TITLE: Use of participant data and biological samples is insufficiently described in participant information leaflets

Emer R. McGrath, MB, PhD[1,2] , Nigel Kirby, MSc[3], Frances Shiely, PhD[4,5,6]

[1] HRB Clinical Research Facility, University of Galway, Galway, Ireland

[2] School of Medicine, University of Galway, Galway, Ireland

[3] Centre for Trials Research, Cardiff University, Neuadd Meirionnydd, Heath Park, Cardiff CF14 4YS, UK

[4] TRAMS (Trials Research and Methodologies Unit), HRB Clinical Research Facility, University College Cork, Ireland.

[5] School of Public Health, University College Cork, Ireland.

[6] HRB Trials Methodology Research Network, University College Cork, Ireland.

Correspondence:

Dr Emer McGrath

HRB Clinical Research Facility

University of Galway

University Road

Galway

Ireland, H91 TK33

T. +353 91 524411

Email: emer.mcgrath@universityofgalway.ie

What is new?

Key findings

PIL descriptions regarding current and future use of participant data and biological samples are often insufficiently detailed or ambiguous. The majority fail to adequately describe how data confidentiality will be maintained, to specify the intended duration of data storage or specify whether data will be used in future studies or shared with other researchers. Many PILs do not specify the purpose for obtaining biological samples, whether samples will be anonymously stored/deidentified or whether participants will be informed of their own individual-level results. Few of the included PILs were related to trials with patient and public involvement (PPI).

In the absence of published guidelines on describing current use and future re-use of participant data and samples to trial participants in PILs, we provide eight suggested recommendations of items that should be included in PILs to improve practice, help trial teams when designing PILs and better meet trial participant knowledge needs.

What this adds to what is known related to methods research within the field of clinical epidemiology

Prior research to date has evaluated the opinions of focus groups and potential trial participants on the sharing and re-use of participant data and biological samples. Our study builds upon this work, evaluating current practices via an in-depth, qualitative assessment of the descriptions of current use and future re-use of data and samples in a large corpus of PILs (n=240) from across the UK and Ireland. We also provide a summary of adequate and inadequate descriptions of these issues to help inform future PIL development.

What is the implication, what should change now?

In the absence of published guidelines on describing current use and future re-use of participant data and samples to trial participants in PILs, we provide eight suggested recommendations of items that should be included in PILs to improve practice and help trial teams when designing PILs.

ABSTRACT

Background and Objectives:

With greater availability of participant data and biobank repositories following clinical trial completion, adequately describing future data and biological sample re-use plans to trial participants is increasingly important. We evaluated how trial teams currently describe current and future use of participant data and biological samples in participant information leaflets (PILs).

Methods: Retrospective qualitative analysis of 240 PILs (182 clinical trials) in Ireland and the UK. Descriptions of data and sample use/re-use were extracted and analysed using a four-stage pragmatic content analysis approach. A recommended list of questions to be addressed by trial teams when designing PILs was developed.

Results: Of the 240 included PILs, 85% specifically mentioned, or directly implied, how confidentiality of participant data would be maintained; 38% were considered by the authors to adequately describe how data confidentiality would be maintained (i.e. the PIL specifically mentioned data deidentification and compliance with data protection regulations); 47% reported the intended duration of data storage (mean 15; SD ±9 years); 40% specified if data would be used in future research studies and 28% stated if data would be shared with other researchers. Of the 117 PILs stating biological samples would be collected from participants, 80% provided a reason for requesting the sample, 66% stated whether stored samples would be deidentified, 21% specified if individual-level results would be made available to participants and 70% specified whether samples may be used for future studies. Of the 73 PILs specifying planned future sample storage, 18% stated the intended duration of storage and 48% specified if samples would be shared with other researchers. A list of eight recommended questions to be addressed by trial

teams when designing PILs were identified, e.g. ‘What is the intended duration of data and sample storage for the current study?’.

Conclusion: PILs often provide insufficient detail regarding plans for current use and future re-use of participants’ data and their biological samples. The majority do not adequately describe plans for maintaining data confidentiality. Best practice approaches to describing data use and re-use in PILs are needed. This will require multi-stakeholder input, including potential trial participants to progress this.

Keywords: participant information leaflets, communication, trials, data use, biological samples

Running Title: Describing data use and biological sample use and re-use in PILs

Tables: 5

Supplementary appendix: supplementary tables 8

Word Count: 3447, abstract 350

1. BACKGROUND:

Participant information leaflets (PILs) are necessary to provide information in order to obtain informed consent from clinical trial participants. PILs are written documents, approximately 15-20 pages in length, which provide detailed information necessary to support a potential trial participant in making an informed choice about participating in a trial. They are similar to 'consent forms' utilized in US trials[1]. PILs should be easy to understand and provide a sufficiently detailed, yet concise, description of relevant study procedures[2]. With the increasing availability of participant data and biobank repositories following completion of trials, there is a growing awareness of the importance of adequately describing future data and sample re-use plans to trial participants[3]. However, in current PILs, descriptions of current and future use of trial participants' data and their biological samples may not be sufficient to meet regulatory requirements or participants' needs. For example, a group consensus document on the sharing and re-use of clinical trial data identified the following participant knowledge needs: sufficient information to obtain explicit informed consent from trial participants, adequate descriptions regarding the sharing and re-use of participant data for research purposes, and a clear description of data de-identification steps, data storage and data sharing plans[3]. In studies and surveys of potential trial participants, the most important concern cited by potential trial participants is the need for appropriate privacy and protection of their data[4]. While general guidelines exist on the development of PILs for a clinical context[5], there is a paucity of dedicated guidelines specific to PIL development for clinical trials[2]. In addition, many of the currently available guidelines for participant information documents focus solely on formatting, literacy and readability[2,5-12], with few to none describing the ideal minimum required content for informative PILs[13].

Current General Data Protection Regulations (GDPR) introduced in 2018, require study sponsors to inform trial participants about how their personal data will be stored and processed[14,15]. However, there are currently no guidelines on the best practice approach to describing current use and future re-use of participant data and biological samples in clinical trial PILs. In fact, knowledge of how trial teams address this within current PILs is not recorded. It is imperative that an evidence-based, standardised approach for describing these issues to trial participants is available, to assist trial teams when designing PILs that meet participants' knowledge needs. To do this, we first need to evaluate how trial teams currently address this issue within PILs.

The aim of this retrospective study was to evaluate how trial teams describe the current use and future re-use of participant data and biological samples in clinical trial PILs. We also identified adequate and inadequate examples of this practice and developed a list of recommended questions to be addressed by trial teams when designing PILs.

2. METHODS:

Data collection and extraction

For a prior study, we collected PILs from trial units in the UK and Ireland[1]. In brief, PILs were collected for a Health Research Board Trials Methodology Research Network funded study. A request for PILs was placed to all Irish Clinical Trials Units (CTUs) and all UK CTUs (n=52) using publicly available contact details in October 2020. CTUs are units which have specific expertise in the design and delivery of clinical trials and typically coordinate the running of several clinical trials at any one time. Thus, they are a diverse sample for sourcing PILs. This request was supplemented by using existing PIL datasets which had been collected for existing

projects led by co-applicants. Collated PIL versions ranged between the years 2003 and 2020. We included PILs from randomised controlled trials (RCTs) of any phase, including PILs designed for adults, minors, parents of minors, legal representatives of individuals lacking capacity to provide informed consent, and adults later regaining capacity to provide informed consent during a trial. If trial units provided more than one PIL for an individual trial, only the most recent PIL version was included. Some trials had a separate PIL and informed consent form (ICF) (n=29), information from these was also included. For the current study, data were extracted on general trial characteristics including trial design, phase, intervention and control arms, and descriptions of current use and future re-use of participant data and biological samples (**Supplemental Table 1**). Microsoft Excel was used for data extraction and recording and SPSS v28.0 was used to quantitatively summarise the data. The Standards for Reporting Qualitative Research were followed in reporting this research[16].

Pragmatic content analysis

For the analyses, we utilised a pragmatic content analysis approach as per Bengtsson et al. consisting of four stages: data decontextualisation, re-contextualisation, categorisation and compilation[17]. The unit of analysis was the individual trial PIL. A meaning unit was defined as the smallest number of sentences and paragraphs required to answer the relevant research question[18]. For the purposes of this study, this included any words, phrases or sentences related to current use or future re-use of participant data and biological samples. For example, "*..your doctor will send basic information about you…to the Trial Office…where it will be securely stored"*. We modified the pragmatic content analysis by including extra validation steps, as described below.

S*tage One – Data decontextualisation and categorisation*

We reviewed 240 PILs and extracted the relevant meaning units. Meaning units were coded and categorised using an inductive process as sequentially more PILs were reviewed. ERM devised the data categorisation framework and then piloted this framework by completing data categorisation on seven PILs. The data categorisation for these seven PILs was then reviewed by FS and NK, to enhance the reliability and consistency of the categorisation process[18]. Any discrepancies in the pilot phase were resolved by group consensus (ERM, FS and NK). Following completion of this pilot phase, four main categories and 27 sub-categories were created. ERM then extracted the relevant meaning units and categorised the data for the remaining 233 PILs. Of note, one sub-category item, '*Was the PIL description of how participant data would be confidentially maintained, adequate?*' required the opinion of the authors. For this, all three authors agreed a definition of 'adequate': as one which mentioned both 1) deidentification of participant data, e.g., via use of a unique trial ID number/code, and 2) compliance with GDPR/Data Protection Act or other similar named regulation. For the subcategory item '*Did the PIL mention if the biological sample would be stored in an anonymised or deidentified format?*', we considered samples stored using a trial ID number and no more than one identifier (e.g. date of birth or participant initials) as a deidentified or anonymised sample. We considered samples stored using two or more identifiers as not deidentified or anonymised, due to the greater risk of sample identification in this setting.

Stage Two – Data re-contextualisation, refinement of data categorisation.

Following completion of data categorisation, ERM re-reviewed and reverse compared all meaning units and data categories with the extracted descriptions from the 240 PILs, to ensure data were accurately captured without loss of important or relevant information. Any ambiguous meaning units were discussed at an online discursive meeting (ERM, FS and NK) and group consensus was reached regarding these. Examples of ambiguous terminology are included in **Supplemental Table 2**.

Stage Three –Validation of data categorisation

Ten percent of included PILs were randomly selected by FS using a random number generator, in line with previously reported rates for source data verification[19]. Two reviewers (FS and NK) then independently validated and verified data categorisation for this subsample of PILs, to further improve reliability and consistency of the data categorisation process. Any discrepancies in data categorisation following the 10% review were resolved by group consensus during an online discursive meeting (ERM, FS and NK). "This process involved a group review of the original descriptions included in the selected PILs, followed by a group discussion and a majority vote to reach a final consensus decision. Of note, <5% of items checked were discrepant between reviewers (1.6% [5/312] of items checked were discrepant between reviewers ERM and FS and 2.6% [8/312)] of items checked were discrepant between reviewers ERM and NK). In addition, for the single sub-category item requiring the reviewer's opinion (i.e. *was the description of maintaining confidentiality of data adequate?*), a subsample of PIL descriptions were provided to a Quality Assurance Manager in a Clinical Trials Unit for further external validation. This included five examples of adequate descriptions, five examples of inadequate

descriptions and three examples of ambiguous terminology. No discrepancies in data categorisation for this item were noted (13/13 examples, 100% agreement).

Stage Four – Data compilation

Data were compiled using a manifest analysis process, a process which aims to remain close to the original text by describing what is stated in the PILs[17]. We described data categories and subcategories quantitatively, using descriptive statistics, and qualitatively, using illustrative quotes from PILs.

Stage Five – Drafting recommendations

Following completion of the above stages, we then identified a list of recommended questions to be addressed by trial teams when designing PILs.

3. RESULTS:

PIL and trial characteristics

A total of 240 available PILs from 182 trials provided information on current use and future re-use of participant data and/or biological samples. All 240 PILs were published in English and 39% (n=94) used graphical illustrations (e.g. images, flow charts and/or tables) to describe the trial. The majority (63%, n=152) of PILs targeted adults, however minors (15%), parent/guardians (12%), legal representatives (4%) and other groups were also represented. Nineteen (8%) of the included PILs were related to trials which included patient and public involvement (PPI). Vulnerable patient populations (defined in table 2, footnote) were included in

34% (n=62) of trials. **Table 1** provides a detailed breakdown of PIL characteristics and their related trials.

Table 1. Characteristics of included participant information leaflets and clinical trials

PIL characteristics (N=240)	**n (%)**
PPI involvement mentioned	19 (8%)
Graphics included	94 (39%)
PIL target population	
Adults	152 (63%)
Minors	37 (15%)
Parent/caregiver (trials of minors)	28 (12%)
Legal representatives	9 (4%)
Participants regaining capacity for informed consent during the trial	14/ (6%)
Trial Characteristics (N=182)	**n (%)**
Drug trial[1]	25 (14%)
Non-drug trial[2]	85 (47%)
Mixed intervention trial[3]	70 (38%)
Vaccine trial	2 (1%)
No. of trial arms	
2	151 (83%)
3	15 (8%)
4	6 (3%)
5	3 (2%)
6	1 (1%)
Other[4]	6 (3%)
Commercial trial[5]	24 (13%)
Vulnerable participant population[6]	62 (34%)
Multiple PILs	22 (12%)
Disease area	
Oncology	34 (19%)
Obstetrics and gynaecology	21 (12%)
Cardiology	19 (10%)
Nephrology	18 (10%)
Neurology	10 (7%)
Respiratory	9 (5%)
Gastroenterology	8 (4%)
Infectious diseases (including covid-19)	8 (4%)
Rheumatology	7 (4%)
Dermatology	4 (2%)
Dentistry	3 (2%)
Endocrinology	3 (2%)

Paediatrics	2 (1%)
Psychiatry, Psychology and Mental health	2 (1%)
Orthopaedics	2 (1%)
Other[7]	32 (18%)

[1]Drug trial: Both the intervention and comparator were drugs.
[2]Non-drug trial: Both the intervention and comparator were non-drugs.
[3]Mixed interventions: The intervention and/or comparator included both drugs and non-drugs.
[4]Other: Trials included varying numbers of intervention arms, e.g. intervention arms with varying drug doses.
[5]Commercial trials: Trial funding was obtained from a commercial company or trial was linked with a commercial company with a potential conflict of interest, e.g., pharmaceutical company providing drugs free of charge.
[6]Vulnerable populations were defined via the local ethics committee and ICH-GCP definitions and included: infants and children aged 17 years and under, pregnant women, institutionalised individuals (prisoners, nursing home residents, or residents of mental health institutions), critically ill/ICU patients/patients on ventilators/patients unable to provide consent, adults aged 60 and over, participants with learning disabilities, participants with dementia, adults with terminal illness, homeless individuals and refugees, adults with mental illness, and members of the armed forces and medical/nursing/dental/ pharmacy students where there was a hierarchy in the trial that could influence the decision to take part voluntarily.
[7]Other: other disease area with <2 trials in that disease category.

Participant information leaflet descriptions of current use of participant data

All 240 PILs provided some description of how participant data would be stored and utilised in the current study. The majority (85%, 204/240) specifically mentioned, or directly implied, that confidentiality of participant data would be maintained. The description of maintenance of confidentiality was deemed to be adequate in 38% (92/240) of PILs. We defined adequate as a PIL description which mentioned both 1) deidentification of participant data, e.g., via use of a unique trial ID number/code, and 2) compliance with GDPR/Data Protection Act or other similar named regulation. Half (49%, 118/240) of PILs specified if a participant's electronic or paper medical records would be accessed during the trial. Forty seven percent (113/240) provided the intended duration of storage of participant data. Of those (n=113), the mean (±standard deviation, SD) duration of storage was 15.2 (±9) years. Thirty-eight percent (92/240) mentioned if participant data would be stored and utilised in compliance with current data protection regulations (**Table 2**). Sample PIL descriptions regarding current use of participant data are detailed in **Supplemental Table 3**.

Table 2. PIL descriptions of current use of participant data

n	n (%) N=240
Maintaining participant confidentiality mentioned or directly implied	204 (85%)
Description of maintenance of confidentiality of participant data	92 (38%)
Participant's GP will be informed of trial participation	
Yes	84 (35%)
No	0 (0%)
Not mentioned	156 (65%)
Participants' electronic/paper medical records will be accessed during the trial	
Yes	118 (49%)
No	0 (0%)
Not mentioned	122 (51%)
If yes, reason for access stated	102 (86%)
Duration of storage of participant data specified	113 (47%)
Minimum duration of data storage, yrs. (mean, standard deviation)	15.2 (9.0)
Minimum duration of data storage, yrs. (median, range)	15 (2-75)

PIL descriptions of future use of participant data

Of the 240 included PILs, 40% (96/240) specified if participant data would be used for future research studies and 28% (68/240) specified whether participant data may be shared with other researchers (**Table 3**). Of the 67 PILs which stated that data would be shared with other researchers, most (93%, 62/67) stated that any data shared would be in an anonymised or deidentified format. Sample PIL descriptions are detailed in **Supplemental Table 4**.

Table 3. PIL descriptions of future use of participant data

Future use of participant data	n (%) N=240
Participant data may be used in future research studies	
Yes	93 (39%)
No	3 (1%)
Not mentioned	144 (60%)
Participants' information may be used for contact about future research	
Yes	9 (4%)
No	17 (7%)

Not mentioned	214 (89%)
Data may be shared with other researchers	
Yes	67 (28%)
No	1 (0.4%)
Not mentioned	172 (72%)
If yes, PIL specified that any data shared would be anonymised or deidentified	62 (93%)

PIL descriptions of current use of biological samples

Of 240 included PILs, 49% (117/240) stated that biological samples would be requested from participants during the study. Of these, 38% (44/117) stated that samples may be used for genetic analysis, 80% (94/117) provided a reason for requesting samples and 66% (77/117) specified if samples would be stored in an anonymised or deidentified format. Only 21% (24/117) specified whether individual-level results of research sample analysis would be made available to participants (**Table 4**). Sample PIL descriptions are detailed in **Supplemental Table 5**.

Table 4. PIL descriptions of current use of biological samples

Current use of biological samples	**n (%) N=240**
Biological samples will be collected from participants during the study	
Yes	117 (49%)
No	117 (49%)
Not mentioned	6 (2%)
Type of sample requested specified	117 (100%)
Specific type of biological sample to be collected	
Blood	92 (79%)
Urine	24 (21%)
Saliva	12 (10%)
Tissue (e.g. from biopsy)	20 (17%)
Stool	4 (3%)
Bone marrow	5 (4%)
Cerebrospinal fluid	8 (7%)
Respiratory secretions	17 (15%)
Skin	2 (2%)
Sample may be used for genetic analysis	

Yes	44 (38%)
No	0 (0%)
Not mentioned	73 (62%)
Reason provided for collection of biological sample	
Yes	94 (80%)
Not mentioned	23 (20%)
Description provided of how sample will be stored	
Yes	81 (69%)
Not mentioned	36 (31%)
Sample will be stored in anonymised/deidentified format	
Yes	72 (62%)
No	5 (4%)
Not mentioned	40 (34%)
Description provided of how sample will be analysed (e.g. location of lab)	
Yes	60 (51%)
Not mentioned	57 (49%)
Individua-llevel results of sample analysis will be provided to participants	
Yes	9 (8%)
No	15 (13%)
Not mentioned	93 (79%)

PIL descriptions of future use of biological samples

117 PILs specified that biological samples would be collected from participants. 62% (73/117) stated that samples would be stored for future research use, 8% (9/117) specified that samples would be destroyed at the end of the current study, while the remaining 30% (35/117) provided no information. Of the 73 PILs which specified planned storage of samples for future research, 75% (55/73) described how samples would be stored, 18% (13/75) specified the intended duration of storage, 81% (59/73) specified if samples would be stored in an anonymised or deidentified format and 48% (35/73) specified whether samples would be shared with other researchers (**Supplemental Table 6)**. Sample PIL descriptions are detailed in **Supplemental Table 7.**

Ambiguous descriptions in PILs

In the included PILs, many descriptions regarding participant data and sample use and re-use were ambiguous and open to interpretation. For example, with respect to how participant data confidentiality would be maintained, one PIL described this as "*data protection regulations provide you with control over your personal data and how it is used…however, some of those rights may be limited in order for the research to be reliable and accurate*". Additional samples of ambiguous quotes are included in **Supplemental Table 8.**

4. DISCUSSION

Summary of findings

We found that PIL descriptions regarding current use and future re-use of participant data and biological samples are often insufficiently detailed or ambiguous. The majority fail to adequately describe how data confidentiality will be maintained, specify the intended duration of data storage or specify whether data will be used in future studies or shared with other researchers. Many PILs do not specify the purpose for obtaining biological samples, whether samples will be anonymously stored/deidentified or whether participants will be informed of their own individual-level results. Only 8% of the included PILs were related to trials with patient and public involvement.

In our sample, over 60% of PILs did not meet our definition of an adequate description of how participant data confidentiality would be maintained. Some descriptions were ambiguous and open to interpretation. Given that 31% of trial participants in a recent survey expressed a concern that their data could be stolen[4], our findings underscore the importance of providing clear and

sufficiently detailed descriptions in PILs regarding how participant data will be safeguarded and confidentially maintained. Otherwise, recruitment and retention to the trial could be impacted.

Important information was often located under incorrect headings within PILs, further limiting clarity and readability for participants. For example, information relevant to future use of anonymised research samples was included under the section "*Will my taking part in this study be confidential?*" and not included in the section "*What will happen to any samples I provide?*". Previous studies evaluating readability of patient-facing documents report they are often inappropriately complex[20-25], with one study noting that 0/154 PILs evaluated met the recommended mean reading age of 12 years or under[20]. We suggest that PILs should use specific headings (e.g. "*Will my data be kept confidential*", "*What will happen to samples I provide*", etc.), to clearly signpost topics for participants, and if necessary, repeat relevant, important information under different headings so that participants don't have to scan multiple sections of the PIL to obtain the information they require.

In our sample, almost 80% of PILs failed to specify whether participants would receive the results of their own individual research sample analysis (either directly or via their GP). Research on the provision of information regarding dissemination of trial results in PILs found that although 74% of PILs provided information on sharing of results with participants, 70% of those were passive, i.e., requiring participants to seek the results.[1] In our study, none of the included PILs specified whether a participant or their GP would be informed if an actionable, abnormal incidental finding was detected during their research sample analysis, e.g., a tumour or clinically relevant laboratory abnormality.A handful of advisory statements are available

providing recommendations on reporting of incidental laboratory abnormalities, genetic anomalies and radiological findings to trial participants[26-29]. An advisory statement from the US Bioethics Commission recommends that informed consent processes for trial participants should clearly describe the types of incidental findings that may occur (e.g. anticipated and unanticipated), plans for disclosing and managing incidental findings (particularly for significant and clinically actionable findings), and should also specify how participants may opt out of receiving results, if desired[26,27]. We recommend that all PILs should clearly describe plans for sharing of individual-level findings, particularly actionable findings, to ensure participants have a full understanding of the implications of consenting to participate in a trial that requires biological samples.

There is currently a lack of available guidelines recommending best practice approaches to describing current and future use of participant data and biological samples in PILs[2]. Consensus guidance from multi-stakeholder task forces, including patient and public representatives, methodologists, researchers, funders and standards development groups, are available which address some of these issues[3,4,13,30-34]. A group consensus document authored by 41 individuals on the sharing and re-use of clinical trial data was published in 2017. This was a consensus-building process occurring over a series of three workshops held over one year (March 2016, October 2016 and March 2017). The process involved a multistakeholder task force. Members included researchers, patient representatives, funding representatives, information technology experts and methodologists. The taskforce was set within the Horizon 2020-funded project CORBEL (Coordinated Research Infrastructures Building Enduring Life-science Services) and coordinated by the European Clinical Research Infrastructure Network. The taskforce

recommended obtaining explicit informed consent from trial participants regarding the sharing and re-use of their data for research purposes (separate to the consent required for trial participation), deidentifying data prior to sharing, and clearly describing de-identification steps to trial participants[3]. The group specifically recommended that trial documents such as PILs should provide detailed descriptions of data storage and sharing plans[3]. In a 2018 survey of 771 current and recent clinical trial participants at three US academic medical centres (surveyed via mail and clinic waiting rooms), the majority (93%) indicated high-to-moderate willingness to share their data with university scientists, provided appropriate data security safeguards were in place[4]. Another study completed an in-person survey of 452 cancer patients attending a US university clinic (predominantly non-Hispanic white patients) and inner city county hospital (predominantly African American patents from lower socio-economic backgrounds) between 2003 and 2004. The majority (95%) of participants were willing to provide biological samples for future research use[30]. In a third study, a representative random sample of 4659 US adults completed an online survey between December 2007 and January 2008 (response rate of 58%). In this study, 60% of potential participants were willing to provide biobank samples for genetic research[31], with data privacy and protection concerns cited as the biggest factor which could limit potential participation[31]. Importantly, only 8% of the included PILs in our review were related to trials with patient and public involvement. Active involvement of patients and public representatives is key in helping trial teams to design 'best practice' future PILs that appropriately address the information needs and concerns of potential trial participants[13,33,35].

Strengths and limitations

The large corpus of PILs (n=240) is a particular strength of this study. The methodology chosen, which included independent verification of our categorisation framework as well as verification for one component by a Quality Assurance manager at a clinical trials unit, adds to the credibility of the findings and recommendations. Furthermore, we included PILs from different disease areas and targeting different trial populations, including adults, minors, legal representatives of individuals lacking capacity to provide informed consent and vulnerable populations, enhancing the generalisability of our findings.

Our study has several limitations. Only PILs from trial units in the UK and Ireland were eligible for inclusion, and so results may not be generalisable to PILs used in other geographic regions. However, with respect to data privacy specifically, there are similar data privacy protection regulations in place in the US, i.e. the Health Insurance Portability and Accountability Act of 1996 and Privacy Rule, 2003. We were only able to analyse PILs from trial units that responded to our request (23/52 clinical trial units, 445 response rate). Patients and trial participants were not contacted to obtain their insights and opinions on how PILs should describe current use and future re-use of participant data and biological samples, and only 8% of the included PILs were related to trials with patient and public involvement.

5. Implications for practice

In the absence of current consensus-based recommendations or published guidelines on describing current use and future re-use of participant data and samples to trial participants in PILs, we recommend the following questions be addressed by trial teams when designing PILs (**Table 5**).

Table 5. Recommended questions to be addressed by trial teams when designing PILs

	Recommended questions
1	Will biological samples be collected from participants? What are the reasons for obtaining samples, and what are the planned specific sampling procedures (e.g. number, timing, frequency of sampling)?
2	Are there any plans for genetic analyses on samples? What are the potential implications of these analyses? (e.g. health insurance risk if there is any breach of data confidentiality, management plan if clinically actionable genetic anomalies are identified including whether genetic counselling and follow-up will be offered to participants and their families)
3	How will participant data and biological samples be confidentially maintained in the current study? What are the plans for data and sample deidentification, secure storage and transfer? Which data protection regulations will these procedures follow?
4	Will participant health records be accessed during the trial, and for what purpose?
5	What is the intended duration of data and sample storage for the current study?
6	Will data and samples be stored for future research use? If so, what are the sample storage plans including use of anonymisation/deidentification of samples (or if not, which identifiers will be used for sample storage e.g. initials, date of birth)? What is the intended duration of data and sample storage for future use?
7	Will data and samples be shared with other researchers for future research (specify yes or no)? If so, will data and samples be shared in an anonymised/deidentified format?
8	Will results of any current and future research sample analysis be made available to participants or their GPs? What types of incidental findings (anticipated and unanticipated) could occur and what are the plans for disclosing and managing incidental findings that do occur? How can participants opt out of receiving such results, if desired.

7. CONCLUSIONS:

PIL descriptions regarding current use and future re-use of participants' data and their biological samples are often insufficiently detailed or ambiguous and most do not adequately describe plans for maintaining confidentiality of participant data. In the absence of published guidelines, we

provide eight suggested questions for trial teams to consider when designing PILs which should improve practice.

LIST OF ABBREVIATIONS:

GDPR, General Data Protection Regulations

PIL, patient information leaflet

PPI, patient and public involvement

RCT, randomised controlled trial

SWAT, study within a trial

TT, trial team

DECLARATIONS:

Ethics approval and consent to participate

Ethical approval was not required for the purposes of this study as it was a retrospective analysis of publicly available PILs.

Consent for publication

All authors agree and consent to manuscript submission and publication.

Availability of data and materials

The dataset used during the current study is available from Dr Frances Shiely on request. There are plans to make the dataset available in the future on the Trial Forge website https://www.trialforge.org/

Competing interests

The authors have no conflicts of interest to declare.

Funding

No funding was received for this study which was completed in part fulfilment of the MSc Clinical Trials degree at University College Cork. However, ERM holds a Health Research

Board of Ireland (CSF-2020-011) career development award. The views expressed in this paper are those of the author(s). None of the funding entities listed had any role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.

Authors' contributions

ERM contributed to the conception and design of the study, the analysis and interpretation of data, drafting the article and revising it critically for important intellectual content and has given final approval of the version submitted.
NK contributed to the conception and design of the study, the acquisition, analysis and interpretation of data, revising the article critically for important intellectual content and has given final approval of the version submitted.
FS contributed to the conception and design of the study, the acquisition, analysis and interpretation of data, revising the article critically for important intellectual content and has given final approval of the version submitted.

Acknowledgements

None.

Authors' information

ERM is an Associate Professor in the School of Medicine, University of Galway and HRB-Clinical Research Facility Galway. FS is Director of Education and a Senior Lecturer in Patient Focused Research and Epidemiology at the HRB Clinical Research Facility and School of Public Health at University College Cork. She is Programme Director for the MSc Clinical Trials (online) and UCC PI for the HRB Trials Methodology Research Network. NK is a Research Fellow and Head of Data Management at the UKCRC registered trials unit Centre for Trials Research in Cardiff.

Tables with captions

Table 1. Characteristics of included participant information leaflets and clinical trials

REFERENCES:

1. Bjorklund M, Shiely F, Gillies K. Information about dissemination of trial results in patient information leaflets for clinicals trials in the UK and Ireland: The what and the when. *PLoS One*. 2022;17(5):e0268898. doi:10.1371/journal.pone.0268898
2. Coleman E, O'Sullivan L, Crowley R, et al. Preparing accessible and understandable clinical research participant information leaflets and consent forms: a set of guidelines from an expert consensus conference. *Research Involvement and Engagement*. 2021/05/18 2021;7(1):31. doi:10.1186/s40900-021-00265-2
3. Ohmann C, Banzi R, Canham S, et al. Sharing and reuse of individual participant data from clinical trials: principles and recommendations. *BMJ Open*. 2017;7(12):e018647. doi:10.1136/bmjopen-2017-018647
4. Mello MM, Lieou V, Goodman SN. Clinical Trial Participants' Views of the Risks and Benefits of Data Sharing. *New England Journal of Medicine*. 2018;378(23):2202-2211. doi:10.1056/NEJMsa1713258
5. Cutts M. Making leaflets clearer for patients. *Medical Writing*. 2015/03/01 2015;24(1):14-19. doi:10.1179/2047480614Z.000000000271
6. United Kingdom Department of Health. Toolkit for producing patient information. Version 2.0. Accessed 15 February, 2023, 2023. https://www.uea.ac.uk/documents/746480/2855738/Toolkit_for_producing_patient_informa tion.pdf.
7. Lampert A, Wien K, Haefeli WE, Seidling HM. Guidance on how to achieve comprehensible patient information leaflets in four steps. *International Journal for Quality in Health Care*. 2016;28(5):634-638. doi:10.1093/intqhc/mzw077
8. Hoffmann T, Worrall L. Designing effective written health education materials: Considerations for health professionals. *Disability and Rehabilitation*. 2004/10/07 2004;26(19):1166-1173. doi:10.1080/09638280410001724816
9. Buck ML. Providing patients with written medication information. *Ann Pharmacother*. Sep 1998;32(9):962-9. doi:10.1345/aph.17455
10. Medicines and Healthcare products Regulatory Agency. Best practice guidance on patient information leaflets. Updated 31 December 2020. Accessed February 10, 2023, 2023. https://www.gov.uk/government/publications/best-practice-guidance-on-patient-information-leaflets
11. Health Research Authority and Medical Research Council UK. Accessed February 15, 2023, https://www.hra-decisiontools.org.uk/consent/docs/PIS-Template_version2.pdf
12. National Office for Research Ethics Committees. PIL Summary Guidance. Accessed February 15, 2023, 2023. https://www.nrecoffice.ie/pil-summary-guidance/
13. Innes K, Cotton S, Campbell MK, Elliott J, Gillies K. Relative importance of informational items in participant information leaflets for trials: a Q-methodology approach. *BMJ Open*. 2018;8(9):e023303. doi:10.1136/bmjopen-2018-023303
14. Clarke N, Vale G, Reeves EP, et al. GDPR: an impediment to research? *Irish Journal of Medical Science*. 2019/11/01 2019;188(4):1129-1135. doi:10.1007/s11845-019-01980-2
15. The European Parliament and Council. Regulation (EU) 2016/679 (General Data Protection Regulation). Accessed 15 February, 2023, https://gdpr-info.eu
16. O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for reporting qualitative research: a synthesis of recommendations. *Acad Med*. Sep 2014;89(9):1245-51. doi:10.1097/acm.0000000000000388

17. Bengtsson M. How to plan and perform a qualitative study using content analysis. *NursingPlus Open*. 2016/01/01/ 2016;2:8-14. doi:https://doi.org/10.1016/j.npls.2016.01.001
18. Graneheim UH, Lundman B. Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. *Nurse Education Today*. 2004/02/01/ 2004;24(2):105-112. doi:https://doi.org/10.1016/j.nedt.2003.10.001
19. Houston L, Probst Y, Humphries A. Measuring Data Quality Through a Source Data Verification Audit in a Clinical Research Setting. *Stud Health Technol Inform*. 2015;214:107-13.
20. O'Sullivan L, Sukumar P, Crowley R, McAuliffe E, Doran P. Readability and understandability of clinical research patient information leaflets and consent forms in Ireland and the UK: a retrospective quantitative analysis. *BMJ Open*. Sep 3 2020;10(9):e037994. doi:10.1136/bmjopen-2020-037994
21. Ménoni V, Lucas N, Leforestier J-F, et al. Readability of the Written Study Information in Pediatric Research in France. *PLOS ONE*. 2011;6(4):e18484. doi:10.1371/journal.pone.0018484
22. Ménoni V, Lucas N, Leforestier JF, et al. The Readability of Information and Consent Forms in Clinical Research in France. *PLOS ONE*. 2010;5(5):e10576. doi:10.1371/journal.pone.0010576
23. Samadi A, Asghari F. Readability of informed consent forms in clinical trials conducted in a skin research center. *J Med Ethics Hist Med*. 2016;9:7.
24. Terblanche M, L. B. Examining the readability of patient-informed consent forms . *Open Access Journal of Clinical Trials*. 2010;2:157-162. doi:https://doi.org/10.2147/OAJCT.S13608
25. Ennis L, Wykes T. Sense and readability: participant information sheets for research studies. *Br J Psychiatry*. Feb 2016;208(2):189-94. doi:10.1192/bjp.bp.114.156687
26. Presidential Commission for the Study of Bioethical Issues. For IRB members: Incidental and secondary findings. . Accessed 12 May, 2023., https://bioethicsarchive.georgetown.edu/pcsbi/sites/default/files/IRB%20Primer%20Incidental%20Findings%2010.30.16.pdf.
27. Presidential Commission for the Study of Bioethical Issues. Anticipate and Communicate: Ethical Management of Incidental and Secondary Findings in the Clinical, Research, and Direct-to-Consumer Contexts. Accessed May 12, 2023., http://bioethics.gov/sites/default/files/FINALAnticipateCommunicate_PCSBI_0.pdf.
28. Kalia SS, Adelman K, Bale SJ, et al. Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update: a policy statement of the American College of Medical Genetics and Genomics. *Genet Med*. Feb 2017;19(2):249-255. doi:10.1038/gim.2016.190
29. Royal College of Radiologists (RCR) and the Scottish Imaging Network. Management of Incidental Findings Detected During Research Imaging. Accessed May 12, 2023, https://www.rcr.ac.uk/system/files/publication/field_publication_files/BFCR%2811%298_ethics.pdf
30. Pentz RD, Billot L, Wendler D. Research on stored biological samples: Views of African American and White American cancer patients. *American Journal of Medical Genetics Part A*. 2006;140A(7):733-739. doi:https://doi.org/10.1002/ajmg.a.31154
31. Kaufman DJ, Murphy-Bollinger J, Scott J, Hudson KL. Public Opinion about the Importance of Privacy in Biobank Research. *The American Journal of Human Genetics*. 2009;85(5):643-654. doi:10.1016/j.ajhg.2009.10.002

32. Trinidad SB, Fullerton SM, Bares JM, Jarvik GP, Larson EB, Burke W. Genomic research and wide data sharing: Views of prospective participants. *Genetics in Medicine*. 2010;12(8):486-495. doi:10.1097/GIM.0b013e3181e38f9e
33. Sanderson SC, Brothers KB, Mercaldo ND, et al. Public Attitudes toward Consent and Data Sharing in Biobank Research: A Large Multi-site Experimental Survey in the US. *The American Journal of Human Genetics*. 2017;100(3):414-427. doi:10.1016/j.ajhg.2017.01.021
34. Garrison NA, Sathe NA, Antommaria AHM, et al. A systematic literature review of individuals' perspectives on broad consent and data sharing in the United States. *Genetics in Medicine*. 2016;18(7):663-671. doi:10.1038/gim.2015.138
35. O'Sullivan L, Carroll TP, Clarke N, et al. Harmonising the human biobanking consent process: an Irish experience. *HRB Open Res*. 2021;4:96. doi:10.12688/hrbopenres.13384.2