MEMORIAL SLOAN KETTERING CANCER CENTER

LIBRARY

MAKE THE CONNECTION

RESEARCH IN REVIEW:

An Analysis of MSK Publications 2016-2020

Memorial Sloan Kettering
Cancer Center

Table of Contents

Director's Prologue

Donna Gibson
Director, Library Services

Welcome to the first five-year report highlighting works published by and associated with Memorial Sloan Kettering (MSK) Cancer Center. This report will showcase the variety of existing metrics used to measure research impact at an article-, journal-, or author-level.

In the period covering 2016 to 2020, the Synapse team retrieved and curated over 25,500 works, and it is with these citations that we hope to share one aspect of a very complex research story. Measuring impact is not a perfect science. The pages that follow will highlight the demographics of our author community, existing metrics, and how research impact is gauged. We will also share interesting facts about publication-related datasets gleaned from a variety of credible resources such as PubMed Central® (PMC).

A paper's citation count is a well-recognized measure (how often a researcher cites other researchers' publications) to reflect relevance and scientific impact; however, there are other measures that can help contribute to a more holistic view of the published research. This report is not designed to make judgement calls on the current bibliometric portfolio, but rather to raise awareness of possible ways to view research contributions. Because research data is continually updated and metrics can change daily, the numbers outlined in this report are only as accurate as the date the data was retrieved. For clarity, we have noted the data retrieval date and source name throughout this report.

A variety of internal (Synapse), public (PubMed®), and proprietary resources such as Scopus from Elsevier and the Web of Science™ Platform from Clarivate were used in the preparation of this report.
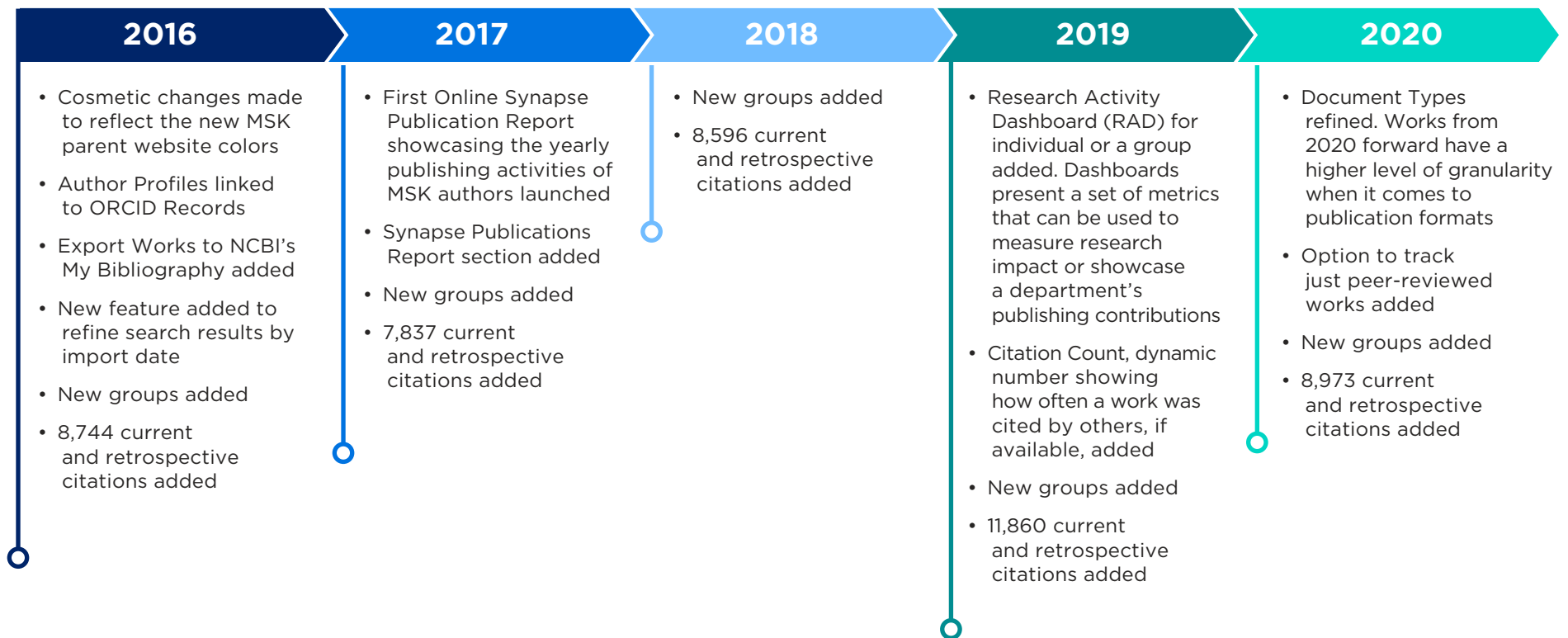
In closing, I'd like to acknowledge and thank two of our Associate Librarians - Jeanine McSweeney in Scholarly Communications, and Anthony Dellureficio in Data Management Services - for their hard work and contributions in making this report a reality.

I hope the information provided will offer the reader a better understanding of ways in which to view research with an emphasis placed on works attributed to MSK.
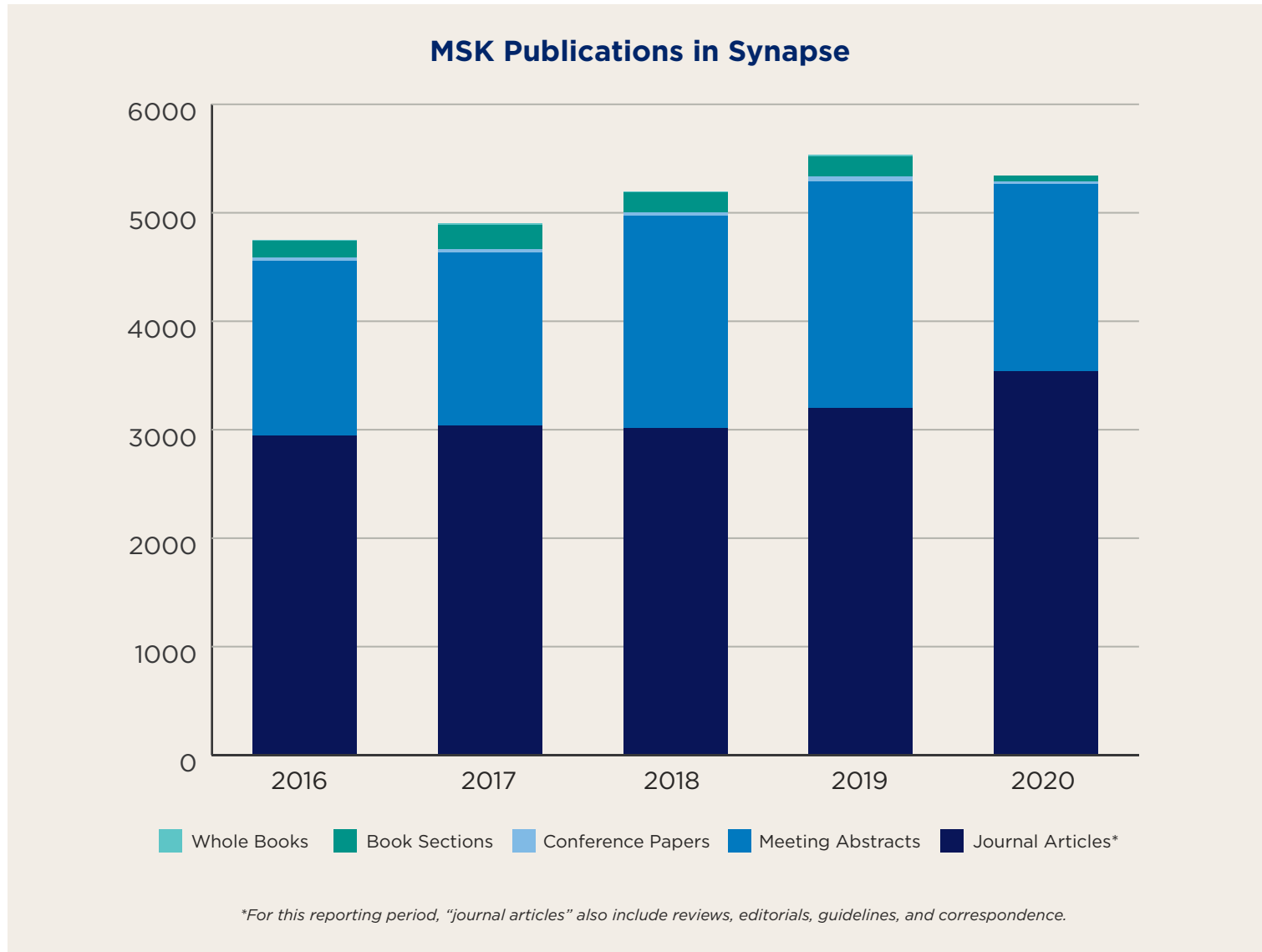
If you have any questions or comments, feel free to [Contact Us!](#)

Synapse Enhancements Timeline

Synapse is a public-facing resource and the authoritative bibliographic database developed and maintained by a skilled team of information professionals at the MSK Library. As of January 2022, Synapse contained over 94,800 references dating back to 1986, embodying both current and former MSK authors. This resource is used by many departments to support their scholarly-related activities, such as monitoring the research being published by their groups and leveraging citation data available in Synapse as the source for any internal bibliographic, grant, or CV applications.
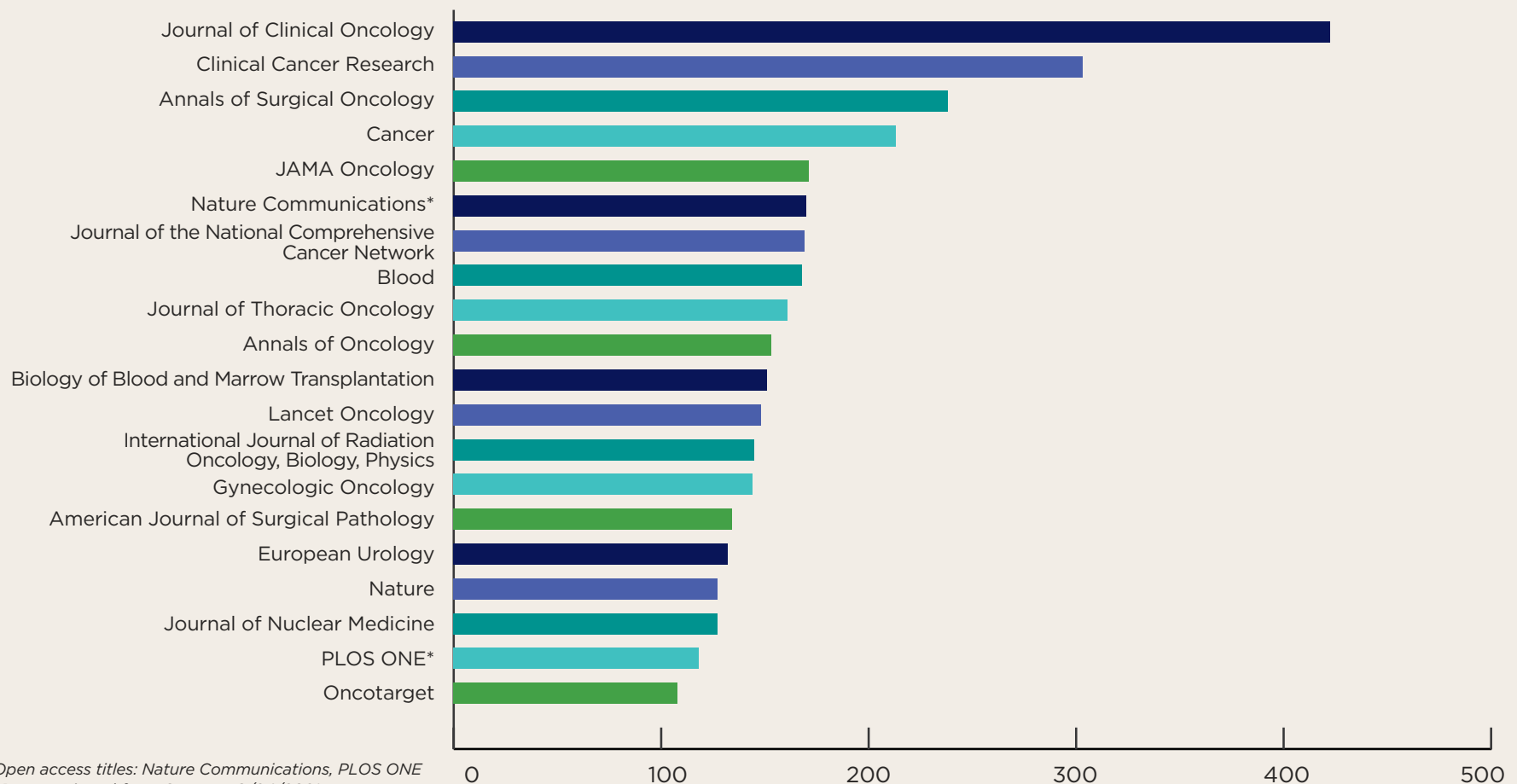
2016

- Cosmetic changes made to reflect the new MSK parent website colors
- Author Profiles linked to ORCID Records
- Export Works to NCBI's My Bibliography added
- New feature added to refine search results by import date
- New groups added
- 8,744 current and retrospective citations added

2017

- First Online Synapse Publication Report showcasing the yearly publishing activities of MSK authors launched
- Synapse Publications Report section added
- New groups added
- 7,837 current and retrospective citations added

2018

- New groups added
- 8,596 current and retrospective citations added

2019

- Research Activity Dashboard (RAD) for individual or a group added. Dashboards present a set of metrics that can be used to measure research impact or showcase a department's publishing contributions
- Citation Count, dynamic number showing how often a work was cited by others, if available, added
- New groups added
- 11,860 current and retrospective citations added

2020

- Document Types refined. Works from 2020 forward have a higher level of granularity when it comes to publication formats
- Option to track just peer-reviewed works added
- New groups added
- 8,973 current and retrospective citations added

Over this 5-year period MSK authors collectively produced an average of **5,140 works per year.**

MSK Publications in Synapse

Legend: Whole Books | Book Sections | Conference Papers | Meeting Abstracts | Journal Articles*

For this reporting period, "journal articles" also include reviews, editorials, guidelines, and correspondence.

Journals

MSK authors published in **2,015 unique journals** between 2016-2020. 76% of those journals had 5 or fewer works published in them, and only 3% had more than 50 works.

Journal Titles with the Highest Number of MSK Papers, 2016-2020

- Journal of Clinical Oncology
- Clinical Cancer Research
- Annals of Surgical Oncology
- Cancer
- JAMA Oncology
- Nature Communications*
- Journal of the National Comprehensive Cancer Network
- Blood
- Journal of Thoracic Oncology
- Annals of Oncology
- Biology of Blood and Marrow Transplantation
- Lancet Oncology
- International Journal of Radiation Oncology, Biology, Physics
- Gynecologic Oncology
- American Journal of Surgical Pathology
- European Urology
- Nature
- Journal of Nuclear Medicine
- PLOS ONE*
- Oncotarget

0   100   200   300   400   500

*Open access titles: Nature Communications, PLOS ONE
**Data retrieved from Synapse 8/24/2021

Impact Factor

MSK authors published in an **average of 892 different journals** each year. This chart shows the majority have an impact factor between 0-10. 50% have an **impact factor between 2-6.**
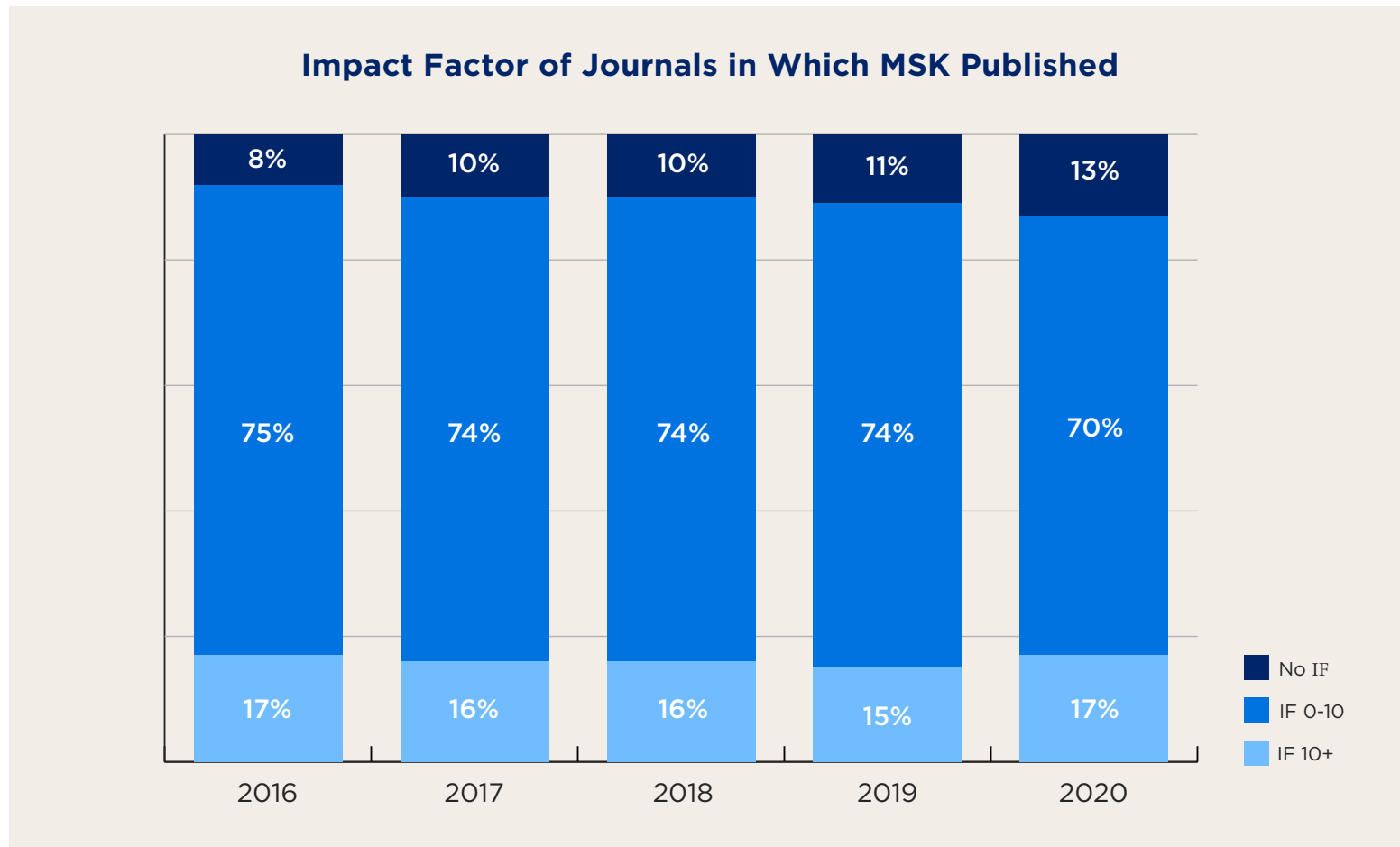
Impact Factor of Journals in Which MSK Published

Year	No IF	IF 0-10	IF 10+
2016	8%	75%	17%
2017	10%	74%	16%
2018	10%	74%	16%
2019	11%	74%	15%
2020	13%	70%	17%

Legend:
- No IF
- IF 0-10
- IF 10+

*For this reporting period, "journal articles" also include reviews, editorials, guidelines, and correspondence.
**Journal Impact factor data retrieved from Journal Citation Reports™ (Clarivate 2021)

Citation Growth

An analysis of the combined citation growth of the top 500 cited papers published by MSK authors over the last decade shows that citation growth tends to peak 3-4 years after publication while continuing to receive a relatively steady number of citations for years after. Another observed trend is that the number of citations for MSK papers appears to show an overall increase over time.


MSK Citation Growth Over Time

Year of Publication

Legend: 2010, 2013, 2016, 2019
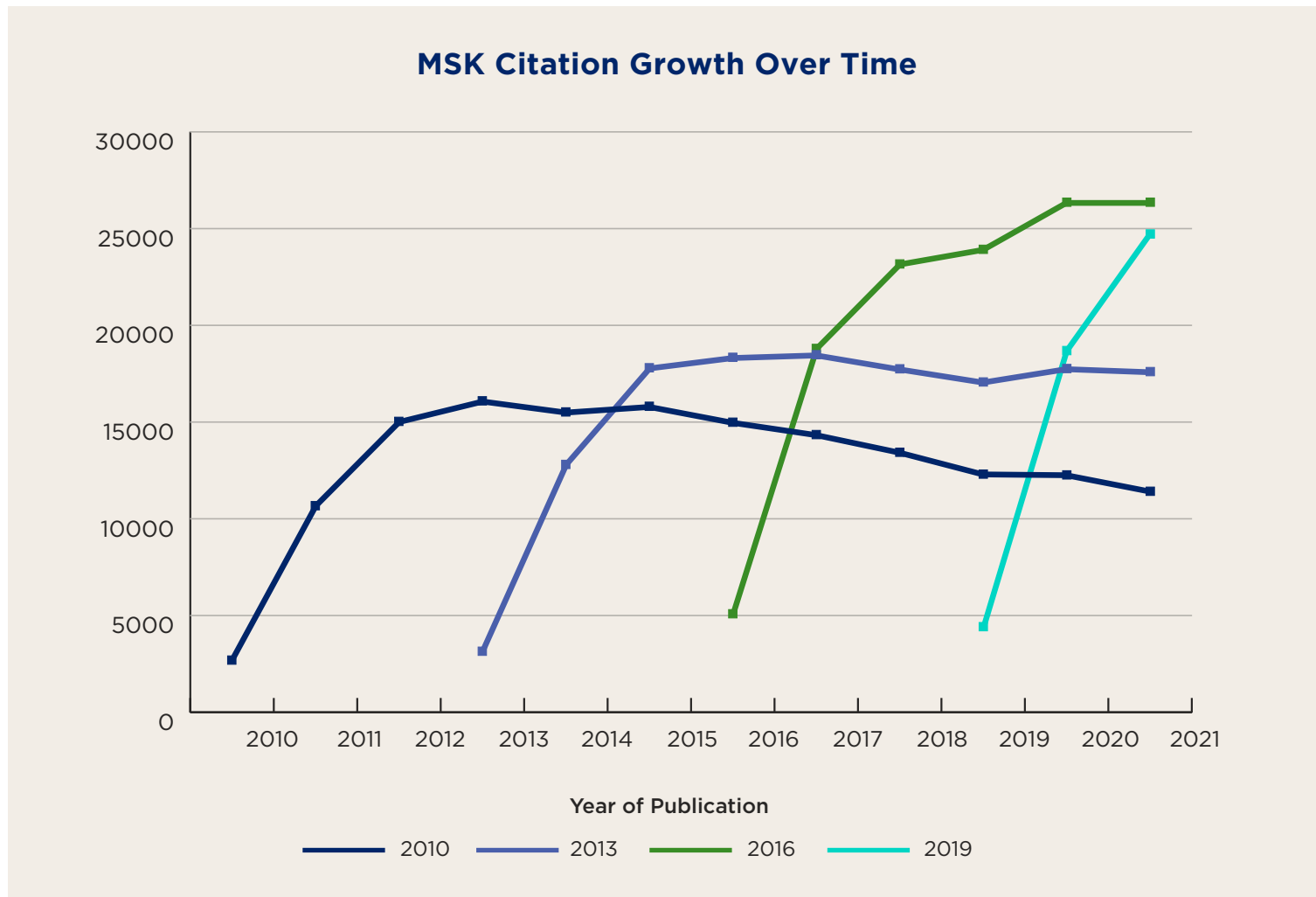
Citation Impact

This chart shows the cumulative citation scores of all MSK papers in the last decade by their publication year. Keeping in mind what we learned in the previous chart, that citation growth peaks 3-4 years after publication and maintains a relatively steady growth rate for years after, we see fewer total citations for more recent years' publications.

Total Citations: All MSK Papers by Publication Year

Data retrieved from Scopus® (Elsevier) 10/15/2021
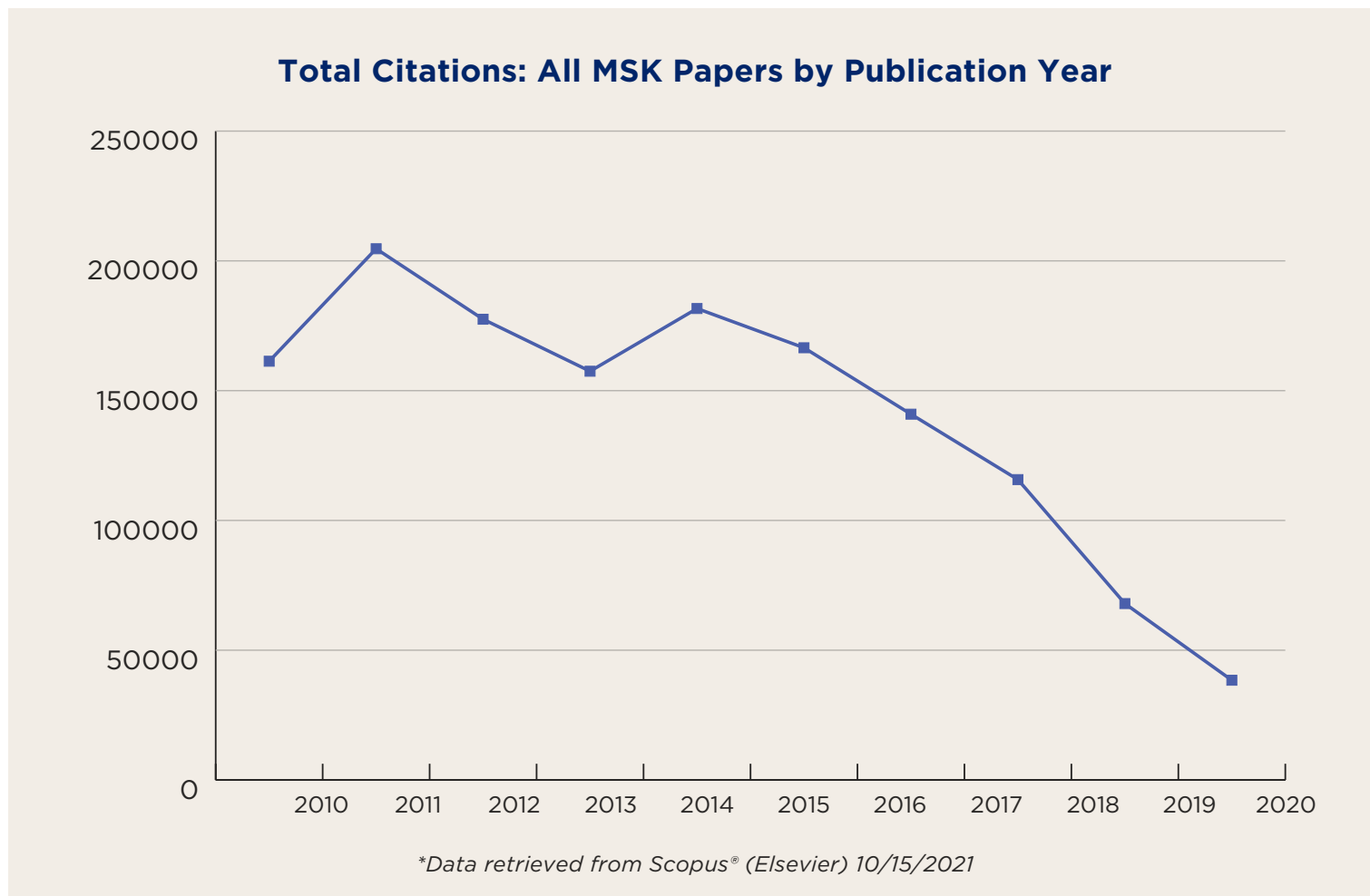
Altmetrics

Unlike citation counts, Altmetric scores tend to peak in the year of publication and show very little growth after that. Looking at the Altmetric scores of the 20 highest scoring papers in each year, we saw an **average growth rate of just 4%** from the original score.

Altmetric Scores: Growth Over Time

Legend: ■ Original Score ■ Current Score
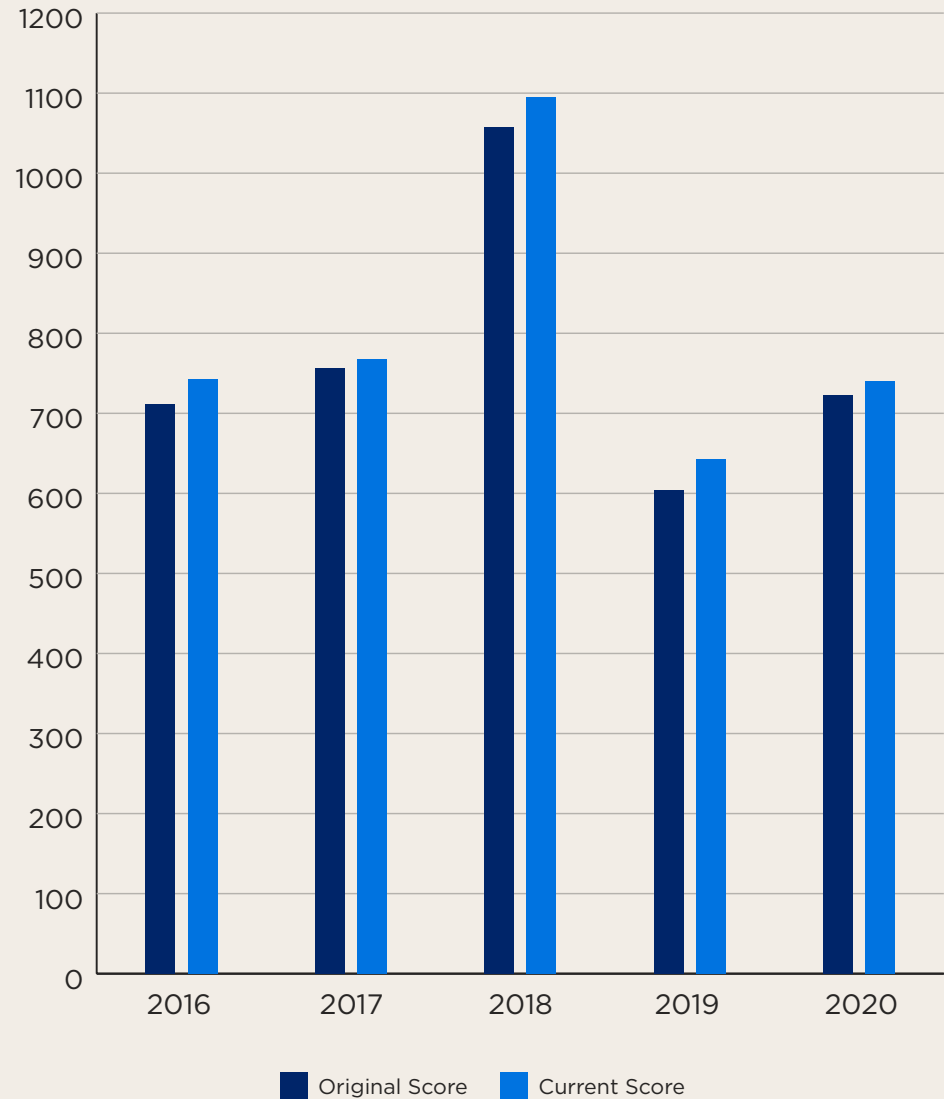
Data retrieved from Altmetric.com 11/17/2021

Citation and Altmetric scores showed a similar pattern of being highest in a small number of prestigious, high-impact journals. Our analysis of the MSK 100 top cited and 100 top Altmetric papers in each year showed that **50% of those papers** were published in just **8 journals.**

Journals Representing 50% of Top Cited and Altmetric Papers, 2016-2020

Journal	Citation score	Altmetric score
New England Journal of Medicine	24	19
Nature	12	6
Science	5	6
Nature Medicine	4	5
JAMA: Journal of the American Medical Association	0	8
JAMA Oncology	0	8
The Lancet Oncology	4	0
Cell	4	0

Legend:
- Citation score
- Altmetric score
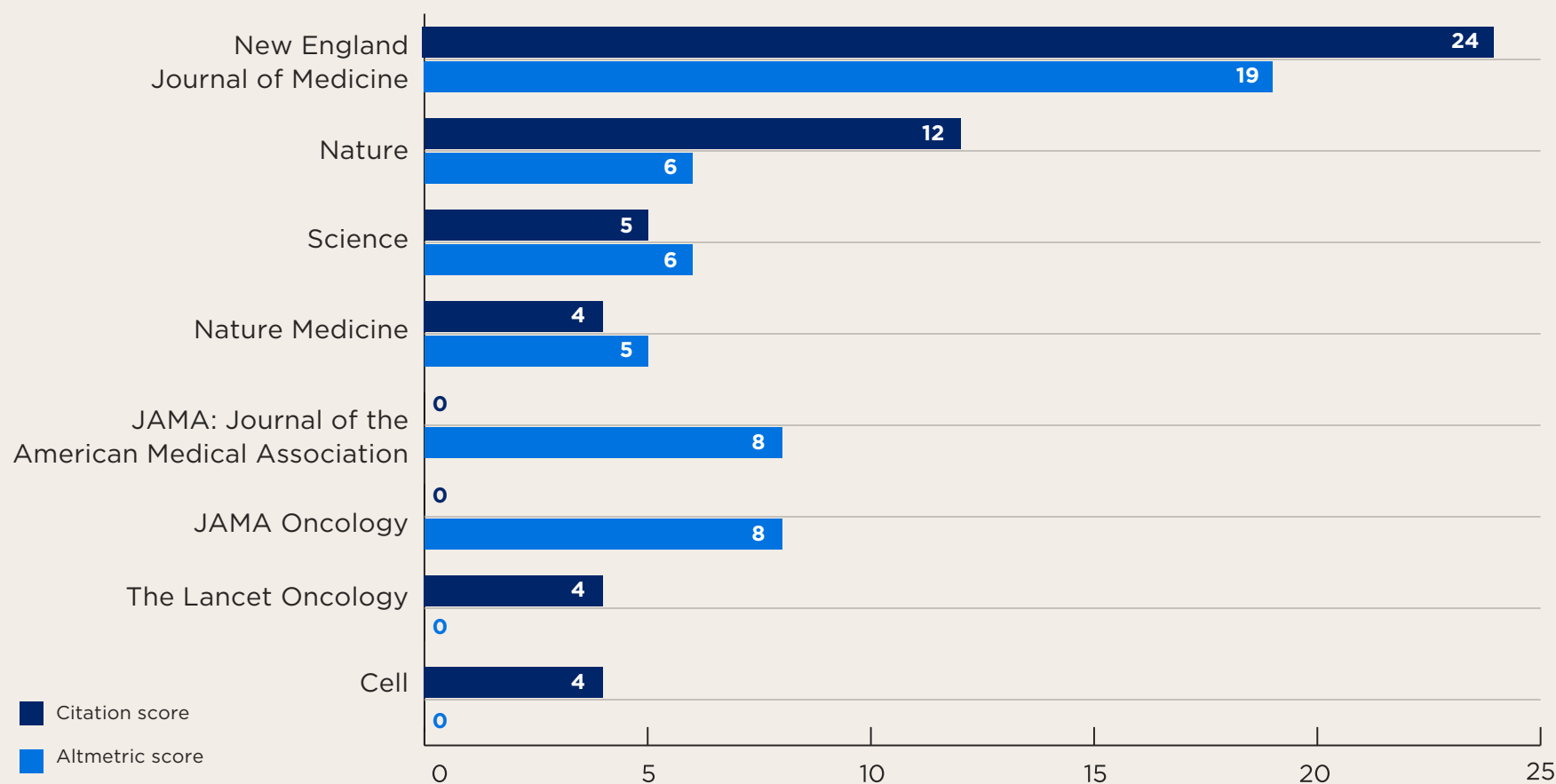
iCite

Developed by the National Institutes of Health (NIH), Office of Portfolio Analysis, iCite provides metrics and visualization information for journal articles that have been included in PubMed. Users type in a PubMed query or upload their selection of PubMed IDs (PMID).

iCite can translate "how Human, Animal, or Molecular/Cellular Biology-oriented each paper is," and uses this information to track and predict citation by clinical articles. Read about how the Approximate Potential to Translate (APT), a machine learning-based estimate of the likelihood that a paper will be cited in later clinical trials/guidelines, is calculated at PLOS Biology.

Hutchins BI, Davis MT, Meseroll RA, Santangelo GM (2019) Predicting translational progress in biomedical research. PLOS Biology 17(10): e3000416. https://doi.org/10.1371/journal.pbio.3000416

Summary:

One of the inherent difficulties in predicting translation progress in biomedical research is the long timespan for most of its pipelines. This produces citation dynamics which are extremely chaotic and challenging for traditional statistical methods. Researchers at NIH decided to see if they could develop Artificial Intelligence/Machine Learning models which could predict which papers were likely to be cited in future clinical trials or guidelines. Fundamentally, they found "knowledge flow trajectories" at the article level which underlay bench-to-bedside translation. Their initial investigation found that "important early steps in translation may appear topically distinct from their clinical descendants." Additionally, they found from citation analysis that post-publication translational article citation tends to plateau in the 5-10 year range, so their assumption was that most knowledge transfer occurs within the first few years after publication. Data from their tests suggested that there may be key markers for prediction, and that if those markers appear early enough, they may be able to generate a machine learning-based model to make early predictions. Using Medical Subject Headings (MeSH) to categorize research by branch (human, animal, or molecular/cell) they were able to further increase the accuracy of predicting future citations in clinical papers. Working backwards through increasingly shorter post-publication data timeframes, the researchers found that at 2 years the models were still surprisingly accurate and within their statistical boundaries for predicting a citation in clinical papers based on early reactions to publications.

A search was conducted in PubMed® to produce a subset of MSK-associated research derived from articles published in Nature, Cell, and Science journals between 2016-2020 (1,541 papers). Using iCite, the PMIDs were entered into this tool with the intention to highlight the potential for bench-to-bedside translation of basic science research at MSK.

Based on the iCite algorithm, these papers had an APT (Approximate Potential to Translate) score of 47.1%. In addition, 31% of the papers have already been cited by a clinical study.

Triangle of Biomedicine Visualizations
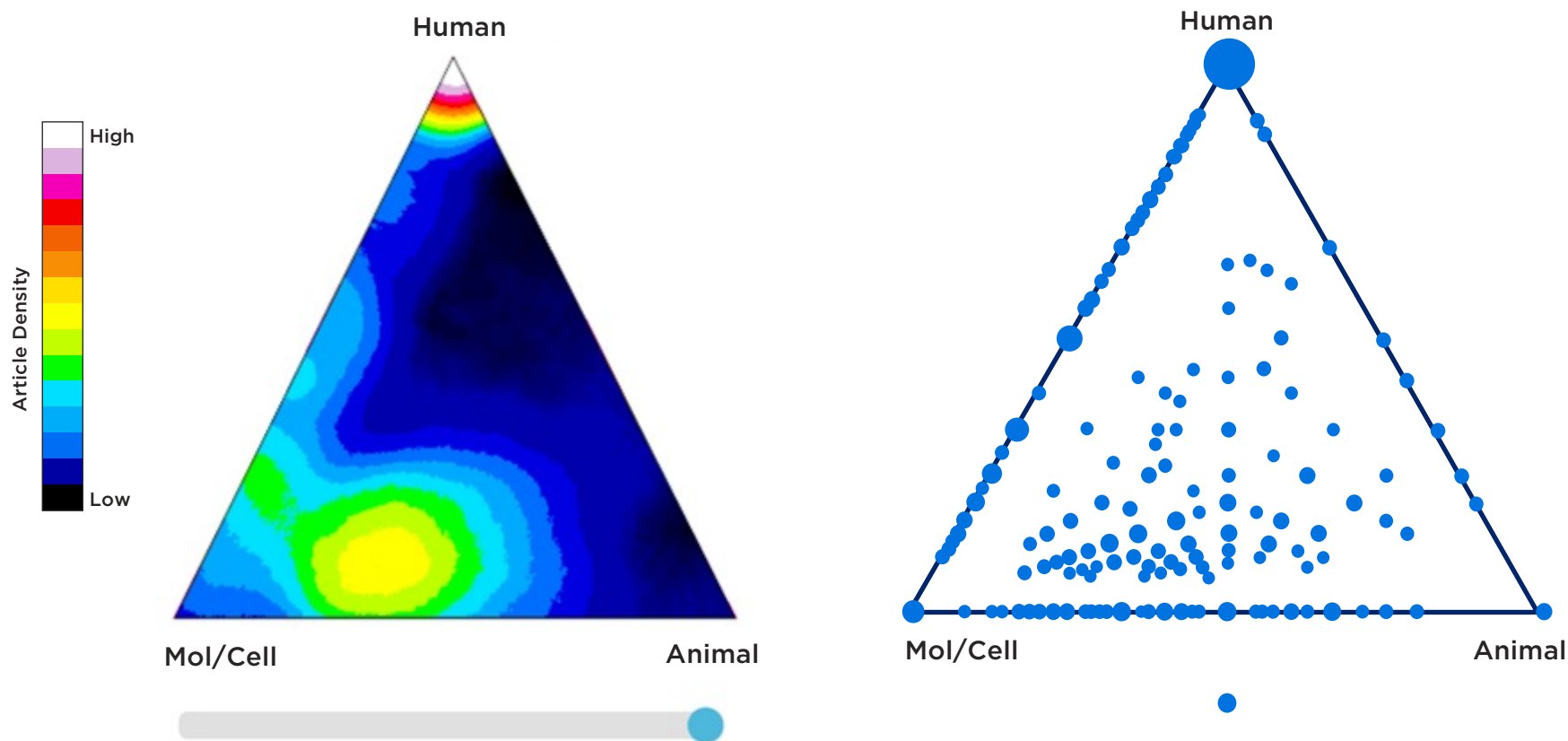
Open Access

Types of Open Access Journals

Gold: fully open access journals with a Creative Commons (CC) license.

Hybrid gold: journals which provide authors the choice of publishing open access and including the Creative Commons license.

Bronze: the publisher has chosen to provide temporary or permanent free access to the publication, but without a license.

Green: published or manuscript version accepted for publication and available on an institutional or other open repository.

There are funders that require researchers to publish in open access journals, and the following may be of interest to MSK authors:

Howard Hughes Medical Institute (HHMI) Open Access to Publications Policy - Effective January 1, 2022, research manuscripts with major contributions from an HHMI lab will need to be published in open access journals and made immediately and freely available. The policy requires that HHMI researchers publish under a CC BY 4.0 license. Additional information can be found here.
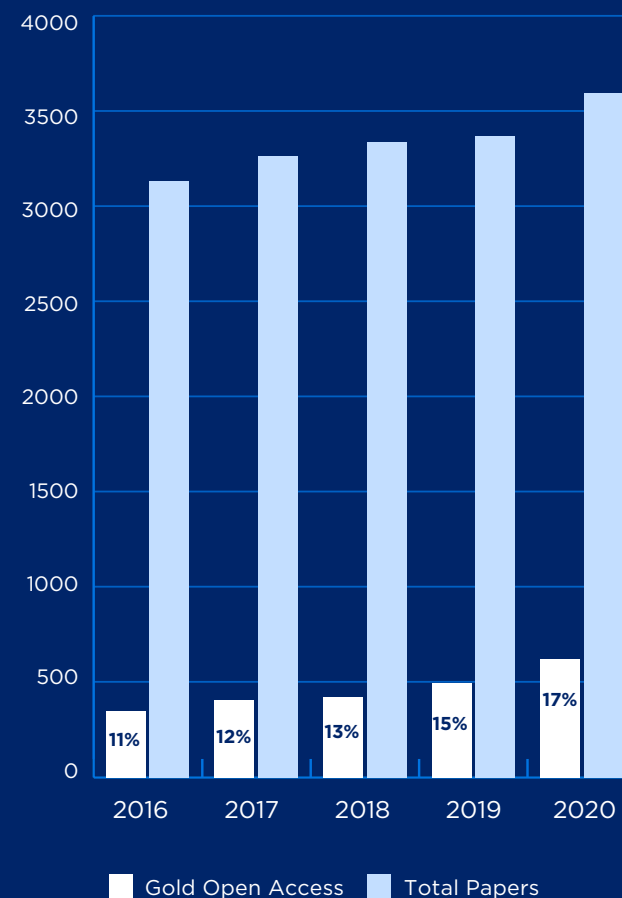
NCI Cancer Moonshot℠ Public Access and Data Sharing Policy

US National Science Foundation Public Access Plan

Also view FAQs for Public Access

The number of papers MSK authors contributed to fully open access journals has **increased slightly, but steadily,** over this 5-year period.
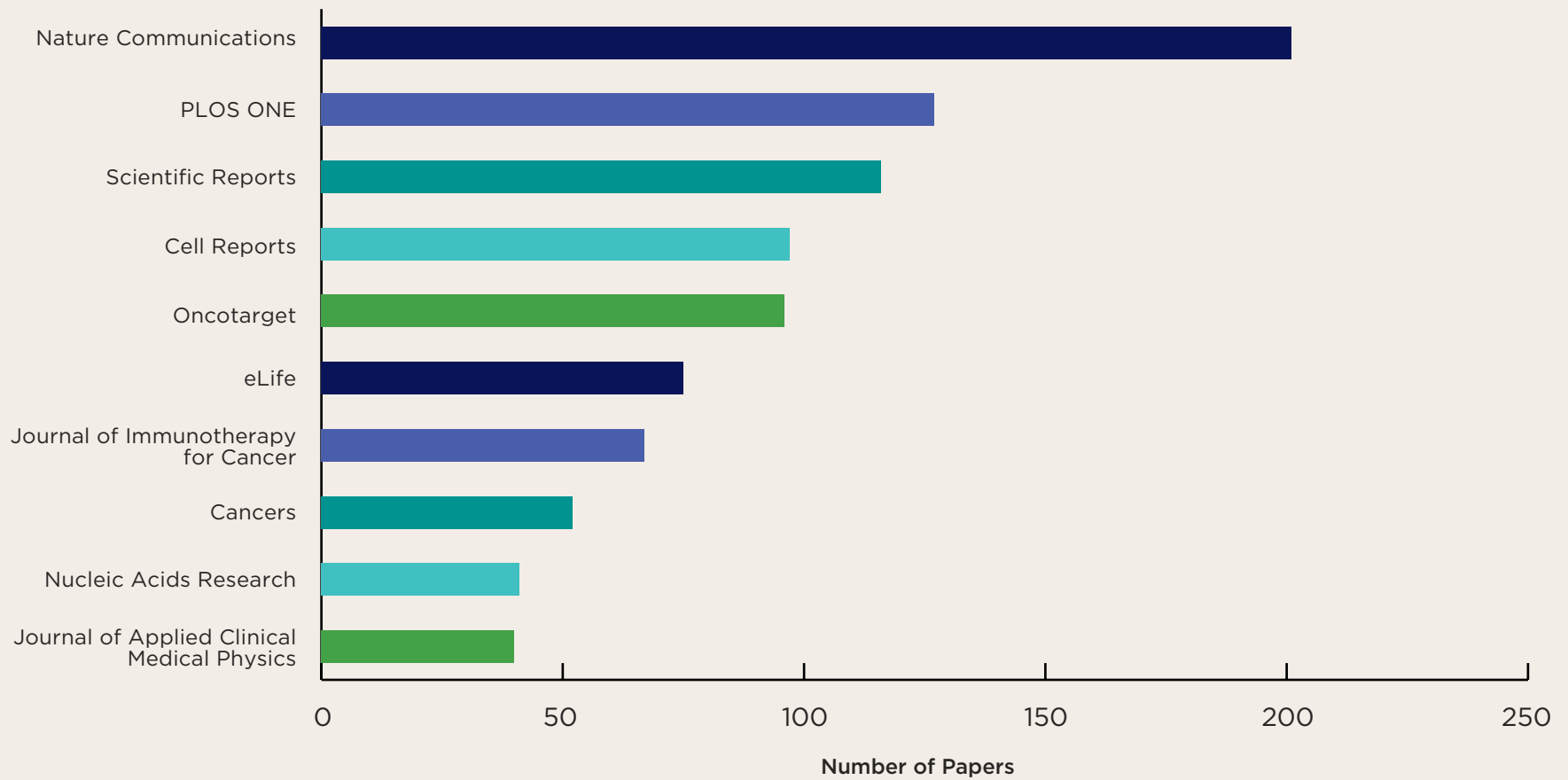
Gold Open Access

Bar chart showing Gold Open Access and Total Papers from 2016 to 2020. Y-axis ranges 0 to 4000. Gold Open Access percentages: 2016: 11%, 2017: 12%, 2018: 13%, 2019: 15%, 2020: 17%.
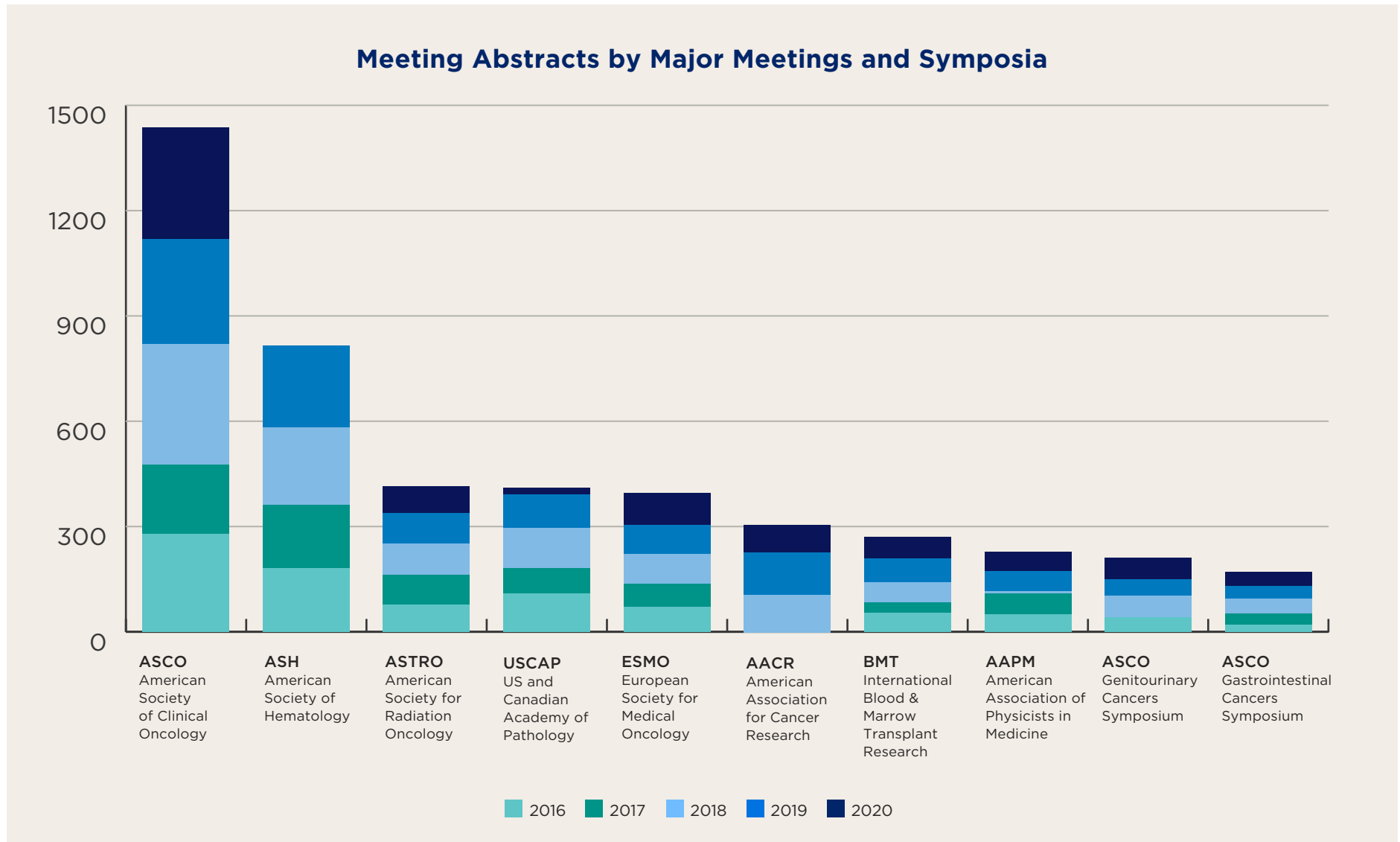
Legend: ☐ Gold Open Access ☐ Total Papers

*Data retrieved from Scopus® (Elsevier) 10/13/2021

These **10 journals** account for **40% of the Gold Open Access papers** published from 2016-2020.

Gold Open Access Journals with the Most MSK Papers, 2016-2020

Number of Papers

Meeting Abstracts

Meeting Abstracts by Major Meetings and Symposia
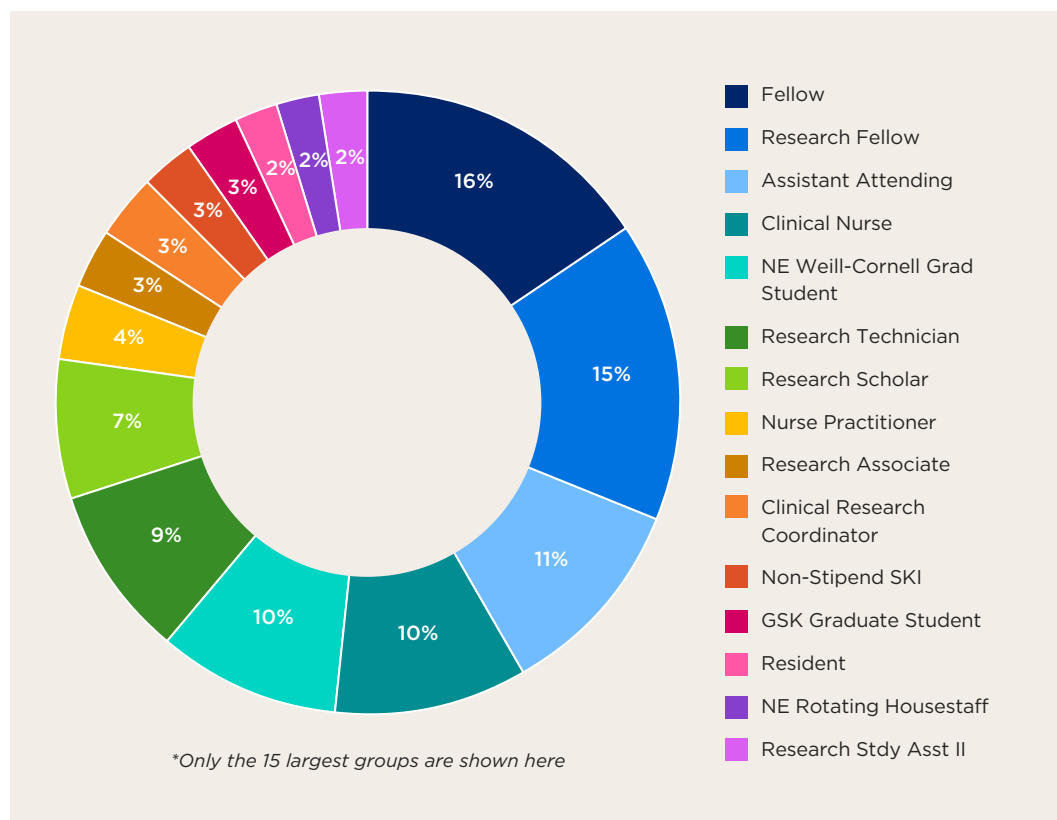

2016	2017	2018	2019	2020

Note: some years are missing data due to inconsistencies with abstracts and their identifying metadata such as author affiliation being reported/indexed in citation databases.
Data retrieved from Synapse 8/24/2021

Over this 5-year period, we added almost **3,000 new author profiles** to Synapse.

Our authors represent contributions across the entire span of MSK's departments and services. Here, we break down the new authors by their job titles.

New Authors Added to Synapse

- Fellow
- Research Fellow
- Assistant Attending
- Clinical Nurse
- NE Weill-Cornell Grad Student
- Research Technician
- Research Scholar
- Nurse Practitioner
- Research Associate
- Clinical Research Coordinator
- Non-Stipend SKI
- GSK Graduate Student
- Resident
- NE Rotating Housestaff
- Research Stdy Asst II

Only the 15 largest groups are shown here

28% of active MSK employees have registered their ORCID iD with the MSK Library via ORCID@MSK.

Having an ORCID iD ensures that your works are associated with you and not someone else - especially someone with a similar name.

Learn more and register for your ORCID iD.

ORCID iD

Legend: ■ # Active Employees with ORCID Profile   ■ % Current Employees
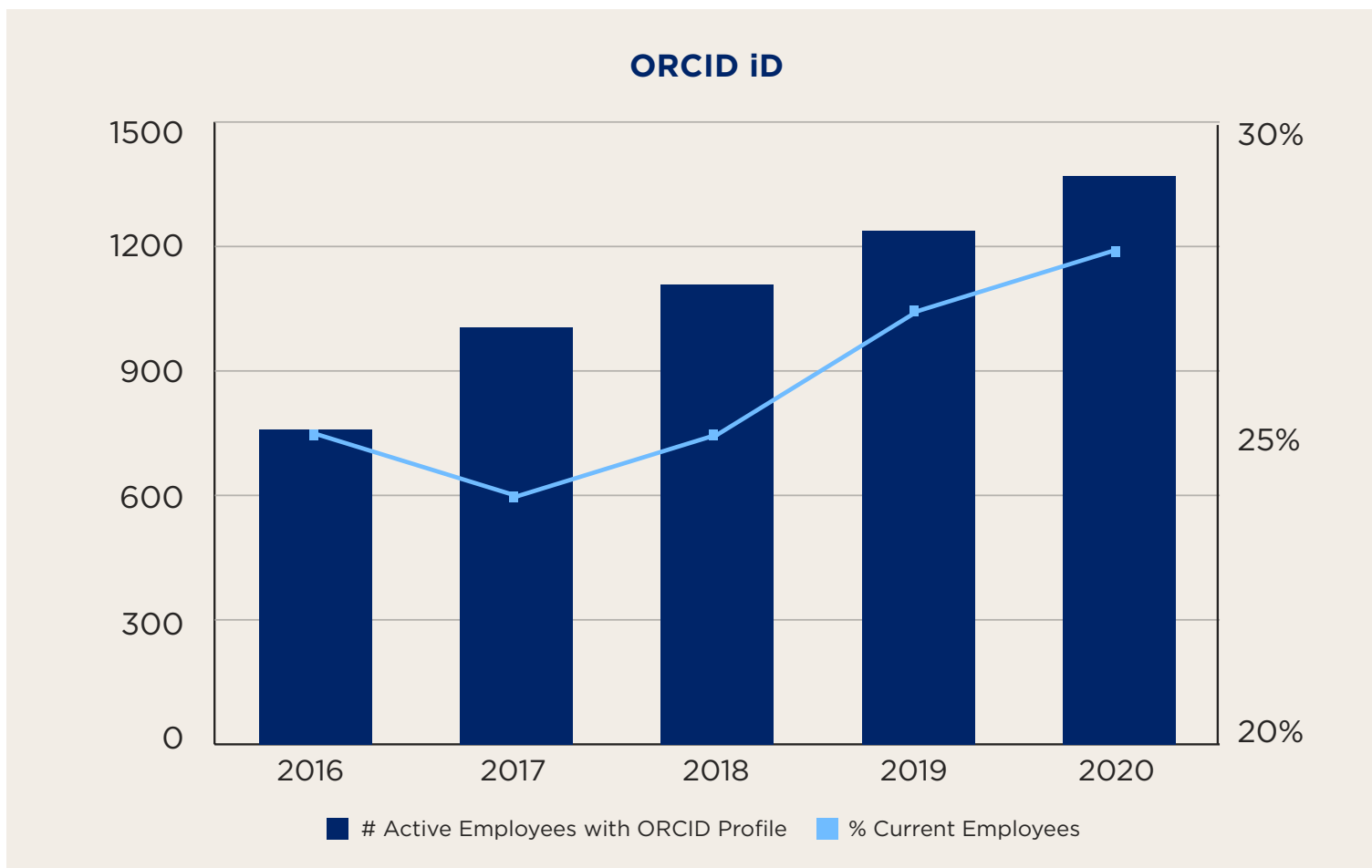
Research Collaborations

VOSviewer is a software tool for constructing and visualizing bibliometric networks. Those networks can include relationships between authors, journals, institutions, or keywords/subjects, and are based on citation information from research publications.

Here, we used the citations of all MSK-authored publications from 2016-2020 and analyzed the author affiliations of all co-authors using the VOSviewer tool. The map on the following page shows the 60 organizations our MSK authors most frequently collaborated with on publications. The larger the circles (nodes), the greater the frequency of occurrence of shared authorships with MSK. The different colors further indicate the clusters of organizations that appeared most frequently together on papers. Looking at the blue group in our map, you will see that publications co-authored with Weill Cornell will also frequently have authors from Columbia, NYU, Rockefeller University, and Mount Sinai. Publications clustered in the purple group frequently include authors from Harvard, Brigham & Womens, Dana Farber, etc.

[Learn more about the VOSviewer tool](#)
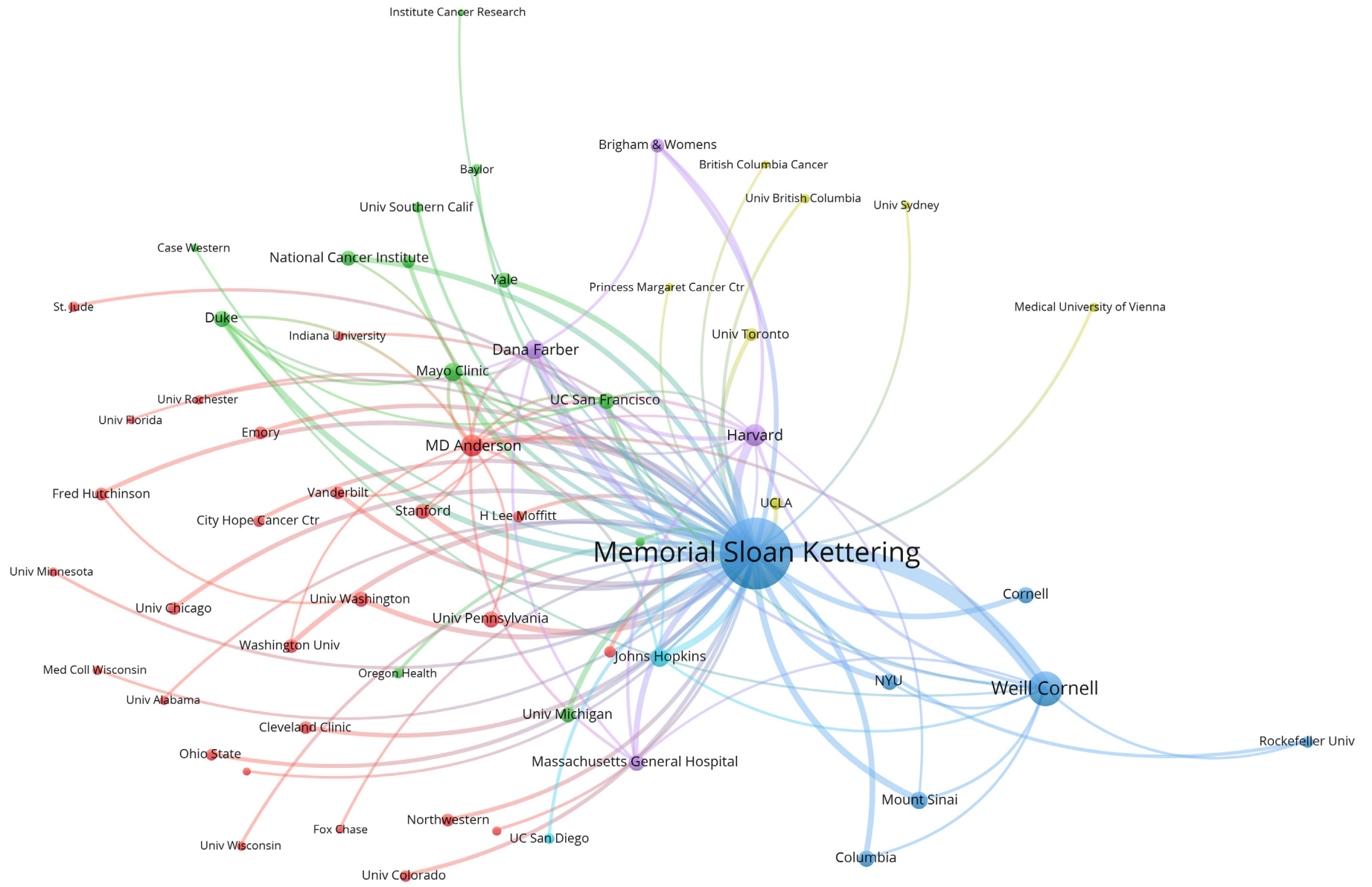
Research Data Management in the Library

The MSK Library launched a new Research Data Management program in 2019 in response to a 2015 joint survey by the MSK and Rockefeller University libraries, global trends in the role of libraries in research data management, and increasing data reporting and sharing requirements from funders and publishers. The program was founded on the principles of promoting the global standards of Findability, Accessibility, Interoperability, and Reuse in MSK research data. In light of NIH's forthcoming Data Management and Sharing Policy (January 2023), staff are designing new services to proactively support MSK researchers throughout the lifecycle of their projects, reduce the administrative burden required to comply with new regulations, and provide educational outreach so MSK can operate at the forefront of best practices in data management.

Our implementation components fall into the following categories:

- Data discovery and library platform integration

- Internal and external relationship building

- Engagement with our researchers

- Maintaining and enhancing our services over time

Thus far, our program has specifically included:

- Launching a data catalog with MSK-specific integrations and enhancements

- Data management planning assistance

- DOI minting support

- Developing strong ties within the institution and the RDM library community

[Click for More Information about the MSK Library Research Data Management Program](#)

Research Data Trends

Given the newness of the MSK Library's Research Data Management program and the fast-paced development of the field in general, this section of the report will take a broader view of research data trends by MSK researchers over the past 10 years, with a specific interest in the intersection of datasets and publications.

Trends in Dataset Submission to Public Repositories by MSK Researchers

These graphs show similar best-fit lines indicating increasing deposits of datasets in repositories by MSK researchers over a 10-year span.

Deposits are in the discipline-specific NCBI Gene Expression Omnibus repository and the discipline-agnostic Figshare repository.

These trends are likely indicative of industry-wide growth and changes related to the role of data in research.

MSK Submissions to the GEO

Figshare Deposits by MSK Researchers
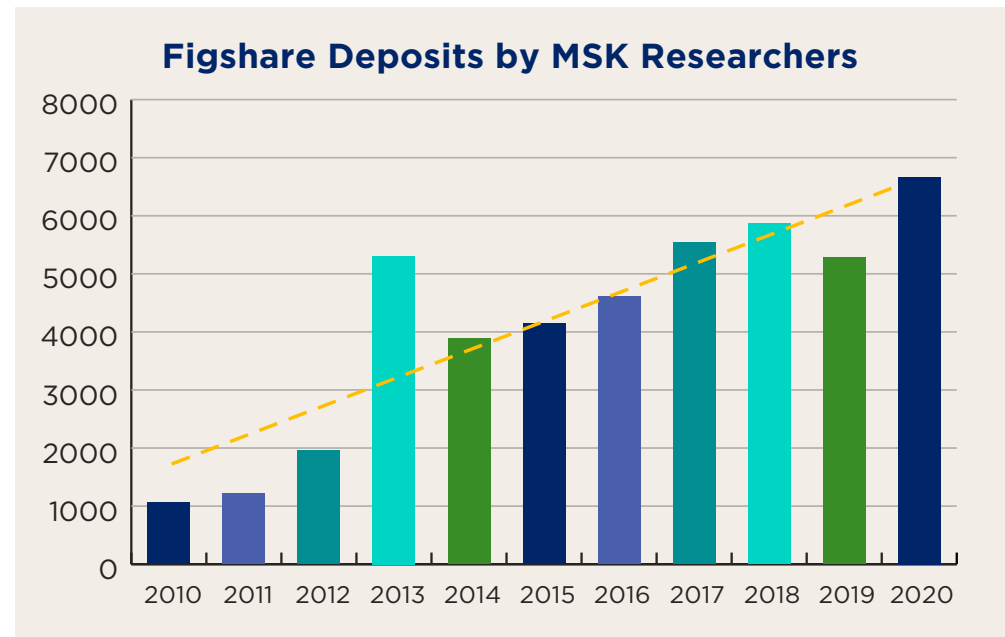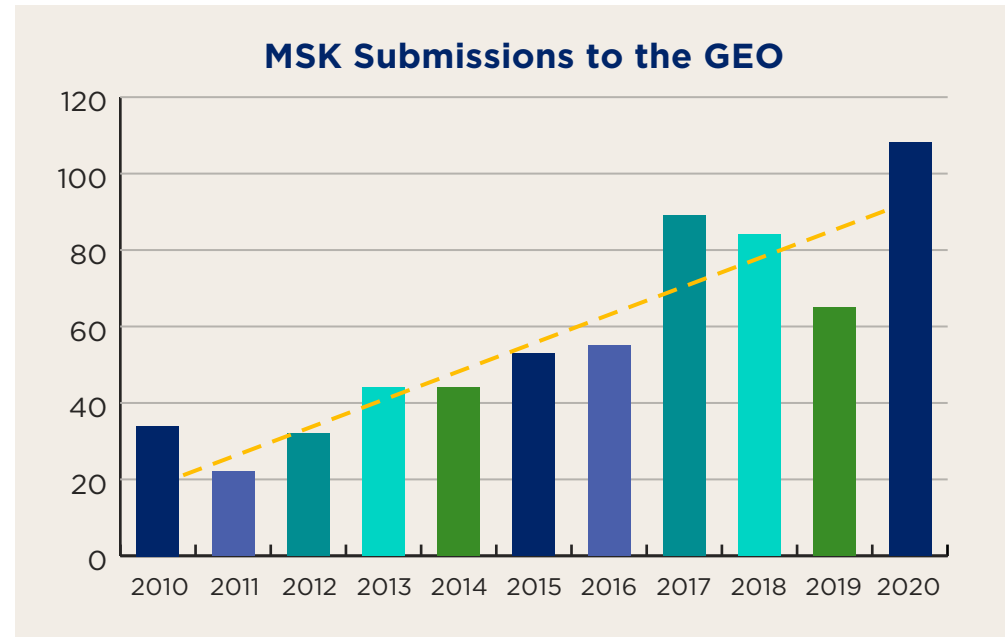
Publications and Associated Data

This graph indicates a steady increase in the percentage of PubMed Central® (PMC) articles by MSK-affiliated authors with associated data identified in the publication.

This is likely due to changing requirements by publishers to include data availability statements, as well as the methods by which publishers and discovery services identify data in publications.

% of MSK Publications in PMC with Associated Data
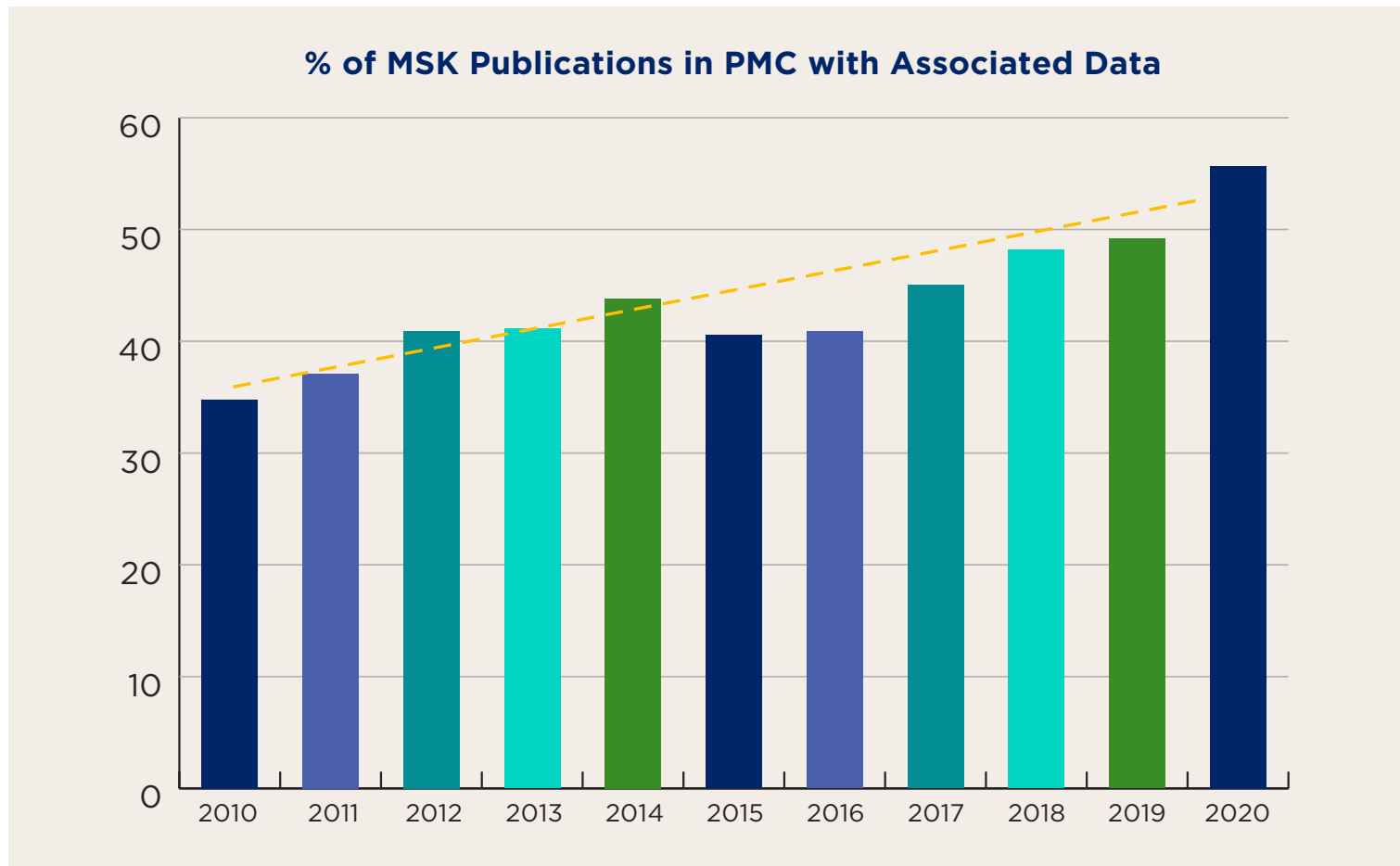
Bibliometrics Resources At-A-Glance

Article-Level Metrics

- [Web of Science™ (Clarivate)](#) – Citation counts, Web of Science™-specific usage counts/metrics

- [Scopus (Elsevier)](#) – Citation counts, PlumX Metrics, Scopus-specific view counts/metrics

- [Google Scholar](#) – Citation counts

- [Synapse (MSK Library)](#) – Citation counts (based on Dimensions data), Altmetrics

- [Dimensions (Digital Science)](#) – Citation counts, Altmetrics

Journal-Level Metrics

- [Journal Citation Reports (Clarivate)](#) – Journal Impact Factor (JIF), Eigenfactor score, and other metrics (based on Web of Science™ data)

- [Scopus Sources (Elsevier)](#) – CiteScore™ metrics, SCImago Journal Rank (SJR), and other metrics (based on Scopus data)

- [Google Scholar Metrics](#) – h5-index, h5-median (based on Google PageRank™ data)

- [Synapse (MSK Library) Publications](#) – Number of MSK-authored works

- [Scimago Journal & Country Rank](#) – H-Index, SCImago Journal Rank (SJR) (based on Scopus and Google PageRank™ data)

Author-Level Metrics

- [Web of Science™ (Clarivate) Researcher Profile](#) – H-Index (based on Web of Science™ data)

- [Scopus (Elsevier) Author Profile](#) – H-Index (based on Scopus data)

- [Google Scholar Author Profile](#) – H-Index, i10-index (based on Google PageRank™ data)

- [Synapse (MSK Library) Author Research Activity Dashboard](#) – H-Index (based on Scopus data)

- [Dimensions (Digital Science) Researchers](#) – Numbers of works, Total citations

Please note that Journal Citation Reports, Scopus, and Web of Science are subscription-based resources that require a personal or institutional subscription to access.