	2104	Dexamethasone in Hospitalized Patients with Covid-19.	The New England journal of medicine	32678530	PMC73835	BACKGRO	2021 Feb 25	['Horby P', 'Lim WS', 'Emberson JR', 'Mafh
3210	43548	Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine.	The New England journal of medicine	33301246	PMC77451	BACKGRO	2020 Dec 31	['Polack FP', 'Thomas SJ', 'Kitchin N', 'Abs
2456	199	A Trial of Lopinavir-Ritonavir in Adults Hospitalized with Severe Covid-19.	The New England journal of medicine	32187464	PMC71214	BACKGRO	2020 May 7	['Cao B', 'Wang Y', 'Wen D', 'Liu W', 'Wang
2248	1062	Remdesivir for the Treatment of Covid-19 - Final Report.	The New England journal of medicine	32445440	PMC72627	BACKGRO	2020 Nov 5	['Beigel JH', 'Tomashek KM', 'Dodd LE', 'M
1481	237	Remdesivir in adults with severe COVID-19: a randomised, double-blind, placebo-controlled, multicentre trial	Lancet (London, England)	32423584	PMC71903	BACKGRO	2020 May 16	['Wang Y', 'Zhang D', 'Du G', 'Du R', 'Zhao J
815	11330	Repurposed Antiviral Drugs for Covid-19 - Interim WHO Solidarity Trial Results.	The New England journal of medicine	33264556	PMC77273	BACKGRO	2021 Feb 11	['Pan H', 'Peto R', 'Henao-Restrepo AM', '
742	195	Safety and Immunogenicity of Two RNA-Based Covid-19 Vaccine Candidates.	The New England journal of medicine	33053279	PMC7583(	BACKGRO	2020 Dec 17	['Walsh EE', 'Frenck RW Jr', 'Falsey AR', 'K
690	127	Triple combination of interferon beta-1b, lopinavir-ritonavir, and ribavirin in the treatment of patients admitt	Lancet (London, England)	32401715	PMC72115	BACKGRO	2020 May 30	['Hung IF', 'Lung KC', 'Tso EY', 'Liu R', 'Chu
585	200	Effect of Convalescent Plasma Therapy on Time to Clinical Improvement in Patients With Severe and Life-thre	JAMA	32492084	PMC7270(	Importanc	2020 Aug 4	['Li L', 'Zhang W', 'Hu Y', 'Tong X', 'Zheng S
579	105	Patients with Cancer Appear More Vulnerable to SARS-CoV-2: A Multicenter Study during the COVID-19 Outbr	Cancer discovery	32345594	PMC73091	The novel	2020 Jun	['Dai M', 'Liu D', 'Liu M', 'Zhou F', 'Li G', 'Ch
552	0	A Randomized Trial of Hydroxychloroquine as Postexposure Prophylaxis for Covid-19.	The New England journal of medicine	32492293	PMC72892	BACKGRO	2020 Aug 6	['Boulware DR', 'Pullen MF', 'Bangdiwala
525	243	Efficacy of Tocilizumab in Patients Hospitalized with Covid-19.	The New England journal of medicine	33085857	PMC7646(	BACKGRO	2020 Dec 10	['Stone JH', 'Frigault MJ', 'Serling-Boyd N
490	397	Remdesivir for 5 or 10 Days in Patients with Severe Covid-19.	The New England journal of medicine	32459919	PMC7377(	BACKGRO	2020 Nov 5	['Goldman JD', 'Lye DCB', 'Hui DS', 'Marks
447	452	SARS-CoV-2 Neutralizing Antibody LY-CoV555 in Outpatients with Covid-19.	The New England journal of medicine	33113295	PMC7646(	BACKGRO	2021 Jan 21	['Chen P', 'Nirula A', 'Heller B', 'Gottlieb R
444	603	Immunogenicity and safety of a recombinant adenovirus type-5-vectored COVID-19 vaccine in healthy adults a	Lancet (London, England)	32702299	PMC7836(	BACKGRO	2020 Aug 15	['Zhu FC', 'Guan XH', 'Li YH', 'Huang JY', 'Ji
441	667	Hydroxychloroquine with or without Azithromycin in Mild-to-Moderate Covid-19.	The New England journal of medicine	32706953	PMC7397(	BACKGRO	2020 Nov 19	['Cavalcanti AB', 'Zampieri FG', 'Rosa RG',
440	275	REGN-COV2, a Neutralizing Antibody Cocktail, in Outpatients with Covid-19.	The New England journal of medicine	33332778	PMC77811	BACKGRO	2021 Jan 21	['Weinreich DM', 'Sivapalasingam S', 'No
439	21977	Safety and efficacy of an rAd26 and rAd5 vector-based heterologous prime-boost COVID-19 vaccine: an interim	Lancet (London, England)	33545094	PMC7852(	BACKGRO	2021 Feb 20	['Logunov DY', 'Dolzhikova IV', 'Shchebly
417	1561	Effect of Hydroxychloroquine in Hospitalized Patients with Covid-19.	The New England journal of medicine	33031652	PMC7556(	BACKGRO	2020 Nov 19	['Horby P', 'Mafham M', 'Linsell L', 'Bell JL
401	1033	Baricitinib plus Remdesivir for Hospitalized Adults with Covid-19.	The New England journal of medicine	33306283	PMC77451	BACKGRO	2021 Mar 4	['Kalil AC', 'Patterson TF', 'Mehta AK', 'To
393	596	Effect of Remdesivir vs Standard Care on Clinical Status at 11 Days in Patients With Moderate COVID-19: A Ranc	JAMA	32821939	PMC7442(	Importanc	2020 Sep 15	['Spinner CD', 'Gottlieb RL', 'Criner GJ', 'A
386	353	Interleukin-6 Receptor Antagonists in Critically Ill Patients with Covid-19.	The New England journal of medicine	33631065	PMC7953(	BACKGRO	2021 Apr 22	['Gordon AC', 'Mouncey PR', 'Al-Beidh F',
381	193	Safety and Efficacy of Single-Dose Ad26.COV2.S Vaccine against Covid-19.	The New England journal of medicine	33882225	PMC8220(	BACKGRO	2021 Jun 10	['Sadoff J', 'Gray G', 'Vandebosch A', 'Car
377	350	Effect of Dexamethasone on Days Alive and Ventilator-Free in Patients With Moderate or Severe Acute Respir	JAMA	32876695	PMC7489(	Importanc	2020 Oct 6	['Tomazini BM', 'Maia IS', 'Cavalcanti AB',
363	389	Tocilizumab in Patients Hospitalized with Covid-19 Pneumonia.	The New England journal of medicine	33332779	PMC77811	BACKGRO	2021 Jan 7	['Salama C', 'Han J', 'Yau L', 'Reiss WG', 'Kr
358	805	Interim Results of a Phase 1-2a Trial of Ad26.COV2.S Covid-19 Vaccine.	The New England journal of medicine	33440088	PMC7821(	BACKGRO	2021 May 13	['Sadoff J', 'Le Gars M', 'Shukarev G', 'Hee
357	1011	Efficacy of the ChAdOx1 nCoV-19 Covid-19 Vaccine against the B.1.351 Variant.	The New England journal of medicine	33725432	PMC7993(	BACKGRO	2021 May 20	['Madhi SA', 'Baillie V', 'Cutland CL', 'Voy
347	228	A Randomized Trial of Convalescent Plasma in Covid-19 Severe Pneumonia.	The New England journal of medicine	33232588	PMC7722(	BACKGRO	2021 Feb 18	['Simonovich VA', 'Burgos Pratx LD', 'Scib
306	160	Early High-Titer Plasma Therapy to Prevent Severe Covid-19 in Older Adults.	The New England journal of medicine	33406353	PMC7793(	BACKGRO	2021 Feb 18	['Libster R', 'Perez Marc G', 'Wappner D',
293	131	Effect of Tocilizumab vs Usual Care in Adults Hospitalized With COVID-19 and Moderate or Severe Pneumonia:	JAMA internal medicine	33080017	PMC7577(	Importanc	2021 Jan 1	['Hermine O', 'Mariette X', 'Tharaux PL', '
280	464	Convalescent plasma in the management of moderate covid-19 in adults in India: open label phase II multicen	BMJ (Clinical research ed.)	33093056	PMC7578(	OBJECTIV	2020 Oct 22	['Agarwal A', 'Mukherjee A', 'Kumar G', 'C
272	126	Effect of Tocilizumab vs Standard Care on Clinical Worsening in Patients Hospitalized With COVID-19 Pneumor	JAMA internal medicine	33080005	PMC7577(	Importanc	2021 Jan 1	['Salvarani C', 'Dolci G', 'Massari M', 'Merl
257	384	Effect of Hydrocortisone on Mortality and Organ Support in Patients With Severe COVID-19: The REMAP-CAP C	JAMA	32876697	PMC7489(	Importanc	2020 Oct 6	['Angus DC', 'Derde L', 'Al-Beidh F', 'Anna
246	613	Effect of Bamlanivimab as Monotherapy or in Combination With Etesevimab on Viral Load in Patients With Mil	JAMA	33475701	PMC7821(	Importanc	2021 Feb 16	['Gottlieb RL', 'Nirula A', 'Chen P', 'Boscia
226	452	Tocilizumab in Hospitalized Patients with Severe Covid-19 Pneumonia.	The New England journal of medicine	33631066	PMC7953(	BACKGRO	2021 Apr 22	['Rosas IO', 'Brau N', 'Waters M', 'Go RC', '

3) Search Term: “Multiple Sclerosis”[Title] AND fingolimod AND RCT
Search Date: 2010
53 results

CITATIONS	Total Subjects	TI	JT	PMID	PMC	AB	DP	
689	1272	A placebo-controlled trial of oral fingolimod in relapsing multiple sclerosis.	The New I	20089952		BACKGRO	2010 Feb 4	['Kappos L', 'Radue EW', "O'Connor P", 'Polman C', 'Hohlfeld
589	1292	Oral fingolimod or intramuscular interferon for relapsing multiple sclerosis.	The New I	20089954		BACKGRO	2010 Feb 4	['Cohen JA', 'Barkhof F', 'Comi G', 'Hartung HP', 'Khatri BO', '
220	1083	Safety and efficacy of fingolimod in patients with relapsing-remitting multiple sclerosis	The Lance	24685276		BACKGRO	2014 Jun	['Calabresi PA', 'Radue EW', 'Goodin D', 'Jeffery D', 'Rammo
105	970	Oral fingolimod in primary progressive multiple sclerosis (INFORMS): a phase 3, random	Lancet (Lo	26827074		BACKGRO	2016 Mar 1	['Lublin F', 'Miller DH', 'Freedman MS', 'Cree BAC', 'Wolinsk
60	138	Randomized trial of vaccination in fingolimod-treated patients with multiple sclerosis.	Neurology	25636714		OBJECTIVI	2015 Mar 3	['Kappos L', 'Mehling M', 'Arroyo R', 'Izquierdo G', 'Selmaj K'
60	1272	Long-term effects of fingolimod in multiple sclerosis: the randomized FREEDOMS exten	Neurology	25795646	PMC44082	OBJECTIVI	2015 Apr 1	['Kappos L', "O'Connor P", 'Radue EW', 'Polman C', 'Hohlfeld
56	1027	Comparison of fingolimod with interferon beta-1a in relapsing-remitting multiple sclerc	The Lance	21571593		BACKGRO	2011 Jun	['Khatri B', 'Barkhof F', 'Comi G', 'Hartung HP', 'Kappos L', 'M
49	215	Trial of Fingolimod versus Interferon Beta-1a in Pediatric Multiple Sclerosis.	The New I	30207920		BACKGRO	2018 Sep 1	['Chitnis T', 'Arnold DL', 'Banwell B', 'Bruck W', 'Ghezzi A', 'G
47	250	Phase II study of oral fingolimod (FTY720) in multiple sclerosis: 3-year results.	Multiple s	20028707		In a 6-mor	2010 Feb	['Comi G', "O'Connor P", 'Montalban X', 'Antel J', 'Radue EW
46	1272	Relapse and disability outcomes in patients with multiple sclerosis treated with fingolin	The Lance	22494956		BACKGRO	2012 May	['Devonshire V', 'Havrdova E', 'Radue EW', "O'Connor P", 'Zh
46	3635	Correlation between brain volume loss and clinical and MRI outcomes in multiple sclero	Neurology	25632085	PMC43391	OBJECTIVI	2015 Feb 2	['Radue EW', 'Barkhof F', 'Kappos L', 'Sprenger T', 'Haring DA
42	1027	Long-term (up to 4.5 years) treatment with fingolimod in multiple sclerosis: results from	Journal of	26111826	PMC48535	OBJECTIVI	2016 May	['Cohen JA', 'Khatri B', 'Barkhof F', 'Comi G', 'Hartung HP', 'M
38	1272	Impact of fingolimod therapy on magnetic resonance imaging outcomes in patients with	Archives c	22751847		OBJECTIVI	2012 Oct	['Radue EW', "O'Connor P", 'Polman CH', 'Hohlfeld R', 'Calab
31	2615	Ophthalmic evaluations in clinical studies of fingolimod (FTY720) in multiple sclerosis.	Ophthalm	23531349		PURPOSE:	2013 Jul	['Zarbin MA', 'Jampol LM', 'Jager RD', 'Reder AT', 'Francis G',
31	1053	Outcomes of switching directly to oral fingolimod from injectable therapies: Results of t	Multiple s	26265273		BACKGRO	2014 Sep	['Fox E', 'Edwards K', 'Burch G', 'Wynn DR', 'LaGanke C', 'Cray
24	147	A randomized, controlled trial of fingolimod (FTY720) in Japanese patients with multiple	Multiple s	22354739		BACKGRO	2012 Sep	['Saida T', 'Kikuchi S', 'Itoyama Y', 'Hao Q', 'Kurosawa T', 'Nag
20	0	Oral fingolimod (FTY720) in relapsing multiple sclerosis: impact on health-related qualit	Multiple s	21727148		BACKGRO	2011 Nov	['Montalban X', 'Comi G', "O'Connor P", 'Gold S', 'de Vera A'
18	147	Fingolimod (FTY720) therapy in Japanese patients with relapsing multiple sclerosis over	BMC neur	24475777	PMC39115	BACKGRO	2014 Jan 2	['Kira J', 'Itoyama Y', 'Kikuchi S', 'Hao Q', 'Kurosawa T', 'Naga
16	110	Multiple Sclerosis-Secondary Progressive Multi-Arm Randomisation Trial (MS-SMART): a	BMJ open	30166303	PMC61194	INTRODU(	2018 Aug 3	['Connick P', 'De Angelis F', 'Parker RA', 'Plantone D', 'Doshi
15	799	Fingolimod effect on gray matter, thalamus, and white matter in patients with multiple	Neurology	29540589		OBJECTIVI	2018 Apr 1	['Gaetano L', 'Haring DA', 'Radue EW', 'Mueller-Lenke N', 'Th
15	0	Polyphenon E, non-futile at neuroprotection in multiple sclerosis but unpredictably hep	Journal of	26298797	PMC46751	OBJECTIVI	2015 Nov 3	['Lovera J', 'Ramos A', 'Devier D', 'Garrison V', 'Kovner B', 'Re
12	157	Efficacy of fingolimod and interferon beta-1b on cognitive, MRI, and clinical outcomes ir	Journal of	29063244	PMC56882	Cognitive	2017 Dec	['Comi G', 'Patti F', 'Rocca MA', 'Mattioli FC', 'Amato MP', 'Ga
11	1053	Impact of a switch to fingolimod versus staying on glatiramer acetate or beta interferons	BMC neur	25424122	PMC42535	BACKGRO	2014 Nov 2	['Calkwood J', 'Cree B', 'Crayton H', 'Kantor D', 'Steingo B', 'B
11	242	Baseline retinal nerve fiber layer thickness and macular volume quantified by OCT in the	Journal of	24051419	PMC39597	BACKGRO	2013 Dec	['Winges KM', 'Werner JS', 'Harvey DJ', 'Cello KE', 'Durbin M
11	783	Onset of clinical and MRI efficacy occurs early after fingolimod treatment initiation in re	Journal of	26645392	PMC47511	To minimi	2016 Feb	['Kappos L', 'Radue EW', 'Chin P', 'Ritter S', 'Tomic D', 'Lublin
10	0	Relapse rates in patients with multiple sclerosis treated with fingolimod: Subgroup anal	Multiple s	27456887	PMC49855	BACKGRO	2016 Jul	['Derfuss T', 'Ontaneda D', 'Nicholas J', 'Meng X', 'Hawker K'
10	0	Comparative safety and efficacy of ozanimod versus fingolimod for relapsing multiple sc	Journal of	31948278		Aim: Ozar	2020 Mar	['Swallow E', 'Patterson-Lomba O', 'Yin L', 'Mehta R', 'Pelleti
10	0	The influence of patient demographics, disease characteristics and treatment on brain v	Multiple s	24812043		BACKGRO	2014 Nov	['Barkhof F', 'de Jong R', 'Sfikas N', 'de Vera A', 'Francis G', 'C
8	1272	Effect of fingolimod on diffuse brain tissue damage in relapsing-remitting multiple sclei	Multiple s	27237768		BACKGRO	2016 May	['De Stefano N', 'Tomic D', 'Radue EW', 'Sprenger T', 'Meier
7	257	Efficacy of fingolimod in patients with highly active relapsing-remitting multiple scleros	Current m	26121423		OBJECTIVI	2015	['Derfuss T', 'Bergvall NK', 'Sfikas N', 'Tomic DL']
7	0	Five-year results from a phase 2 study of oral fingolimod in relapsing multiple sclerosis.	Multiple s	24293455		We prese	2014 Jun	['Izquierdo G', "O'Connor P", 'Montalban X', 'von Rosenstie
7	272	Fingolimod therapy in early multiple sclerosis: an efficacy analysis of the TRANSFORMS	CNS neuro	24684973	PMC64931	AIMS: The	2014 May	['Agius M', 'Meng X', 'Chin P', 'Grinspan A', 'Hashmonay R']
7	181	Efficacy and safety of fingolimod in Hispanic patients with multiple sclerosis: pooled clir	Advances	25245812		INTRODU(	2014 Oct	['Chinea Martinez AR', 'Correale J', 'Coyle PK', 'Meng X', 'Te
6	0	Consistent control of disease activity with fingolimod versus IFN beta-1a in paediatric-o	Journal of	31467033	PMC69528	BACKGRO	2020 Jan	['Deiva K', 'Huppke P', 'Banwell B', 'Chitnis T', 'Gartner J', 'Kr
6	281	Long-term results from a phase 2 extension study of fingolimod at high and approved dc	Journal of	26338810		Fingolimo	2015 Dec	['Montalban X', 'Comi G', 'Antel J', "O'Connor P", 'de Vera A

4) Search Term: “Rheumatoid arthritis”[Title] AND methotrexate[Title] AND RCT
Search Date: 2000
435 results

CITATIONS	Total Subjects	TI	JT	PMID	PMC	AB	DP	
572	428	Infliximab and methotrexate in the treatment of rheumatoid arthritis. Anti-Tumor Necrosis Fa	The New I	11096166		BACKGRO	2000 Nov	['Lipsky PE', 'van der Heijde DM', 'St Clair EW', 'Fu
393	0	The PREMIER study: A multicenter, randomized, double-blind clinical trial of combination ther	Arthritis a	16385520		OBJECTIVI	2006 Jan	['Breedveld FC', 'Weisman MH', 'Kavanaugh AF', '
362	686	Therapeutic effect of the combination of etanercept and methotrexate compared with each tr	Lancet (Lo	15001324		BACKGRO	2004 Feb 2	['Klareskog L', 'van der Heijde D', 'de Jager JP', 'Go
359	271	Adalimumab, a fully human anti-tumor necrosis factor alpha monoclonal antibody, for the trea	Arthritis a	12528101		OBJECTIVI	2003 Jan	['Weinblatt ME', 'Keystone EC', 'Furst DE', 'Morela
325	632	A comparison of etanercept and methotrexate in patients with early rheumatoid arthritis.	The New I	11096165		BACKGRO	2000 Nov	['Bathon JM', 'Martin RW', 'Fleischmann RM', 'Tes
285	619	Radiographic, clinical, and functional outcomes of treatment with adalimumab (a human anti-t	Arthritis a	15146409		OBJECTIVI	2004 May	['Keystone EC', 'Kavanaugh AF', 'Sharp JT', 'Tanner
224	0	Combination of infliximab and methotrexate therapy for early rheumatoid arthritis: a randomi	Arthritis a	15529377		OBJECTIVI	2004 Nov	['St Clair EW', 'van der Heijde DM', 'Smolen JS', 'M
219	673	Comparison of tocilizumab monotherapy versus methotrexate monotherapy in patients with n	Annals of	19297346	PMC37475	BACKGRO	2010 Jan	['Jones G', 'Sebba A', 'Gu J', 'Lowenstein MB', 'Calv
217	465	The efficacy and safety of rituximab in patients with active rheumatoid arthritis despite metho	Arthritis a	16649186		OBJECTIVI	2006 May	['Emery P', 'Fleischmann R', 'Filipowicz-Sosnowsk
211	958	Tofacitinib versus methotrexate in rheumatoid arthritis.	The New I	24941177		BACKGRO	2014 Jun 1	['Lee EB', 'Fleischmann R', 'Hall S', 'Wilkinson B', 'E
202	399	Tofacitinib (CP-690,550) in combination with methotrexate in patients with active rheumatoid	Lancet (Lo	23294500		BACKGRO	2013 Feb 9	['Burmester GR', 'Blanco R', 'Charles-Schoeman C'
196	359	Double-blind randomized controlled clinical trial of the interleukin-6 receptor antagonist, tocil	Arthritis a	16947782		OBJECTIVI	2006 Sep	['Maini RN', 'Taylor PC', 'Szechinski J', 'Pavelka K',
179	274	Comparison of methotrexate monotherapy with a combination of methotrexate and etanercep	Lancet (Lo	18635256		BACKGRO	2008 Aug 2	['Emery P', 'Breedveld FC', 'Hall S', 'Durez P', 'Char
178	0	Tofacitinib (CP-690,550) in patients with rheumatoid arthritis receiving methotrexate: twelve-r	Arthritis a	23348607		OBJECTIVI	2013 Mar	['van der Heijde D', 'Tanaka Y', 'Fleischmann R', 'K
174	652	Effects of abatacept in patients with methotrexate-resistant active rheumatoid arthritis: a ranc	Annals of	16785475		BACKGRO	2006 Jun 2	['Kremer JM', 'Genant HK', 'Moreland LW', 'Russel
155	982	Certolizumab pegol plus methotrexate is significantly more effective than placebo plus metho	Arthritis a	18975346		OBJECTIVI	2008 Nov	['Keystone E', 'Heijde Dv', 'Mason D Jr', 'Landewe
150	133	Golimumab, a human antibody to tumour necrosis factor {alpha} given by monthly subcutaneo	Annals of	19066176	PMC26745	OBJECTIVI	2009 Jun	['Keystone EC', 'Genovese MC', 'Klareskog L', 'Hsia
146	1196	Tocilizumab inhibits structural joint damage in rheumatoid arthritis patients with inadequate r	Arthritis a	21360490		OBJECTIVI	2011 Mar	['Kremer JM', 'Blanco R', 'Brzosko M', 'Burgos-Varg
146	304	A randomised, double-blind, parallel-group study to demonstrate equivalence in efficacy and	Annals of	23687260	PMC37866	OBJECTIVI	2013 Oct	['Yoo DH', 'Hrycaj P', 'Miranda P', 'Ramiterre E', 'Pi
139	165	Efficacy and safety of abatacept or infliximab vs placebo in ATTEST: a phase III, multi-centre, ra	Annals of	18055472	PMC25648	OBJECTIVI	2008 Aug	['Schiff M', 'Keiserman M', 'Codding C', 'Songcharc
136	632	Etanercept versus methotrexate in patients with early rheumatoid arthritis: two-year radiograp	Arthritis a	12115173		OBJECTIVI	2002 Jun	['Genovese MC', 'Bathon JM', 'Martin RW', 'Fleisch
136	619	Efficacy and safety of certolizumab pegol plus methotrexate in active rheumatoid arthritis: the	Annals of	19015207	PMC26745	BACKGRO	2009 Jun	['Smolen J', 'Landewe RB', 'Mease P', 'Brzezicki J',
121	686	Comparison of etanercept and methotrexate, alone and combined, in the treatment of rheuma	Arthritis a	16572441		OBJECTIVI	2006 Apr	['van der Heijde D', 'Klareskog L', 'Rodriguez-Valv
116	0	A randomized comparative effectiveness study of oral triple therapy versus etanercept plus m	Arthritis a	22508468	PMC40361	OBJECTIVI	2012 Sep	['Moreland LW', "O'Dell JR", 'Paulus HE', 'Curtis JR
114	507	A phase IIb dose-ranging study of the oral JAK inhibitor tofacitinib (CP-690,550) versus placebo	Arthritis a	22006202		OBJECTIVI	2012 Apr	['Kremer JM', 'Cohen S', 'Wilkinson BE', 'Connell G
114	299	Intensive treatment with methotrexate in early rheumatoid arthritis: aiming for remission. Co	Annals of	17519278	PMC2111E	BACKGRO	2007 Nov	['Verstappen SM', 'Jacobs JW', 'van der Veen MJ',
110	419	Treatment of rheumatoid arthritis with anakinra, a recombinant human interleukin-1 receptor	Arthritis a	11920396		OBJECTIVI	2002 Mar	['Cohen S', 'Hurd E', 'Cush J', 'Schiff M', 'Weinblatt
110	14	Adalimumab with or without methotrexate in juvenile rheumatoid arthritis.	The New I	18716298		BACKGRO	2008 Aug 2	['Lovell DJ', 'Ruperto N', 'Goodman S', 'Reiff A', 'Ju
105	140	Phase II study of tofacitinib (CP-690,550) combined with methotrexate in patients with rheuma	Arthritis c	21584942		OBJECTIVI	2011 Aug	['Tanaka Y', 'Suzuki M', 'Nakamura H', 'Toyoizumi S
105	0	Predictors of joint damage in patients with early rheumatoid arthritis treated with high-dose n	Arthritis a	16508926		OBJECTIVI	2006 Mar	['Smolen JS', 'Van Der Heijde DM', 'St Clair EW', 'E
104	0	Very early treatment with infliximab in addition to methotrexate in early, poor-prognosis rheu	Arthritis a	15641102		OBJECTIVI	2005 Jan	['Quinn MA', 'Conaghan PG', "O'Connor PJ", 'Karin
102	0	Evidence of radiographic benefit of treatment with infliximab plus methotrexate in rheumatoi	Arthritis a	15818697		OBJECTIVI	2005 Apr	['Smolen JS', 'Han C', 'Bala M', 'Maini RN', 'Kalden
101	637	Golimumab, a human anti-tumor necrosis factor alpha monoclonal antibody, injected subcutar	Arthritis a	19644849		OBJECTIVI	2009 Aug	['Emery P', 'Fleischmann RM', 'Moreland LW', 'Hsi
100	201	Maintenance, reduction, or withdrawal of etanercept after treatment with etanercept and me	Lancet (Lo	23332236		BACKGRO	2013 Mar 2	['Smolen JS', 'Nash P', 'Durez P', 'Hall S', 'Ilivanova
99	259	Sustained improvement over two years in physical function, structural damage, and signs and s	Arthritis a	15077287		OBJECTIVI	2004 Apr	['Maini RN', 'Breedveld FC', 'Kalden JR', 'Smolen J

Conclusion

This python program offers simplicity, speed and organization. It caters to the needs of a researcher tasked with accumulating data for a meta-analysis or a systematic review but can also be used as a tool by everyone to effortlessly browse the biomedical literature and obtain high value results. There is no doubt that a search in the browser by standard means can prove to be superior in terms of result accuracy especially in the derived measures (citations and total participants number), but when time is of essence automation can prove to be preferable.

Improvements can be made in terms of search customization and result processing. As an example, this tool could be expanded to browse other NCBI's libraries as well and provide data for genes and proteins which could be even more robust owing to the fact of the standardization of values in contrast with the human-made randomness of the medical literature.

References

[1] Glass, G. V. (1976). Primary, Secondary, and Meta-Analysis of Research. *Educational Researcher*, 3-8. doi:10.3102/0013189X005010003

[2] Haidich, A. B. (2010, Dec). Meta-analysis in medical research. *Hippokratia, 14(Suppl 1),* 29-37. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3049418/

[3] Cock, P. J., Antao, T., Chang, J. T., Chapman, B. A., Cox, C. J., Dalke, A., de Hoon, M. J. (2009, June). Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics, 25*(11), 1422-1423. doi:10.1093/bioinformatics/btp163

[4] Sayers, E. (2010). *Entrez Programming Utilities Help.* Retrieved from National Center for Biotechnology Information (US): https://www.ncbi.nlm.nih.gov/books/NBK25501/