Strategies for maximizing IPD retrieval in IPDMA: A mixed method study

Background:

Individual Participant Data (IPD) meta-analysis is regarded as the ideal approach for providing evidence on intervention effect estimation because it can derive standardized outcome definitions and use a consistent analysis method. Our previous study has summarized the methodological and reporting quality of published IPDMA (BMJ 2021;372:n736). However, the current practices and perceived best strategies in IPD data retrieval are still unclear.

Objectives:

The aim of this study is to explore the perceived and practical strategies in IPD data retrieval.

Methods

A cross-sectional survey was conducted. Email addresses of IPDMA authors were identified through PUBMED. An e-questionnaire with 32 questions, related to the authors' demographic, their views and practices on different strategies in conducting IPDMA, were created based on literature. Qualitative email interviews were then conducted to gather in-depth information about their barriers and perceived strategies to maximize the IPD retrieval.

Descriptive statistics, inferential statistics and linear regression models were used to analyze the data collected from the e-questionnaire while content analysis was undertaken for the qualitative interview. Ethics approval was granted by the university IRB.

Results

151 respondents were mostly male (62%), aged≥35 (80.1%), academic staff (69.5%), with >5-year experiences (80.1%). The mean successful rate of retrieving the IPD data was 67.2%. 64.2% included their primary studies in the IPDMA and 44.4% provided authorship for primary study authors as incentive.

Email was the most common methods to contact study author (90.1%) and share data (67.5%), it was also ranked the most effective way in requesting IPD.

Linear regression models revealed that those (i) aged≥65, (ii) were academic staff as the first contact person, (iii) had primary study included in the IPDMA, had significant higher successful rate.

Two main categories were identified from the email interview qualitative responses – (i) '*Contributors of successful IPD retrieval*' and (ii) '*Reasons for failed IPD retrievals*'. (See Tables)

Conclusion

It is revealed that senior academics working in the field with primary data themselves had higher chance to retrieve the IPD data. Providing authorship of the IPDMA publication to the primary study author may also help to improve the successful rate.

[Patient, public and/or healthcare consumer involvement: Nil]

Table 1: Participant characteristics (n = 151)

Characteristics	Frequency (%)	Median (IQR)
Gender		
Female	52 (34.4)	
Male	94 (62.3)	
Prefer not to say/Non-binary	5 (3.3)	
Age (years)		
34 or younger	30 (19.9)	
35 – 49	69 (45.7)	
50 – 64	43 (28.5)	
65 or older	9 (6.0)	
No. of years since completion of highest profession degree		
5 or less	44 (29.1)	
6 – 10	32 (21.2)	
11 – 15	26 (17.2)	
15 – 20	16 (10.6)	
20 or more	33 (21.9)	
Primary place of employment		
Academic institution	125 (82.8)	
Hospital	14 (9.3)	
Non-profit organisation	6 (4.0)	
Private industry	5 (3.3)	
Government	1 (0.7)	
Job title		
Professor	55 (36.4)	
Associate Professor	20 (13.2)	
Assistant Professor	19 (12.6)	
Research Fellow or Postdoctoral Fellow	18 (11.9)	
Lecturer or Senior Lecturer	11 (7.3)	
Research Associate or Research Assistant or Project administrator	8 (5.3)	
Clinician	6 (4.0)	
Graduate student	4 (2.6)	
Others	10 (6.6)	
No. of research grants or funding received (three years before the IPD requests)		
1 – 3	49 (32.5)	
4 – 6	23 (15.2)	
7 – 10	8 (2.3)	
More than 10	19 (12.6)	
None	52 (34.4)	
Likelihood of conducting another IPA meta-analysis in the next 3 years (‘0’ being ‘Extremely Unlikely’ and ‘10’ being ‘Extremely Likely’)		7 (3 – 10)

Table 2: IPD retrieval (n = 151)

	Frequency (%)
Methods used to contact primary study authors *	
Phone	25 (16.6)
Email	136 (90.1)
Post mail	7 (4.6)
Fax	2 (1.3)
In-person meeting (face-to-face)	49 (32.5)
Video-conferencing	30 (19.9)
Academic platforms (e.g. ResearchGate)	23 (15.2)
Personal social media	13 (8.6)
Data sharing repository	0 (0)
Others	27 (17.9)
Method first used to contact primary study authors	
Phone	1(0.7)
Email	120 (79.5)
Post mail	1 (0.7)
Fax	1 (0.7)
In-person meeting (face-to-face)	11 (7.3)
Video-conferencing	1 (0.7)
Academic platforms (e.g. ResearchGate)	5 (3.3)
Personal social media	0 (0)
Data sharing repository	0 (0)
Others	11 (7.3)
Preferred method for contacting primary study authors	
Phone	1 (0.7)
Email	93 (61.6)
Post mail	0 (0)
Fax	0 (0)
In-person meeting (face-to-face)	7 (4.6)
Video-conferencing	0 (0)
Academic platforms (e.g. ResearchGate)	6 (4.0)
Personal social media	1 (0.7)
Data sharing repository	2 (1.3)
Others	11 (7.3)
No preferred method, all methods used simultaneously	30 (19.9)
If the corresponding author's email address is invalid*	
Attempt to contact other authors	105 (69.5)
Search for the updated email address	103 (68.2)
Take it as a failed IPD retrieval attempt	8 (5.3)
Others	25 (16.6)
Provision of incentives for primary study authors*	
No	42 (27.8)
Yes, co-authorship	67 (44.4)
Yes, financial	2 (1.3)
Yes, acknowledgement	33 (21.9)
Yes, others	8 (5.3)
No. of attempts to contact primary study authors before participants stopped trying	

1	11 (7.3)
2	38 (25.2)
3	56 (37.1)
4	9 (6.0)
5 or more	37 (24.5)
Time interval between contacts	
Less than 2 weeks	12 (7.9)
2 weeks to less than 1 month	64 (42.4)
1 month to less than 3 months	39 (25.8)
3 months to less than 6 months	7 (4.6)
6 months or more	5 (3.3)
Shortest duration to receive the IPD from primary study authors	
Less than 2 weeks	75 (49.7)
2 weeks to less than 1 month	28 (18.5)
1 month to less than 3 months	23 (15.2)
3 months to less than 6 months	12 (7.9)
6 months to less than 1 year	6 (4.0)
1 year to less than 2 years	3 (2.0)
2 years to less than 3 years	2 (1.3)
3 years or more	2 (1.3)
Longest duration to receive the IPD from primary study authors	
Less than 2 weeks	8 (5.3)
2 weeks to less than 1 month	8 (5.3)
1 month to less than 3 months	29 (19.2)
3 months to less than 6 months	18 (11.9)
6 months to less than 1 year	30 (19.9)
1 year to less than 2 years	29 (19.2)
2 years to less than 3 years	17 (11.3)
3 years or more	12 (7.9)
Provide purpose of IPD requests	
No	1 (0.7)
Yes (purpose described in a written format with a protocol)	93 (61.6)
Yes (purposed described in a written format without a protocol)	53 (35.1)
Yes (purpose was described verbally)	4 (2.6)
Further information provided upon IPD request*	
Data sharing agreement	86 (57.0)
Data analysis plan	71 (47.0)
Plans pertaining to confidentiality and data storage	85 (56.3)
No further information provided	31 (20.5)
Reasons for declined IPD requests*	
Never received a response	115 (76.2)
Data no longer available	78 (51.7)
Data will be used for further analysis	27 (17.9)
Lack of participants' informed consent	26 (17.2)
Lack of resources in preparing data for sharing	50 (33.1)
Unsure of local laws and/or employer/research funder policy on data sharing	34 (22.5)

Primary study author did not give any reasons for declining request	46 (30.5)
Other reasons	24 (15.9)
IPD data sharing methods*	
Email	102 (67.5)
Data repository website/Data sharing community	44 (29.1)
Secure File Transfer Protocol	67 (44.4)
Cloud drive	26 (17.2)
Others	14 (9.3)
Format of data being shared*	
CSV	72 (47.7)
MS Excel	107 (70.9)
SPSS	64 (42.4)
Stata	37 (24.5)
Others	18 (11.9)
Resources used to conduct IPD meta-analysis*	
The Cochrane Handbook	85 (56.3)
Textbooks	44 (29.1)
Methodological journal articles	117 (77.5)
Other resources	22 (14.6)
Reporting guideline used*	
QUOROM	12 (7.9)
PRISMA	96 (63.6)
PRISMA-IPD	83 (55.0)
Did not follow any guidelines	9 (6.0)
Funding	
No	68 (45.0)
Yes, non-commercial	77 (51.0)
Yes, commercial	3 (2.0)
Yes, both commercial and non-commercial	3 (2.0)

*Participants could select more than one option

Table 3: Perspectives towards IPD retrieval

	Median (IQR)	Frequency (%)
Effectiveness of IPD request method (n = 148*) (Ranking 1 to 9)		
Phone	4 (3 – 5)	
Email	1 (1 – 2)	
Post mail	6 (5 – 8)	
Fax	8 (6 – 9)	
In-person meeting (face-to-face)	4 (2 – 6)	
Video conferencing	4 (3 – 6)	
Academic platform (e.g. ResearchGate)	6 (4 – 7)	
Personal social media	7 (5 – 8)	
Data sharing repository	5 (2.75 – 8)	
Having published studies on the topic of interest improves success of IPD retrieval from primary study authors		
Yes		128 (84.8)
No		23 (15.2)
Ethical and reasonable to give co-authorship to primary authors as an incentive		
Yes		106 (70.2)
No		45 (29.8)

*Of the 151 participants, only 148 valid responses obtained.

Table 3: Effect of participants' characteristics on success rate of IPD retrieval

	Mean (SD)	Regression Coefficient (95% CI)	*p*-value
Age			
34 years or younger	56.3 (26.4)	Ref	N.A.
35 – 49 years	67.6 (26.8)	6.83 (-4.34 – 18.0)	0.229
50 – 64 years	71.1 (23.5)	7.70 (-4.70 – 20.1)	0.222
65 years or older	81.8 (14.5)	19.9 (0.86 – 39.0)	**0.041**
Job title of person who contacted the primary study authors			
Research Associate or Research Assistant or Project administrator	48.1 (23.2)	Ref	N.A.
Graduate student or Undergraduate student	48.6 (26.1)	7.1 (-16.4 – 30.6)	0.551
Research Fellow or Postdoctoral Fellow	66.9 (27.7)	20.9 (0.610 – 41.1)	**0.044**
Others/Cannot recall	67.5 (30.1)	22.2 (0.778 – 43.6)	**0.042**
Faculty member (Lecturer/Senior Lecturer/ Adjunct/Assistant/Associate/Full Professor)	71.3 (23.1)	22.2 (3.44 – 40.9)	**0.021**
Conducted primary study included in the IPD meta-analysis			
No	59.5 (27.7)	Ref	N.A.
Unsure	64.1 (29.2)	2.7 (-17.5 – 22.9)	0.791
Yes	71.1 (24.1)	9.4 (0.176 – 18.7)	**0.046**

Table 4: IPDMA Interview Findings

Category	Sub-category	Participants' quotes
Contributors of successful IPD retrieval	Researcher's credibility and legitimacy	"It may also be easier for a researcher who is well established in the relevant field to make contact, again as this would indicate legitimacy/credibility." (P4)
		"…In order to inspire trust, there must be a demonstrated ability to properly handle the requested data – with a clear, concise, complete protocol regarding the following key issues: data processing, analytical plan, information security. Expertise in the field may also be helpful in inspiring trust…" (P7)
	Being part of relevant research networks	"We will use the established research network." (P11)
		"To have a network of people ready to help (see above), then to have some leaders in the field may help to build such network." (P1)
	Use email communication	"We preferred e-mail communication. E-mail addresses are relatively easy to find for most study authors, particularly if their publications are relatively recent. E-mail communication is a formal (written) method of communication that can be legally archived and reserved for future reference. Any valid agreement regarding international data sharing would require written communication, even if this was not in the legal form of a data sharing agreement." (P7)
		"Most authors that are dealing with my research interest are based globally. Email is a generally accepted approach of contact since most academics and/or therapists check their inboxes within a more or less regular time frame." (P5)
	Provide adequate information about the study and expected outcomes	"A good protocol including update trial search, description of the trials eligible, strategy for analysis (subset, subgroup analyses…), list of steering committee/advisory board member information on the collaborative group structure and the strategy for publication, a time schedule for the project." (P1)

		"Send copy of protocol" (P8)
	Undertake a collaborative approach	"Something that has been helpful is to write the study protocol as an article, get the input and feedback from the primary data holders on all aspects of the project. This gives them a chance to contribute to the project, gain trust and engage in the collaboration with the other study teams, and makes that everybody feel that they are part of the adventure (rather than just being asked to provide their data)." (P9)
		"Investigator meeting was a good way to motivate the investigators, we use to organize them in parallel with scientific meeting and/or to cover part of the cost of the meeting for the investigator (no charge for the meeting, help for hotel or travel)." (P1)
Reasons for failed IPD retrievals	Inability to contact primary study authors	"Sometimes, email addresses no longer work, authors have moved to different institutions/have left academia etc." (P4)
		"author(s) had died" (P5)
	Unwillingness to prepare or share datasets	"it can take a lot of (non-funded) time and effort to provide anonymized IPD data that is fit for purpose." (P3)
		"I have also encountered the situation where researchers are working on an unpublished dataset but do not wish to share it because they want to publish their findings first... On another occasion, a researcher essentially said the data do still exist but that it would be too much hassle to obtain them from storage etc." (P4)
	Communication barrier with primary study authors	"Primary author does not speak English; we used a colleague who is a native speaker of author's language, but this did not solve the issue of simply failing to understand what IPD is." (P8)
		"They were from nations whose first language is not English and this might have been a problem for the author with whom I corresponded…" (P2)