ORIGINAL ARTICLE

Development of a repository of individual participant data from randomized controlled trials of therapists delivered interventions for low back pain

S.W. Hee[1], M. Dritsaki[2], A. Willis[1], M. Underwood[1], S. Patel[1]

1 Division of Health Sciences, Warwick Medical School, University of Warwick, Coventry, UK
2 Oxford Clinical Trials Research Unit, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, UK

Correspondence
Siew Wan Hee
E-mail: S.W.Hee@warwick.ac.uk

Conflicts of interest
The authors declare that there are no conflicts of interest.

Abstract

Background: Individual patient data (IPD) meta-analysis of existing randomized controlled trials (RCTs) is a promising approach to achieving sufficient statistical power to identify sub-groups. We created a repository of IPD from multiple low back pain (LBP) RCTs to facilitate a study of treatment moderators. Due to sparse heterogeneous data, the repository needed to be robust and flexible to accommodate millions of data points prior to any subsequent analysis.
Methods: We systematically identified RCTs of therapist delivered intervention for inclusion to the repository. Some were obtained through project publicity. We requested both individual items and aggregate scores of all baseline characteristics and outcomes for all available time points. The repository is made up of a hybrid database: entity-attribute-value and relational database which is capable of storing sparse heterogeneous datasets. We developed a bespoke software program to extract, transform and upload the shared data.
Results: There were 20 datasets with more than 3 million data points from 9328 participants. All trials collected covariates and outcomes data at baseline and follow-ups. The bespoke standardized repository is flexible to accommodate millions of data points without compromising data integrity. Data are easily retrieved for analysis using standard statistical programs.
Conclusions: The bespoke hybrid repository is complex to implement and to query but its flexibility in supporting datasets with varying sets of responses and outcomes with different data types is a worthy trade off. The large standardized LBP dataset is also an important resource useable by other LBP researchers.
Significance: A flexible adaptive database for pain studies that can easily be expanded for future researchers to map, transform and upload their data in a safe and secure environment. The data are standardized and harmonized which will facilitate future requests from other researchers for secondary analyses.

1. Introduction

Globally, low back pain (LBP) is one of the leading causes of years lived with disability, and in developed countries it is the leading contributor to the burden of disability adjusted life years in young adults (Hoy et al., 2014; Global Burden of Disease Study 2013 Collaborators, 2015). Therapist-delivered interventions—non-drug, non-surgical approaches—to the treatment

of LBP are widely used. There is good evidence that several therapist-delivered treatment approaches are effective. There is also evidence that some of these treatments are cost-effective, e.g. offering a course of manual therapy, including spinal manipulation, comprising up to nine sessions over a period of up to 12 weeks (Savigny et al., 2009). The average effect size of therapists delivered interventions for LBP is typically modest.

One approach for improving outcomes is to identify treatment moderators, baseline characteristics, that predict the greatest benefits or least effectiveness from an intervention for an individual with LBP. To test for a modest interaction between a moderator and a treatment, a randomized controlled trial (RCT) needs at least 503 participants (see, (Gurung et al., 2015) for a full description of this power calculation). Mistry and colleagues reported that most RCTs are too small to reliably identify subgroups (Mistry et al., 2014). Therefore, many of these sub-group analyses were severely underpowered. These data are not suitable for meta-analysis for treatment moderation. Individual participant data (IPD) meta-analysis of RCTs will allow meta-analysis of potential treatment moderators and has the potential to provide adequate statistical power to identify sub-groups who may benefit most from particular treatment options. To achieve this, data from relevant trials need to be assembled and merged into a single useable dataset. Ideally, the dataset structure should also allow additional trial datasets to be added as they become available to facilitate future research.

Clinical trial datasets can be stored in a flat file tabular format such as Microsoft Excel, which typically uses rows to represent a participant record and columns to represent variables captured on case report forms (CRF). Tabular formats are useful for small datasets and have the advantage of being intuitive, relatively simple to create and machine-readable. However, they can be susceptible to excessive growth with each patient record requiring a new row to be inserted and an additional column for every variable. Large numbers of columns can quickly accumulate when clinical and non-clinical items are measured across multiple time points.

Relational databases provide a more robust and efficient solution for larger datasets. This model allows data to be stored and connected in individual tables. Repeating data groups can be separated into their own table and joined back to the main domain using relationships, thus reducing the data redundancy problem associated with flat files.

The rules governing a relational database are specified in a schema that can be complex and time consuming to design. The repository relies on data from multiple RCTs and is frequently altered to accommodate new discoveries requiring a more flexible solution. We describe the process of collating data from multiple RCTs. We also describe the process of developing a hybrid database that is flexible and robust for storing multiple datasets to facilitate current analyses and the addition of future datasets.

2. Methods

2.1 Collation of trials

We systematically identified all RCTs of therapist delivered treatments for LBP up until September 2011. We have described this process in detail elsewhere (Mistry et al., 2014; Gurung et al., 2015). We searched this dataset to identify unique trials with >179 participants. We started with an original lower limit of 200 for the sample size. Allowing for some loss to follow-up, a trial of 200 participants will have 90% statistical power to identify a standardized mean difference of 0.5 between two treatment groups. Any individual trials smaller than this are likely to be seriously underpowered for their primary outcome. Upon screening the trials there were many that obtained a final sample size of just under 200; typically these were studies aiming for around 200 participants that fell short of the final target. We therefore revised our inclusion criteria to >179 participants.

From a practical perspective of approaching trial investigators, our inclusion criteria yielded a manageable number of trials to approach; large trials (those of thousands of participants) and small trials (<100 participants) each create a similar amount of work to collate. We also obtained data from trials that were not on our original list, as investigators became aware of our project. Although these trials had smaller sample sizes than our target studies, we decided to include them to add power to our analysis. As the primary purpose of this project was to identify sub-groups rather than main treatment effects, the omission of some small trials from our dataset is unlikely to have materially affected our conclusions.

Between 2011 and 2012, each investigator was invited to participate and share their data with us to the standardized LBP repository via email. If a response was not received within 6–8 weeks, a reminder email was sent. For those interested a

personalized data sharing agreement was created and sent to the investigator to review and sign. Once the signed document was received by the University, the investigator was provided with details on how to securely send their data to us. We used the University of Warwick secure file transfer service. Investigators were advised that any datasets sent to us needed to be anonymized and encrypted.

On receipt of the data, the statistician (SWH) and health economist (MD) queried them before mapping and transforming the original data to the standard for the repository. Details of the mapping and transformation procedures are described below (Section 2.2).

Data integrity is vital to the repository. To check that the mapping and transformation procedures were done correctly, the repository data were routinely checked against the original datasets. To achieve this, at each time point (baseline and all follow-ups), a random sample of data was extracted and manually cross checked against the source data. Any inconsistencies were flagged and, if required, the mapping and transformation instructions were amended. This process was repeated until the data were deemed to have been transformed correctly, i.e. zero error.

2.2 Bespoke database

Our bespoke database is a hybrid of an entity-attribute-value (EAV) model and a fixed schema relational model. This design is commonly used in clinical trial database management systems (Brandt et al., 2002) as it provides the flexibility of storing sparse heterogeneous data while enforcing high data integrity. A detailed technical description of the system architecture and our approach to data transfer can be found in the Supporting Information (Method S1).

The basic components of the database can be broken down into four main tables. Fig. 1 shows a simplification of three of these tables. The 'Subject' table stores the participant's original identifier (ID) and a unique identifier generated by the system (Fig. 1C). The 'Object' table stores a reference to the 'Subject' table, a unique object identifier, and a key value representing either a single CRF or a group of repeating questions that have been separated into a child object and then re-joined to a parent using a parent/child relationship (Fig. 1D). This approach to storing repeating data groups is an interpretation of the EAV with classes and relationships (EAV/CR) (Nadkarni et al., 1999). The 'Attribute' table is used

to simply store a list of all of the repository's variables. The 'EAV' table (Entity, Attribute and Value) stores references to the related 'Object' and 'Attribute' tables, as well as the actual value (Fig. 1E). Thus, a complete record can typically be recreated by selecting a collection of rows from the EAV table that have the same object identifier.

A simplification of how two original tabular source datasets are stored in the EAV model is shown in Fig. 1. In Fig. 1A, Trial A collected gender and the EuroQol five dimensions questionnaire (EQ5D) (EuroQol Group, 1990) and EuroQol visual analogue scale (EQ-VAS) of health state at baseline (labelled, EQ1, EQ2, EQ3, EQ4, EQ5 and EQ6, respectively) and at the first follow-up (labelled, EQ1_1, EQ2_1, EQ3_1, EQ4_1, EQ5_1 and EQ6_1), where the first follow-up was 4 weeks post-randomization. The coded values 1 and 2 for the variable 'sex' represent male and female, respectively. In Fig. 1B, Trial B collected gender, age, and the Roland Morris Disability Questionnaire (RMDQ) (Roland and Morris, 1983) at baseline (labelled, RDQ_0) and 3-month follow-up (labelled, RDQ_3mo).

For each participant, the repository generated a unique ID as seen in Fig. 1C. For each domain occurrence, a row is created in the 'Object' table (Fig. 1D). For example, for subject #011 (equivalently, unique ID #1000001), one row was created for Demographics and two rows for EQ5D. The demographic data were only recorded once, at baseline per participant, hence, only one row was required. As the EQ5D and EQ-VAS were collected at both baseline and at 4-week follow-up, a row was required for each time point.

A row is then created for each populated cell in the 'EAV' table (Fig. 1E). One row is also created for each time point that the item was collected at. Rows are only created for populated cells from the original data source. For example, five EQ5D rows and five rows for their corresponding follow-up were created for subject #011 to capture the values recorded at baseline. Note that the labels for all of the variables (attributes) in the 'EAV' table follow the repository standard. In addition, original values are transformed, if necessary, to the repository standard. In the standardized repository, male and female values are represented numerically by codes 1 and 2, respectively. In our example, the values 'M' and 'F' for the variable sex from trial B were transformed to 1 and 2, respectively, in the 'EAV' table. For mapping and transformation details, please refer to Supporting Information Method S1 Section 2.2.2.

A Original data source from trial A (tabular formatted)

ID	Sex	EQ1	EQ2	EQ3	EQ4	EQ5	EQ6	EQ1_1	EQ2_1	EQ3_1	EQ4_1	EQ5_1	EQ6_1
011	2	2	2	2	2	2	.	1	1	2	2	1	74
021	1	1	2	1	2	1	60	.	.	.	.	.	45

B Original data source from trial B (tabular formatted)

ID	Gender	Age	RDQ_0	RDQ_3mo
A1001	M	50	13	7
A1002	M	28	.	9
A1003	F	.	18	10

C Subject

Unique subject ID	Original Subject ID
1000001	011
1000002	021
1000003	A1001
1000004	A1002
1000005	A1003

D Object

Object ID	Unique subject ID	Domain
1	1000001	Demographics
2	1000001	EQ5D
3	1000001	EQ5D
4	1000002	Demographics
5	1000002	EQ5D
6	1000002	EQ5D
7	1000003	Demographics
8	1000003	RMDQ
9	1000003	RMDQ
10	1000004	Demographics
11	1000004	RMDQ
12	1000005	Demographics
13	1000005	RMDQ
14	1000005	RMDQ

E EAV

Object ID	Attribute	Value
1	SEX	2
2	EQ5D1	2
2	FU	0
2	EQ5D2	2
2	FU	0
2	EQ5D3	2
2	FU	0
2	EQ5D4	2
2	FU	0
2	EQ5D5	2
2	FU	0
3	EQ5D1	1
3	FU	4
3	EQ5D2	1
3	FU	4
3	EQ5D3	2
3	FU	4
3	EQ5D4	2
3	FU	4
3	EQ5D5	1
3	FU	4
3	EQ5D6	74
3	FU	4
4	SEX	1
5	EQ5D1	1
5	FU	0
5	EQ5D2	2
5	FU	0
5	EQ5D3	1
5	FU	0
5	EQ5D4	2
5	FU	0
5	EQ5D5	1
5	FU	0
5	EQ5D6	60
5	FU	0
6	EQ5D6	45
6	FU	4
7	SEX	1
7	AGE	50
8	RMDQ	13
8	FU	0
9	RMDQ	7
9	FU	13
10	SEX	1
10	AGE	28
11	RMDQ	9
11	FU	13
12	SEX	2
13	RMDQ	18
13	FU	0
14	RMDQ	10
14	FU	13

Figure 1 Sample of tabular clinical data in an EAV table. (A) and (B) Examples of original clinical data from two trials in a tabular format, (C) the 'Subject' table with a new unique ID for each participant, (D) the 'Object' table with an instance of a domain per participant for every derived tabular record, and (E) the 'EAV' table with a row for each populated cell and a row for the follow-up time point where applicable.

2.3 Healthcare resource-use dataset

The mapping and transformation of healthcare resource-use data were more challenging because the different types of resources used across RCTs did not conform to any standard. However, each question and answer in a typical healthcare resource-use questionnaire could be broken into eight parts: the recall period (e.g. 3-month follow-up), the type of resource (e.g. visit to physiotherapist), the reason for using the resource (e.g. LBP), the location of the resource (e.g. community healthcare centre), the unit of measurement (e.g. home visit), the quantity, the cost or expenses incurred, and the payer (e.g. national health care system).

Fig. 2 shows a simplified version of a typical healthcare resource-use questionnaire (Patel et al., 2016). In this example, participants were asked to record all of the healthcare resources they had used at the 3-month follow-up time point (Fig. 2A). The answers provided were stored in a tabular format that used 12 columns to capture all responses to the five questions (Fig. 2B). By using this format, the number of columns required to accommodate the data would grow in line with the maximum number of responses provided by any individual. Fig. 2C shows a view of the repository healthcare resources data generated and pivoted from the EAV/CR model. This view displays the eight standard repository healthcare resource-use attributes and an additional attribute called 'Text' that is used to capture any comments that were written on the CRFs.

The process for creating the transformed healthcare resource-use data involves splitting the original questions into a number of derived parts that will map to the standard attributes. In our example, question 1 asked how many times the participant had consulted their primary care doctor for any reason in the last 3 months. Thus, the recall period was set to '13' (because the repository standard stores the time point in unit weeks), the type of resource was 'GP', the reason for using the resource was 'Any condition', the location of the resource was 'Primary Care Clinic', the unit of measurement was 'Visit', and the payer was 'Public Health Service'. All of these values were derived solely from the information contained in the original question, as opposed to the value of the variable. Only the attribute 'quantity' was directly mapped to the original variable's value. For question 3, both quantity and cost were directly mapped to the original variables' values. Mapping and transformation details for healthcare resource-use data are presented in Supporting Information Method S1 Section 2.2.3.

2.4 Using the repository data

Data in the EAV/CR format are not suitable for analysis because of their fragmented structure. Thus, data from the same domain are pieced together and pivoted so that the dataset for each domain resembles a long format tabular structure. Technical details of the extraction, transformation, loading and outputting the repository data are in Supporting Information Method S1 Sections 2.2.4 and 2.25. For example, suppose we would like to view and analyse demographics data from our simplistic example in Fig. 1. We can extract data from Object ID #1, #4, #7, #10 and #12 (Fig. 1D) and their corresponding data from tables 'Subject' (Fig. 1C) and 'EAV' (Fig. 1E), joined by their unique subject ID and object ID. The extracted data are presented in a tabular format for analysis, as seen in Fig. 3A.

Fig. 3B and C show the long format tabular structure for data captured at various time points for EQ5D and RMDQ, respectively. End users would need to combine data from different domains, e.g. demographics and RMDQ for analysis to investigate potential variables that moderate RMDQ outcomes. Note that the LBP repository database is more detailed than the one we present here for illustration. In the repository information such as the name of the trial is stored in a fixed table (see, Supporting information Fig. S1) and is pieced together with the domain data into a long tabular structure as seen in Fig. 3.

3. Results

3.1 Identification of trials

Our initial search yielded 658 hits. After exclusions, we identified 42 unique trials datasets that met our entry criteria. These authors were contacted for their trial data. We obtained no response from 15 corresponding authors, six datasets were no longer available and for seven datasets we were unable to conclude negotiations prior to freezing our database. We were able to satisfactorily import data from 14 of these trials. We also included data from five smaller trials offered to us by researchers who were aware of this project (Fig. 4). The final dataset included data from 20 datasets (one trial had a feasibility study prior to the main trial) and a total of 9328 participants (Patel et al., 2016).

A Sample of questions and answers at 3-month follow-up.

Public healthcare professionals	No. of visits	
1. In the last 3 months, how many times have you consulted your doctor or another doctor for any reason?	*Pri1*	
2. In the last 3 months, how many times have you consulted the NHS physiotherapist for low back pain?	*Pri2*	

Private healthcare professionals	No. of visits	Cost (£)
3. In the last 3 months, how many times have you consulted any private physiotherapist for low back pain?	*nPriv1*	*cPriv1*

Medication

4. In the last 3 months, list the medicines prescribed by your doctor for your low back pain.

Medicine prescribed		No. of prescriptions
i.	*pmed1*	*nmed1*
ii.	*pmed2*	*nmed2*

5. In the last 3 months, list the medicines or treatments that you personally bought over-the-counter to treat your low back pain.

Aid/medicine bought		Cost (£)
i.	*bmed1*	*cmed1*
ii.	*bmed2*	*cmed2*

B Sample of data in a tabular database.

Subject ID	Pri1	Pri2	nPriv1	cPriv1	pmed1	nmed1	pmed2	nmed2	bmed1	cmed1	bmed2	cmed2
100	1	1	1	40	Ibuprofen	1						
101	0	5	0	0					Mattress	325		
102			3	120					Mattress	299	Walking stick	5.65

C The view of the healthcare resource-use data.

ID	FU	RP	Type	Text	Reason	Location	Unit	Quantity	Cost	Payer
100	13	13	GP		Any	PRI	Visit	1		PHS
100	13	13	Physio		LBP	COMM	Visit	1		PHS
100	13	13	Physio		LBP	PTE	Visit	1	40	IND
100	13	13	NSAID	Ibuprofen	LBP	N/A	Px	1		PHS
101	13	13	GP		Any	PRI	Visit	0		PHS
101	13	13	Physio		LBP	COMM	Visit	5		PHS
101	13	13	Physio		LBP	PTE	Visit	0		IND
101	13	13	Aid	Mattress	LBP		Item		325	IND
102	13	13	Physio		LBP	PTE	Visit	3	120	IND
102	13	13	Aid	Mattress	LBP		Item		299	IND
102	13	13	Aid	Walking stick	LBP		Item		5.65	IND

Figure 2 Sample healthcare resource-use data. (A) A simplified healthcare resource-use questionnaire. (B) Sample of healthcare resource-use data in a tabular format. (C) The view of the original source data generated and pivoted from the EAV/CR table. FU, follow-up; RP, recall period; GP, primary care doctor; Any, any reason; PRI, primary care clinic; PHS, public health service; Physio, physiotherapist; LBP, low back pain; COMM, community clinic; PTE, private clinic; IND, individual; NSAID, non-steroidal anti-inflammatory drugs; Px, prescription; Aid, aids and adaptations.

A Demographics

Object ID	Unique subject ID	Original Subject ID	Domain	Trial	Sex	Age
1	1000001	011	Demographics	Trial A	2	.
4	1000002	021	Demographics	Trial A	1	.
7	1000003	A1001	Demographics	Trial B	1	50
10	1000004	A1002	Demographics	Trial B	1	28
12	1000005	A1003	Demographics	Trial B	2	.

B EQ5D

Object ID	Unique subject ID	Original Subject ID	Domain	Trial	FU	EQ5D1	EQ5D2	EQ5D3	EQ5D4	EQ5D5	EQ5D6
2	1000001	011	EQ5D	Trial A	0	2	2	2	2	2	.
3	1000001	011	EQ5D	Trial A	4	1	1	2	2	1	74
5	1000002	021	EQ5D	Trial A	0	1	2	1	2	1	60
6	1000002	021	EQ5D	Trial A	4	.	.	.	.	.	45

C RMDQ

Object ID	Unique subject ID	Original Subject ID	Domain	Trial	FU	RMDQ
8	1000003	A1001	RMDQ	Trial B	0	13
9	1000003	A1001	RMDQ	Trial B	13	7
11	1000004	A1002	RMDQ	Trial B	13	9
13	1000005	A1003	RMDQ	Trial B	0	18
14	1000005	A1003	RMDQ	Trial B	13	10

Figure 3 Output of data from the (A) Demographics, (B) EuroQol five dimensions questionnaire (EQ5D) and (C) Roland Morris Disability Questionnaire (RMDQ) domains in long format tabular structure, based on the sample data shown in Fig. 1. FU, follow-up; EQ5D1, EQ5D2, EQ5D3, EQ5D4 and EQ5D5 are items 1–5 of EQ5D; EQ5D6, EuroQol visual analogue scale; and RMDQ, the sum of all ticked items from RMDQ.

The characteristics of the included studies are summarized in Supporting information Table S1. Initial examination of the data showed that no two trials studied identical interventions. Even the usual care arms of the included studies are likely to differ according to jurisdiction, site of recruitment and age of the study. To make meaningful comparisons we needed to broadly pool the interventions into groups our analyses.

Considering the potential mechanisms through which the participant characteristics might affect the outcome, we decided to pool interventions that might under other circumstances appear rather heterogeneous. In particular, the decision to include several superficially different interventions as 'passive physiotherapy' may surprise some readers. Our view, however, is that these are very distinctly different from active exercise based interventions, or those working primarily through a psychological approach. Essentially they all consist of an assessment, whatever reassurance and education is provided as part of the treatment session, plus whatever modality is being offered; be it massage/mobilization/manipulation or needling. We consider these to be conceptually sufficiently close in their mode of action that it is unlikely there will be distinctions in how the potential moderators included in our analyses might affect the outcomes. They are, however, distinctly different from active physical and psychological interventions in how treatment moderation might operate. The American Pain Society/American College of Physicians guidelines use a broadly similar approach to group non-pharmacological interventions (Chou and Huffman, 2007).

We first identified the control interventions and classified them as either usual care or sham control. We then split our active interventions into three broad categories; active physical (exercise and graded activity), passive physical (individual physiotherapy, manipulation and acupuncture) and psychological (advice/education and psychological therapy) (Patel et al., 2016). Fig 5 shows a network of the treatment subtypes from trials in the repository.

3.2 Repository database

Data in the repository were captured in the same granularity as the original data provided to us.
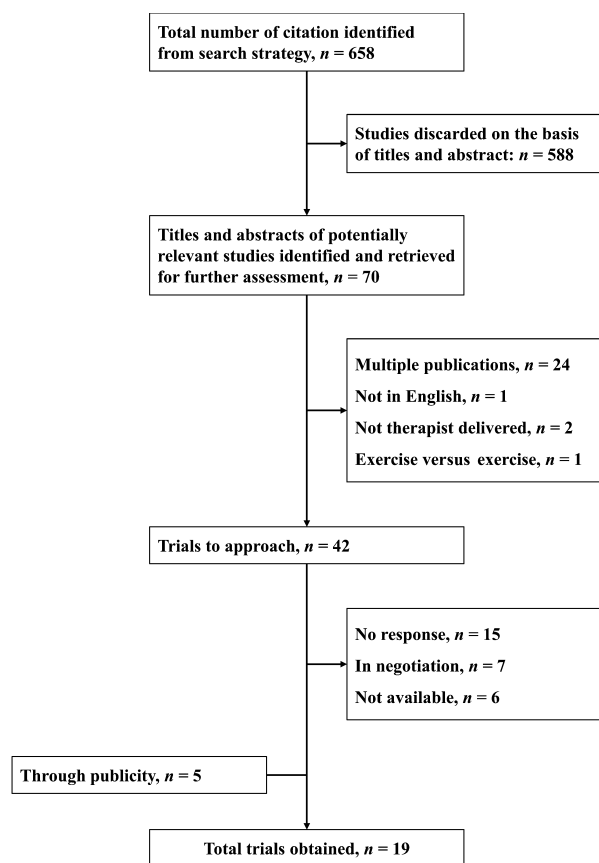
Figure 4 Identification of trials.

Where available, we included both the individual items and the summary score for each questionnaire (domain) at all possible time points when they were collected. This proves to be helpful when only certain items of a questionnaire may be of interest to future researchers. For example, the visual analogue scale (VAS) is usually a standalone instrument that asks the participant to describe their average or worst pain at the present time. The participant marks their pain level on the VAS line, which is visually presented as either a horizontal or vertical line with 'no pain' at one end and 'worst possible pain' at the other end. Similar questions are found in the chronic pain grade scale (CPG) which has two dimensions and one of them is the pain intensity scores (von Korff et al., 1992). Three items contribute to the pain intensity score. Each RCT is unique and no single questionnaire was used by all trials. Thus, future researchers will still be able to pool data from the VAS and the equivalent items from CPG pain intensity items if they were interested to analyse the effect of treatment on pain.

Table 1 presents the list of patient reported outcome measurements stored in the repository alongside demographics and medical history collected at baseline, and healthcare resource-use data at subsequent follow-ups.

All of the trials were able to provide information on sex and age. Other demographics information, such as ethnicity, smoking status, employment status and body mass index (BMI), were given by some trials but were not routinely collected by others (see, Table S2 in the Supporting information for a comprehensive summary). Seven RCTs across three countries provided healthcare resource-use data (Patel et al., 2016).

Other data fields in the repository included a variety of established patient reported outcomes on pain related physical disability (CPG disability, FFbHR, ODI, PDI, PSFS, RMDQ and Troublesome), pain (CPG pain intensity and VAS), and health related quality of life (SF-12/36 and EQ5D3L) (see Table 1 for list of abbreviations). Of note was there was no common instrument that was used by all of the trials (see Table 1).

Supporting information Table S2 shows the response rates for various instruments of interest by treatment group at each possible time point. Most of the RCTs collected outcomes at short- (within 2/3 months post-randomization) and mid-term (6 months post-randomization). Two trials collected outcomes beyond 12 months post-randomization. The follow-up time for trials that collected data 3 months post-randomization or entry to the trial was stored in the repository as 13 weeks. However, one RCT had specifically mentioned in their protocol to collect data at 12 weeks and thus this was stored in the repository, as per protocol (Patel et al., 2016).

4. Discussion

We created a purpose-built repository to store IPD from multiple RCTs and developed the software and procedures to morph heterogeneous datasets. This infrastructure was necessary for our primary objective, identifying treatment moderators for LBP.

The repository and ETL software was intended to provide a solution for standardizing and storing heterogeneous and sparse datasets and output into friendly formatted datasets. Creating the hybrid system based on a relational and EAV model was a relatively simple process. However, creating the bespoke software and standardizing processing solutions was more challenging as it was not possible to predefine classes and attributes until data sharing agreements
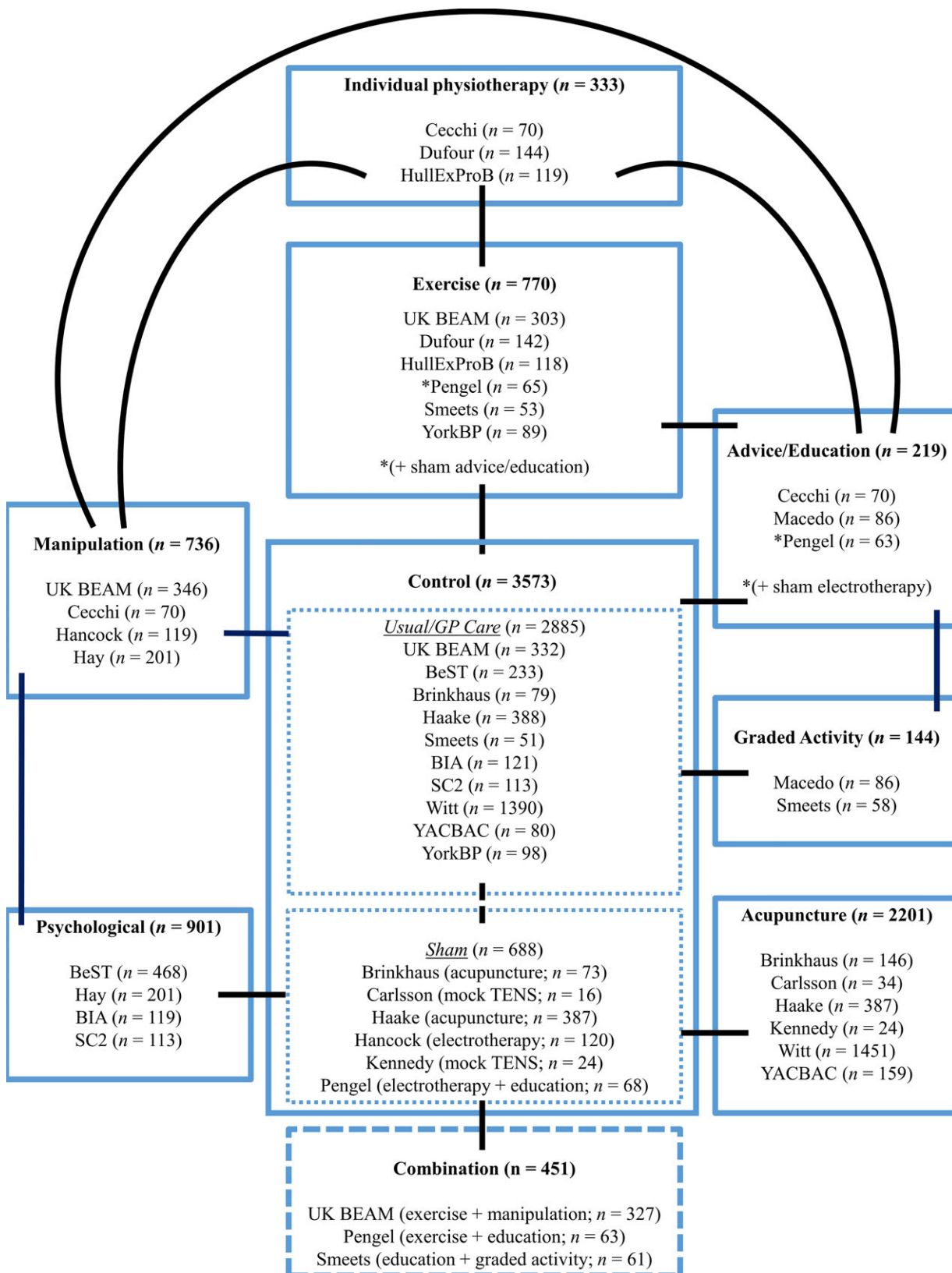
Figure 5 Summary of all included trials.

Table 1 List of patient reported outcome measurements and the number of datasets that collected such information.

Abbreviation	List of measurements	No. of datasets ($m = 20$)	Reference
ABPS	Aberdeen Back Pain Scale	2	Ruta et al., (1994)
ALBPSQ	Acute Low back Pain Screening Questionnaire	3	Linton and Halldén, (1998)
BBQ	Back Beliefs Questionnaire	2	Symonds et al., (1996)
BDI	Beck Depression Inventory	1	Beck et al., (1961, 1979)
CES-D	The Centre for Epidemiologic Studies Depression Scale	1	Radloff, (1977)
CPG	Chronic Pain Grade	6	von Korff et al., (1992)
CSQ	Coping Strategies Questionnaire	2	Rosenstiel and Keefe, (1983)
DASS-21	Depression Anxiety Stress Scales	1	Henry and Crawford, (2005)
DRAM	Distress and Risk Assessment Method	3	Main et al., (1992)
EQ5D3L	EQ-5D-3L	8	EuroQol Group, (1990)
FABQ	Fear-Avoidance Beliefs Questionnaire	6	Waddell et al., (1993)
FFbHR	Hannover Functional Ability Questionnaire for Measuring Back Pain-Related Functional Limitations (Funktionsbeeintrachtigung durch Ruckenschmerzen)	3	Kohlmann and Raspe, (1996)
HADS	Hospital Anxiety and Depression Scale	1	Snaith, (2003)
IPAQ	International Physical Activity Questionnaire	1	Craig et al., (2003)
LSI	Lumbar Spine Instability Questionnaire	1	Cook et al., (2006)
MPQ-SF	McGill Pain Questionnaire (short form)	2	Melzack, (1987)
MSPQ	Modified Somatic Perception Questionnaire	4	Main, (1983)
MZDI	Modified Zung Depression Index	4	Main et al., (1992)
ODI	Oswestry Disability Index	1	Fairbank et al., (1980)
VAS	Visual Analogue Scale (average/worst)	9	
PASS-20	Pain Anxiety Symptoms Scale	1	McCracken and Dhingra, (2002)
PDI	Pain Disability Index	1	Tait et al., (1990)
PRSS	Pain Related Self Statement	2	Flor et al., (1993)
PSEQ	Pain Self-Efficacy Questionnaire	4	Nicholas, (2007)
PSFS	Patient Specific Functional Scale	3	Stratford et al., (1995)
RMDQ	Roland-Morris Disability Questionnaire	15	Roland and Morris, (1983)
SES	Pain Experience Scale (Schmerzempfindungsskala)	1	Geissner, (1995)
SF-12	12-item Short Form Health Survey	3	Ware et al., (2002)
SF-36	36-item Short Form Health Survey	9	Ware et al., (2000)
TSK	Tampa Scale for Kinesiophobia	5	Vlaeyen et al., (1995)
Troublesome	Troublesomeness	5	Parsons et al., (2006)
WHODAS 2.0	World Health Organization Disability Assessment Schedule 2.0	1	Üstün et al., (2010)

were in place. Our approach avoided the need to pre-specify all variables of interest and follow-up time points before constructing the database. It also allows us to easily add new datasets in a relatively straightforward manner as they become available.

Seven RCTs across three countries provided healthcare resource-use data. The lack of standardization in the recording of healthcare resource-use items for health economic analyses increased the complexity in harmonizing data from multiple sources. Nevertheless, by splitting each healthcare resource-use item into eight distinct parts, we were able to standardize and harmonize the data from different centres and countries.

Two major strengths of our database are the rigour with which the data were checked and cleaned, and its arrangement, which will allow more

trials and variables to be added easily. By including mainly trials with >179 participants, we have included higher quality trials. We have not, however, sought to include all trials on an individual intervention. The database is not currently suitable for comparing the main effects of different treatments. Additional work is needed to identify and upload smaller trials that would contribute to any such analysis and an updated search for relevant studies is also needed.

We were limited by the nature of the available trial data. The heterogeneous nature of the interventions, which were often poorly characterized, the heterogeneous populations studied, and the variable choices in outcome selection will limit the comparisons that can be made using the current database. This in marked contrast to an IPD database of a drug

intervention used on a well-defined population with a clear outcome, e.g. death. In the long-term, researchers should use intervention taxonomies when reporting their trials to allow for better replication and pooling of trials (Schulz et al., 2010). Notwithstanding these challenges this database is an important step forward in back pain research.

The research community is in favour of having a central repository of data collected for IPD meta-analyses (Tudur Smith et al., 2014). They alluded to the many advantages of standardizing, safeguarding and storing such data centrally. There are other researchers who are attempting to create secure central databases of previously collected IPD but their work is ongoing (Tudur Smith et al., 2011). Our pooled dataset, with standardized and harmonized data, will provide an excellent resource for back pain researchers. Although primarily developed for our own work on the identification of sub-groups, there are many potential uses to which such a database could be used, e.g. secondary analyses. However, any use is dependent on the agreement of the data custodians who donated the original datasets.

Acknowledgements

Collaborators

Author contributions

References

Beck, A.T., Ward, C.H., Mendelson, M., Mock, J., Erbaugh, J. (1961). An inventory for measuring depression. *Arch Gen Psychiatr* 4, 561–571.

Beck, A.T., Rush, A.J., Shaw, B.F., Emery, G. (1979). *Cognitive Therapy of Depression.* (New York: Guilford Press).

Brandt, C.A., Morse, R., Matthews, K., Sun, K., Deshpande, A.M., Gadagkar, R., Cohen, D.B., Miller, P.L., Nadkarni, P.M. (2002). Metadata-driven creation of data marts from an EAV-modeled clinical research database. *Int J Med Inform* 65, 225–241.

Chou, R., Huffman, L.H. (2007). Nonpharmacologic therapies for acute and chronic low back pain: A review of the evidence for an American Pain Society/American College of Physicians Clinical Practice Guideline. *Ann Intern Med* 147, 492–504.

Cook, C., Brismée, J.-M., Sizer, P.S. Jr (2006). Subjective and objective descriptors of clinical lumbar spine instability: A Delphi study. *Man Ther* 11, 11–21.

Craig, C.L., Marshall, A.L., Sjöström, M., Bauman, A.E., Booth, M.L., Ainsworth, B.E., Pratt, M., Ekelund, U., Yngve, A., Sallis, J.F., Oja, P. (2003). International physical activity questionnaire: 12-country reliability and validity. *Med Sci Sports Exerc* 35, 1381–1395.

EuroQol Group (1990). EuroQol – a new facility for the measurement of health-related quality of life. *Health Policy* 16, 199–208.

Fairbank, J., Couper, J., Davies, J., O'brien J. (1980). The Oswestry low back pain disability questionnaire. *Physiotherapy* 66, 271–273.

Flor, H., Behle, D.J., Birbaumer, N. (1993). Assessment of pain-related cognitions in chronic pain patients. *Behav Res Ther* 31, 63–73.

Geissner, E. (1995). [The Pain Perception Scale–a differentiated and change-sensitive scale for assessing chronic and acute pain]. *Die Rehabilitation* 34, XXXV–XLIII.

Global Burden of Disease Study 2013 Collaborators (2015). Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990–2013: A systematic analysis for the Global Burden of Disease Study 2013. *Lancet* 386, 743–800.

Gurung, T., Ellard, D., Mistry, D., Patel, S., Underwood, M. (2015). Identifying potential moderators for response to treatment in low back pain: A systematic review. *Physiotherapy* 101, 243–251.

Henry, J.D., Crawford, J.R. (2005). The short-form version of the Depression Anxiety Stress Scales (DASS-21): Construct validity and normative data in a large non-clinical sample. *Br J Clin Psychol* 44, 227–239.

Hoy, D., March, L., Brooks, P., Blyth, F., Woolf, A., Bain, C., Williams, G., Smith, E., Vos, T., Barendregt, J., Murray, C., Burstein, R., Buchbinder, R. (2014). The global burden of low back pain: Estimates from the Global Burden of Disease 2010 study. *Ann Rheum Dis* 73, 968–974.

Kohlmann, T. and Raspe, H. (1996). Hannover Functional Questionnaire in ambulatory diagnosis of functional disability caused by backache. *Die Rehabilitation* 35, I–VIII.

von Korff, M., Ormel, J., Keefe, F.J., Dworkin, S.F. (1992). Grading the severity of chronic pain. *Pain* 50, 133–149.

Linton, S.J., Halldén, K. (1998). Can we screen for problematic back pain? A screening questionnaire for predicting outcome in acute and subacute back pain. *Clin J Pain* 14, 209–215.

Main, C.J. (1983). The modified somatic perception questionnaire (MSPQ). *J Psychosom Res* 27, 503–514.

Main, C.J., Wood, P.L.R., Hollis, S., Spanswick, C.C., Waddell, G. (1992). The distress and risk assessment method: A simple patient classification to identify distress and evaluate the risk of poor outcome. *Spine* 17, 42–52.

McCracken, L.M., Dhingra, L. (2002). A short version of the Pain Anxiety Symptoms Scale (PASS-20): Preliminary development and validity. *Pain Res Manage* 7, 45–50.

Melzack, R. (1987). The short-form McGill pain questionnaire. *Pain* 30, 191–197.

Mistry, D., Patel, S., Hee, S.W., Stallard, N., Underwood, M. (2014). Evaluating the quality of subgroup analyses in randomized controlled trials of therapist-delivered interventions for nonspecific low back pain: A systematic review. *Spine* 39, 618–629.

Nadkarni, P.M., Marenco, L., Chen, R., Skoufos, E., Shepherd, G., Miller, P. (1999). Organization of heterogeneous scientific data using the EAV/CR representation. *J Am Med Inform Assoc* 6, 478–493.

Nicholas, M.K. (2007). The pain self-efficacy questionnaire: Taking pain into account. *Eur J Pain* 11, 153–163.

Parsons, S., Carnes, D., Pincus, T., Foster, N., Breen, A., Vogel, S., Underwood, M. (2006). Measuring troublesomeness of chronic pain by location. *BMC Musculoskelet Disord* 7, 34.

Patel, S., Hee, S.W., Mistry, D., Jordan, J., Brown, S., Dritsaki, M., Ellard, D., Friede, T., Lamb, S.E., Lord, J., Madan, J., Morris, T., Stallard, N., Tysall, C., Willis, A., Underwood, M. (2016). Identifying back pain subgroups: developing and applying approaches using individual patient data collected within clinical trials. Programme Grants for Applied Research 4.

Radloff, L.S. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Appl Psychol Meas* 1, 385–401.

Roland, M., Morris, R. (1983). A study of the natural history of back pain: Part I: Development of a reliable and sensitive measure of disability in low-back pain. *Spine* 8, 141–144.

Rosenstiel, A.K., Keefe, F.J. (1983). The use of coping strategies in chronic low back pain patients: Relationship to patient characteristics and current adjustment. *Pain* 17, 33–44.

Ruta, D.A., Garratt, A.M., Wardlaw, D., Russell, I.T. (1994). Developing a valid and reliable measure of health outcome for patients with low back pain. *Spine* 19, 1887–1896.

Savigny, P., Watson, P., Underwood, M. (2009). Early management of persistent non-specific low back pain: Summary of NICE guidance. *BMJ* 338, b1805.

Schulz, R., Czaja, S.J., McKay, J.R., Ory, M.G., Belle, S.H. (2010). Intervention Taxonomy (ITAX): Describing essential features of interventions. *Am J Health Behav* 34, 811–821.

Snaith, R.P. (2003). The hospital anxiety and depression scale. *Health Qual Life Outcomes* 1, 29.

Stratford, P., Gill, C., Westaway, M., Binkley, J. (1995). Assessing disability and change on individual patients: A report of a patient specific measure. *Physiother Can* 47, 258–263.

Symonds, T.L., Burton, A.K., Tillotson, K.M., Main, C.J. (1996). Do attitudes and beliefs influence work loss due to low back trouble? *Occup Med* 46, 25–32.

Tait, R.C., Chibnall, J.T., Krause, S. (1990). The pain disability index: Psychometric properties. *Pain* 40, 171–182.

Tudur Smith, C., Dwan, K., Clarke, M., Riley, R., Altman, D., Williamson, P. (2011). Feasibility of establishing a central repository for the individual participant data from research studies. *Trials* 12, A56.

Tudur Smith, C., Dwan, K., Altman, D.G., Clarke, M., Riley, R., Williamson, P.R. (2014). Sharing individual participant data from clinical trials: An opinion survey regarding the establishment of a central repository. *PLoS ONE* 9, e97886.

Üstün, T.B., Kostanjsek, N., Chatterji, S., Rehm, J. (2010). *Measuring health and disability: Manual for WHO Disability Assessment Schedule (WHODAS 2.0)*. (Geneva: World Health Organization).

Vlaeyen, J.W.S., Kole-Snijders, A.M.J., Boeren, R.G.B., van Eek, H. (1995). Fear of movement/(re)injury in chronic low back pain and its relation to behavioral performance. *Pain* 62, 363–372.

Waddell, G., Newton, M., Henderson, I., Somerville, D., Main, C.J. (1993). A Fear-Avoidance Beliefs Questionnaire (FABQ) and the role of fear-avoidance beliefs in chronic low back pain and disability. *Pain* 52, 157–168.

Ware, J.E., Kosinski, M., Dewey, J.E. (2000). *How to score version 2 of the SF-36 health survey*. (Lincoln, RI: QualityMetric Incorporated.

Ware, J.E., Kosinski, M., Turner-Bowker, D.M., Gandek, B. (2002). *How to score version 2 of the SF-12 health survey (with a supplement documenting version 1)*. (Lincoln, RI: QualityMetric Incorporated).

Supporting Information

Additional Supporting Information may be found online in the supporting information tab for this article:

Table S1. Summary of the trials in the low back pain standard repository.

Table S2. Comprehensive summary of the number of trials (m) and participants (n) for each outcome by follow-up time point and treatment arm.

Method S1. Technical details of the bespoke hybrid database.

Figure S1. Entity-relationship diagram, depicting the fixed schema with the sub-scheme entity-attribute-value (EAV) tables.

Figure S2. XML instructions to map the original clinical data to the equivalent repository attributes and transform the original values to the repository standard.

Figure S3. XML instructions to map the original healthcare resource-use data to the equivalent repository attributes and to transform the original values to the repository standard.

Figure S4. Sample of XML instructions to map more than one time point of the original healthcare resource-use data to the equivalent repository attributes and standard.