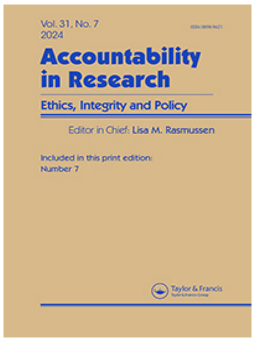
Author/s:
Hamilton, DG;Page, MJ;Everitt, S;Fraser, H;Fidler, F

Title:
Cancer researchers’ experiences with and perceptions of research data sharing: Results of a cross-sectional survey

Date:
2025-01-01

Citation:
Hamilton, D. G., Page, M. J., Everitt, S., Fraser, H. & Fidler, F. (2025). Cancer researchers’ experiences with and perceptions of research data sharing: Results of a cross-sectional survey. Accountability in Research, 32 (4), pp.530-557. https://doi.org/10.1080/08989621.2024.2308606.



Accountability in Research
Ethics, Integrity and Policy

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/gacr20

Cancer researchers’ experiences with and perceptions of research data sharing: Results of a cross-sectional survey

Daniel G. Hamilton, Matthew J. Page, Sarah Everitt, Hannah Fraser & Fiona Fidler

To cite this article: Daniel G. Hamilton, Matthew J. Page, Sarah Everitt, Hannah Fraser & Fiona Fidler (01 Feb 2024): Cancer researchers’ experiences with and perceptions of research data sharing: Results of a cross-sectional survey, Accountability in Research, DOI: 10.1080/08989621.2024.2308606

To link to this article: https://doi.org/10.1080/08989621.2024.2308606

ACCOUNTABILITY IN RESEARCH, 2024
https://doi.org/10.1080/08989621.2024.2308606

Taylor & Francis
Taylor & Francis Group

Cancer researchers' experiences with and perceptions of research data sharing: Results of a cross-sectional survey

Daniel G. Hamilton [a,b], Matthew J. Page [c], Sarah Everitt[d], Hannah Fraser [a], and Fiona Fidler [a,e]

[a]MetaMelb Research Group, School of BioSciences, University of Melbourne, Melbourne, Australia; [b]Melbourne Medical School, Faculty of Medicine, Dentistry & Health Sciences, University of Melbourne, Melbourne, Australia; [c]Methods in Evidence Synthesis Unit, School of Public Health & Preventive Medicine, Monash University, Melbourne, Australia; [d]Sir Peter MacCallum Department of Oncology, University of Melbourne, Melbourne, Australia; [e]School of History & Philosophy of Sciences, University of Melbourne, Melbourne, Australia

ABSTRACT

Background: Despite wide recognition of the benefits of sharing research data, public availability rates have not increased substantially in oncology or medicine more broadly over the last decade. Methods: We surveyed 285 cancer researchers to determine their prior experience with sharing data and views on known drivers and inhibitors. Results: We found that 45% of respondents had shared some data from their most recent empirical publication, with respondents who typically studied non-human research participants, or routinely worked with human genomic data, more likely to share than those who did not. A third of respondents added that they had previously shared data privately, with 74% indicating that doing so had also led to authorship opportunities or future collaborations for them. Journal and funder policies were reported to be the biggest general drivers toward sharing, whereas commercial interests, agreements with industrial sponsors and institutional policies were the biggest prohibitors. We show that researchers' decisions about whether to share data are also likely to be influenced by participants' desires. Conclusions: Our survey suggests that increased promotion and support by research institutions, alongside greater championing of data sharing by journals and funders, may motivate more researchers in oncology to share their data.

ARTICLE HISTORY
Received 20 September 2023
Accepted 18 January 2024

KEYWORDS
data sharing; survey; oncology; cancer; open science

Introduction

Cancer is a major health burden and leading cause of premature death worldwide. Based on estimates from the World Health Organisation in

---

CONTACT Daniel G. Hamilton hamilton.d@unimelb.edu.au MetaMelb Research Group, School of BioSciences, University of Melbourne, Room 602, West Wing, Arts West Building 148, Grattan St, Melbourne, Victoria 3010, Australia

Supplemental data for this article can be accessed online at https://doi.org/10.1080/08989621.2024.2308606

2019, cancer was one of the top three leading causes of premature death in all but 6 of the 183 countries assessed (Bray et al. 2021). Furthermore, based on recent trends, cancer is expected to overtake cardiovascular disease as the leading cause of death in most countries this century (Bray et al. 2021). It is therefore no surprise that approximately USD $24.5 billion was directed toward cancer-related research between 2016 and 2020 globally, with almost three quarters of this sum being invested in pre-clinical research (McIntosh et al. 2023). However, despite the large investment of public and philanthropic funds into cancer research worldwide, significant concerns about the reliability of published findings have been raised following the repeated inability to replicate the key results of promising pre-clinical research (Begley and Lee 2012; Errington et al. 2021; Prinz, Schlange, and Asadullah 2011).

Irreproducible and irreplicable medical research is a major cause of research waste and has been previously estimated to cost the United States alone more than USD $28 billion each year (Freedman, Cockburn, and Simcoe 2015). One of the major contributors to research waste in medicine is the unavailability of important research information, including the underlying data (Chan et al. 2014). In oncology, there are many examples of how sharing and using publicly available research data has advanced scientific knowledge and improved patient outcomes. One such example includes a series of pioneering prospective individual participant data (IPD) meta-analyses performed in the 1980s and 1990s by the Early Breast Cancer Trialists' Collaborative Group (EBCTCG), which greatly reshaped our understanding of the safety and efficacy of many treatment regimens for breast cancer (Kawahara et al. 2018). However, despite the oncology research community valuing greater access to research data, previous audits of the literature have shown that public sharing of research data is uncommon. For example, a study by Walters et al. (2019) observed that only 2% of 194 sampled cancer-related articles published between 2014 and 2018 shared all data required to reproduce their findings publicly. Related work by Hamilton et al. (2022) found that 16% of cancer-related articles published in 2019 shared some data publicly, but less than 1% shared in accordance with widely accepted guidance on how to ensure data are maximally findable, accessible, interoperable, and reusable to others (i.e., the FAIR Guiding Principles (Wilkinson et al. 2016, 2019)).

Given the low frequency with which data are shared in oncology (Hamilton et al. 2022; Walters et al. 2019), medicine in general (Hamilton et al. 2023) and most scientific disciplines more broadly (Serghiou et al. 2021), many surveys of researchers, as well as other key stakeholders (e.g., funders, publishers, research participants) have been performed to improve our understanding of drivers and inhibitors of data sharing (Fecher et al. 2015; Hamilton et al. 2023; Howe et al. 2018; Thoegersen and Borlund 2021; Zuiderwijk et al. 2020). For example, a systematic review and thematic analysis of surveys of researchers' attitudes toward data sharing by Zuiderwijk et al. (2020) showed that numerous factors play a role in researchers' decisions to share or withhold their data. The most frequently identified drivers of sharing related to personal and intrinsic motivations (e.g., positive attitudes toward the Open Science movement) and academic benefits and recognition (e.g., increased impact of the research). On the other hand, commonly observed inhibitors related to navigating relevant legislations and regulations (e.g., protecting confidentiality), time and resource constraints, and concerns about negative impacts on academic performance (e.g., loss of ability to publish further analyses, misuse or misinterpretation of shared data). The findings of these reviews have been substantiated by contemporary surveys of researchers in other fields such as psychology (Houtkoop et al. 2018), dentistry (Spallek et al. 2019), and food science and technology (Melero and Navarro-Molina 2020), as well as authors of systematic reviews (Nguyen et al. 2023) and clinical trials (Tan et al. 2021; Tannenbaum et al. 2018).

Much of the previous research has directly or indirectly surveyed medical and health researchers (Bosserdt, Hamm, and Dewey 2019; Campbell et al. 2002; Genevieve et al. 2014; Nguyen et al. 2023; Rathi et al. 2012, 2014; Spallek et al. 2019; Tan et al. 2021; Tannenbaum et al. 2018; Tenopir et al. 2011, 2015). Some studies also happen to include subpopulations of cancer researchers (Campbell et al. 2002; Tan et al. 2021). However, no such study has directly engaged cancer researchers. This study addresses this gap, and specifically surveys cancer researchers about their prior experiences with sharing data and their views on previously identified drivers and inhibitors of data sharing. We also take this opportunity to explore whether factors previously proposed to moderate respondents' data sharing behaviors (e.g., career stage (Campbell et al. 2002; Tenopir et al. 2015), collecting genomics data (Zuiderwijk et al. 2020), studying non-human participants (Tenopir et al. 2015), performing qualitative research (Houtkoop et al. 2018)) were associated with prior experience sharing data and support of the practice.

Materials and methods

Survey design

We used a cross-sectional survey design to capture cancer researchers' experiences with, and views on sharing research data. The survey contained 32 questions (20 main and 12 auxiliary) separated into seven sections.

- Section 1: Collected relevant demographic information from participants including: country of residence, gender, research sector, cancer research area, whether respondents regularly work with human omics data, the type of research participants they typically include in their studies, number of publications, and clinical, research and grant reviewing experience.

- Section 2: Outlined our definition of research data, which was defined as quantitative (countable), or qualitative (textual, visual, or audio) information collected at the level of the subject of research (e.g., human participant, animal, cell) that would enable others to verify or regenerate any published findings, including data visualizations.
- Section 3: Characterized details on respondents' most recent published article for which they were the person responsible for the data (e.g., did they collect information from human participants, who owned the data, were any data shared publicly, and if so, why).
- Section 4: Characterized respondents' general experiences with sharing data and code publicly and privately, as well as whether private requests for data had resulted in authorship opportunities or future collaborations.
- Section 5: Characterized respondents' knowledge about current sharing rates in oncology and the FAIR principles (Wilkinson et al. 2016; Wilkinson et al. 2019).
- Section 6: Characterized respondents' perceptions of the value of data sharing, confidence in their archival practices and the usefulness of research data management training.
- Section 7: Examined the influence of various medical research stakeholders on respondents' data sharing practices. This section also characterized respondents' views on Houtkoop and colleagues' (2018) "fear- and non-fear-related barriers" to sharing, where "fear" refers to negative anticipated consequences, such as being scooped, secondary users misinterpreting shared data, or the detection of errors.

The survey included single- and multi-answer multiple-choice questions. Where possible, we used previously validated instruments relevant to an oncology setting to measure key demographics (e.g., the International Cancer Research Partnership's (ICRP) Common Scientific Outline (CSO) classification system for cancer research area). Furthermore, to allow comparison with relevant research, as well as to avoid extensive piloting phases, we adapted some definitions, terminologies, and questions from previous surveys in the medical, social and life sciences (refer to Table S1 for further information). The complete survey transcript is publicly available on the project's Open Science Framework page (Hamilton 2023).

Sampling frame

The target sample for this research were corresponding authors of cancer-related scholarly publications indexed in Clarivate Analytics' Web of Science Core Collection. To identify potential survey participants, we systematically searched Web of Science on 1 December 2020, for cancer-related articles indexed after 1 January 2020, using the following pragmatic search strategy: (SU=Oncology AND PY = 2020) AND (LA=English AND DT=Article). In appreciation of previous research reporting that almost a quarter of corresponding author e-mail addresses become invalid within a year of being published, we selected this date range to minimize the chance that contact addresses had expired. Search records (article WoSID and DOI) and associated meta-data (corresponding author name, first affiliated country, and e-mail address) were extracted and imported (Wren et al. 2006) into R for cleaning (e.g., missing data checks, e-mail format checks), and de-duplication. If multiple corresponding authors were listed, the first reported address was used.

Survey piloting and distribution

The survey was pilot tested by ten active cancer researchers of varying research backgrounds. Piloters provided feedback on the language, flow, and time taken to complete the survey. Following completion of the survey piloting phase, 7,600 randomly selected corresponding authors were invited to participate. Survey invitations were sent in three phases: a test phase (Phase I) to ensure the survey worked as expected (14 December 2020; $N =$ 100 invitations), a preliminary phase (Phase II) to estimate the likely response rate (5 January 2021; $N = 400$ invitations), and the main phase (Phase III; 9 February 2021; $N = 7{,}100$ invitations). Throughout the course of the project, three (Phase I and II) and five (Phase III) follow up reminder e-mails were sent at 4-week intervals to those who had not completed the survey (and not opted-out of future e-mail reminders) following the initial invitation. Qualtrics Solutions' Online Survey Software (Qualtrics, Provo, UT) was used to create and host the survey. The Qualtrics Mailing service was used in all phases to send out personalized e-mail invitations asking researchers to participate in a survey to share their views on research data sharing. E-mails in the first two phases that were reported to have bounced were checked and replaced if another address could be sourced (bounced e-mails were not investigated in Phase III due to resource limitations). Several features were also incorporated into the survey to enhance understanding of the content (e.g., tooltips), minimize data entry errors (e.g., response validation criteria) and prevent or identify possibly fraudulent submissions (e.g., personalized links). The use of personalized links to the survey also allowed tracking of response rates and enabled sending of reminder e-mails.

Statistical analysis

Given the exploratory nature of the project, we aimed to obtain at least 380 responses to our main question of interest which asked whether respondents agreed or disagreed with the statement “My cancer research community should promote and facilitate the sharing of research data” to ensure a 95% confidence interval width less than 10% using the modified Wilson method for calculation of simultaneous confidence intervals for multinomial proportions (Brown, Cai, and DasGupta 2001). Consequently, assuming a conservative response rate of 5% (Christian and Hoover 2017; Ho et al. 2013), we invited 7,600 corresponding authors in total.

Simple descriptive statistics have been used to analyze all quantitative survey questions, with categorical data presented as counts and percentages, and continuous data presented as median and interquartile ranges (IQR). For simplicity, we report the cumulative percentage of positive response options for the questions utilizing 5-point and 7-point Likert scales (e.g., the sum of “somewhat agree,” “agree” and “strongly agree” responses). Simple unadjusted risk ratios and 95% confidence intervals were also calculated using the epitools package (Aragon 2020) in R to explore whether gender, research area, research experience, typical research subjects, data analysis strategy, experience working with omics data, clinical experience and grant reviewing experience were associated with experience sharing data and support for data sharing. In a further analysis, multiple logistic regression models were generated using the glm function in R to estimate adjusted odds ratios and 95% confidence intervals for some predictors (e.g., grant reviewing experience) while controlling for possible confounding effects of other variables (e.g., research experience).

Data cleaning, visualization and analysis was performed in R (R Foundation for Statistical Computing, Vienna, Austria, v4.2.1). We did not impute missing data, nor adjust for potential non-representativeness of the sample. Missing responses have been omitted from all reported results. De-identified data, code, and materials are publicly available on the project’s OSF page (Hamilton 2023). The findings of this study are also reported in accordance with the Checklist for Reporting of Survey Studies (CROSS) guidelines (Table S2) (Sharma et al. 2021).

Ethics approval

The study protocol was reviewed and approved by the University of Melbourne’s LNR 2D Human Ethics Committee (Project ID: 2021-

14496-17081-3) prior to survey piloting and recruitment. All participants provided informed consent prior to taking part in the survey. The protocol for this study was not publicly registered prior to data collection.

Results

Survey overview

Our Web of Science search retrieved 45,701 records. Of these, 45,257 listed a corresponding author's e-mail address, 37,405 of which were unique e-mail addresses. Of the random sample of 7,600 invitations that were sent, 341 bounced (4.5%) leaving 7,259 successful invitations. In the period between invitation and deactivation, 317 unique participants entered the survey, of which 313 consented to participate. Of the 313 participants that consented, 249 completed the survey, 36 started but did not finish, and 28 started the survey but did not provide any responses, giving a response rate of 317/7259 (4.4%) and a completion rate of 249/317 (79%). The median completion time was 13 minutes (IQR: 9 to 23 minutes). The number of responses collected at each stage of the study is shown in Figure S1. We also note that response rates ranged between 1 and 18% for the 20 countries with more than 50 invitees, with the poorest response rate occurring among researchers affiliated with Chinese institutions (0.7%, $N = 15/2259$) and the best for those affiliated with Australian institutions (18%, $N = 22/119$) (refer to Table S3 for further details). Information on response rates according to the journal and publisher are also available on the project's OSF page (Hamilton 2023).

Participant characteristics

Details about both the invitees and the participants are reported in Table 1. Slightly more than half of participants were men (55%, $N = 155/283$), and most were affiliated with institutions located in Europe and Central Asia (35%, $N = 100/285$), North America (28%, $N = 81/285$), and East Asia and the Pacific (20%, $N = 56/285$). Most respondents conducted their research in non-industry settings (e.g., academic institutions, hospitals, government agencies) (89%, $N = 250/281$), typically included human participants in their studies (88%, $N = 246/280$) and did not routinely work with human genomic data (74%, $N = 207/281$). Respondents had a median of 12 years research experience (IQR: 7 to 20 years) and had coauthored a median of 40 academic publications (IQR: 15 to 100 publications). Approximately half the respondents had reviewed for a major research funder within the past three

Table 1. Characteristics of the surveyed and invited participants.

	Respondents		Invitees	
	N	%	N	%
Gender*				
Male	155	55%	4048	60%
Female	121	43%	2645	40%
Prefer not to say	7	2%	-	-
Location of respondent				
North America	81	28%	1730	23%
Latin America & Caribbean	10	4%	164	2%
Europe & Central Asia	100	35%	1913	25%
Middle East & North Africa	16	6%	305	4%
Sub-Saharan Africa	5	2%	27	0%
East Asia & Pacific	56	20%	3305	43%
South Asia	17	6%	156	2%
Research sector				
Non-industry	250	89%	-	-
Industry	3	1%	-	-
Both industry and non-industry	10	4%	-	-
Independent researcher	18	6%	-	-
Cancer research area				
Biology	47	17%	-	-
Aetiology	8	3%	-	-
Prevention	10	4%	-	-
Detection, diagnosis and prognosis	46	17%	-	-
Treatment	87	31%	-	-
Control, survivorship and outcomes	79	29%	-	-
Research participants typically studied†				
Human participants	246	88%	-	-
Animals (in vivo)	60	21%	-	-
Cells and tissues from humans or animals	88	31%	-	-
Bacteria or viruses	12	4%	-	-
Other	1	0%	-	-
Routinely works with omics data				
Yes	74	26%	-	-
No	207	74%	-	-
Research experience (years)				
Median (IQR)	12	7 to 20	-	-
Number of academic publications				
Median (IQR)	40	15 to 100	-	-
Grant reviewing experience^				
Yes	145	52%	-	-
No	136	48%	-	-
Clinical experience				
Yes	153	54%	-	-
No	128	46%	-	-

*Genderize.io was used to guess the gender of invitees based on their first name. (Note: 907 were unable to be classified.)†Percentages do not add to 100% (multiple answers possible). ^Reviewed for amajor research funder within the previous three years.

years (52%, $N = 145/281$), and had experience working with patients in a clinical setting (54%, $N = 153/281$).

Data sharing practices

We found that 45% of respondents ($N = 110/243$) reported that some of the research data from their most recent publication were shared publicly (see Table 2). For context, of these respondents, most stated that their study

Table 2. Data sharing outcomes from participants' most recent empirical publication where they were the data guarantor.

	N	%
Some data shared publicly (*N* = 243)		
Yes	110	45%
No	133	55%
Reasons why researchers who shared data publicly did so (*N* = 107)*		
Journal requirement	69	64%
To promote transparency and reproducibility	57	53%
To increase the impact of the research	42	39%
Funder requirement	32	30%
Academic benefits or recognition	30	28%
Institutional/research group culture	28	26%
Institutional requirement	14	13%
Other	3	3%
Data location (*N* = 107)*		
Article supplementary materials	66	62%
Online data repository	36	34%
Institutional or personal website	9	8%
Within the published article	51	48%
Paper in a data journal	21	20%
Other	4	4%
For researchers who did not share data, were they able to share (*N* = 131)*		
Yes	55	42%
No	32	24%
Unsure	44	34%
Reasons why researchers who could share or were unsure did not (*N* = 95)*^		
To protect the confidentiality of participants	38	40%
Lack of time to prepare data for sharing	25	26%
Concerns about misuse or misinterpretation	24	25%
To protect my ability to publish further analyses	23	24%
Never considered sharing before	21	22%
Lack of training in how to share data	20	21%
Unsure of employer or funder policy	18	19%
To protect commercially sensitive information	10	11%
Concerns about not being attributed	6	6%
Other	15	16%
Reason for inability to share (*N* = 31)*		
Institutional review board restrictions	8	26%
Patient consent was not obtained	7	23%
Anonymity could not be guaranteed	5	16%
Stakeholder interests	3	10%
Funder/advisor did not allow	1	3%
Other	7	23%
Would share data during peer review with the following party (*N* = 126)*		
A peer reviewer	28	22%
A journal editor	40	32%
A journal-appointed intermediary	15	12%
Would not share with any of the above	43	34%

*Total number of respondents are different due to participants leaving the survey between questions. ^Percentages do not add to 100% (multiple answers possible).

collected information from human participants (88%, $N = 212/242$), and used quantitative (51%, $N = 124/241$) or mixed methods (42%, 102/241) approaches to analyze the data. When asked who owned the data, 51% stated their research institution did ($N = 121/237$), 15% reported that they owned the data ($N = 36/237$) and 13% did not know ($N = 31/237$). The remaining respondents reported that the data were owned by their collaborator or department (each 6%, $N = 15$), the funder (3%, $N = 7$) or the publisher (2%, $N = 4$).

When asked why they shared data, multiple combinations of reasons were given. The top five reasons included: journal requirements (64%, $N = 69$), to promote transparency and reproducibility (53%, $N = 57$), to increase the impact of the research (39%, $N = 42$), funder requirements (30%, $N = 32$) and academic benefits or recognition (28%, $N = 30$). Data were shared via multiple avenues, including article supplementary materials (62%, $N = 66$), within the article itself (e.g., presented in a table) (48%, $N = 51$), online data repositories (34%, $N = 36$), data papers (20%, $N = 21$) and institutional or personal websites (8%, $N = 9$). Furthermore, for the respondents who reported that data were posted on online repositories, 77% ($N = 27/35$) indicated that data were associated with a persistent and unique identifier, 11% ($N = 4/35$) claimed data had a license outlining its terms of use, and 71% ($N = 25/35$) stated they had included documentation to describe the data to facilitate interpretation and re-use (i.e., "meta-data").

For the respondents who reported that data from their most recent empirical publication were not shared, 42% ($N = 55/131$) stated that they could have shared data but chose not to, and the remainder stated that they were either unable to share data (24%, $N = 32/131$) or were unsure whether they were able to (34%, $N = 44/131$). For respondents who were able to share data publicly, but chose not to, the most frequently stated reason for withholding data was to protect participants' confidentiality (40%, $N = 38$). Additionally, approximately a quarter of respondents stated that a lack of time to prepare data (26%, $N = 25$), concerns about misuse or misinterpretation (25%), and the desire to protect their ability to publish further analyses (24%, $N = 23$) influenced their decision to not share data. There were also 21 respondents (22%) who indicated that they had never considered sharing data before. In contrast, of the respondents who claimed they were unable to share data, institutional review board restrictions (26%, $N = 8/31$), lack of participant consent (23%, $N = 7/31$) and the inability to guarantee participant confidentiality (16%, $N = 5/31$) were cited as the top three reasons that prevented sharing. Lastly, when all respondents who did not share data publicly were asked whether they would be willing to organize private access to data during peer review (e.g., to allow the journal to verify the authenticity of the data), 22% ($N = 28/126$) stated they would organize access for peer reviewers, 32% ($N = 40/126$) only with journals editors, and 12% ($N = 15/126$) would only share with a journal-appointed intermediary if asked.

Prior experiences with, and reasons for sharing data were similar to the findings observed for participants' most recent publication (see Table 3).

Table 3. Respondents' prior experience with sharing their research data and code.

	N	%
Some data shared publicly ($N = 227$)		
Yes	102	45%
No	104	46%
Can't remember	21	9%
Reason for sharing data publicly ($N = 102$)*		
Journal requirement	68	67%
To promote transparency and reproducibility	57	56%
To increase the impact of the research	47	46%
Funder requirement	33	32%
Academic benefits or recognition	31	30%
Institutional/research group culture	28	27%
Institutional requirement	11	11%
Other	4	4%
Shared data privately ($N = 225$)		
Yes	73	32%
No (Declined all requests)	30	13%
No (Never been asked)	118	52%
Other	4	2%
Sharing data privately resulted in authorship or collaboration opportunities ($N = 73$)		
Yes	54	74%
No	19	26%
Obtained data privately ($N = 225$)		
Yes	63	28%
No (All requests have been declined)	10	4%
No (Never asked)	151	67%
Other	1	0%
Obtaining data privately resulted in authorship or collaboration opportunities ($N = 63$)		
Yes	53	84%
No	10	16%
Code sharing experience ($N = 225$)		
Publicly shared	31	14%
Privately shared	29	13%
Both publicly and privately shared	16	7%
Never shared	145	64%
Other	4	2%

*Percentages do not add to 100% (multiple answers possible).

Additionally, we asked respondents about their prior experience sharing and requesting data privately. A third of respondents reported that they had shared data in response to a private request in the past (32%, $N = 73/225$), with 74% also ($N = 54/73$) indicating that sharing data privately had led to authorship opportunities or future collaborations for them. In contrast, 63 of 225 respondents (28%) reported that they had successfully obtained data from another researcher in the past via a private request. Of the 63 respondents who had obtained data privately in the past, 53 (84%) added that it had resulted in a joint research paper or future collaborations with the person who shared the data.

Attitudes towards data sharing

Four out of five respondents (80%, $N = 205/257$) agreed that their cancer research community should promote and facilitate the sharing of research

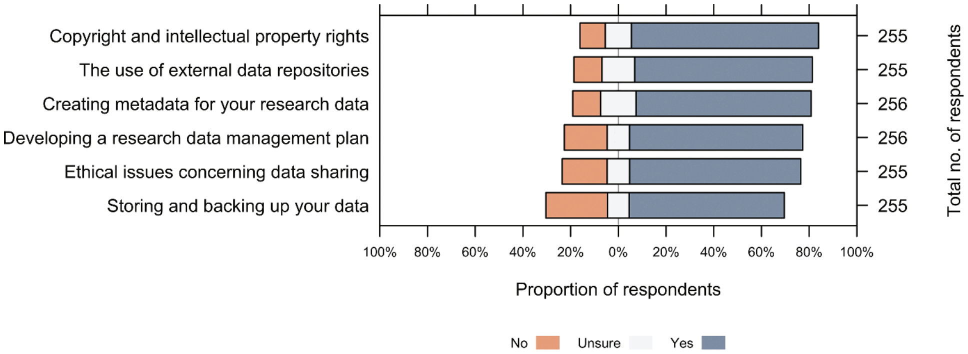


Figure 1. Perceived usefulness of training in six general research data management areas.

data. Furthermore, when asked to estimate what percentage of cancer-related research published in 2019 had some of its data shared publicly, the median guess was 29% (IQR: 18% to 50%). While we found strong support for data sharing among cancer researchers, only 17% were familiar with the FAIR principles ($N = 43/259$). However, most respondents reported that they probably (50%, $N = 128/257$), or definitely (33%, 85/257) could regenerate their most recent study's main findings five years from now if required. Despite high confidence in their archival practices, 65% to 78% of respondents reported that they thought further training in research data management would be useful to them (Figure 1). The top three areas of training deemed useful were copyright and intellectual property rights (78%, $N = 200/255$), the use of external data repositories (75%, $N = 190/255$) and the creation of metadata (73%, $N = 188/256$).

What drives data sharing?

Respondents were asked about whether, and to what extent, seven different factors influenced their current data sharing behaviors (Figure 2a). Factors frequently reported to have no influence on participants' data sharing behaviors included formal instructions received while in training (44%, $N = 100/227$) and attitudes and practices of former advisors and mentors (35%, $N = 78/225$). In contrast, the most common influencers of respondents toward sharing their data were policies of journals (60%, $N = 142/236$) and funding agencies (53%, $N = 118/222$). On the other hand, the most frequently reported influencers away from sharing data were intellectual property and commercial interests (55%, $N = 111/201$), agreements with industrial sponsors (38%, $N = 62/165$), and the policies of the institution (37%, $N = 86/234$).

Respondents were also asked whether study participants' views on data sharing could influence their decision to share data from their next human research project (Figure 2b). We observed that participants' desires to share

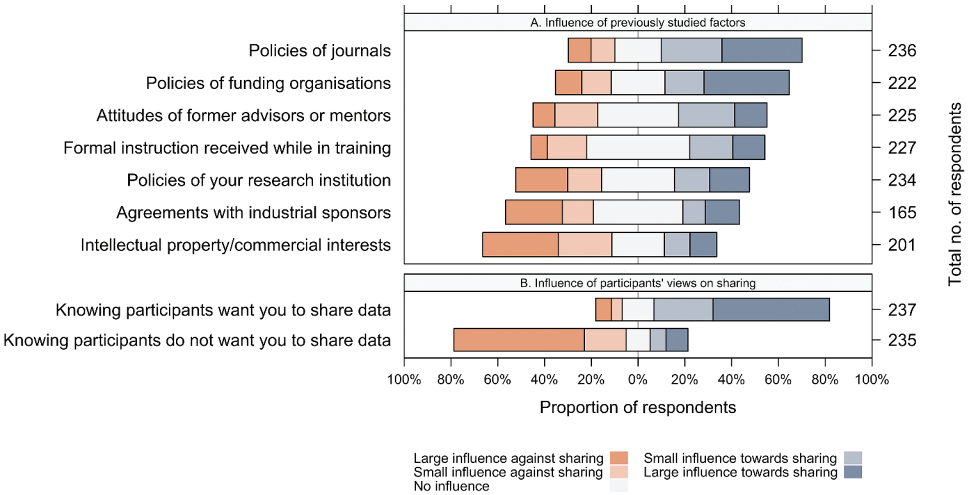


Figure 2. Influencers of respondents' data sharing behaviors.

or withhold research data had a stronger influence on respondents' data sharing behaviors than any of the other factors examined. Specifically, 75% ($N = 178/237$) of respondents stated they would be more likely to share data if they were informed most participants wanted them to, whereas 74% ($N = 173/235$) reported they would be more likely to refrain from sharing data if that was the preference of most participants.

Why NOT share research data?

The three most frequently cited barriers preventing respondents from sharing their data publicly were a preference to share it privately instead (i.e., the desire to maintain control over their intellectual property) (66%, $N = 162/247$), time and resource burdens (58%, $N = 143/246$), and a lack of training in how to share data online (56%, $N = 137/245$) (see Figure 3). Two thirds of respondents (67%, $N = 165/247$) disagreed with the statement that that their data were not important enough to share. Almost half also disagreed with the statements that no suitable repository exists for their data (49%, $N = 120/246$), and that they do not have the rights to share (45%, $N = 110/247$).

Respondents were also asked whether, and to what extent, commonly investigated concerns prevent them, as well as other cancer researchers, from sharing their data publicly (Figure 4). The two different sets of questions were asked to help detect the presence of social desirability biases. The results showed that while respondents often agreed that many of these concerns prevented them from sharing their data (Figure 4a), they also indicated that other researchers are more strongly affected by these concerns than they themselves are (Figure 4b). For example, the fear that others may misinterpret their data (66%, $N = 160/244$), use it for unintended purposes

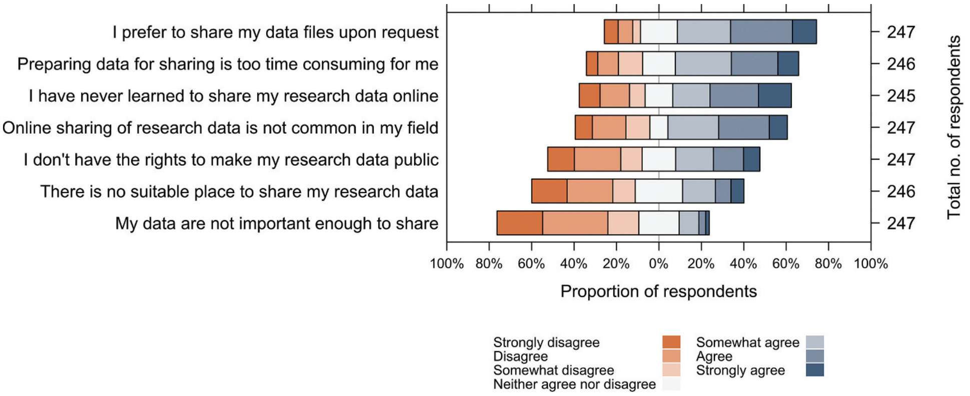


Figure 3. Respondents' perceptions on the extent to which non-fear-related barriers prevent them from sharing data publicly.

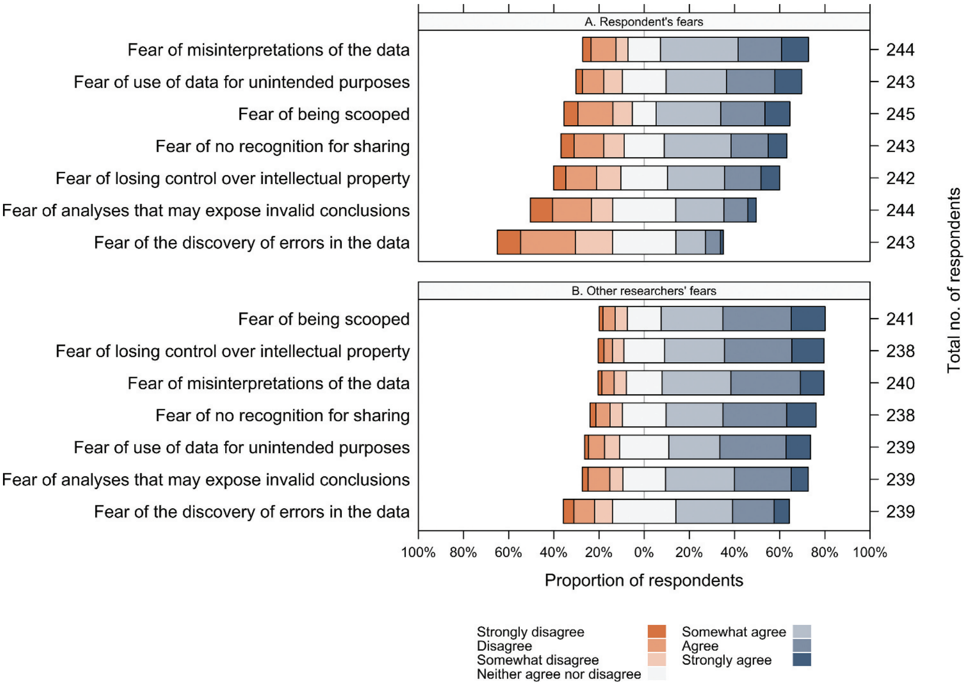


Figure 4. Respondents' perceptions on the extent to which fear-related barriers prevent them (A) and other cancer researchers (B) from sharing data publicly.

(60%, $N = 146/243$), and fear of the loss of the ability to publish further analyses (59%, $N = 145/245$), lack of recognition (54%, $N = 132/243$) and loss of control over intellectual property (50%, $N = 120/242$) were the five most common barriers for respondents. In contrast, when the same concerns were evaluated for other researchers, the top five barriers remained the same. However, excepting concerns about misinterpretation of shared data (72%, $N = 172/240$), and the use of share data for unintended purposes (63%, $N =$

150/239), respondents indicated that same barriers affected other researchers much more strongly than themselves.

What factors moderated data sharing behaviours and attitudes?

The association between key demographic characteristics and data sharing behaviors were also explored (Table 4). We found that respondents in the upper quartile for research experience (i.e., >20 years' experience) were more than twice as likely to have prior experience sharing data than those in the lower quartile (<8 years' experience) (RR: 2.43, 95% CI: 1.58 to 3.75, $p < 0.0001$). However, greater research experience did not appear to be associated with a greater likelihood of sharing data from respondents' most recent publication. Respondents who typically studied human research participants, as well as those who do not routinely work with human genomic data, were noted to be 48 to 60% less likely to have shared data from both their most recent publication and prior research. We also noted that respondents who used qualitative analyses in their most recent study were 26% less likely to share data alongside the subsequent article than those who used quantitative analyses (RR: 0.74, 95% CI: 0.32 to 1.68, $p = 0.431$). However, the low precision limits our ability to interpret this finding conclusively.

We found that researchers with experience working with cancer patients in a clinical setting did not appear to be more or less likely to share data than those who do not. However, experience reviewing grants for a major research funding organization within the last three years was associated with an increased likelihood of having shared data previously (RR: 2.12, 95% CI: 1.53 to 2.94, $p < 0.0001$). The relationship between grant reviewing experience and data sharing remained significant when controlling for research experience. Lastly, in contrast to the predictors of data sharing behaviors, the only factor that was statistically significantly associated with increased support for the promotion and facilitation of sharing of research data was routinely working with human genomics data (RR: 1.21, 95% CI: 1.08 to 1.34, $p = 0.006$) (Table S4).

Disussion

Principal findings of the survey

In this study we surveyed cancer researchers to determine their experience with and views on sharing research data. We found close to half of respondents reported that they had shared some research data, and there was very high support for the practice, particularly among cancer researchers who regularly worked with human genomics data. However, despite high self-reported levels of sharing and support for the practice, as well as confidence

Table 4. Unadjusted risk ratios and 95% confidence intervals for the association between participants' characteristics and the data sharing outcome from their most recent publication and their prior experience sharing data publicly.

	Shared data from most recent publication				Prior experience sharing data publicly			
	Frequency (%)	Risk ratio	95% CI	P value	Frequency (%)	Risk ratio	95% CI	P value
Gender								
Female	34/97 (35%)	reference	-	-	38/94 (40%)	reference	-	-
Male	71/138 (51%)	1.47	1.07 to 2.01	0.013	68/131 (52%)	1.28	0.96 to 1.73	0.089
Research area*								
Cancer biology	31/40 (78%)	reference	-	-	32/38 (84%)	reference	-	-
Cancer control & outcomes	23/68 (34%)	0.44	0.30 to 0.63	<0.0001	23/63 (37%)	0.43	0.30 to 0.62	<0.0001
Cancer treatment	34/78 (44%)	0.56	0.42 to 0.76	<0.0001	29/74 (39%)	0.47	0.34 to 0.64	<0.0001
Cancer detection & prognosis	15/38 (39%)	0.51	0.33 to 0.78	0.001	20/36 (56%)	0.66	0.48 to 0.91	0.007
Research experience								
Quartile 1 (0–7 years)	23/55 (42%)	reference	-	-	17/56 (30%)	reference	-	-
Quartile 2 (8–12 years)	24/63 (38%)	0.91	0.58 to 1.42	0.680	24/63 (38%)	1.25	0.76 to 2.08	0.375
Quartile 3 (13–20 years)	32/70 (46%)	1.09	0.73 to 1.64	0.663	35/65 (54%)	1.77	1.12 to 2.80	0.009
Quartile 4 (>20 years)	30/52 (58%)	1.38	0.93 to 2.04	0.101	34/46 (74%)	2.43	1.58 to 3.75	<0.0001
Typical research participants								
Human participants	87/215 (40%)	reference	-	-	89/204 (44%)	reference	-	-
Non-human participants	22/26 (85%)	2.09	1.66 to 2.63	<0.0001	20/26 (77%)	1.76	1.36 to 2.29	0.001
Data analysis								
Quantitative analysis	56/124 (45%)	reference	-	-	NA	NA	NA	NA
Qualitative analysis	4/12 (33%)	0.74	0.32 to 1.68	0.431	NA	NA	NA	NA
Mixed methods	49/102 (48%)	1.06	0.80 to 1.41	0.666	NA	NA	NA	NA
Works with omics data								
No	67/175 (38%)	reference	-	-	58/162 (36%)	reference	-	-
Yes	43/67 (64%)	1.68	1.29 to 2.17	<0.0001	52/69 (75%)	2.10	1.65 to 2.69	<0.0001
Clinical experience								
No	52/104 (50%)	reference	-	-	52/105 (50%)	reference	-	-
Yes	58/138 (42%)	0.84	0.64 to 1.11	0.218	58/126 (46%)	0.93	0.71 to 1.22	0.597
Grant reviewing experience								

(*Continued*)

Table 4. (Continued).

	Shared data from most recent publication				Prior experience sharing data publicly			
	Frequency (%)	Risk ratio	95% CI	P value	Frequency (%)	Risk ratio	95% CI	P value
No	40/110 (36%)	reference	-	-	31/105 (30%)	reference	-	-
Yes	70/132 (53%)	1.46	1.09 to 1.96	0.010	79/126 (63%)	2.12	1.53 to 2.94	<0.0001

*Data from respondents working in the fields of cancer etiology and cancer prevention are not reported

in their data archival procedures, most respondents were unaware of, or unfamiliar with, the FAIR principles (Wilkinson et al. 2016, 2019). Journal and funder requirements, as well as the desire to promote transparency and reproducibility, were consistently identified as positive influences on sharing practices, whereas commercial interests, agreements with industry sponsors, and institutional policies were identified as influences against sharing. Our results showed that researchers' decisions about whether to share data or not are also likely to be influenced by their participants desires. Additionally, several barriers to sharing were identified, most of which related to potential negative impacts on academic performance.

Whilst the rate of self-reported data sharing in this survey is consistent with other surveys of researchers, it is substantially higher than prevalence estimates derived through audits of the published literature. For example, our observed self-reported data sharing prevalence of 45% was more than two times higher than the recent estimate of 16% derived from an audit of cancer-related papers published two years prior to the conduct of the survey (Hamilton et al. 2022). This same discrepancy between researcher- and literature-derived data sharing estimates has been repeatedly observed in other areas of research. Some examples include, clinical trials in radiology (34% versus 0%) (Bosserdt, Hamm, and Dewey 2019; Wright et al. 2020), systematic reviews (28% versus 4%) (Nguyen et al. 2022, 2023; Page et al. 2022), as well as psychology research (45% versus 2%) (Houtkoop et al. 2018; Hardwicke et al. 2021). This likely reflects a social desirability effect (or non-response biases), resulting in the overrepresentation of positive views on sharing. Consequently, it is important not to be too optimistic about data sharing practices based on self-report values.

Our observation that 80% of respondents agreed that cancer researchers should promote and facilitate the sharing of research data is highly consistent with previous surveys of clinical triallists (77–88%) (Rathi et al. 2014; Tan et al. 2021), and researchers in dentistry (100%) (Spallek et al. 2019) and psychology (77%) (Houtkoop et al. 2018). However, this too might be impacted by the same social desirability bias as the previous result.

Furthermore, despite high support for sharing, we note that familiarity with the FAIR principles in our sample was low (17%), and almost half the prevalence observed in a survey of 820 life sciences researchers using the same question in the same year (30%) (Simons et al. 2021). Low awareness of the FAIR principles among respondents was also reflected by the low number of respondents using data repositories (34%) and providing clear licensing conditions for their data (11%), both of which are recommended by the FAIR principles. The infrequent use of practices adhering to the FAIR principles is consistent with previous audits of the oncology literature (Hamilton et al. 2022), as well as published systematic reviews (Nguyen et al. 2022; Page et al. 2022), dental research (Spallek et al. 2019), veterinary epidemiological research (Meyer et al. 2021), and COVID-19-related research (Zuo et al. 2021) to name a few. Low awareness of practices consistent with the FAIR principles may also explain the high appetite for further training in concepts related to copyright and intellectual property rights, and the use of external data repositories among our respondents, as well as participants of previous surveys (Melero and Navarro-Molina 2020).

Implications of the research

A summary of our top five observed drivers and barriers to sharing in comparison to other relevant surveys performed since 2016 (i.e., since the inception of the FAIR principles (Wilkinson et al. 2016)) are reported in Table 5. When looking at this body of research, three key drivers repeatedly rank among the top three reasons why researchers shared data previously: the promotion of open science, and to comply with journal requirements and funder requirements. In contrast, the top three barriers in our study included: concerns about maintaining participants' confidentiality, time and resource constraints, and concerns about inappropriate use of shared data. These findings raise the following implications for oncology research and medical research more broadly.

Firstly, our findings suggest that further training, education, and support of cancer researchers in their research data collection and archival practices, as well as greater promotion of data sharing by cancer research institutions, could result in major advances in the culture of research data sharing in oncology. Some examples of how this can be achieved include incorporating into education and training materials information on the importance of open science practices and benefits of sharing data, such as accelerating research on rare cancers (Denton et al. 2022) and reported citation advantages associated with sharing cancer microarray data (Piwowar et al. 2007). Additionally, changes to hiring and promotion criteria that move away from the use of journal-based metrics to assess researchers (e.g., endorsement of the San Francisco Declaration on Research Assessment) might serve to

Table 5. Comparison of key findings with surveys in the post-FAIR era (i.e., post-2016).

Study	Discipline	N	Top five reasons for sharing data	Top five identified barriers to sharing data
The current study	Oncology	285	1. Journal requirements 2. Promotion of open science 3. Increased impact 4. Funder requirements 5. Academic benefits	1. Confidentiality concerns 2. Time and resource constraints 3. Concerns about inappropriate use 4. Fear of scooping 5. Never considered sharing
Nguyen et al. (2023)	Systematic reviews	417	1. Promotion of open science 2. Journal requirements 3. Cultural norms 4. Increased impact 5. Academic benefits	1. Not a common practice 2. Time and resource constraints 3. Preferences to share on request 4. Fear of scooping 5. Concerns about inappropriate use
Tan et al. (2021)	Clinical trials	281	1. Promotion of open science 2. Journal requirements 3. Funder requirements 4. Institution requirements 5. Increased impact	1. Legal constraints 2. Confidentiality concerns 3. Concerns about inappropriate use 4. Time and resource constraints 5. Fear of scooping
Melero and Navarro-Molina (2020)	Food science and technology	101	1. Funder requirements 2. Cultural norms 3. Institution requirements 4. Promotion of open science 5. Increased impact	1. Legal constraints 2. Loss of control of intellectual property 3. Rights to share 4. Commercialisation 5. Concerns about inappropriate use
Bosserdt et al. (2019)	Clinical trials (radiology)	132	NA	1. Loss of control of intellectual property 2. Confidentiality concerns 3. No location to share 4. Legal constraints 5. Time and resource constraints
Spallek et al. (2019)	Dental research	42	NA	1. Concerns about inappropriate use 2. Confidentiality concerns 3. Fear of scooping 4. Concerns about lack of recognition 5. Time and resource constraints
Houtkoop et al. (2018)	Psychology	600	NA	1. Not a common practice 2. Preferences to share on request 3. Time and resource constraints 4. Lack of training 5. Concerns about inappropriate use

incentivize sharing as well as help alleviate concerns regarding loss of academic productivity (e.g., being "scooped"). Concerns, particularly those regarding being scooped, are also likely to be more pronounced among researchers working in disciplines that require greater resources (e.g., genomics (Mohamed, Awad, and Mosa 2016)), are highly competitive, and whose ongoing work is well-publicized (Laine 2017). Further training and support on how to develop research protocols and data management plans that maximize data availability, specifically addressing areas such as participant consent, data anonymization procedures, data licensing, and how to use repositories to make data available to the public, or to restricted groups, would directly address three most common barriers above. Such changes would also assist the half of respondents that stated they could not share data due to institutional review board (IRB) restrictions and lack of patient consent. Additionally, training could help minimize confusion surrounding ownership of research data which can lead to tensions between researchers and institutions (Minutillo, Panza, and Mauri 2021), which are expected to become more commonplace as institutions become more protective of research that may have commercial value (Yang et al. 2021).

Secondly, our findings demonstrate that both journals and funders have a strong influence over cancer researchers' data sharing behaviors. With respect to policymaking among publishers of medical and cancer research, encouragingly we see increased institution of policies requiring researchers to share data (Hamilton et al. 2020; Kim et al. 2020; Resnik et al. 2019; Vasilevsky et al. 2017). We also note recent empirical research which demonstrates that these policies, while not perfect, result in significantly higher rates of data sharing than alternative policies such as those that "encourage" sharing, or no policy at all (Hamilton et al. 2022, 2023). Furthermore, like the publishers, medical research funders also have a unique and powerful ability to redress the current state of poor incentives to share data, with increasing numbers of funders now requiring awardees to share research data (DeVito, French, and Goldacre 2018; Gaba et al. 2020; Whitlock et al. 2019). In addition to instituting mandatory sharing policies, other policies are also gradually being incorporated into grant review processes. For example, some funders now consider applicants' track record in sharing, as well as studies performed by unrelated groups using shared data, in funding decisions. This process has already been implemented by some funders such as Alex's Lemonade Stand which is a major funder of pediatric cancer research in the United States (Byrd et al. 2020). Others also mandate and check compliance with awardee's data management plans. For example, the US National Institutes of Health (NIH), which is the largest public funder of biomedical research globally, now requires all applicants to submit a data management and sharing plan, expecting that "researchers will maximise the appropriate sharing of scientific data" (NIH 2020). The policy also states that plans will be reviewed to ensure they appropriately consider these factors, and that noncompliance may be met with termination of awards and could also affect future funding decisions – practices that bring the funder in line with other health funding initiatives like the European commission's Horizon Europe program (Bjaalie et al. 2022).

Consequently, mirroring previous successes seen in other areas of research, such as those that work with genomics data (Paltoo et al. 2014), continued adoption of progressive policies by both journals and funders, concerning data availability would greatly improve the culture of sharing data in oncology. For example, for journals this could include requiring data sharing statements, introducing mechanisms into editorial workflows to check compliance with data sharing policies, and disincentivising authors depositing data into short-term storage locations such as journal supplementary materials. For funders, this might include requiring detailed data management plans, allowing applicants to budget for reasonable costs associated with retrieving or archiving data, making data management plans publicly available, and highlighting projects that generated data that has been highly reused.

Strengths and limitations

We note several strengths of our study. Our survey achieved a high completion rate, with more than three quarters of participants who began the survey completing it. The use of multiple survey functions (e.g., personalized invitations, tooltips, response validation criteria) gives us confidence that we have been able to elicit honest and informed responses, social desirability biases notwithstanding. We also did not see any substantial differences between invited and responding participants by participant gender. Additionally, we note similar participant distributions by research area when compared to Hamilton et al's audit of a random sample of oncology papers in 2019 (Hamilton et al. 2022).

Nevertheless, our study also had some limitations. Our decision to use Clarivate Analytics' Web of Science to identify potential participants may have resulted in a pool of authors that are not truly representative of all cancer researchers. While we employed multiple strategies to minimize non-response rates, including not using the term "survey" in the e-mail subject, keeping the median survey length under 15 minutes, as well as sending multiple reminders to non-responders, our overall response rate, while in line with our original conservative estimates and other similar surveys (Houtkoop et al. 2018), was lower than anticipated. It is likely that this was a consequence of performing the research during the COVID-19 pandemic (e.g., lockdowns (Cooper 2021) and survey fatigue (de Koning et al. 2021)). It is also possible that the decision not to identify and filter out articles published in journals generally considered to be predatory may have also negatively impacted our response rate, with research indicating that papermilled and retracted articles are more likely to use single-purpose non-institutional e-mail accounts than their counterparts (Bimler 2022; Liu and Chen 2021). For this reason, outside of our exploration of factors moderating respondents' data sharing behaviors and attitudes, we have chosen not to report confidence intervals for this study and have refrained from making inferential statements. Additionally, our response rate was particularly low among researchers affiliated with institutions in Asia (e.g., China (1%), Korea (2%) and Japan (3%). However, appreciating previous research by Tenopir et al. (2015) that demonstrated negligible differences in data sharing practices across geographic regions, we do not think the underrepresentation of these groups would have substantially impacted most of our findings.

Conclusion

Our survey has shown strong support for sharing data amongst a sample of cancer researchers, particularly among those that regularly work with genomics data, but low knowledge of best practices. Journal and funder requirements were consistently identified as major drivers toward researchers sharing their data, and commercial interests, agreements with industry sponsors and institutional policies against it. The protection of participants' confidentiality, time and resources required to prepare data for sharing, and concerns about misuse or misinterpretation of shared data were also identified as key motivators against sharing. We recommend that increased promotion and support by research institutions, for example, reform in hiring and promotion criteria, as well as further education and training on research data management, in combination with greater championing of data sharing by journals and funders, be done to motivate more researchers in oncology to share their data.

Acknowledgments

We thank Simine Vazire, Raquel Ashton and Steve Kambouris for their advice on the survey design, and Aaron Willcox and David Wilkinson for their assistance with the data cleaning and analysis.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

DGH is supported by an Australian Commonwealth Government Research Training Programme Scholarship.

Authorship statement

DGH conceived and designed the study, collected data, performed the formal analysis, curated the data and wrote and prepared the original draft. All authors contributed to the methodology, interpreted the results, contributed to writing the manuscript, approved the final version, and had final responsibility for the decision to submit for publication. DGH is guarantor. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Data availability statement

De-identified data, study materials and analytic code are publicly available on the Open Science Framework under a under a Creative Commons Zero v1.0 Universal (CC0 1.0) license (DOI: 10.17605/OSF.IO/E6UV3).

ORCID

Daniel G. Hamilton http://orcid.org/0000-0001-8104-474X
Matthew J. Page http://orcid.org/0000-0002-4242-7526
Hannah Fraser http://orcid.org/0000-0003-2443-4463
Fiona Fidler http://orcid.org/0000-0002-2700-2562

References

Aragon, T. J. 2020. “Epitools: Epidemiology Tools.” R Package Version 0.5-10.1. https://CRAN.R-project.org/package=epitools.

Begley, C. G., and M. E. Lee. 2012. “Raise Standards for Preclinical Cancer Research.” *Nature* 483 (7391): 531–533. https://doi.org/10.1038/483531a.

Bimler, D. 2022. “Better Living Through Coordination Chemistry: A Descriptive Study of a Prolific Papermill That Combines Crystallography and Medicine.” *Research Square*. https://doi.org/10.21203/rs.3.rs-1537438/v1. 2022 Apr 15.

Bjaalie, J. G., C. Goble, S.-A. Sansone, R. Nakamura, and M. E. Martone. 2022. “Perspectives on Data Sharing and the New NIH Policy from the European Union.” *Harvard Data Science Review* 4 (3). https://doi.org/10.1162/99608f92.bcd0b999.

Bosserdt, M., B. Hamm, and M. Dewey. 2019. “Clinical Trials in Radiology and Data Sharing: Results from a Survey of the European Society of Radiology (ESR) Research Committee.” *European Radiology* 29 (9): 4794–4802. https://doi.org/10.1007/s00330-019-06105-y.

Bray, F., M. Laversanne, E. Weiderpass, and I. Soerjomataram. 2021. “The Ever-Increasing Importance of Cancer as a Leading Cause of Premature Death Worldwide.” *Cancer* 127 (16): 3029–3030. https://doi.org/10.1002/cncr.33587.

Brown, L. D., T. T. Cai, and A. DasGupta. 2001. “Interval Estimation for a Binomial Proportion.” *Statistical Science* 16 (2): 101–117. https://doi.org/10.1214/ss/1009213286.

Byrd, J. B., A. C. Greene, D. Venkatesh Prasad, X. Jiang, and C. S. Greene. 2020. “Responsible, Practical Genomic Data Sharing That Accelerates Research.” *Nature Reviews Genetics* 21 (10): 615–629. https://doi.org/10.1038/s41576-020-0257-5.

Campbell, E. G., B. R. Clarridge, M. Gokhale, L. Birenbaum, S. Hilgartner, N. A. Holtzman, and D. Blumenthal. 2002. “Data Withholding in Academic Genetics. Evidence from a National Survey.” *JAMA* 287 (4): 473–480. https://doi.org/10.1001/jama.287.4.473.

Chan, A.W., F. Song, A. Vickers, T. Jefferson, K. Dickersin, P. C. Gøtzsche, H. M. Krumholz, D. Ghersi, and H. Bart van der Worp. 2014. "Increasing Value and Reducing Waste: Addressing Inaccessible Research." *Lancet* 383 (9913): 257–266. https://doi.org/10.1016/S0140-6736(13)62296-5.

Christian, H., and G. A. Hoover. 2017. "How Prevalent is Academic Misconduct in Management Research?" *Journal of Business Research* 80 (November): 73–81. https://doi.org/10.1016/j.jbusres.2017.07.003.

Cooper, M. April 30th, 2021. Half the Planet is on Lockdown, but Not Every U.S. State Is, Even After Alabama Issues an Order. New York Times. https://www.nytimes.com/2020/04/03/world/coronavirus-news-updates.html.

de Koning R., A. Egiz, A. C. Ciuculete, S. Z. Y. Ooi, N. D. A. Bankole, J. Erhabor, G. Higginbotham, et al. 2021. "Survey Fatigue During the COVID-19 Pandemic: An Analysis of Neurosurgery Survey Response Rates." *Frontiers in Surgery 8*. https://www.frontiersin.org/articles/10.3389/fsurg.2021.690680.

Denton, N., A. E. Mulberg, M. Molloy, S. Charleston, D. C. Fajgenbaum, E. D. Marsh, and P. Howard. 2022. "Sharing is Caring: A Call for a New Era of Rare Disease Research and Development." *Orphanet Journal of Rare Diseases* 17 (1): 389. https://doi.org/10.1186/s13023-022-02529-w.

DeVito, N. J., L. French, and B. Goldacre. 2018. "Noncommercial Funders' Policies on Trial Registration, Access to Summary Results, and Individual Patient Data Availability." *JAMA* 319 (16): 1721–1723. https://doi.org/10.1001/jama.2018.2841.

Errington, T. M., M. Mathur, C. K. Soderberg, A. Denis, N. Perfito, E. Iorns, and B. A. Nosek. 2021. "Investigating the Replicability of Preclinical Cancer Biology." Edited by Renata Pasqualini and Eduardo Franco. *eLife* 10 (December): e71601. https://doi.org/10.7554/eLife.71601.

Fecher, B., S. Friesike, M. Hebing, and R. S. Phillips. 2015. "What Drives Academic Data Sharing?" *PLoS ONE* 10 (2): e0118053. https://doi.org/10.1371/journal.pone.0118053.

Freedman, L. P., I. M. Cockburn, and T. S. Simcoe. 2015. "The Economics of Reproducibility in Preclinical Research." *PLoS Biology* 13 (6): e1002165. https://doi.org/10.1371/journal.pbio.1002165.

Gaba, J. F., M. Siebert, A. Dupuy, D. Moher, F. Naudet, and D. Koletsi. 2020. "Funders' Data-Sharing Policies in Therapeutic Research: A Survey of Commercial and Non-Commercial Funders." *PLOS ONE* 15 (8): e0237464. https://doi.org/10.1371/journal.pone.0237464.

Genevieve, P.-K., D. E. Zinner, E. G. Campbell, C. Borgman, and C. Borgman. 2014. "Codifying Collegiality: Recent Developments in Data Sharing Policy in the Life Sciences." *PloS ONE* 9 (9): e108451. https://doi.org/10.1371/journal.pone.0108451.

Hamilton, D. G. 2023. "What Do Cancer Researchers Think About Data Sharing?" *Open Science Framework*. https://doi.org/10.17605/OSF.IO/y4qnh.

Hamilton, D. G., S. Everitt, M. J. Page, and F. Fidler. 2023. "What Do Australians Affected by Cancer Think About Oncology Researchers Sharing Research Data: A Cross-Sectional Survey." medRxiv. https://doi.org/10.1101/2023.05.01.23289334.

Hamilton, D. G., H. Fraser, R. Hoekstra, and F. Fidler. 2020. "Journal Policies and Editors' Opinions on Peer Review." *eLife* 9 (November): e62529. https://doi.org/10.7554/eLife.62529.

Hamilton, D. G., K. Hong, H. Fraser, A. Rowhani-Farid, F. Fidler, and M. J. Page. 2023. "Prevalence and Predictors of Data and Code Sharing in the Medical and Health Sciences: Systematic Review with Meta-Analysis of Individual Participant Data." *BMJ* 382 (July): e075767. https://doi.org/10.1136/bmj-2023-075767.

Hamilton, D. G., M. J. Page, S. Finch, S. Everitt, and F. Fidler. 2022. "How Often Do Cancer Researchers Make Their Data and Code Available and What Factors are Associated with Sharing?" *BMC Medicine* 20 (1): 438. https://doi.org/10.1186/s12916-022-02644-2.

Hardwicke, T. E., R. T. Thibault, J. E. Kosie, J. D. Wallach, M. C. Kidwell, and J. P. A. Ioannidis. 2021. "Estimating the Prevalence of Transparency and Reproducibility-Related Research Practices in Psychology (2014–2017)." *Perspectives on Psychological Science* March, 1745691620979806. https://doi.org/10.1177/1745691620979806.

Ho, R. C.-M., K.-K. Mak, R. Tao, L. Yanxia, J. R. Day, and F. Pan. 2013. "Views on the Peer Review System of Biomedical Journals: An Online Survey of Academics from High-Ranking Universities." *BMC Medical Research Methodology* 13 (1): 74. https://doi.org/10.1186/1471-2288-13-74.

Houtkoop, B. L., C. Chambers, M. Macleod, D. V. M. Bishop, T. E. Nichols, and E.-J. Wagenmakers. 2018. "Data Sharing in Psychology: A Survey on Barriers and Preconditions." *Advances in Methods and Practices in Psychological Science* 1 (1): 70–85. https://doi.org/10.1177/2515245917751886.

Howe, N., E. Giles, D. Newbury-Birch, and E. McColl. 2018. "Systematic Review of Participants' Attitudes Towards Data Sharing: A Thematic Synthesis." *Journal of Health Services Research & Policy* 23 (2): 123–133. https://doi.org/10.1177/1355819617751555.

Kawahara, T., M. Fukuda, K. Oba, J. Sakamoto, and M. Buyse. 2018. "Meta-Analysis of Randomized Clinical Trials in the Era of Individual Patient Data Sharing." *International Journal of Clinical Oncology* 23 (3): 403–409. https://doi.org/10.1007/s10147-018-1237-z.

Kim, J., S. Kim, H.-M. Cho, J. Hwa Chang, and S. Young Kim. 2020. "Data Sharing Policies of Journals in Life, Health, and Physical Sciences Indexed in Journal Citation Reports." *PeerJ* 8 (October): e9924. https://doi.org/10.7717/peerj.9924.

Laine, H. 2017. "Afraid of Scooping – Case Study on Researcher Strategies Against Fear of Scooping in the Context of Open Science." *Data Science Journal* 16:29. https://doi.org/10.5334/dsj-2017-029.

Liu, X., and X. Chen. 2021. "Authors' Noninstitutional Emails and Their Correlation with Retraction." *Journal of the Association for Information Science and Technology* 72 (4): 473–477. https://doi.org/10.1002/asi.24419.

McIntosh, S. A., F. Alam, L. Adams, I. S. Boon, J. Callaghan, I. Conti, E. Copson, V. Carson, M. Davidson, H. Fitzgerald, et al. 2023. "Global Funding for Cancer Research Between 2016 and 2020: A Content Analysis of Public and Philanthropic Investments." *The Lancet Oncology* 24 (6): 636–645. https://doi.org/10.1016/S1470-2045(23)00182-1.

Melero, R., and C. Navarro-Molina. 2020. "Researchers' Attitudes and Perceptions Towards Data Sharing and Data Reuse in the Field of Food Science and Technology." *Learned Publishing* 33 (2): 163–179. https://doi.org/10.1002/leap.1287.

Meyer, A., C. Faverjon, M. Hostens, A. Stegeman, and A. Cameron. 2021. "Systematic Review of the Status of Veterinary Epidemiological Research in Two Species Regarding the FAIR Guiding Principles." *BMC Veterinary Research* 17 (1): 270–270. https://doi.org/10.1186/s12917-021-02971-1.

Minutillo, A., G. Panza, and M. C. Mauri. 2021. "Retraction Note: Musical Practice and BDNF Plasma Levels as a Potential Marker of Synaptic Plasticity: An Instrument of Rehabilitative Processes." *Neurological Sciences* 42 (7): 3067–3067. https://doi.org/10.1007/s10072-021-05283-2.

Mohamed, H., M. Awad, and K. A. Mosa. 2016. "Limited Resources of Genome Sequencing in Developing Countries: Challenges and Solutions." *Applied & Translational Genomics* 9 (June): 15–19. https://doi.org/10.1016/j.atg.2016.03.003.

National Institutes of Health. 2020, October 29th. NOT-OD-21-013: Final NIH Policy for Data Management and Sharing. Accessed July 20th, 2023. https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-013.html.

Nguyen, P.-Y., R. Kanukula, J. E. McKenzie, Z. Alqaidoom, S. E. Brennan, N. R. Haddaway, D. G. Hamilton, S. Karunananthan, S. McDonald, D. Moher, et al. 2022. "Changing Patterns in Reporting and Sharing of Review Data in Systematic Reviews with Meta-Analysis of the Effects of Interventions: Cross Sectional Meta-Research Study." *BMJ* 379 (November): e072428. https://doi.org/10.1136/bmj-2022-072428.

Nguyen, P.-Y., J. E. McKenzie, D. G. Hamilton, D. Moher, P. Tugwell, F. M. Fidler, N. R. Haddaway, et al. 2023. "Systematic Reviewers' Perspectives on Sharing Review Data, Analytic Code, and Other Materials: A Survey." *Cochrane Evidence Synthesis and Methods* 1 (2): e12008. https://doi.org/10.1002/cesm.12008.

Page, M. J., P.-Y. Nguyen, D. G. Hamilton, N. R. Haddaway, R. Kanukula, D. Moher, and J. E. McKenzie. 2022. "Data and Code Availability Statements in Systematic Reviews of Interventions Were Often Missing or Inaccurate: A Content Analysis." *Journal of Clinical Epidemiology* 147 (July): 1–10. https://doi.org/10.1016/j.jclinepi.2022.03.003.

Paltoo, D. N., L. L. Rodriguez, M. Feolo, E. Gillanders, E. M. Ramos, J. L. Rutter, S. Sherry, et al. 2014. "Data Use Under the NIH GWAS Data Sharing Policy and Future Directions." *Nature Genetics* 46 (9): 934–938. https://doi.org/10.1038/ng.3062.

Piwowar, H. A., R. S. Day, D. B. Fridsma, and J. Ioannidis. 2007. "Sharing Detailed Research Data is Associated with Increased Citation Rate." Edited by John Ioannidis *PloS One* 2 (3): e308. https://doi.org/10.1371/journal.pone.0000308.

Prinz, F., T. Schlange, and K. Asadullah. 2011. "Believe It or Not: How Much Can We Rely on Published Data on Potential Drug Targets?" *Nature Reviews Drug Discovery* 10 (9): 712–712. https://doi.org/10.1038/nrd3439-c1.

Rathi, V. K., K. Dzara, C. P. Gross, I. Hrynaszkiewicz, S. Joffe, H. M. Krumholz, K. M. Strait, and J. S. Ross. 2012. "Sharing of Clinical Trial Data Among Trialists: A Cross Sectional Survey." *BMJ* 345 (nov20 3): e7570–e7570. https://doi.org/10.1136/bmj.e7570.

Rathi, V. K., K. M. Strait, C. P. Gross, I. Hrynaszkiewicz, S. Joffe, H. M. Krumholz, K. Dzara, and J. S. Ross. 2014. "Predictors of Clinical Trial Data Sharing: Exploratory Analysis of a Cross-Sectional Survey." *Trials* 15 (1): 384. https://doi.org/10.1186/1745-6215-15-384.

Resnik, D. B., M. Morales, R. Landrum, M. Shi, J. Minnier, N. A. Vasilevsky, and R. E. Champieux. 2019. "Effect of Impact Factor and Discipline on Journal Data Sharing Policies." *Accountability in Research* 26 (3): 139–156. https://doi.org/10.1080/08989621.2019.1591277.

Serghiou, S., D. G. Contopoulos-Ioannidis, K. W. Boyack, N. Riedel, J. D. Wallach, J. P. A. Ioannidis, and L. Bero. 2021. "Assessment of Transparency Indicators Across the Biomedical Literature: How Open is Open?" *PLoS Biology* 19 (3): e3001107. https://doi.org/10.1371/journal.pbio.3001107.

Sharma, A., N. Tran Minh Duc, T. Luu Lam Thang, N. Hai Nam, N. Sze Jia, K. Said Abbas, N. Tien Huy, A. Marušić, C. L. Paul, J. Kwok, et al. 2021. "A Consensus-Based Checklist for Reporting of Survey Studies (CROSS)." *Journal of General Internal Medicine* 36 (10): 3179–3187. https://doi.org/10.1007/s11606-021-06737-1.

Simons, N., G. Goodey, M. Hardeman, C. Clare, S. Gonzales, D. Strange, G. Smith, et al. 2021. "The State of Open Data 2021." *Digital Science*. https://doi.org/10.6084/M9.FIGSHARE.17061347.V1.

Spallek, H., S. M. Weinberg, M. Manz, S. Nanayakkara, X. Zhou, and L. Johnson. 2019. "Perceptions and Attitudes Toward Data Sharing Among Dental Researchers." *JDR Clinical & Translational Research* 4 (1): 68–75. https://doi.org/10.1177/2380084418790451.

Tan, A. C., L. M. Askie, K. Elizabeth Hunter, A. Barba, R. John Simes, and A. Lene Seidler. 2021. "Data Sharing—Trialists' Plans at Registration, Attitudes, Barriers and Facilitators: A Cohort Study and Cross-Sectional Survey." *Research Synthesis Methods* 12 (5): 641–657. https://doi.org/10.1002/jrsm.1500.

Tannenbaum, S., J. S. Ross, H. M. Krumholz, N. R. Desai, J. D. Ritchie, R. Lehman, G. M. Gamble, J. Bachand, S. Schroter, T. Groves, et al. 2018. "Early Experiences with Journal Data Sharing Policies: A Survey of Published Clinical Trial Investigators." *Annals of Internal Medicine* 169 (8): 586. https://doi.org/10.7326/M18-0723.

Tenopir, C., S. Allard, K. Douglass, A. Umur Aydinoglu, W. Lei, E. Read, M. Manoff, M. Frame, C. Neylon, and C. Neylon. 2011. "Data Sharing by Scientists: Practices and Perceptions." *PLoS ONE* 6 (6): e21101. https://doi.org/10.1371/journal.pone.0021101.

Tenopir, C., E. D. Dalton, S. Allard, M. Frame, I. Pjesivac, B. Birch, D. Pollock, K. Dorsett, and P. van den Besselaar. 2015. "Changes in Data Sharing and Data Reuse Practices and Perceptions Among Scientists Worldwide." *PloS ONE* 10 (8): e0134826. https://doi.org/10.1371/journal.pone.0134826.

Thoegersen, J. L., and P. Borlund. 2021. "Researcher Attitudes toward Data Sharing in Public Data Repositories: A Meta-Evaluation of Studies on Researcher Data Sharing." *Journal of Documentation* 78 (7): 1–17. https://doi.org/10.1108/JD-01-2021-0015.

Vasilevsky, N. A., J. Minnier, M. A. Haendel, and R. E. Champieux. 2017. "Reproducible and Reusable Research: Are Journal Data Sharing Policies Meeting the Mark." *PeerJ* 5 (April): e3208. https://doi.org/10.7717/peerj.3208.

Walters, C., Z. J. Harter, C. Wayant, V. Nam, M. Warren, J. Chronister, D. Tritz, and M. Vassar. 2019. "Do Oncology Researchers Adhere to Reproducible and Transparent Principles? A Cross-Sectional Survey of Published Oncology Literature." *British Medical Journal Open* 9 (12): e033962. https://doi.org/10.1136/bmjopen-2019-033962.

Whitlock, E. P., K. M. Dunham, K. DiGioia, E. Lazowick, T. C. Gleason, and D. Atkins. 2019. "Noncommercial US Funders' Policies on Trial Registration, Access to Summary Results, and Individual Patient Data Availability." *JAMA Network Open* 2 (1): e187498. https://doi.org/10.1001/jamanetworkopen.2018.7498.

Wilkinson, M. D., M. Dumontier, I. Jan Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, et al. 2016. "The FAIR Guiding Principles for Scientific Data Management and Stewardship." *Scientific Data* 3 (1). https://doi.org/10.1038/sdata.2016.18.

Wilkinson, M. D., M. Dumontier, S.-A. Sansone, L. O. B. da Silva Santos, M. Prieto, D. Batista, P. McQuilton, T. Kuhn, P. Rocca-Serra, M. Crosas, et al. 2019. "Evaluating FAIR Maturity Through a Scalable, Automated, Community-Governed Framework." *Scientific Data* 6 (1): 174. https://doi.org/10.1038/s41597-019-0184-5.

Wren, J. D., J. E. Grissom, and T. Conway. 2006. "E-Mail Decay Rates Among Corresponding Authors in MEDLINE." *EMBO Reports* 7 (2): 122–127. https://doi.org/10.1038/sj.embor.7400631.

Wright, B. D., V. Nam, J. Nolan, A. L. Johnson, T. Braaten, D. Tritz, and M. Vassar. 2020. "An Analysis of Key Indicators of Reproducibility in Radiology." *Insights into Imaging* 11 (1): 65. https://doi.org/10.1186/s13244-020-00870-x.

Yang, J., H.-L. Intan, L. F. Bruno, W.-V. Sacha, and Z. Hao. 2021. "Measuring Global Patenting of Universities and Public Research Institutes." In *Harnessing Public Research for Innovation in the 21st Century: An International Assessment of Knowledge Transfer Policies*, edited by A. Arundel, S. Wunsch-Vincent, and S. Athreye, Intellectual Property, Innovation and Economic Development. 80–138. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781108904230.012.

Zuiderwijk, A., R. Shinde, W. Jeng, and F. Sudzina. 2020. "What Drives and Inhibits Researchers to Share and Use Open Research Data? A Systematic Literature Review to Analyze Factors Influencing Open Research Data Adoption." *PLoS ONE* 15 (9): e0239283. https://doi.org/10.1371/journal.pone.0239283.

Zuo, X., Y. Chen, L. Ohno-Machado, and X. Hua. 2021. "How Do We Share Data in COVID-19 Research? A Systematic Review of COVID-19 Datasets in PubMed Central Articles." *Briefings in Bioinformatics* 22 (2): 800–811. https://doi.org/10.1093/bib/bbaa331.