

Data sharing in cancer research: A qualitative study exploring community members' preferences

Elizabeth A. Johnston^{1,2,3}, Xanthia E. Bourdaniotis¹, Susannah K. Ayre^{1,2}, Leah Zajdlewicz¹,
Vanessa L. Beesley^{4,5,6}, Belinda C. Goodwin^{1,7,8}

¹ Cancer Council Queensland, Fortitude Valley, Queensland, Australia.

² School of Exercise and Nutrition Sciences, Queensland University of Technology, Kelvin Grove, Queensland, Australia.

³ Population Health Program, QIMR Berghofer Medical Research Institute, Herston, Queensland, Australia.

⁴ Psychedelic Medicine and Supportive Care Lab, QIMR Berghofer Medical Research Institute, Herston, Queensland, Australia.

⁵ School of Psychology, The University of Queensland, St Lucia, Queensland, Australia.

⁶ School of Nursing, Queensland University of Technology, Kelvin Grove, Queensland, Australia

⁷ Centre for Health Research, University of Southern Queensland, Springfield, Queensland, Australia.

⁸ School of Population and Global Health, University of Melbourne, Carlton, Victoria, Australia

Corresponding Author: Elizabeth Johnston, Viertel Cancer Research Centre, Cancer Council Queensland, 553 Gregory Terrace, Fortitude Valley, QLD, Australia 4006.

Telephone: +61 7 3634 5162. Email: e23.johnston@qut.edu.au. ORCIDiD:

<https://orcid.org/0000-0002-9486-5704>

Acknowledgments: The authors thank all individuals who participated in a workshop for this study.

Conflict of Interest Statement: The authors declare that there are no conflicts of interest.

Funding Statement: This research was funded by Cancer Council Queensland.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

Data Availability Statement: The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Ethical Approval Statement: The study was approved by the Human Research Ethics Committee of the University of Southern Queensland (ETH2023-0140). All participants provided written informed consent prior to participation.

Author contributions:

EAJ: conceptualization, methodology, investigation, formal analysis, visualization, writing – original draft, writing – review and editing

XEB: project administration, data curation, investigation, formal analysis, writing – review and editing

SKA: conceptualization, methodology, investigation, formal analysis, writing – review and editing

LZ: project administration, writing – review and editing

VLB: supervision, writing – review and editing

BCG: conceptualization, funding acquisition, methodology, supervision, writing – review and editing

Email Addresses:

Elizabeth: e23.johnston@qut.edu.au

Xanthia: xanthiaboutdaniotis@cancerqld.org.au

Susannah: susannahayre@cancerqld.org.au

Leah: leahzajdlewicz@cancerqld.org.au

Vanessa: Vanessa.Beesley@qimrberghofer.edu.au

Belinda: belindagoodwin@cancerqld.org.au

1 **Abstract**

2 Advancements in cancer treatment and survivorship rely on participation in research and
3 access to health records. This study explored preferences for data access and sharing in 14
4 workshops with 42 community members, most of whom were a cancer survivor or carer.
5 Various scenarios for data access and sharing were presented and discussed, with participants'
6 preferences summarized using descriptive statistics. Reasons underlying these preferences
7 were identified through a thematic analysis of workshop transcripts. Most participants
8 indicated a willingness for researchers to use their self-reported data and current health
9 records for a specific research project (86%). Many were also willing for their self-reported
10 data and current (62%) or all future (44%) health records to be shared with other researchers
11 for use in other studies if made aware of this. Willingness to consent to data access and
12 sharing data in cancer research was influenced by: (i) the potential for data sharing to
13 advance medical discoveries and benefit people impacted by cancer in the future, (ii)
14 transparency around researchers' credibility and their intentions for data sharing, (iii) level of
15 ownership and control over data sharing, and (iv) protocols for privacy and confidentiality in
16 data sharing. Based on these themes, we present practical strategies for optimizing data
17 access and sharing in cancer research.

18 **Keywords**

19 Clinical cancer research, community outreach, ethical considerations, translational research

20 **Introduction**

21 With cancer a leading cause of mortality worldwide (1), there is an ongoing need for
22 research-driven innovations in cancer treatment and survivorship (2). Achieving this requires
23 researchers to collect, store, and analyze a wide range of data, including demographic, health,
24 psychosocial, behavioral, and genetic information (3). Collecting these data from a large,
25 representative samples in target populations is important for generating valid research outputs
26 that can inform future interventions (4). Further, using these data in other research projects or
27 sharing with other researchers reduces participant burden, prevents duplication of work, and
28 supports mutual resource use (5). Thus, there is a need to not only optimize community
29 participation in cancer research, but also facilitate efficient data collection and sharing.

30 The collection, use, storage, and sharing of human research data is tightly regulated
31 through legislative and institutional policies and guidelines to protect the privacy and well-
32 being of research participants. For example, in Australia, researchers are required to adhere to
33 the principles and guidelines outlined in the National Statement on Ethical Conduct in
34 Human Research in order to receive ethical approval and funding from research institutions
35 and funding bodies (6). In most instances, these guidelines require researchers to disclose
36 information about the purpose of the research to prospective participants, including what their
37 participation will involve, and obtain their voluntary consent to participate, which if feasible,
38 may be withdrawn at any stage (6). Ethical approval also requires that personal information is
39 stored securely and kept confidential (6). Such regulation serves to protect individuals with
40 respect to the use of their data but can be perceived as a barrier to processes such as data
41 sharing and linkage (7).

42 When consumers are considering taking part in medical research, review evidence
43 indicates that a key determinant of research participation is how their data will be collected,

44 stored, and accessed (8). Some studies have investigated consumers' preferences for data
45 sharing within the oncology setting, particularly in the United States, identifying a willingness
46 for data sharing beyond the immediate research project but a need for further disclosure
47 around sharing practices (9–11). However, little is known about the principles that underly
48 consumers' preferences for how their data are used in cancer research and how studies could
49 be designed to optimize data collection and sharing. This paper explores community
50 members' preferences for use of their self-report data and health records in cancer research. It
51 also examines the reasons underlying consumers' preferences for data access and sharing in
52 cancer research and provides practical recommendations for designing and conducting cancer
53 research studies. Understanding and applying consumers' preferences for data access and
54 sharing will support research integrity and may optimise community participation in cancer
55 research.

56 **Methods**

57 *Context and setting*

58 Data were collected for this study during workshops held to co-design a population-
59 based cancer survivorship study in Queensland, Australia, with methods described elsewhere
60 (12). Briefly, co-design research actively involves consumers in developing, designing,
61 implementing, and evaluating new products or services (13). Consumer involvement in
62 research can contribute to better study outcomes, including higher enrolment and retention
63 rates (14). The current study adopted a pragmatic approach to co-design with community
64 members participating in workshops to design study invitation materials and a qualitative
65 survey tool for collecting detailed information on the supportive care needs and experiences
66 of people affected by cancer (12). The session activities and visual materials used in the
67 workshops were developed by the broader study team, including cancer survivorship

68 researchers, medical oncologists, and executive and senior leaders in a cancer support
69 organisation and a population-based registry. This paper reports on findings from the final
70 activity in the workshop that was completed in 14 of the 15 workshops ($n=42$ community
71 members from a total sample of 44) (see **Figure 1**). Ethical approval for the study was
72 obtained from the University of Southern Queensland Human Research Ethics Committee
73 (ref: ETH2023-0140).

74 **Figure 1. Flowchart of participant recruitment and selection for the co-design**
75 **workshops**

76 ^a Invalid/ineligible responses to the workshop invitation were identified based on a
77 combination of factors (e.g., in the past duplicated IP addresses with different names,
78 invalid postcodes or phone numbers, replicated responses in a short period of time, unusual
79 completion times have been indicative of phishing attempts). Where validity could not be
80 determined from the response, it was investigated further through phone and/or email
81 contact with the respondent.

82 *Participants and recruitment*

83 Participant recruitment was undertaken from October to December 2023 using digital
84 and printed flyers distributed via networks associated with Cancer Council Queensland or the
85 broader research team. To support the recruitment of priority populations, such as culturally
86 and/or linguistically diverse (CALD) groups, the research team submitted study information
87 to a health consumers network for inclusion in their e-newsletter (15). As data collection
88 advanced, recruitment was supplemented through snowball sampling, with workshop
89 participants invited to share recruitment flyers with friends, family members, and colleagues.
90 Community members were eligible to participate in the co-design workshops if they were
91 aged 18 years or older, English-speaking, and residing in Queensland, Australia. Participants

92 included people with a personal experience of cancer (either as someone diagnosed with
93 cancer or a carer for someone with cancer), as well as community members who did not have
94 a personal experience of cancer. The latter group was included as current cancer incidence
95 rates means it is likely that everyone will be impacted by cancer at some stage in their life,
96 either as a patient or carer (16).

97 Participants were allocated to a workshop based on their nominated availability, with no
98 more than five registered participants per workshop. Recruitment continued until a diverse
99 sample had been achieved and the research questions had been adequately explored,
100 determined by the authors through concurrent data collection and analysis. Due to the large
101 number of online registrations for the interviews, participants were purposively sampled
102 based on their demographic characteristics, including gender, ethnicity, Indigenous status, and
103 geographical location, as well as their experience with cancer (i.e., survivor or carer) to
104 ensure that diverse perspectives were represented.

105 *Data collection*

106 Workshops were facilitated by two female researchers with undergraduate or
107 postgraduate degrees in health science fields and training in qualitative data collection. The
108 facilitators had no prior relationship with workshop participants. At the start of each
109 workshop, the facilitators introduced themselves, including their role in the research team and
110 their academic background. Workshops were conducted as either online (n=9), in-person
111 (n=1), or hybrid (i.e., online and in-person) (n=4) sessions using Microsoft Teams. In-person
112 participants attended the session at one of two not-for-profit organisations, where participants
113 were provided with the relevant materials (e.g., pen, paper). Participants attending online
114 were asked to source these materials themselves Workshops were audio-recorded and
115 transcribed using Microsoft Teams. After completing the workshop, participants received a

116 voucher valued at AU\$100.00 for their time (approx. 120 minutes). Data for the current study
117 were drawn from the final activity of the workshops, when participants were given a series of
118 hypothetical scenarios developed by the research team to prompt discussion about
119 preferences for data collection and sharing in cancer research (**Table 1**). Participants were
120 advised that in the scenarios, ‘self-report data’ refers to information about themselves that
121 they provide to researchers (e.g., needs and experiences), while ‘health records’ refers to
122 information about themselves that researchers gather from registries or medical documents
123 (e.g., cancer type and stage). The facilitator then provided a short explanation for each
124 scenario (see italics in **Table 1**). For scenarios that involved data sharing, participants were
125 informed that their information would be kept confidential. The lead facilitator asked each
126 participant to verbally indicate which scenario (i.e., level of data sharing) they felt
127 comfortable with and discuss the reasoning behind their choice. This often resulted in further
128 discussion among group members around data sharing in cancer research.

129 *Data analysis*

130 Descriptive statistics were used to summarize participant characteristics and
131 preferences for data access and sharing based on the scenarios presented. Workshop
132 transcripts were analyzed using codebook thematic analysis to identify recurring patterns in
133 the data. As described by Braun and Clarke, codebook thematic analysis is a structured
134 approach to coding that conceptualises themes as topic summaries of a central concept and is
135 distinct from their reflexive approach to thematic analysis (17). First, two members of the
136 research team (XB, EJ) reviewed workshop transcripts to familiarise themselves with the
137 data. Second, transcripts were coded inductively by one researcher (XB) based on the words
138 used by participants to describe their data sharing preferences and reasons underlying their
139 preferences. These codes were documented in a coding framework alongside representative
140 participant quotes. Third, initial themes were generated by grouping codes together that

Table 1. Preferences for data access and sharing in cancer research among community members participating in workshops to co-design a population-based cancer research study.

| Scenarios used to prompt discussion about use of data in cancer research [†] | Total N=42 | Personal experience of cancer | | No personal experience of cancer N=8 |
|--|------------------|-------------------------------|---------------|--|
| | | Diagnosis N=18 | Carer N=16 | |
| Where do you sit on the scale? | N (%) | N (%) | N (%) | N (%) |
| 1. I am <u>okay</u> with self-report data that I provide directly to researchers being used for a specific research project. <i>This is the same as what you consented to for this workshop.</i> | 42 (100%) | 18 (100%) | 16 (100%) | 8 (100%) |
| 2. I am <u>okay</u> with #1 + my current health records being used for a specific research project. <i>This means you would not have to self-report information that can be accessed directly from your health record, such as cancer type and stage.</i> | 36 (86%) | 14 (78%) | 15 (94%) | 7 (88%) |
| 3. I am <u>okay</u> with #2 + my self-report data and my current health records being shared with other researchers for other projects if I am made aware of them. <i>Your information would be kept confidential, but your information could be used to provide data for other projects that you are made aware of.</i> | 26 (62%) | 10 (56%) | 11 (69%) | 5 (63%) |
| 4. I am <u>okay</u> with #3 + future health records being shared with other researchers for other projects if I am made aware of them. <i>Similar to the previous option, but you are providing consent for ongoing collection of information about you from your health record that you will be made aware of.</i> | 18 (43%) | 6 (33%) | 7 (44%) | 5 (63%) |

[†] Scenarios are not mutually exclusive and build on each other (i.e., participants who selected Scenario 3 were also comfortable with Scenarios 1 and 2).

1 addressed the same concept (e.g., the need for privacy and confidentiality). Within each
2 theme, initial sub-themes were developed to capture the various perspectives and nuances
3 expressed by participants for each concept. Fourth, themes and sub-themes were further
4 developed through discussion with a second researcher (EJ) and documented using an audit
5 trail. Finally, each theme was assigned a title to summarize the key concept that it
6 represented.

7 **Results**

8 The characteristics of the 42 workshop participants are shown in **Table 2**. Participants
9 had a median age of 43 years (range 23-79 years) and 31 (74%) were female. Thirty-four
10 (81%) people had a personal experience of cancer, either as someone diagnosed with cancer
11 ($n=18$; 43%) or as a carer of someone diagnosed with cancer ($n=16$; 38%). Eleven (26%)
12 people were living in a regional or remote area and 19 (45%) lived in an area of low to
13 medium socioeconomic status. Nine (21%) people were born overseas, 4 (10%) spoke a
14 language other than English at home, and 3 (7%) identified as Aboriginal and/or Torres Strait
15 Islander.

16 **Table 2.** Characteristics of participants in the 14 co-design workshops ($n=42$)

| Characteristic | Co-design workshops, N (%) ^a |
|--|---|
| Age (years) | |
| Median (range) | 43 (23-79) |
| Gender | |
| Female | 31 (74%) |
| Male | 11 (26%) |
| Born overseas | |
| No | 31 (74%) |
| Yes | 9 (21%) |
| Not reported | 2 (5%) |
| Ethnicity | |
| Caucasian | 34 (81%) |
| Asian | 5 (12%) |
| Aboriginal and/or Torres Strait Islander | 3 (7%) |
| Language spoken at home | |
| English only | 36 (86%) |

| | |
|--|----------|
| Other | 4 (9%) |
| Unknown | 2 (5%) |
| Geographic remoteness (ARIA) | |
| Major city | 31 (74%) |
| Regional or remote | 11 (26%) |
| Area-level socioeconomic status (SEIFA)^b | |
| High | 23 (55%) |
| Medium | 13 (31%) |
| Low | 6 (14%) |
| Personal experience with cancer | |
| Cancer survivor | 16 (38%) |
| Cancer caregiver | 16 (38%) |
| Both | 2 (5%) |
| Neither | 8 (19%) |
| Time since diagnosis | |
| <1 year | 4 (10%) |
| 1-5 years | 6 (14%) |
| >5 years | 7 (17%) |
| Unknown | 1 (2%) |
| Not applicable | 24 (57%) |
| Duration of caregiving | |
| <1 year | 7 (17%) |
| 1-5 years | 5 (12%) |
| >5 years | 5 (12%) |
| Unknown | 1 (2%) |
| Not applicable | 24 (57%) |

17 **Abbreviations:** ARIA, Accessibility/Remoteness Index of Australia; SEIFA, Socio-Economic Indexes for Areas

18 ^a Number of participants and percentage of sample unless otherwise stated.

19 ^b Higher scores indicate higher relative socioeconomic advantage and lower relative socioeconomic
20 disadvantage in general. High = deciles 7-10, medium = deciles 4-6, low = deciles 1-3.

21

22 *Preferences for data access and sharing in cancer research*

23 Preferences for data access and sharing in cancer research are shown in **Table 1**. All
24 participants agreed with the baseline scenario of providing self-report data directly to
25 researchers for a specific research project, as they had done for the workshops reported in this
26 paper. In addition to providing self-report data, thirty-six (86%) would be willing to grant
27 researchers access to their *current* health records for a specific research project. Twenty-six
28 (62%) would be willing for their de-identified data and *current* health records from the
29 original research project to be shared with other researchers for other projects if they were

30 made aware of them. Less than half (n=18; 44%) would be willing for their de-identified
31 *future* health records to be shared with other researchers for other projects if they were made
32 aware of them. Of the 26 people who would be willing for their self-report data to be shared
33 with other researchers for other projects, 5 participants (19%) felt they would not need to be
34 made aware of how their information was being used (i.e., receive information about the
35 other research projects and other researchers). These participants included 3 cancer survivors,
36 1 cancer carer, and 1 person with no personal history of cancer.

37 *Reasons underlying preferences for data access and sharing in cancer research*

38 Four themes were identified, representing key principles that underpinned
39 participants' willingness to share their self-report data and health records in cancer research.
40 These themes were: (i) the potential for data sharing to advance cancer research and benefit
41 people impacted by cancer in the future, (ii) transparency around researchers' credibility and
42 their intentions for data sharing, (iii) level of ownership and control over data sharing, and
43 (iv) protocols for privacy and confidentiality (**Table 3**).

44 *Theme 1: Potential for data sharing to advance cancer research and benefit people impacted* 45 *by cancer in the future*

46 This theme related to the purpose and outcomes of the proposed research and how that
47 influenced participants' willingness to share data. Participants expressed a willingness to
48 share their self-report data and health records as part of a cancer research study if sharing this
49 information would advance cancer research discoveries and help other people impacted by
50 cancer. Participants emphasised the importance of research that "makes life better" and
51 "makes a difference" for people affected by cancer in the future:

52 **Table 3.** Themes and sub-themes identified in a thematic analysis of community members’ preferences for data access and sharing in cancer research and
 53 reasons underlying those preferences.

| Themes | Sub-themes |
|---|---|
| Potential for data sharing to advance cancer research and benefit people impacted by cancer in the future | <ul style="list-style-type: none"> • Desire to contribute to new research discoveries and help other people impacted by cancer. • Personal beliefs regarding the value of research and importance of data. • New perspective on data sharing after cancer experience. • Conflict between personal preferences for limited data sharing and contributing to “the greater good”. |
| Transparency around researchers’ credibility and their intentions for data sharing | <ul style="list-style-type: none"> • Need for transparency around researchers’ intentions for data sharing prior to providing consent. • Concerns that data could be sold to third parties or used for commercial purposes. • Lower need for transparency around who is using the data if data sharing is anonymous and/or used to advance medical research. |
| Level of ownership and control over data sharing | <ul style="list-style-type: none"> • Need to retain ownership and control over data, particularly health records. • Unclear why full health record would be relevant to cancer research. • Hesitant to consent to researchers accessing future health record due to uncertainty and potential risk. • Limited need for ownership and control over what data is shared if data sharing is de-identified, opportunity to opt-out in the future, and/or used to benefit others or advance medical research. • Preference for data sharing influenced by previous experiences of data sharing in research studies. |
| Protocols for privacy and confidentiality | <ul style="list-style-type: none"> • Personal information needs to be de-identified when shared and not potentially re-identifiable. • Concerns about privacy due to recent data breaches in the commercial space. • Less concerned about privacy due to the volume of personal information publicly available online. • Clear communication about protocols for privacy and confidentiality could encourage data sharing. |

54

55 “If my data and my experience and my records help someone research something that
56 makes life better for people that are coming up behind me, I’m all for that.” (P16,
57 *Workshop J, personal cancer diagnosis*)

58 “I’ve been incredibly active in this space... I’m a consumer reviewer for [cancer
59 research grants] ... Would I give my data to big pharma to make an expensive drug
60 that not many people can afford? No. Would I give my data to someone who is
61 working towards the benefit of all? Yes. I love translational research. It makes a
62 difference for the end users.” (P42, *Workshop A, carer*)

63 Personal beliefs regarding the value of research and importance of data encouraged
64 data sharing:

65 “[Data] is important, and it could help shape the way cancer is treated in the future if
66 [researchers] have more data. The more data [researchers] have, the more knowledge
67 they have about the situation.” (P7, *Workshop M, no personal experience of cancer*)

68 For some participants, their experience of cancer changed their perspective on data
69 sharing. Having gained a better understanding of the importance of cancer research they
70 indicated they would be more willing to share their personal information:

71 “I probably used to be on the other end of the scale and didn’t really share much,
72 particularly online. This cancer experience sort of flipped that for me. That’s my
73 ‘why’... I would probably say yes to scenario 4 [future health records shared with
74 other researchers for other projects]. I know based on my own experience with my
75 son, the drug that saved him has never been used before (that we are aware of) in a
76 child as young as him or his type of cancer.” (P35, *Workshop C, carer*)

77 Some participants expressed a conflict between their personal preference for limited
78 data sharing and being more open to data sharing for “the greater good”:

79 “I guess I have some uneasiness around [sharing] future health records... but when I
80 heard [workshop participant] reiterate how important clinical trials are, I hesitated and
81 I thought, maybe the greater good is more important.” (P31, *Workshop D, carer*)

82 “As each [scenario] came up, my risk adverse nature kicked in. Thank you to the other
83 participants – you’ve opened me up. I’m trying to let my walls down.” (P32,
84 *Workshop D, carer*)

85 *Theme 2: Transparency around researchers’ credibility and their intentions for data sharing*
86 *in cancer research*

87 This theme addressed participants’ desire to know who is conducting the research and
88 how their data would be used. Participants valued transparency around researchers’ intentions
89 for data use prior to providing consent for data sharing:

90 “I do like to know where the information is going. I think an important part of consent
91 is being aware of [how the data will be used].” (P35, *Workshop C, carer*)

92 “Every specific research [project that uses my data], I need to be aware of... What are
93 the intentions of the project?” (P21, *Workshop H, no personal experience with cancer*)

94 Participants expressed concerns that their data could be sold to third parties or used
95 for commercial purposes rather than “public good”. Participants indicated that the level of
96 transparency around researchers’ credibility and intentions would influence their willingness
97 to share data as part of cancer research:

98 “I would need to be aware of who [the researchers] are. It could change my mind...
99 Some people come to the table with terrific credibility. Others you may not know
100 about. [You may be giving] the chance for someone to sell your information because

101 we all know that happens. [I would] probably need to be aware of who they are, rather
102 than just ‘researchers’, for me to be comfortable with that.” (P4, *Workshop O, carer*)

103 “There are a number of researchers that are looking to privatize their ideas. They are
104 looking for public money, but they are looking to commercialize their results. I’m a
105 believer in public funds for public good. I wouldn’t give my data to any old
106 researcher. My answer is quite nuanced. Is it public or private? What is the intent? Is
107 the research going to be siloed? Is it going to see the light of day?” (P42, *Workshop A,*
108 *carer*)

109 On the other hand, some participants indicated a lower need for transparency around
110 data sharing, provided privacy and confidentiality were maintained or the data were used to
111 advance medical research:

112 “[I would be comfortable] as long as [my data] is still kind of contained, it's just
113 [accessible to] researchers, it is not out to the world. I don't think I'd need to know the
114 specific projects, but it would be nice to know when it is being used.” (P34, *Workshop*
115 *C, carer*)

116 “If I wasn’t made aware [of who was using the data and how], I’d be a little bit
117 hesitant, but still be happy to give it if it’s medical research and it’s for a cure and I
118 can help in any way.” (P27, *Workshop F, carer*)

119 *Theme 3: Level of ownership and control over data sharing in cancer research*

120 This theme related to participants’ considerations about what data they would be
121 willing to share as part of cancer research and their level of autonomy over this. Participants
122 expressed a need to retain ownership and control over their data, particularly their health
123 records. Participants also wanted the opportunity to choose their level of data sharing for each
124 project:

125 “I would look at everything on a case-by-case basis. I wouldn’t mind [researchers]
126 using my health records but only if it’s for something that I considered fully before I
127 opted in.” (P15, Workshop J, personal cancer diagnosis and carer)

128 “I want to have ownership and control of my own data. I’m [name]. I’m not my body.”
129 (P28, Workshop F, no personal experience with cancer)

130 “I think you could [provide the option] ‘I’m willing to be contacted for future research
131 projects’ [on the consent form]. You can give people the option to say, I do want to keep
132 contributing, if there is something else in future, I just want to have a sphere of
133 influence around that.” (P32, Workshop D, carer)

134 Some participants were unclear why all the information in their health record would be
135 relevant to cancer research. This limited their willingness to share this data, specifically
136 records not related to their cancer diagnosis:

137 “I’ve opted out of my health record because I don’t feel it is anybody’s business
138 knowing my vaccination status or any of my past history.” (P25, Workshop G, personal
139 cancer diagnosis)

140 “If it was going to cancer research, the only records that would be relevant in my mind
141 is the cancer-related records. To throw in a lung condition ... I think it would be more
142 confusing. Not my full records. A lot of it wouldn’t be relevant at all.” (P18, Workshop
143 I, personal cancer diagnosis/ carer)

144 Other participants were hesitant to grant researchers access to their future health
145 information, mostly due to uncertainty around their future situation and the risk of personal
146 and sensitive information being shared:

147 “I guess I have some uneasiness around future health records because I don’t yet
148 know what my future health issues will be. If it was something around fertility, that
149 might be something quite private, something that is sensitive to me. Maybe I might
150 not feel comfortable with that.” (P31, Workshop D, carer)

151 “I don't know what future health problems I'm going to have and how I'm going to
152 react emotionally to them yet, so I don't know if I want to put myself in a position to
153 give out all my data for a future issue... [sharing] data I currently have now, very
154 comfortable. But I don't really know about the future if things will change.” (P5,
155 Workshop N, carer)

156 Some participants expressed a limited need for ownership and control over their data
157 if other conditions were met. These included de-identification, opportunity to opt-out, and the
158 data being used “for good”:

159 “If it is anonymous and being used to help or prevent something in the future, I
160 personally don’t see anything wrong with it.” (P41, Workshop A, carer)

161 “I'm pretty liberal with data sharing as long as there is a way to opt out should I need
162 to.” (P9, Workshop L, carer)

163 “If my data can help someone in the future not suffer as much or lead to a cure or
164 better treatment, something that improves outlook for future patients, I'd be all for
165 that.” (P16, Workshop J, personal cancer diagnosis)

166 Other participants indicated that their preference for data sharing was influenced by
167 the degree of data sharing that they have previously consented to as part of a research study:

168 “My health records have already been out in the world. I have been in clinical trials to
169 help future research for cancer. So, whatever I can do to help.” (P24, *Workshop G,*
170 *personal cancer diagnosis*)

171 “I’ve participated in another study where I took medications and I consented to my
172 medical records [being used] for that study only... that is okay, definitely not just for
173 any project.” (P19, *Workshop I, carer*)

174 *Theme 4: Protocols for privacy and confidentiality of data sharing in cancer research*

175 This theme addressed participants’ desire to know how their data would be shared
176 with researchers and how it would be stored, protected and reported. Privacy and
177 confidentiality were important to most participants, including their personal information
178 being de-identified when shared and not potentially re-identifiable:

179 “I wouldn’t want my name and date of birth and address and [phone] number being
180 shared with everyone. If I’m just an entity in the system, a ‘male in their 40s that had
181 this’ and ‘this was my treatment’, I’m totally fine.” (P16, *Workshop J, personal*
182 *cancer diagnosis*)

183 “As long as it [indicates] that my future health records will be shared in a non-
184 identifiable manner... it has to be non-identifiable.” (P33, *Workshop D, carer*)

185 Some participants expressed concerns about privacy due to recent data breaches in the
186 commercial space, emphasising the importance of ensuring secure data storage and sharing
187 practices:

188 “Is there any way the data can get out to the public? [This is] the one thing I need to
189 ask in light of certain companies having problems with security or going down, like

190 [company name]. How do we know the software protecting this information is safe
191 enough?" (P20, Workshop H, personal cancer diagnosis)

192 "The number one question is can it be hacked? Can it be scammed? If I'm giving you
193 all my data, it's lovely that you tell me it's going to be fine, but I was with [company
194 name] and [company name] and look what happened." (P39, Workshop B, personal
195 cancer diagnosis)

196 Others were less concerned about privacy due to the volume of personal information
197 that is publicly available online:

198 "If you don't want to share your data for good, then the likelihood is that you will get
199 your data hacked anyway, and your data may get lost. Data is just data." (P41,
200 Workshop A, carer)

201 "As far as I'm concerned, people know everything about you anyway." (P14,
202 Workshop K, no personal experience with cancer)

203 Participants agreed that clear communication with prospective research participants
204 about protocols for privacy and confidentiality could encourage data sharing: "I think
205 the important thing is to explain what checks and balances and protocols are in place
206 around people's private information, particularly in the health area. I think it would be
207 important it was clearly pointed out in communication to the people involved in the
208 survey. That would be the number one priority." (P8, Workshop M, personal cancer
209 diagnosis)

210 "Unless you reassure them [prospective research participants], they are going to
211 default to scenario one [no data sharing]. Not because they don't believe in the
212 usefulness of their information for research but because they don't have the
213 information about what's in place [to protect their data]... You need to get your

214 wording right up to do date about whatever the latest is for integrity in sharing data.”

215 (*P39, Workshop B, personal cancer diagnosis*)

216 **Discussion**

217 This study provides important information on the types of data that community
218 members may be willing to share in cancer research, with whom, and for what purpose. In
219 general, most community members are willing to share their self-report data and health
220 records in a cancer research study, and many are willing to provide consent for this
221 information to be shared with other researchers for other studies that they are made aware of.
222 For example, community members were happy to consent to receiving information about
223 future studies where their data could be used as they arose and to decide on data sharing on a
224 case-by-case basis. In general, community members were less willing to share their future
225 health records due to uncertainty around their future health and the potential for sensitive
226 information to be disclosed. As this is a small qualitative study, we were unable to further
227 explore preferences for data access and sharing in cancer research based on participant
228 characteristics.

229 While we identified four themes underlying community members’ willingness to share
230 their data in cancer research, the sub-themes presented within each theme reveal that these
231 themes are intertwined and should be viewed together. For instance, if data sharing was
232 anonymous and the research findings would benefit other people impacted by cancer in the
233 future, then community members indicated a lower need for transparency around who was
234 using their data and a lower need for ownership and control over the projects their data were
235 used in. A reoccurring narrative throughout the themes was that clear communication around
236 the potential benefits of the research, what data would be shared with whom and how this
237 would be undertaken, and the protocols in place to ensure anonymity, could encourage

238 community members to consent to data sharing in cancer research. This highlights the
239 importance of designing study invitation materials with community members to optimise the
240 readability, relevance, and acceptability of these materials (12).

241 The findings presented in this paper expand on previous studies investigating data
242 sharing preferences among people diagnosed with cancer. Consistent with our findings,
243 several studies from the United States have shown that cancer patients are generally willing
244 to share their health data to improve care, both in general and for themselves (9–11). In those
245 studies, patients have also reported a need for transparency and control regarding how, when,
246 and why their data are used, and the importance of measures that protect their privacy and
247 security (9–11). The current study demonstrates the applicability of these findings in the
248 Australian setting, not only with cancer survivors, but also people caring for someone with
249 cancer. A novel finding of this study is that the potential for research to benefit people
250 affected by cancer in the future is a key motivation for consenting to data sharing. However,
251 willingness to share data is often contingent on a combination of factors, such as the
252 credibility of the research team and assurance of data anonymity, particularly in the context
253 of increased cybersecurity concerns and breaches. By understanding these nuanced
254 preferences, researchers can better design study protocols that align with the values and
255 expectations of the community, thereby enhancing participation in cancer research.

256 Progressing data sharing in cancer research will require broad engagement with a
257 range of stakeholders. In Australia, while the current data landscape includes many
258 population-based datasets that are relevant to cancer research, there are several barriers to
259 data sharing (18). These include siloed datasets with data custodians bound by restrictive
260 legislation and approval processes for data access that can take years to negotiate and execute
261 (18). These challenges are long-standing, with a systematic review published ten years ago
262 reporting twenty potential barriers to data sharing in public health across technical,

263 motivational, economic, political, legal, and ethical spheres (19). While addressing these
264 barriers requires extensive consultation with, and commitment from, stakeholders and data
265 custodians, findings from our study demonstrate community support for greater data sharing
266 in cancer research and contribute to community engagement efforts to establish protocols for
267 advancing data sharing in cancer research.

268 Based on our findings, there are several steps that researchers can take to support data
269 sharing in cancer research, despite the system-level barriers. For example, researchers can
270 emphasize on study recruitment materials or during debrief sessions how participants’
271 involvement in the research, including sharing data with other researchers for other projects,
272 may benefit people impacted by cancer in the future and contribute to new medical
273 discoveries, and how their anonymity will be maintained in data sharing and reporting.
274 Researchers can also provide prospective participants with links to information about the
275 research team and examples of how their previous work has been implemented in practice. If
276 offering participants the option to consent to their data being used by other researchers for
277 other projects, assure community members that they will be provided with information about
278 each project and the researchers involved, and will have the to confirm consent or opt-out of
279 sharing their data. Finally, if asking community members to consent to their full health record
280 being used for a cancer research study, provide information on why non-cancer related
281 records are relevant to the proposed research.

282 *Strengths and Limitations*

283 This study included a diverse sample of community members, including people living
284 in rural areas and those with English as a second language. A key limitation of this study is
285 that participants were community members who had already consented to providing self-
286 report data for a cancer research study. Therefore, findings from this study may not represent

287 the views of community members in general, particularly those who prefer to not engage in
288 research and those who experience geographical, technological, or language barriers to
289 research participation; groups that are hard to reach but likely important to include in
290 population-based cancer research studies. A large proportion of the workshop participants
291 were female, however, there is limited evidence to suggest that gender influences data sharing
292 preferences for health research in general (20,21). While the themes identified in this study
293 were discussed by participants with and without a personal experience of cancer, the
294 preferences for data access and sharing reported by the small number of people with no
295 personal experience of cancer could change if they were to be diagnosed, or care for
296 someone, with cancer (22).

297 Groupthink is a commonly cited limitation of group-based qualitative research, where
298 members of a group seek cohesion and conformity in decision-making rather than diversity
299 and different perspectives (23). While our workshops included group discussion around
300 reasons underlying preferences for data access and sharing, the preferences reported by
301 participants in nearly all the workshops were varied. Some participants did indicate that the
302 group discussion challenged their own perspective on data sharing, however, few participants
303 changed their preferred scenario after participating in the group discussion.

304 *Conclusions*

305 Overall, these findings provide community support for improved data sharing in
306 cancer research, a priority for advancing cancer control. Considerations underlying
307 community members' willingness to share data in cancer research studies included the
308 potential for data sharing to advance medical discoveries and benefit others impacted by
309 cancer in the future, transparency around researchers' intentions for data sharing and their
310 credibility, level of ownership and control over data sharing, and protocols for privacy and

311 confidentiality. Findings provide several practical strategies for researchers to use when
312 designing cancer research studies, including emphasizing on study invitation materials how
313 participants' involvement in the research will benefit people impacted by cancer in the future,
314 explaining why access to participants' full medical record is relevant for the proposed study,
315 and providing opportunities for people to consent to being contacted about future research
316 studies where their data could be used. Incorporating community preferences into the design
317 and conduct of cancer research studies has the potential to optimise community participation
318 and the applicability and translation of research findings into practice.

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2021;71(3):209–49.
2. Elmore LW, Greer SF, Daniels EC, Saxe CC, Melner MH, Krawiec GM, et al. Blueprint for cancer research: Critical gaps and opportunities. *CA Cancer J Clin*. 2021;71(2):107–39.
3. Cancer Australia. A National Cancer Data Strategy for Australia [Internet]. [cited 2024 Mar 1]. Available from: <https://www.canceraustralia.gov.au/publications-and-resources/cancer-australia-publications/national-cancer-data-strategy-australia>
4. Glasgow RE, Emmons KM. How can we increase translation of research into practice? Types of evidence needed. *Annu Rev Public Health*. 2007;28:413–33.
5. Green AK, Reeder-Hayes KE, Corty RW, Basch E, Milowsky MI, Dusetzina SB, et al. The project data sphere initiative: Accelerating cancer research by sharing data. *The Oncologist*. 2015;20(5):464–e20.
6. NHMRC. National Statement on Ethical Conduct in Human Research 2023 [Internet]. 2023 [cited 2024 Feb 29]. Available from: <https://www.nhmrc.gov.au/about-us/publications/national-statement-ethical-conduct-human-research-2023>
7. Andrew NE, Sundararajan V, Thrift AG, Kilkenny MF, Katzenellenbogen J, Flack F, et al. Addressing the challenges of cross-jurisdictional data linkage between a national clinical quality registry and government-held health data. *Aust N Z J Public Health*. 2016;40(5):436–42.
8. Kirkby HM, Calvert M, Draper H, Keeley T, Wilson S. What potential research participants want to know about research: A systematic review. *BMJ Open*. 2012;2(3):e000509.
9. Raj M, Ryan K, Amara PS, Nong P, Calhoun K, Trinidad MG, et al. Policy preferences regarding health data sharing among patients with cancer: Public deliberations. *JMIR Cancer*. 2023;9:e39631.
10. Jones RD, Krenz C, Griffith KA, Spence R, Bradbury AR, De Vries R, et al. Patient experiences, trust, and preferences for health data sharing. *JCO Oncol Pract*. 2022;18(3):e339–50.
11. Beckjord EB, Rechis R, Nutt S, Shulman L, Hesse BW. What do people affected by cancer think about Electronic Health Information Exchange? Results from the 2010 LIVESTRONG Electronic Health Information Exchange Survey and the 2008 Health Information National Trends Survey. *J Oncol Pract*. 2011;7(4):237–41.
12. Ayre SK, Johnston EA, Bourdaniotis XE, Zajdlewicz L, Beesley VL, Pole JD, et al. From many voices, one question: Community co-design of a population-based qualitative cancer research study. medRxiv; 2024.

13. Vargas C, Whelan J, Brimblecombe J, Allender S. Co-creation, co-design, co-production for public health – A perspective on definition and distinctions. *Public Health Res Pract.* 2022;32(2).
14. Domecq JP, Prutsky G, Elraiyah T, Wang Z, Nabhan M, Shippee N, et al. Patient engagement in research: A systematic review. *BMC Health Serv Res.* 2014;14(1):89.
15. Health Consumer Queensland. Health Consumers Queensland [Internet]. 2024. Available from: <https://www.hcq.org.au/>
16. Cancer Australia. Cancer in Australia statistics [Internet]. [cited 2024 Jul 4]. Available from: <https://www.canceraustralia.gov.au/impacted-cancer/what-cancer/cancer-australia-statistics>
17. Braun V, Clarke V. Conceptual and design thinking for thematic analysis. *Qual Psychol.* 2022;9(1):3–26.
18. Cancer Council Australia. Developing a National Data Strategy for Cancer: A Report for Discussion [Internet]. Cancer Council Australia; 2023. Available from: <https://www.cancer.org.au/assets/pdf/developing-national-data-strategy-for-cancer>
19. van Panhuis WG, Paul P, Emerson C, Grefenstette J, Wilder R, Herbst AJ, et al. A systematic review of barriers to data sharing in public health. *BMC Public Health.* 2014;14:1144.
20. Sørnum H, Eg R, Presthus W. A gender perspective on GDPR and information privacy. *Procedia Comput Sci.* 2022;196:175–82.
21. Aitken M, McAteer G, Davidson S, Frostick C, Cunningham-Burley S. Public preferences regarding data linkage for health research: A discrete choice experiment. *Int J Popul Data Sci.* 3(1):429.
22. FeldmanHall O, Mobbs D, Evans D, Hiscox L, Navrady L, Dalgleish T. What we say and what we do: The relationship between real and hypothetical moral choices. *Cognition.* 2012;123(3):434–41.
23. Cyr J. The pitfalls and promise of focus groups as a data collection method. *Sociol Methods Res.* 2016;45(2):231–59.

