Title: Enhancing the quality and reproducibility of research: how to work effectively with medical and data librarians

Running Title: Working with librarians to enhance research

Kaitlin Fender Throgmorton, MLIS [1]
Natalia Festa, MD, MBA [2]
Michelle Doering, MLS [3]
Christopher R. Carpenter, MD, MSc, AGSF [4]
Thomas M. Gill, MD [2]

1. Harvey Cushing/John Hay Whitney Medical Library, Yale University, New Haven, Connecticut
2. Department of Internal Medicine, Yale School of Medicine, New Haven, Connecticut
3. Bernard Becker Medical Library, Washington University School of Medicine in St. Louis, St. Louis, Missouri
4. Mayo Clinic Department of Emergency Medicine, Rochester, Minnesota

Disclosure: Some information in this paper was first presented during a Clin-STAR Webinar on February 3, 2023.

Funding:

- Clin-STAR Coordinating Center, Grant/Award Number: U24AG065204

23 • National Institute on Aging (NIA), Training Grant (T32), Grant/Award Number:
24 T32AG019134
25 • Yale Claude D. Pepper Older Americans Independence Center, Grant/Award Number:
26 P30AG021342

27

28 Address correspondence to:
29 Kaitlin Fender Throgmorton (kaitlin.throgmorton@yale.edu)
30 Cushing/Whitney Medical Library
31 Yale University
32 P.O. Box 208014
33 New Haven, CT 06520-8014

34

35 Word count, manuscript: 1,500
36 Table(s): 1; Figure(s): 1
37 References: 20
38 Keywords: Medical Librarians, Team Science, Reproducibility, Research Quality

Key points:

1. To enhance rigor and reproducibility, clinician-investigators should engage librarians early in their research to determine whether further collaboration would be beneficial. In extended collaborations, clinician-investigators and librarians should establish roles, templates, and timelines to work together most effectively and beneficially.
2. Librarian co-authorship in research, especially on systematic reviews, may contribute to higher quality information retrieval and lower risk of bias.
3. Some librarians now specialize in data acquisition, processing, management, and analysis. These highly trained individuals, often called data librarians, may augment clinician-investigators' success in data-intensive research projects.

Why does this paper matter?

Collaborating with librarians can substantially enhance the quality of clinician-investigator research, and many clinician-investigators can access a wide array of information services, including assistance with systematic reviews and data-intensive projects, through libraries affiliated with their institution. We share information about these two types of services as well as guidance for clinician-investigators on how to work effectively with medical and data librarians.

39

Introduction

Aging research increasingly utilizes information sources that vary in quality, user-friendliness, and accessibility. As information available to aging researchers increases in volume and complexity (**Figure 1**), partnerships with librarians are essential to the informed use and interpretation of datasets and the medical literature. As experts in the acquisition, assessment, and synthesis of existing information, librarians contribute to the quality [1, 2] and reproducibility [3] of research projects.

During the February 2023 Clinician-Scientists Transdisciplinary Aging Research (Clin-STAR) Webinar,[4] two clinician-investigator and librarian dyads discussed their collaborative experiences and highlighted opportunities to incorporate librarians into aging research projects. In this commentary, we discuss the role and expertise of medical and data librarians and how clinician-investigators can facilitate an effective longitudinal research collaboration by incorporating librarians into team-science projects. While titles and job responsibilities vary in librarianship, we define medical librarians as professionals providing information services, such as literature searching, within medical contexts. We define data librarians as professionals who offer specialized information services, such as identifying and assessing data sources for research objectives. We also provide specific examples about how the librarians from each dyad, a medical librarian and a data librarian, have contributed to the research of their partnering clinician-investigator.

Cross-cutting Themes

While clinician-investigators should be familiar with evidence synthesis, data management, and statistical fundamentals,[5] many clinical projects would benefit from the expertise of librarians, especially in effectively navigating the literature and data sources.[3,6] We describe key considerations for identifying and collaborating with librarians on research projects.

Assembling a Research Team

Understanding the unique skills and training of librarians

Librarians often hold master's degrees in library and information science. Although their training and experience can vary considerably, many possess specialized skills, with expertise in evidence synthesis, data science, bioinformatics, geospatial information systems (GIS), web design, application development, and more. Institutional expectations for librarian services can also differ with focal points on consultation, instruction, resource acquisition, student support, research, and a variety of custom services.

Identifying librarian collaborators

Many medical organizations, including academic institutions and hospitals, have libraries onsite. Librarian allocation varies widely; a single librarian may support an entire campus, or a group of librarians may support a single department. While the number of librarians and available services differs across institutions, many libraries participate in and contribute to the research life cycle, managing and synthesizing necessary evidence. Clinician-investigators should seek to understand the predominant working models and range of library services available at their institution.

If librarians are available at your current institution, select the librarian who is the best fit for your project or discipline and meet with them to learn more about how you might work together. If you are uncertain or the librarian isn't the right fit, you should contact the library for assistance.

Building Effective Clinician-Investigator—Librarian Collaborations

Establish partnerships that can accommodate librarians' competing priorities

To support the research of a broad user base, librarians have many competing priorities. Therefore, clinician-investigators should clearly communicate expectations regarding the objectives, approach, and deadlines for a project. To make the best use of librarian time, clinician-investigators should provide well-formulated research questions—allowing the librarian to fully engage with the project, identify useful resources and strategies, and determine whether additional consultations, or an ongoing collaboration, should be pursued.

When working with librarians, clinician-investigators should take a team-science approach and invite librarians to serve as co-authors on manuscripts stemming from their collaborations. At some institutions, this may be an expectation or requirement for the librarian. At other institutions, librarians may have publication and presentation requirements for promotion. Regardless of institutional requirements, many librarians are experienced researchers who are passionate about knowledge production and preservation. Collegial relationships between clinician-investigators and librarians can be intellectually rich and mutually beneficial, as illustrated in the applied examples below.

Identify how to best incorporate librarian expertise at each stage of the project life cycle

Although clinician-investigators and librarians both conduct research, each will have unique areas of expertise, specialized skills, and particular vocabularies associated with their work. Librarians are unlikely to have deep clinical expertise, and clinician-investigators may not be information experts. Translating across this divide may take some time and effort. To do this successfully, we recommend the following best practices:

- Communicate early and often, accommodating time for clarifying questions
- Provide examples of prior studies that can serve as potential templates for the current project
- Be teachable and consider the entire process a learning experience
- Start small and build to larger and more complex projects over time

Leveraging Librarian Expertise: Two Applied Examples

To demonstrate the range of librarian contributions to clinician-investigator research, we present two applied examples – one from each clinician-investigator and librarian dyad. The first describes how a medical librarian co-authored systematic reviews with a clinician-investigator, and the second discusses how a data librarian processed open government data for statistical analyses by the clinician-investigator, leading to several co-authored original research reports.

Applied Example: Evidence Synthesis with a Medical Librarian (Dyad 1)

High quality systematic reviews and scoping reviews[7] require a thorough and reproducible evidentiary search. The first dyad developed an *Academic Emergency Medicine* series, Evidence Based Diagnostics, which provided a meta-analysis of history, physical exam, labs, and imaging

pertinent for a common emergency department diagnosis or syndrome such as dementia [8] or falls.[9] The medical librarian engaged in this series from the outset and developed a refined diagnostic research search filter that improved traditional PubMed Clinical Queries.[10] Following the Enhancing the Quality and Transparency of health Research (EQUATOR) Network's Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies (PRISMA-DTA) requirements,[11] the medical librarian co-authored diagnostic systematic reviews for the series that incorporated geriatric components, using a wide range of subject headings and synonyms for older adults that included aged, frail elderly, geriatric, elder, nonagenarian, octogenarian, or centenarian, among many others.[8,9] Table 1 highlights medical librarian involvement in evidence synthesis.

As part of a larger team working across four institutions, the first dyad also collaborated on scoping reviews for the Geriatric Emergency Care Applied Research 2.0 – Advancing Dementia Care (GEAR 2.0-ADC) Network. These scoping reviews identified stakeholder priorities for emergency department research involving persons living with dementia, including detection,[12] communication/decision-making,[13] best practices,[14] and care transitions.[15] The GEAR 2.0-ADC team selected seven databases to ensure discovery of on-topic medical and psychological literature and devised a methodology for search engine strategy peer-review based on existing recommendations. In addition to designing and reproducibly reporting geriatric-centric peer-reviewed search strategies for each GEAR 2.0-ADC topic, the medical librarians collaborated with clinician-investigators to efficiently utilize results in a multi-institutional citation management platform and update search results prior to the pre-planned consensus conference.

Applied Example: Data Acquisition and Processing with a Data Librarian (Dyad 2)

Prior research suggests that many investigators are unaware that some librarians may be able to help them identify and effectively use unique data sources.[16] Data librarians' expertise[17] can be instrumental to effectively preparing data for advanced analysis and reporting on data sourcing and methodology.

In the second dyad, the data librarian was embedded in the research process from the outset of a series of projects, first advising the clinician-investigator on data sourcing and later managing most aspects of data processing, including data retrieval and management, and writing Python code to clean, merge, and restructure data for statistical analyses conducted by the clinician-investigator. To date, this ongoing collaboration has led to publication of three original research reports.[18-20]

Table 1 highlights the common types of data librarian involvement in research. To illustrate, we provide specific examples of how the second dyad navigated each of the five areas:

- *Data project planning*
    - During the early stages of the project, the clinician-investigator consulted with the data librarian who advised on data sourcing, quality assessment, analysis, and dissemination methodologies.
- *Data acquisition*
    - The data librarian was well-versed in accessing open, public data and assisted the clinician-investigator with finding viable alternatives to expensive data sources, reducing total project costs.

- *Data management*
    - The data librarian provided guidance on how best to manage, organize, and document data throughout the project.
- *Data processing and analysis*
    - The data librarian wrote scripts in Python to clean, merge, and re-structure datasets into analysis-ready formats. Public data can be poorly organized and documented, and combining disparate sources of publicly available data can present unique challenges that may compound the underlying limitations of any individual dataset. The data librarian guided early data quality assessment and data processing, which was essential to the success of the research project.
- *Data sharing*
    - The data librarian advised on reproducibility best practices and contributed to data sourcing and methodology sections in manuscripts describing the project's processes.

Conclusion

To enhance the quality and reproducibility of aging research, clinician-investigators should consider collaborating with librarians. In this commentary, we discussed how librarians complement and contribute to the work of clinician-investigators and illustrated how clinician-investigators can initiate a productive, enriching collaboration with a librarian. By improving the process of acquiring, evaluating, managing, and synthesizing information – both literature and data – librarians can be invaluable collaborators in team-science approaches to aging research.

Acknowledgements

We would like to extend our gratitude to program manager Andrea Sherman for her assistance with coordinating our Clin-STAR webinar; she disseminated notes from our planning meeting, coordinated our schedules and files, ensured that the webinar went smoothly, and archived all associated materials, which we built upon for this manuscript.

Author Contributions

All authors meet the criteria for authorship as stated by the International Committee of Medical Journal Editors.

Conflicts of Interest

Throgmorton, KF, None; Festa, N, None; Doering, M, None; Carpenter, CR, None; Gill, TM, None.

Funding & Sponsor's Role

The work is supported by a Clinician-Scientists Transdisciplinary Aging Research (Clin-STAR) Coordinating Center grant U24AG065204. Dr. Festa is supported by National Institute on Aging (NIA) training grant T32AG019134. This work was conducted, in part, at the Yale Claude D. Pepper Older Americans Independence Center (P30AG021342).

Disclaimer: The contents of this manuscript are the sole responsibility of the authors and do not represent the views of the NIH or other funders.

References

1. Aamodt M, Huurdeman H, Strømme H. Librarian Co-Authored Systematic Reviews are Associated with Lower Risk of Bias Compared to Systematic Reviews with Acknowledgement of Librarians or No Participation by Librarians. *Evid Based Libr Inf Pract*. 2019;14(4):103-127. doi:10.18438/eblip29601

2. Rethlefsen ML, Farrell AM, Osterhaus Trzasko LC, Brigham TJ. Librarian co-authors correlated with higher quality reported search strategies in general internal medicine systematic reviews. *J Clin Epidemiol*. 2015;68(6):617-626. doi:10.1016/j.jclinepi.2014.11.025

3. Federer L, Foster ED, Glusker A, Henderson M, Read K, Zhao S. The Medical Library Association Data Services Competency: a framework for data science and open science skills development. *J Med Libr Assoc*. 2020;108(2):304-309. doi:10.5195/jmla.2020.909

4. Gill TM, Carpenter C, Doering M, Festa N, Throgmorton KF. Clin-STAR Webinar: Enhancing Your Research: How to Collaborate Effectively with Medical and Data Librarians. Clin-STAR. Published February 3, 2023. Accessed July 21, 2023. https://clin-star.org/clin-star-webinar-librarian-2023

5. Lee J, Kamdar BB, Bergstrom J, Murphy TE, Gill TM. Modeling success: How to work effectively with your biostatistician. *J Am Geriatr Soc*. 2022;70(8):2449-2454. doi:10.1111/jgs.17888

6. Brackett A, Batten J. Ensuring rigor in systematic reviews: Part 7, critical appraisal of systematic review quality. *Heart Lung J Crit Care*. 2022;53:32-35. doi:10.1016/j.hrtlng.2022.01.008

7. Grant MJ, Booth A. A typology of reviews: an analysis of 14 review types and associated methodologies. *Health Inf Libr J*. 2009;26(2):91-108. doi:10.1111/j.1471-1842.2009.00848.x

8. Carpenter CR, Banerjee J, Keyes D, et al. Accuracy of Dementia Screening Instruments in Emergency Medicine: A Diagnostic Meta-analysis. *Acad Emerg Med*. 2019;26(2):226-245. doi:10.1111/acem.13573

9. Carpenter CR, Avidan MS, Wildes T, Stark S, Fowler SA, Lo AX. Predicting Geriatric Falls Following an Episode of Emergency Department Care: A Systematic Review. *Acad Emerg Med*. 2014;21(10):1069-1082. doi:10.1111/acem.12488

10. Fowler SA, Yaeger LH, Rogers M, Carpenter CR. Validating a search filter for diagnostic sensitivity and specificity. Presented at: 2013 Medical Library Association Annual Meeting and Exhibition; 2013; Boston, Massachusetts. https://digitalcommons.wustl.edu/becker_pubs/37

11. McInnes MDF, Moher D, Thombs BD, McGrath TA, Bossuyt PM, and the PRISMA-DTA Group. Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies: The PRISMA-DTA Statement. *JAMA*. 2018;319(4):388-396. doi:10.1001/jama.2017.19163

12. Nowroozpoor A, Dussetschleger J, Perry W, et al. Detecting Cognitive Impairment and Dementia in the Emergency Department: A Scoping Review. *J Am Med Dir Assoc*. 2022;23(8):1314.e31-1314.e88. doi:10.1016/j.jamda.2022.03.019

13. Carpenter CR, Leggett J, Bellolio F, et al. Emergency Department Communication in Persons Living With Dementia and Care Partners: A Scoping Review. *J Am Med Dir Assoc*. 2022;23(8):1313.e15-1313.e46. doi:10.1016/j.jamda.2022.02.024

14. Dresden SM, Taylor Z, Serina P, et al. Optimal Emergency Department Care Practices for Persons Living With Dementia: A Scoping Review. *J Am Med Dir Assoc*. 2022;23(8):1314.e1-1314.e29. doi:10.1016/j.jamda.2022.05.024

15. Gettel CJ, Falvey JR, Gifford A, et al. Emergency Department Care Transitions for Patients With Cognitive Impairment: A Scoping Review. *J Am Med Dir Assoc*. 2022;23(8):1313.e1-1313.e13. doi:10.1016/j.jamda.2022.01.076

16. Tenopir C, Rice NM, Allard S, et al. Data sharing, management, use, and reuse: Practices and perceptions of scientists worldwide. *PLoS ONE*. 2020;15(3):e0229003. doi:10.1371/journal.pone.0229003

17. Goldman J, Muilenburg J, Schorr AN, Ossom-Williamson P, Uribe-Lacy CJ. Trends in Research Data Management and Academic Health Sciences Libraries. *Med Ref Serv Q*. 2023;42(3):273-293. doi:10.1080/02763869.2023.2218776

18. Festa N, Throgmorton KF, Davis-Plourde K, et al. Assessment of Regional Nursing Home Preparedness for and Regulatory Responsiveness to Wildfire Risk in the Western US. *JAMA Netw Open*. 2023;6(6):e2320207. doi:10.1001/jamanetworkopen.2023.20207

19. Festa N, Throgmorton KF, Heaphy N, Canavan M, Gill TM. Association of Nursing Home Exposure to Hurricane-Related Inundation With Emergency Preparedness. *JAMA Netw Open*. 2023;6(1):e2249937. doi:10.1001/jamanetworkopen.2022.49937

20. Festa N, Heaphy NM, Throgmorton KF, Canavan M, Gill TM. Evaluating California nursing homes' emergency preparedness for wildfire exposure. *J Am Geriatr Soc*. 2023;71(3):895-902. doi:10.1111/jgs.18142

LEGENDS

Figure 1: Number of Research Works* About Older Adults from 1993 to 2022. Bibliographic data used to create the figure were extracted from the Scopus database with targeted queries and adjacency search techniques as described below (**Supplementary File S1**), then plotted using Python 3 and seaborn (**Supplementary File S2**). Additional details are provided in Appendices 1 and 2.

The blue line represents the total number of research works about older adults in the thirty-year sample and includes works tagged in the publication's title, abstract, or keyword fields as about geriatrics or older adults, including more than a dozen synonyms for these terms. Within the total, the green dotted line represents works tagged as systematic reviews or scoping reviews, and the orange dashed line represents works tagged as data-intensive research, using terms such as: data, datasets, data bank, data mining, metadata, data analysis, or statistical methods used in data analysis.

*Works refer to the citation records contained in the Scopus database and can include journal articles, abstract reports, books, book chapters, business articles, conference papers, conference reviews, data papers, editorials, erratum, letters, multimedia, notes, press releases, reports, retractions, reviews, and short surveys.

Supplementary File S1: Search Strategies Used to Create Figure 1. These search strategies were used to extract data from Scopus used to create Figure 1.

311

312 **Supplementary File S2:** Python Code Used to Create Figure 1. This code was written to

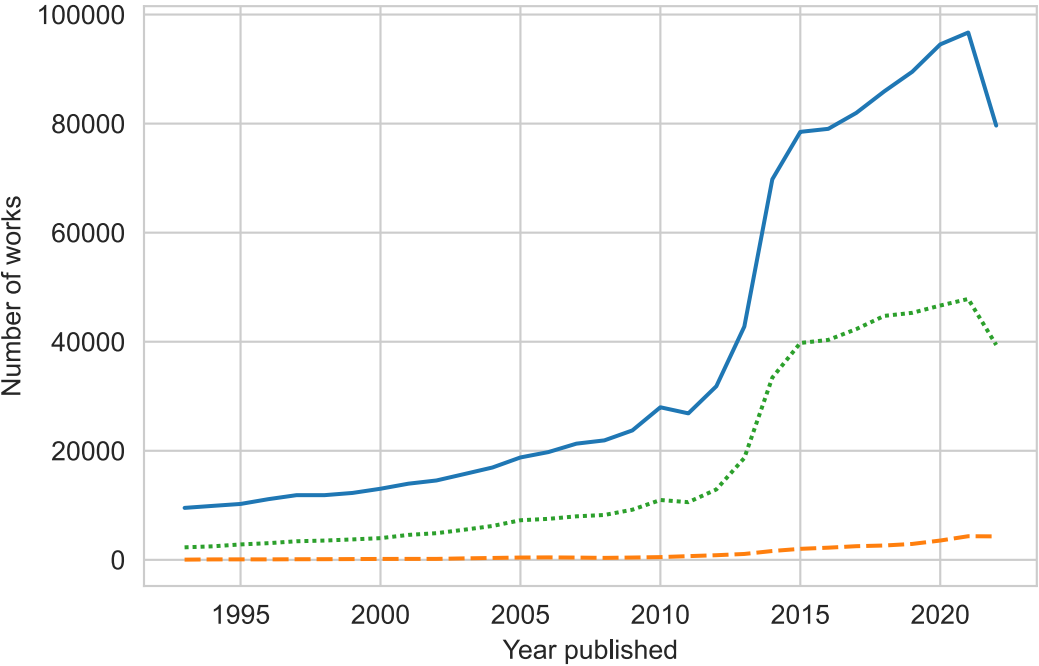
313 transform data and create the visualization for Figure 1.

314 **Table 1. Common Types of Medical and Data Librarian Involvement in Research**

Research Type	Research Stage	Librarian Type	Common Types of Involvement
Evidence Synthesis Research	Before screening articles for literature reviews	Medical	• Select subject headings and synonyms • Run the search in an appropriate database using strategies to narrow the search to a manageable volume for screening • Search clinical trial registries if the literature review is for a grant proposal
Evidence Synthesis Research	Before submitting protocol for systematic and scoping reviews	Medical	• Summarize search plan for protocol • Select subject headings and synonyms for the key concepts in the research question • Run the search using field tags through multiple databases, including non-academic literature resources to reduce publication bias • Provide access to screening tools • Report the search methods for the manuscript
Any	Planning	Data	• Advise and train on best practices for acquiring, processing, analyzing, disseminating, and sharing data
Data-Intensive Research	Data Collection & Acquisition	Data	• Source datasets needed for research • Identify suitable open, public data available for free • Navigate controlled data acquisition protocols (e.g., data use agreements) • Use library budget funds to purchase data • Consult on how to store, organize, and secure data

Data-Intensive Research	Data Processing & Analysis	Data	• Process data, such as by performing data cleaning, data merging, data transformation, data harmonization, etc. • Train on relevant data tools, such as programming languages (e.g., Python, R, etc.) and software
Any	Data Management	Data	• Train on funder policy and data management principles, such as the National Institutes of Health's (NIH) new Data Management and Sharing Policy and the FAIR (findable, accessible, interoperable, reusable) principles • Review and provide feedback on data management (and sharing) plans • Coordinate data management support across relevant departments, such as institutional review boards, compliance, information technology (IT), and computing services
Any	Data Sharing	Data	• Advise and train on reproducibility best practices, including on related policy, appropriate mechanisms, institutional requirements and infrastructure, data licensing, and data repositories

315


100000
80000
60000
40000
20000
0
Number of works
1995
2000
2005
2010
2015
2020
Year published

Supplementary File S1

Search strategies used to extract data for Figure 1.

Research works about older adults:

(((TITLE-ABS-KEY (elderly OR elderlies OR centenarian* OR nonagenarian* OR octogenarian* OR septuagenarian* OR sexagenarian* OR geriatric* OR senium* OR "aged adult" OR "aged adults" OR "aged patient" OR "aged patients")) OR (TITLE-ABS-KEY (senior* w/1 citizen*)) OR (TITLE-ABS-KEY ((older OR frail*) w/2 (hospitalized OR hospitalised OR elder* OR patient* OR person* OR inpatient* OR people OR adult*))) OR (SRCTITLE (gerontolog* OR geriatric* OR ageing OR aging OR "Older adults")) AND PUBYEAR > 1992 AND PUBYEAR < 2023)

Within research works about older adults, works about systematic and scoping reviews:

(((TITLE-ABS-KEY ((review w/2 (scoping OR systematic)) OR metaanalysis OR meta-analysis))) AND ((TITLE-ABS-KEY (elderly OR elderlies OR centenarian* OR nonagenarian* OR octogenarian* OR septuagenarian* OR sexagenarian* OR geriatric* OR senium* OR "aged adult" OR "aged adults" OR "aged patient" OR "aged patients")) OR (TITLE-ABS-KEY (senior* W/1 citizen*)) OR (TITLE-ABS-KEY ((older OR frail*) W/2 (hospitalized OR hospitalised OR elder* OR patient* OR person* OR inpatient* OR people OR adult*))) OR (SRCTITLE (gerontolog* OR geriatric* OR ageing OR aging OR "Older adults"))) AND PUBYEAR > 1991 AND PUBYEAR < 2023)

Within research works about older adults, works about data-intensive research:

(((TITLE-ABS-KEY (dataset* OR data-set* OR data OR databank OR datamining OR metadata OR (data W/3 analysis))) OR (TITLE-ABS-KEY (factual-database OR statistic* OR estimation-technique* OR life-tables OR quality-adjusted-life-years OR area-under-curve OR confidence-intervals OR likelihood-functions OR linear-model* OR logistic-model* OR nomogram* OR partin-table* OR proportional-hazards-model* OR monte-carlo-method OR geographic-mapping OR spatial-regression OR binomial-distribution* OR chi-square-distribution OR normal-distribution OR poisson-distribution OR stochastic-process* OR markov-chain* OR clustering-algorithm)) OR(TITLE-ABS-KEY (analysis w/2 (data OR spatial OR cluster OR classification OR area OR actuarial OR variance OR multivariate OR latent-class OR scaling OR small-area OR discriminant OR factor OR matched-pair OR component OR regression OR least-squares OR survival OR path)))) AND ((TITLE-ABS-KEY (elderly OR elderlies OR centenarian* OR nonagenarian* OR octogenarian* OR septuagenarian* OR sexagenarian* OR geriatric* OR senium* OR "aged adult" OR "aged adults" OR "aged patient" OR "aged patients")) OR (TITLE-ABS-KEY (senior* W/1 citizen*)) OR (TITLE-ABS-KEY ((older OR frail*) W/2 (hospitalized OR hospitalised OR elder* OR patient* OR person* OR inpatient* OR people OR adult*))) OR (SRCTITLE (gerontolog* OR geriatric* OR ageing OR aging OR "Older adults"))) AND PUBYEAR > 1991 AND PUBYEAR < 2023)

Supplementary File S2

Python Code Used to Create Figure 1.

To review search strategies that created the Scopus CSVs used below, see Supplementary File S1.

In [1]:
```
from platform import python_version
import pandas as pd
import matplotlib
import matplotlib.pyplot as plt
import seaborn as sns

#see versions in next cell and citations for libraries at end of notebook
```

In [2]:
```
print("Python version:", python_version())
print("pandas version:", pd.__version__)
print("matplotlib version:", matplotlib.__version__)
print("seaborn version:", sns.__version__)
```

```
Python version: 3.9.7
pandas version: 1.4.3
matplotlib version: 3.5.1
seaborn version: 0.11.2
```

In [3]:
```
#load in data

#geriatrics research in general
scopus_main = pd.read_csv("Scopus-10-Analyze-Year_geriatrics_research-CLEAN.csv")

#scoping and systematic reviews in geriatrics
scopus_reviews = pd.read_csv("Scopus-10-Analyze-Year_geriatrics_SRMetaScopingReviews-CLEAN.csv")

#data-intensive research in geriatrics
scopus_data = pd.read_csv("Scopus-10-Analyze-Year_geriatrics_data_broad_statistical-CLEAN.csv")

#concat (merge) all above into one dataframe
scopus_all = pd.concat([scopus_main, scopus_reviews, scopus_data],axis=1)
```

In [4]:
```
#drop redundant columns
scopus_all = scopus_all.drop(columns=["YEAR_reviews", "YEAR_data_broad"])
```

In [5]:
```
#rename columns to more readable names for visualization
scopus_all = scopus_all.rename(columns={"YEAR_all": "Year published",
                                        "COUNT_all": "All works",
                                        "COUNT_reviews": "Scoping and systematic reviews",
                                        "COUNT_data_broad": "Data-intensive research"})
```

In [6]:
```
#reset index so each row is the year for the counts (important for later visualization)
scopus_all = scopus_all.set_index("Year published")
```

In [7]:
```
#view info about dataframe, ensure there are 30 years of data included
scopus_all.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 30 entries, 2022 to 1993
Data columns (total 3 columns):
 #   Column                          Non-Null Count  Dtype
---  ------                          --------------  -----
 0   All works                       30 non-null     int64
 1   Scoping and systematic reviews  30 non-null     int64
 2   Data-intensive research         30 non-null     int64
dtypes: int64(3)
memory usage: 960.0 bytes
```

In [8]:
```
#view dataframe before visualization
scopus_all
```

Out[8]:

Year published	All works	Scoping and systematic reviews	Data-intensive research
2022	79633	4301	39406
2021	96704	4331	47841
2020	94524	3539	46635
2019	89506	2916	45298
2018	85910	2628	44739
2017	81961	2502	42312
2016	79039	2230	40329
2015	78484	2017	39771
2014	69782	1620	33412
2013	42777	1077	18679
2012	31835	847	12908
2011	26860	677	10574
2010	27978	502	10973
2009	23734	423	9193
2008	21907	358	8226
2007	21305	415	7979
2006	19762	449	7516
2005	18776	421	7260
2004	16933	343	6207
2003	15737	268	5528
2002	14556	173	4888
2001	13967	173	4575
2000	13038	164	3977
1999	12268	140	3746
1998	11861	120	3534
1997	11861	109	3407
1996	11131	86	3057
1995	10249	87	2825
1994	9893	80	2476
1993	9527	36	2293

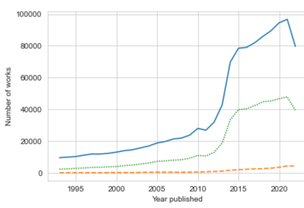
In [9]:
```
#visualize data with a seaborn lineplot and add title
sns.set_style("whitegrid")
sns.lineplot(data=scopus_all).set(title="Number of Research Works About Older Adults - 1993-2022", ylabel="Numb
```

Out[9]:
```
[Text(0.5, 1.0, 'Number of Research Works About Older Adults - 1993-2022'),
 Text(0, 0.5, 'Number of works')]
```

In [10]:
```
#same visualization without title and without legend
#as needed per JAGS author guidelines
#final fig:
sns.set_style("whitegrid")
sns.lineplot(data=scopus_all,legend=False).set(ylabel="Number of works")


#uncomment to save figures to file
#plt.savefig("older_adult_research_no_legend_high_res.pdf",dpi=600)
#plt.savefig("older_adult_research_no_legend_high_res.png",dpi=600)
```

Out[10]:
```
[Text(0, 0.5, 'Number of works')]
```



Citations for libraries used:

- The pandas development team. pandas-dev/pandas: Pandas (v2.1.4). Zenodo. 2023. doi:10.5281/zenodo.10304236
- Hunter JD. Matplotlib: A 2D Graphics Environment. Computing in Science & Engineering. 2007;9(3): 90-95. doi:10.1109/MCSE.2007.55.
- Waskom ML. seaborn: statistical data visualization. Journal of Open Source Software. 2021;6(60): 3021. doi:10.21105/joss.03021.

Note:

- This document was created with Juypter Notebook.

In [ ]:
```

```