

RESEARCH

Open Access

Predictors of clinical trial data sharing: exploratory analysis of a cross-sectional survey

Vinay K Rathi¹, Kelly M Strait², Cary P Gross³, Iain Hrynaszkiewicz⁴, Steven Joffe⁵, Harlan M Krumholz^{2,6,7}, Kristina Dzara³ and Joseph S Ross^{2,3,7*}

Abstract

Background: A number of research funders, biomedical journals, pharmaceutical companies, and regulatory agencies have adopted policies advocating or mandating that clinical trialists share data with external investigators. We therefore sought to determine whether certain characteristics of trialists or their trials are associated with more unfavorable perceptions of data sharing. To date, no prior research has addressed this issue.

Methods: We conducted an exploratory analysis of responses to a cross-sectional, web-based survey. The survey sample consisted of trialists who were corresponding authors of clinical trials published in 2010 or 2011 in one of six general medical journals with the highest impact factors in 2011. The following key characteristics were examined: trialists' academic productivity and geographic location, trial funding source and size, and the journal in which it was published. Main outcome measures included: support for data sharing in principle, concerns with data sharing through repositories, and reasons for granting or denying requests. Chi-squared tests and Fisher's exact tests were used to assess statistical significance.

Results: Of 683 potential respondents, 317 completed the survey (response rate 46%). Both support for data sharing and reporting of specific concerns with sharing data through repositories exceeded 75%, but neither differed by trialist or trial characteristics. However, there were some significant differences in explicit reasons to share or withhold data. Respondents located in Western Europe more frequently indicated they have or would share data in order to receive academic benefits or recognition when compared with respondents located in the United States or Canada (58 versus 31%). In addition, respondents who were the most academically productive less frequently indicated they have or would withhold data in order to protect research subjects when compared with less academically productive respondents (24 versus 40%), as did respondents who received industry funding when compared with those who had not (24 versus 43%).

Conclusions: Respondents indicated strong support for data sharing overall. There were few notable differences in how trialists viewed the benefits and risks of data sharing when categorized by trialists' academic productivity and geographic location, trial funding source and size, and the journal in which it was published.

Keywords: Data sharing, Clinical trial, Data repository

* Correspondence: joseph.ross@yale.edu

²Center for Outcomes Research and Evaluation, Yale-New Haven Hospital, 1 Church Street, Suite 200, New Haven, CT 06510, USA

³Section of General Internal Medicine and the Robert Wood Johnson Foundation Clinical Scholars Program, Department of Medicine, Yale University School of Medicine, P.O. Box 208093, New Haven, CT 06520, USA
Full list of author information is available at the end of the article

Background

A number of major clinical research funders [1-3] and biomedical journals [4-7] have adopted policies supporting or mandating clinical trial data sharing in an effort to maximize the reproducibility and reliability of the results of clinical trials and maximize return on investment in research. Sharing data can increase its value in a number of ways, including by allowing external investigators to investigate secondary hypotheses, aiding future trial design, and providing training resources for students and fellows [8]. Clinical researchers most commonly share data either by depositing it in repositories, with terms of access defined by maintaining organizations, or by granting personal requests for the data on their own terms. Secondary users of trial data, such as the Cochrane Collaboration, are now advocating for stronger data sharing policies as a means to safeguard against selective reporting of outcomes and improve the medical evidence base for clinical decision-making [9-17]. These growing calls for clinical trial data sharing have prompted the European Medicines Agency [18] and several companies in the pharmaceutical industry [19-21] to commit to far-reaching data access policies for the first time. It now appears that many policymakers are no longer contemplating the need to share data, but rather the best way to do so [22]. Indeed, the United States Food and Drug Administration has recently proposed to share de-identified clinical trial data for the first time, albeit on a limited basis and such that product identity would not be disclosed, only therapeutic class [23].

Recently, we conducted a survey of clinical trialists to inform these ongoing data sharing efforts. Despite previous research finding clinical researchers to be amongst the least likely to share data [24], we found strong support for clinical trial data sharing, both in principle and in practice [25]. Most, but not all, respondents to our survey agreed that the clinical research community should promote and facilitate data sharing, and the majority of respondents who were required or requested to share data had done so. However, respondents identified a number of practical concerns with sharing trial data. Because trialists represent a diverse community of investigators with respect to funding sources, research aims and scope, and academic aspirations, they may vary in their concerns or support for data sharing. For instance, trialists receiving industry funding and investigators who do not publish frequently may both be less inclined to share data, albeit for different reasons. Trialists receiving industry funding may withhold data because of restrictions on their control over the data [26-29] or potential conflicts with their research funders, whereas investigators who do not publish frequently may withhold data in an effort to optimize their own academic productivity. Similarly, trialists conducting either very large or small clinical studies may be less inclined to share data for

different reasons, as trialists conducting bigger studies may be concerned with receiving adequate return on investment from large-scale undertakings in data collection, and trialists conducting smaller studies may be concerned with protecting patient confidentiality. On the other hand, trialists who published their clinical studies in journals advocating for data sharing and those located in Western Europe may be more inclined to share data in accordance with journal data sharing policies [5,30], and a cultural shift towards data sharing in their scientific community [18], respectively. To date, no research has addressed this issue.

Accordingly, using the data collected during the previously described survey, we conducted a secondary analysis to explore whether concerns and support for data sharing varied among trialists according to professional background and aspects of the trials they conducted. In particular, we examined the following key characteristics, which we anticipated may influence trialists' perceptions of data sharing: trialists' academic productivity, geographic location, and trial characteristics including funding source, size, and the journal in which it was published. We expect these findings will help inform policy-makers by identifying groups of trialists who may be more resistant to new requirements and regulations on data sharing, or who may have specific concerns about data sharing that could be addressed, in order to facilitate data sharing.

Methods

Study sample and design

The study sample, design, and data have been reported in detail [25]. In brief, we assembled a sample of clinical trialists ($n = 683$) using an Ovid Medline search to identify the first corresponding author named in each clinical trial published in 2010 or 2011 in one of six general medical journals with the highest impact factors in 2011 (Journal Citation Reports, Thomson Reuters; *New England Journal of Medicine*, *Lancet*, *Journal of the American Medical Association*, *Annals of Internal Medicine*, *PLoS Medicine*, and *The BMJ*). We also obtained from the article the following information from the original article: trialists' geographic location and affiliation, trial funding source and size, and the journal in which the study was published. This information was used to compare survey respondents and non-respondents. After soliciting participation from our study sample via email and telephone, a total of 317 corresponding authors completed an anonymous web-based survey, yielding a 46% response rate. Participation was voluntary and included an opportunity to win one of five \$100 gift certificates from Amazon.com. Approval from the Yale University School of Medicine Human Research Protection Program was obtained before conducting the study (HIC Protocol# 1207010491), and

consent was considered to be implied when participants completed the online survey.

Survey instrument development and domains

The development and domains of the survey instrument have been fully described [25]. Briefly, the design of our 38-item survey instrument (Additional file 1) was informed by a review of the literature [8,31-39] and discussions among the authors and with experts. The survey was conducted between July and September of 2012 using the Qualtrics™ (Provo, Utah, United States) online platform. The survey instrument assessed support for and prevalence of data sharing, concerns with and reasons for data sharing, and respondent characteristics. Respondents were first asked whether they were required by their research funder to deposit data from their published study in a repository and if they had done so. Next, respondents were asked whether they supported requirements for data sharing in principle, through data repositories, and in response to personal requests. Respondents were then asked about concerns with sharing data through repositories, experiences with receiving and making personal data sharing requests, reasons for granting or declining personal requests, and their beliefs on the right of first use of trial data. Finally, the following self-reported sociodemographic and professional information was obtained: age, sex, primary employer, academic rank (if applicable), geographic location, academic productivity, and funding status.

Main outcome measures

We examined four main outcome measures to explore differences in support for and concerns about data sharing among clinical trialists. First, we examined support for data sharing in principle, ascertained through the statement 'The clinical research community should promote and facilitate clinical trial data sharing'; responses were categorized as 'Agree' or 'Disagree'. Second, we examined general concerns with sharing data through repositories, ascertained through one of the following three questions, depending upon respondents' indicated experience with sharing their study data through a repository: 'If you had been required to share the deidentified data from this study through a data repository, would you have any of the following concerns?', 'Since sharing the deidentified data from this study through a data repository, have you had any of the following concerns?', or 'In anticipation of sharing the deidentified data from this study through a data repository, do you have any of the following concerns?'. Responses were categorized as 'Any Concern' or 'No Concerns' and those with concerns were then prompted to select any of the following items as applicable: concerns related to investigator or funder interests, the protection of research subjects, appropriate data use, or other concerns.

Third, we examined general reasons for sharing data in response to personal requests, ascertained through one of the following two questions, depending on respondents' indicated experience with sharing their study data in response to personal requests: 'For what general reason(s) did you share the study data?' or 'For what general reason(s) would you be most likely to share the study data?'. Those with reasons for sharing data were then prompted to select any of the following items as applicable: administrative requirements, promotion of open science, academic benefits or recognition, or other reasons. Lastly, we examined general reasons for withholding data in response to personal requests. This was ascertained through one of the following two questions, depending on respondents' indicated experience with withholding their study data in response to personal requests: 'For what general reason(s) did you not share the study data?' or 'For what general reason(s) would you be most likely to not share the study data?'. Those with reasons for withholding data were then prompted to select any of the following items as applicable: protect investigator or funder interests, protect research subjects, ensure appropriate data use, or other reasons.

Main independent variables

We explored five key characteristics of trialists or their trials in an effort to differentiate support for and concerns about data sharing among clinical trialists: trialists' academic productivity and geographic location, trial funding source and size, and the journal in which it was published. These characteristics were selected based on a review of the literature, which, among other reasons, identified trialists' motivation to publish [8,33,35], the protection of proprietary interests [32], ineffective journal-mandated policies [35,40], and concern for patient confidentiality [32] as potential barriers to clinical trial data sharing. These characteristics were identified prior to conducting our exploratory analysis.

Trialists' academic productivity was assessed by asking respondents how many research articles they had published over the past three years and what percent overall job effort they had devoted to research activities during the 2011 to 2012 fiscal year. Our findings were consistent regardless of how academic productivity was defined, and so results are only presented when defining academic productivity by the number of research articles published over the last three years (≤ 10 articles, 11 to 25 articles, or > 25 articles). Trialists' geographic location was categorized as United States or Canada, Western Europe, or elsewhere. Trial funding source was categorized as solely funded by government sources, solely funded by for-profit industry sources or mixed funding (which includes industry funding), or solely funded by non-profit sources (such as charitable foundations, professional associations, and

universities). Trial size was divided by quartile; of these, the middle two groups were then combined (≤ 239 study subjects, 240 to 2,016 study subjects, and $\geq 2,017$ study subjects). Lastly, the journal in which the trial was published was assigned as one of the following: *New England Journal of Medicine*, *Lancet*, *Journal of the American Medical Association*, *Annals of Internal Medicine*, *PLoS Medicine*, and *The BMJ*. Among these journals, *Annals of Internal Medicine* [5], *The BMJ* [30], and *PLoS Medicine* [41] have advocated for data sharing, and the *New England Journal of Medicine* has recently begun requiring authors to post clinical trial protocols [42].

Statistical analysis

The analyses were not pre-specified at the time of survey administration. Rather, they were conducted with the purpose of generating hypotheses for future research efforts seeking to inform the development of data sharing policies. As described previously [25], we used chi-square tests for categorical variables (author employer and geographic location, trial funder, and journal) and the Kruskal-Wallis test for continuous variables (trial enrollment) to compare respondents and non-respondents. We used descriptive statistics to summarize our main outcome measures overall, and by the key characteristics of trialists and their trials described above. Chi-squared tests and Fisher's exact tests were used to assess statistical significance. In order to account for multiple comparisons of the survey responses for the five main independent variables of interest, we applied the Bonferroni correction, specifying a *P*-value threshold of 0.01 or less to be considered a statistically significant difference. Survey data were analyzed by KMS using SAS version 9.3 (SAS Institute Inc., Cary, North Carolina, United States).

Results

There were 317 corresponding authors of clinical trials who responded to our survey. The majority of respondents were between 50 and 64-years-old (50%) and male (77%, Table 1). A minority of respondents published 10 articles or fewer over the past three years (22%), with roughly equal numbers publishing between 11 and 25 articles (37%), and more than 25 articles (41%) during the same period. More than half of our respondents were located in United States or Canada (53%), with over one third located in Western Europe (36%), and the remaining located elsewhere (11%). Nearly half of all respondents received industry funding for their study (48%), whereas a much smaller group was supported solely by non-profit funding sources (14%). With respect to trial size, a quarter of respondents were corresponding authors of trials with approximately 240 subjects or fewer, and another quarter with approximately 2,000 subjects or more. The majority of respondents

Table 1 Respondent and trial characteristics

	Respondents (%)
Age	
≤49 years	126 (40)
50 to 64 years	159 (50)
≥ 65 years	31 (10)
Sex	
Male	243 (77)
Female	73 (23)
Trialist academic productivity (research articles published in past three years)	
1 to 10 articles	71 (22)
11 to 25 articles	117 (37)
>25 articles	129 (41)
Trialist geographic location	
United States or Canada	167 (53)
Western Europe	113 (36)
Other	37 (11)
Trial funding source	
Government	120 (38)
Industry or mixed funding	152 (48)
Other	45 (14)
Trial size	
≤239 subjects	80 (25)
240 to 2,016 subjects	158 (50)
≥2,017 subjects	79 (25)
Journal in which trial was published	
<i>NEJM</i>	113 (36)
<i>Lancet</i>	70 (22)
<i>JAMA</i>	43 (14)
<i>Annals of Internal Medicine</i>	24 (8)
<i>The BMJ</i>	53 (16)
<i>PLoS Medicine</i>	14 (4)

JAMA, *Journal of the American Medical Association*; *NEJM*, *New England Journal of Medicine*.

published their trial in *New England Journal of Medicine* (36%), *Lancet* (22%), or *Journal of the American Medical Association* (14%). Approximately one fourth of respondents (28%) published their trial in one of the three journals advocating for data sharing, which included *Annals of Internal Medicine* (16%), *The BMJ* (8%), and *PLoS Medicine* (4%). Survey respondents did not differ from non-respondents with respect to corresponding author location or affiliation, trial enrollment, or journal in which the article was published. However, trial funders did differ between respondents and non-respondents; corresponding authors of trials funded solely by government sources responded more often than corresponding

authors of trials funded solely by industry or by mixed funding sources (data not shown, provided in prior publication [25]).

Support for data sharing

We found no significant differences in support for data sharing in principle between respondents categorized by trialists' academic productivity and geographic location, trial funding source and size, and the journal in which it was published (Table 2), as rates of support ranged between 81 and 100%.

Concerns with sharing data through repositories

Overall, the majority of respondents (76%) reported at least one experiential or hypothetical concern with sharing data from their published study through repositories (Table 3). However, we found no significant differences in overall concern about sharing data through repositories between respondents categorized by trialists' academic productivity and geographic location, trial funding source

and size, and the journal in which it was published, as rates of overall concern ranged between 67 and 84%. In regards to specific concerns with sharing data through repositories, most respondents (65%) identified appropriate data use as a concern, noting issues such as the prevention of misleading secondary analyses and misinterpretation of data. As with overall concern, we found no significant differences in specific concerns with sharing data through repositories between respondents categorized by trialist and trial characteristics.

Reasons for sharing data

The majority of respondents (78%) identified the promotion of open science as an experiential or hypothetical reason for sharing data from their published study (Table 4). We found no significant differences in reasons for sharing data between respondents categorized by trialists' academic productivity and geographic location, trial funding source and size, and the journal in which it was published, with one exception. When respondents were asked if they

Table 2 Respondents' support of clinical trial data sharing, stratified by trialist and trial characteristics

	'The clinical research community should promote and facilitate clinical trial data sharing'			P value
	Respondents	Agree (%)	Disagree (%)	
Overall	317	278 (88)	39 (12)	
Trialist academic productivity (number of research articles published in past three years)				0.22
1 to 10 articles	71	66 (93)	5 (7)	
11 to 25 articles	117	103 (88)	14 (12)	
>25 articles	129	109 (85)	20 (16)	
Trialist geographic location				0.17*
United States or Canada	167	141 (84)	26 (16)	
Western Europe	113	104 (92)	9 (8)	
Other	37	33 (89)	4 (11)	
Trial funding source				0.07*
Government	120	108 (90)	12 (10)	
Industry or mixed funding	152	127 (84)	25 (16)	
Other	45	43 (96)	2 (4)	
Trial size				0.71
≤239 subjects	80	69 (86)	11 (14)	
240 to 2,016 subjects	158	141 (89)	17 (11)	
≥2,017 subjects	79	68 (86)	11 (14)	
Journal in which the trial was published				0.01*
<i>NEJM</i>	113	92 (81)	21 (19)	
<i>Lancet</i>	70	61 (87)	9 (13)	
<i>JAMA</i>	43	36 (84)	7 (16)	
<i>Annals of Internal Medicine</i>	24	23 (96)	1 (4)	
<i>The BMJ</i>	53	52 (98)	1 (2)	
<i>PLoS Medicine</i>	14	14 (100)	0 (0)	

*Fisher exact test. *JAMA*, *Journal of the American Medical Association*; *NEJM*, *New England Journal of Medicine*.

Table 3 Respondents' concerns about sharing clinical trial data through a data repository, stratified by trialist and trial characteristics

	Concerns about sharing clinical trial data through a data repository										
	Respondents	No concern (%)	Any concern (%)	<i>P</i> value	Concerns related to investigator or funder interests (%)	<i>P</i> value	Concerns related to the protection of research subjects (%)	<i>P</i> value	Concerns related to appropriate data use (%)	<i>P</i> value	Other concerns (%)
Overall	317	77 (24)	240 (76)		129 (41)		91 (29)		205 (65)		35 (11)
Trialist academic productivity (number of research articles published in past three years)				0.32		0.09		0.59		0.18	
1 to 10 articles	71	15 (21)	56 (79)		21 (30)		22 (31)		47 (66)		7 (10)
11 to 25 articles	117	25 (21)	92 (79)		52 (44)		36 (31)		82 (70)		10 (9)
>25 articles	129	37 (29)	92 (71)		56 (43)		33 (26)		76 (59)		18 (14)
Trialist geographic location				0.88		0.93		0.65		0.70	
United States or Canada	167	41 (25)	126 (75)		68 (41)		45 (27)		111 (66)		21 (13)
Western Europe	113	26 (23)	87 (77)		45 (40)		36 (32)		72 (64)		11 (10)
Other	37	10 (27)	27 (73)		16 (43)		10 (27)		22 (59)		3 (8)
Trial funding source				0.18		0.01		0.34		0.52	
Government	120	35 (29)	85 (71)		39 (33)		32 (27)		75 (63)		16 (13)
Industry or mixed funding	152	30 (20)	122 (80)		75 (49)		42 (28)		103 (68)		13 (9)
Other	45	12 (27)	33 (73)		15 (33)		17 (38)		27 (60)		6 (13)
Trial size				0.35		0.10		0.40		0.84	
≤239 subjects	80	24 (30)	56 (70)		26 (33)		24 (30)		52 (65)		9 (11)
240 to 2,016 subjects	158	34 (22)	124 (78)		64 (41)		49 (31)		104 (66)		16 (10)
≥2,017 subjects	79	19 (24)	60 (76)		39 (49)		18 (23)		49 (62)		10 (13)
Journal in which the trial was published				0.29*		0.57		0.81*		0.31	
<i>NEJM</i>	113	26 (23)	87 (77)		48 (42)		30 (27)		79 (70)		13 (12)
<i>Lancet</i>	70	11 (16)	59 (84)		31 (44)		18 (26)		48 (69)		5 (7)
<i>JAMA</i>	43	14 (33)	29 (67)		16 (37)		15 (35)		26 (60)		5 (12)
<i>Annals of Internal Medicine</i>	24	7 (29)	17 (71)		8 (33)		8 (33)		14 (58)		4 (17)
<i>The BMJ</i>	53	16 (30)	37 (70)		18 (34)		17 (32)		32 (60)		8 (15)
<i>PLoS Medicine</i>	14	3 (21)	11 (79)		8 (57)		3 (21)		6 (43)		0 (0)

*Fisher exact test. *JAMA*, *Journal of the American Medical Association*; *NEJM*, *New England Journal of Medicine*.

Table 4 Respondents' reasons for sharing clinical trial data*

	Reasons for sharing clinical trial data*							
	Respondents	Administrative requirements (%)	P value	Promote open science (%)	P value	Academic benefits or recognition (%)	P value	Other reasons (%)
Overall	317	55 (17)		248 (78)		133 (42)		56 (18)
Trialist academic productivity (number of research articles published in past three years)			0.64		0.26		0.94	
1 to 10 articles	71	10 (14)		58 (82)		31 (44)		8 (11)
11 to 25 articles	117	20 (17)		95 (81)		49 (42)		21 (18)
>25 articles	129	25 (19)		95 (74)		53 (41)		27 (21)
Trialist geographic location			0.02		0.46		<0.0001	
United States or Canada	167	32 (19)		135 (81)		52 (31)		35 (21)
Western Europe	113	12 (11)		86 (76)		65 (58)		18 (16)
Other	37	11 (30)		27 (73)		16 (43)		3 (8)
Trial funding source			0.62		0.26		0.28	
Government	120	24 (20)		99 (83)		45 (38)		17 (14)
Industry or mixed funding	152	24 (16)		113 (74)		65 (43)		28 (18)
Other	45	7 (16)		36 (80)		23 (51)		11 (24)
Trial size			0.36		0.82		0.48	
≤239 subjects	80	18 (23)		64 (80)		33 (41)		14 (18)
240 to 2,016 subjects	158	24 (15)		124 (78)		71 (45)		22 (14)
≥2,017 subjects	19	13 (16)		60 (76)		29 (37)		20 (25)
Journal in which the trial was published			0.27**		0.55		0.46	
<i>NEJM</i>	113	17 (15)		87 (77)		46 (41)		22 (19)
<i>Lancet</i>	70	15 (21)		51 (73)		33 (47)		14 (20)
<i>JAMA</i>	43	6 (14)		33 (77)		14 (33)		8 (19)
<i>Annals of Internal Medicine</i>	24	8 (33)		20 (83)		13 (54)		2 (8)
<i>The BMJ</i>	53	8 (15)		46 (87)		20 (38)		7 (13)
<i>PLoS Medicine</i>	14	1 (7)		11 (79)		7 (50)		3 (21)

*If respondent had not been asked to share data they were asked to share their hypothetical reasons for sharing. **Fisher exact test. *JAMA*, *Journal of the American Medical Association*; *NEJM*, *New England Journal of Medicine*.

have or would share data from their published study in order to receive academic benefits or recognition, their responses differed significantly based on geographic location ($P < 0.001$). Respondents located in Western Europe responded affirmatively most frequently (58%), as compared to respondents located in the United States or Canada (31%), and elsewhere (43%).

Reasons for withholding data

The majority of respondents (74%) identified ensuring appropriate data use as an experiential or hypothetical reason for withholding data from their published study (Table 5), noting specific concerns such as mistrust of the data requester's intent, data not appropriate for the requested purpose, and the potential for misinterpretation and misleading secondary analyses. We found no significant differences in reasons for withholding data between respondents categorized by trialists' academic productivity and geographic location, trial funding source and size, and the journal in which it was published, with three exceptions. When respondents were asked if they have or would withhold data from their published study in order to protect research subjects, their responses differed significantly based on trialist academic productivity ($P = 0.01$), trial funding source ($P = 0.003$), and journal of trial publication ($P < 0.001$). Respondents who were most academically productive (>25 articles published over the past three years) responded affirmatively least frequently (24%), as compared to respondents who published 1 to 10 articles (41%), and 11 to 25 articles (40%). Respondents who received industry funding also responded affirmatively least frequently (24%), as compared to respondents who received government funding (42%), and non-profit funding (44%). In contrast, authors of trials published in *Annals of Internal Medicine* (67%) and *The BMJ* (47%) responded affirmatively more frequently when compared with respondents who published in *New England Journal of Medicine* (27%), *Lancet* (26%), *Journal of the American Medical Association* (35%), and *PLoS Medicine* (21%).

Discussion

In our survey assessing clinical trialists' views on data sharing (the first study of this scope) we conducted exploratory analyses to determine whether certain characteristics of investigators and their trials were associated with more favorable perceptions of sharing. No prior study has examined how perceptions of data sharing may differ among individuals in this diverse community of clinical trial investigators. We found few notable differences in how respondents viewed the benefits and risks of data sharing. When respondents were categorized by trialists' academic productivity and geographic location, trial funding source and size, and the journal in which it was published, respondents consistently indicated strong

support for data sharing. Our findings suggest that, among authors of trials published in general medical journals with high impact factors, no particular group of trialists is more likely than others to be resistant to new requirements and regulations on data sharing.

However, while trialists were generally in agreement about the benefits of data sharing, we found several specific differences in their perceptions of data sharing that merit further discussion. With respect to academic productivity, respondents who published most often (>25 articles over the past three years) cited the protection of research subjects as a reason for withholding data less often. Investigators achieving such a high level of academic productivity are likely to have substantial experience handling data, and this finding may reflect their comfort and familiarity with implementing measures to ensure data de-identification. Future data sharing initiatives may be optimized for less experienced trialists by offering educational materials on best practice for data de-identification [31], and making ethical consultation available when there is no explicit consent for data sharing [43].

Respondents who conducted studies funded by industry cited the protection of research subjects as a reason for withholding data less often. These respondents were more often concerned with protecting investigator or funder interests when sharing data through a repository, a finding which lay just inside the region of statistical significance. These findings may in part reflect respondents' lack of perceived ownership of data or access to data from industry-funded trials [26-29]. Furthermore, these findings suggest that avoiding potential conflicts with corporate research sponsors is of primary importance to industry-funded trialists. Future data sharing policies may address this concern and secure engagement from trialists involved in industry-funded studies by ensuring robust data access, ownership and stewardship agreements and, where applicable, copyright or other licensing agreements [44].

There were also interesting patterns in the reasons respondents noted for sharing data. More than half of respondents located in Western Europe cited the potential for academic benefits or recognition as a reason for sharing data, whereas respondents located in the United States or Canada cited this reason less often. This finding suggests that the professional desirability of data sharing depends on cultural norms within scientific communities, and the community of trialists based in Western Europe has already begun to foster a culture in which investigators recognize that their academic aspirations may be furthered by sharing data. In the near future, data sharing policies which ensure co-authorship rights to data originators [8] may help incentivize more trialists to share data and align their professional interests with those of the public good. However, academic institutions and

Table 5 Respondents' reasons for withholding clinical trial data*

	Reasons for withholding clinical trial data*							
	Respondents	Protect investigator or funder interests (%)	P value	Protect research subjects (%)	P value	Ensure appropriate data use (%)	P value	Other reasons (%)
Overall	317	121 (38)		107 (34)		233 (74)		37 (12)
Trialist academic productivity (number of research articles published in past three years)			0.32		0.01		0.31	
1 to 10 articles	71	22 (31)		29 (41)		55 (77)		3 (4)
11 to 25 articles	117	45 (38)		47 (40)		89 (76)		18 (15)
>25 articles	129	54 (42)		31 (24)		89 (69)		16 (12)
Trialist geographic location			0.72		0.79		0.40	
United States or Canada	167	66 (40)		57 (34)		123 (74)		22 (13)
Western Europe	113	43 (38)		36 (32)		86 (76)		14 (12)
Other	37	12 (32)		14 (38)		24 (65)		1 (3)
Trial funding source			0.65		0.003		0.92	
Government	120	43 (36)		50 (42)		89 (74)		13 (11)
Industry or mixed funding	152	62 (41)		37 (24)		112 (74)		17 (11)
Other	45	16 (36)		20 (44)		32 (71)		7 (16)
Trial size			0.45		0.08		0.75	
≤239 subjects	80	26 (33)		35 (44)		58 (73)		7 (9)
240 to 2,016 subjects	158	62 (39)		50 (32)		119 (75)		20 (13)
≥2,017 subjects	79	33 (42)		22 (28)		56 (71)		10 (13)
Journal in which the trial was published			0.11		0.0008**		0.86	
<i>NEJM</i>	113	50 (44)		30 (27)		85 (75)		15 (13)
<i>Lancet</i>	70	22 (31)		18 (26)		49 (70)		10 (14)
<i>JAMA</i>	43	11 (26)		15 (35)		34 (79)		4 (9)
<i>Annals of Internal Medicine</i>	24	11 (46)		16 (67)		17 (71)		3 (13)
<i>The BMJ</i>	53	19 (36)		25 (47)		37 (70)		5 (9)
<i>PLoS Medicine</i>	14	8 (57)		3 (21)		11 (79)		0 (0)

*If respondent had not been asked to share data they were asked to share their hypothetical reasons for not sharing. **Fisher exact test. *JAMA*, *Journal of the American Medical Association*; *NEJM*, *New England Journal of Medicine*.

promotions committees will need to begin rewarding trialists for creating data and sharing it with other investigators if the global culture of academic medicine is to shift from 'publish or perish' to more open science [45,46]. Some funding agencies, including the National Science Foundation, have already begun to consider data as a research product when evaluating investigators [47].

Limitations

There are several limitations to consider when interpreting our study. First, our analyses were exploratory and conducted *post-hoc* (after the main survey results were analyzed), and were not powered to detect differences between groups of respondents. Second, with respect to the overall survey there are several important considerations concerning the generalizability of our results including social desirability response bias [48,49], a sample comprised entirely of trialists published in the general medical journals with the highest impact factors, and an overall response rate which was lower than that of some web-based surveys of clinical trial investigators [27,29,50], although higher than other such studies [51]. Furthermore, we did not explicitly define 'data' in our survey, and respondents may thus have interpreted this concept to mean the level of data that they would be most comfortable sharing (for example, data sufficient to remake published tables versus all participant level data). Because of these issues, our findings may overestimate support for and willingness to share data in the clinical trial community. Finally, our study was also limited by the scope of our survey, as we omitted important questions about whether funders have explicitly prohibited data sharing, experiences negotiating data ownership with others, sharing in the context of respondents' other trials, and the sharing of other study materials (such as intervention manuals, analysis scripts, and output files) [52] in order to reduce response burden.

Conclusions

In our survey of clinical trialists' perceptions and experiences with data sharing, we found few notable differences among subgroups of trialists in how they viewed the benefits and risks of data sharing. When respondents were categorized by trialists' academic productivity and geographic location, trial funding source and size, and the journal in which it was published, strong support for data sharing remained. These findings suggest that no particular group of trialists is more likely to be resistant to new requirements and regulations on data sharing, as the clinical trials research community appears to have adopted a more favorable outlook on sharing. However, support for data sharing among trialists in principle may not translate to practice, as research funders and other stakeholders often figure prominently in the decision of

whether or not to share. A better understanding of the context in which trialists conduct research may thus be helpful in identifying and addressing specific, practical barriers to sharing in order to align public and private interests and fully realize the promise of more open science.

Additional file

Additional file 1: Survey distributed to potential respondents.

Abbreviations

NEJM: *New England Journal of Medicine*; *JAMA*: *Journal of the American Medical Association*; *SAS*: *Statistical Analysis System*.

Competing interests

Drs. Gross, Krumholz, and Ross receive research support from Medtronic, Inc. and Johnson & Johnson to develop methods to promote data sharing. Drs. Gross and Ross report that they are members of a scientific advisory board for FAIR Health, Inc. Dr. Krumholz reports that he chairs a scientific advisory board for UnitedHealthcare. Dr. Joffe reports that he is a paid member of a data monitoring committee for Genzyme/Sanofi.

Authors' contributions

All authors read and approved the final version of the manuscript. VKR helped conceive and design the study, draft the manuscript, and acquire the data. KMS performed the statistical analysis. CPG, IH, SJ, and HMK participated in the analysis and interpretation of data, and critically revised the manuscript for important intellectual content. KD helped acquire the data and perform statistical analysis. JSR was responsible for the conception and design of the study, participated in the analysis and interpretation of data, helped draft the manuscript and acquire the data.

Acknowledgments

Data access and responsibility: all authors had full access to all the data in the study and Dr. Ross takes responsibility for the integrity of the data and the accuracy of the data analysis. Data sharing: requests for statistical code and dataset can be made to the corresponding author at joseph.ross@yale.edu. The dataset is publicly accessible via the Dryad data repository at <http://datadryad.org/resource/doi:10.5061/dryad.6544v>.

This study was not supported by any external grants or funds, although Mr. Rathi received support from the Yale University School of Medicine Medical Student Research Fellowship. Dr. Krumholz is supported by a National Heart Lung Blood Institute Cardiovascular Outcomes Center Award (grant number: 1U01HL105270-02). Dr. Ross is supported by the National Institute on Aging (grant number: K08 AG032886) and by the American Federation for Aging Research through the Paul B Beeson Career Development Award Program.

Author details

¹Yale University School of Medicine, 333 Cedar Street, New Haven, CT 06510, USA. ²Center for Outcomes Research and Evaluation, Yale-New Haven Hospital, 1 Church Street, Suite 200, New Haven, CT 06510, USA. ³Section of General Internal Medicine and the Robert Wood Johnson Foundation Clinical Scholars Program, Department of Medicine, Yale University School of Medicine, P.O. Box 208093, New Haven, CT 06520, USA. ⁴Nature Publishing Group, The Macmillan Building, 4 Crinan St, London N1 9XW, UK. ⁵Department of Medical Ethics and Health Policy, University of Pennsylvania Perelman School of Medicine, 3401 Market Street, Suite 320, Philadelphia, PA 19104, USA. ⁶Section of Cardiovascular Medicine and the Robert Wood Johnson Foundation Clinical Scholars Program, Department of Medicine, Yale University School of Medicine, P.O. Box 208088, New Haven, CT 06520, USA. ⁷Department of Health Policy and Management, Yale University School of Public Health, P.O. Box 208034, New Haven, CT 06520, USA.

Received: 19 December 2013 Accepted: 22 September 2014
Published: 2 October 2014

References

1. Bill and Melinda Gates Foundation What we do. [http://www.gatesfoundation.org/what-we-do]
2. National Institutes of Health (NIH): Final NIH statement on sharing research data. 2003, [http://grants.nih.gov/grants/guide/notice-files/NOT-OD-03-032.html]
3. Medical Research Council (MRC) MRC policy on research data-sharing. [http://www.mrc.ac.uk/research/research-policy-ethics/data-sharing/policy/]
4. Groves T: **BMJ policy on data sharing.** *BMJ* 2010, **340**:c564.
5. Laine C, Goodman SN, Griswold ME, Sox HC: **Reproducible research: moving toward research the public can really trust.** *Ann Intern Med* 2007, **146**:450–453.
6. Nature Publishing Group Authors and References, Policies, Availability of data and materials. [http://www.nature.com/authors/policies/availability.html]
7. **Trials. About Trials, Editorial Policies, Data and materials release.** [http://www.trialsjournal.com/about#trials]
8. Vickers AJ: **Whose data set is it anyway? Sharing raw data from randomized trials.** *Trials* 2006, **7**:15.
9. Gotzsche PC: **Why we need easy access to all data from all clinical trials and how to accomplish it.** *Trials* 2011, **12**:249.
10. The Cochrane Collaboration: **Press Release: Cochrane Collaboration statement on access to clinical trial data.** 2011, [http://www.cochrane.org/features/clinical-trials-statement-press-release]
11. Ross JS, Gross CP, Krumholz HM: **Promoting transparency in pharmaceutical industry-sponsored research.** *Am J Public Health* 2012, **102**:72–80.
12. Ross JS, Lehman R, Gross CP: **The importance of clinical trial data sharing: toward more open science.** *Cir Cardiovasc Qual Outcomes* 2012, **5**:238–240.
13. Gotzsche PC, Jorgensen AW: **Opening up data at the European Medicines Agency.** *BMJ* 2011, **342**:d2686.
14. Chan AW: **Access to clinical trial data.** *BMJ* 2011, **342**:d80.
15. Godlee F, Loder E: **Missing clinical trial data: setting the record straight.** *BMJ* 2010, **341**:c5641.
16. Wieseler B, McGauran N, Kaiser T: **Still waiting for functional EU Clinical Trials Register.** *BMJ* 2011, **342**:d3834.
17. Chan AW, Altman DG: **Identifying outcome reporting bias in randomised trials on PubMed: review of publications and survey of authors.** *BMJ* 2005, **330**:753.
18. Hampton T: **European drug agency works to improve transparency, but skepticism remains.** *JAMA* 2012, **308**:850–851.
19. Krumholz HM, Ross JS: **A model for dissemination and independent analysis of industry data.** *JAMA* 2011, **306**:1593–1594.
20. Loder E: **Liberating Clinical Trial Data.** 2012, [http://blogs.bmj.com/bmj/2012/10/09/elizabeth-loder-liberating-clinical-trial-data/]
21. Thomas K: **Glaxo Opens Door to Data on Research.** New York, NY: New York Times; 2012 [http://www.nytimes.com/2012/10/11/business/glaxo-opens-door-to-data-on-its-research.html]
22. Mello MM, Franer JK, Wilenzick M, Teden P, Bierer BE, Barnes M: **Preparing for responsible sharing of clinical trial data.** *N Engl J Med* 2013, **369**:1651–1658.
23. **Availability of Masked and De-identified Non-Summary Safety and Efficacy Data; Request for Comments.** [https://federalregister.gov/a/2013-13083]
24. Piwowar HA: **Who shares? Who doesn't? Factors associated with openly archiving raw research data.** *PLoS One* 2011, **6**:e18657.
25. Rathi V, Dzara K, Gross CP, Hrynaskiewicz I, Joffe S, Krumholz HM, Strait KM, Ross JS: **Sharing of clinical trial data among trialists: a cross sectional survey.** *BMJ* 2012, **345**:e7570.
26. Gotzsche PC, Hrobjartsson A, Johansen HK, Haahr MT, Altman DG, Chan AW: **Constraints on publication rights in industry-initiated clinical trials.** *JAMA* 2006, **295**:1645–1646.
27. Lundh A, Krogsboll LT, Gotzsche PC: **Access to data in industry-sponsored trials.** *Lancet* 2011, **378**:1995–1996.
28. Lundh A, Krogsboll LT, Gotzsche PC: **Sponsors inverted question mark participation in conduct and reporting of industry trials: a descriptive study.** *Trials* 2012, **13**:146.
29. Rochon PA, Sekeres M, Hoey J, Lexchin J, Ferris LE, Moher D, Wu W, Kalkar SR, Van Laethem M, Gruneir A, Gold J, Maskalyk J, Streiner D, Taback N, Chan AW: **Investigator experiences with financial conflicts of interest in clinical trials.** *Trials* 2011, **12**:9.
30. Godlee F: **Clinical trial data for all drugs in current use.** *BMJ* 2012, **345**:e7304.
31. Hrynaskiewicz I, Altman DG: **Towards agreement on best practice for publishing raw clinical trial data.** *Trials* 2009, **10**:17.
32. Kirwan JR: **Making original data from clinical studies available for alternative analysis.** *J Rheumatol* 1997, **24**:822–825.
33. Reidpath DD, Allotey PA: **Data sharing in medical research: an empirical investigation.** *Bioethics* 2001, **15**:125–134.
34. Piwowar HA, Day RS, Fridsma DB: **Sharing detailed research data is associated with increased citation rate.** *PLoS One* 2007, **2**:e308.
35. Savage CJ, Vickers AJ: **Empirical study of data sharing by authors publishing in PLoS journals.** *PLoS One* 2009, **4**:e7078.
36. Blumenthal D, Campbell EG, Anderson MS, Causino N, Louis KS: **Withholding research results in academic life science. Evidence from a national survey of faculty.** *JAMA* 1997, **277**:1224–1228.
37. Blumenthal D, Campbell EG, Gokhale M, Yucel R, Clarridge B, Hilgartner S, Holtzman NA: **Data withholding in genetics and the other life sciences: prevalences and predictors.** *Acad Med* 2006, **81**:137–145.
38. Campbell EG, Clarridge BR, Gokhale M, Birenbaum L, Hilgartner S, Holtzman NA, Blumenthal D: **Data withholding in academic genetics: evidence from a national survey.** *JAMA* 2002, **287**:473–480.
39. Walport M, Brest P: **Sharing research data to improve public health.** *Lancet* 2011, **377**:537–539.
40. Alsheikh-Ali AA, Qureshi W, Al-Mallah MH, Ioannidis JP: **Public availability of published research data in high-impact journals.** *PLoS One* 2011, **6**:e24357.
41. **PLoS Medicine Editorial and Publishing Policies - 13. Sharing of Materials, Methods and Data.** [http://www.plosmedicine.org/static/policies.action#sharing]
42. **New England Journal of Medicine Author Center: Supplementary Appendix.** [http://www.nejm.org/page/author-center/supplementary-appendix]
43. Hrynaskiewicz I, Norton ML, Vickers AJ, Altman DG: **Preparing raw clinical data for publication: guidance for journal editors, authors, and peer reviewers.** *BMJ* 2010, **340**:c181.
44. Hrynaskiewicz I, Cockerill MJ: **Open by default: a proposed copyright license and waiver agreement for open access research and data in peer-reviewed journals.** *BMC Res Notes* 2012, **5**:494.
45. Piwowar HA, Becich MJ, Bilofsky H, Crowley RS: **Towards a data sharing culture: recommendations for leadership from academic health centers.** *PLoS Med* 2008, **5**:e183.
46. Ross JS, Krumholz HM: **Ushering in a new era of open science through data sharing: the wall must come down.** *JAMA* 2013, **309**:1355–1356.
47. Piwowar H: **Altmetrics: Value all research products.** *Nature* 2013, **493**:159.
48. Fowler FJ: *Survey Research Methods.* Newbury Park CA: Sage; 1993.
49. Sudman S, Bradburn NM: *Asking Questions: A Practical Guide to Questionnaire Design.* San Francisco, CA: Jossey-Bass; 1982.
50. Haahr MT, Hrobjartsson A: **Who is blinded in randomized clinical trials? A study of 200 trials and a survey of authors.** *Clin Trials* 2006, **3**:360–365.
51. Schroll JB, Bero L, Gotzsche PC: **Searching for unpublished data for Cochrane reviews: cross sectional study.** *BMJ* 2013, **346**:f2231.
52. Crutzen R, Peters GJ, Abraham C: **What about trialists sharing other study materials?** *BMJ* 2012, **345**:e8352.

doi:10.1186/1745-6215-15-384

Cite this article as: Rathi et al.: Predictors of clinical trial data sharing: exploratory analysis of a cross-sectional survey. *Trials* 2014 **15**:384.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

