DATA SHARING

CATRIN TUDUR SMITH, UNIVERSITY OF LIVERPOOL

CAT1@LIV.AC.UK

INDIVIDUAL PARTICIPANT DATA (IPD)

Patient Number	Treatment	Survival Time (Days)	Status	Age	Sex	Stage
1	E	44	Dead	67	m	IV
2	E	54	Dead	64	m	III
3	E	67	Alive	55	f	III
4	C	43	Dead	79	f	IV
5	C	70	Alive	62	m	IV
6	E	88	Dead	60	f	IV
7	C	99	Alive	57	m	III
8	C	45	Dead	66	m	III
9	E	90	Alive	59	f	III
10	C	23	Dead	53	m	IV

BENEFITS OF SHARING DATA

- encourages scientific enquiry and debate
- promotes innovation and potential new data uses
- leads to new collaborations between data users and data creators
- maximises transparency and accountability
- enables scrutiny of research findings
- encourages the improvement and validation of research methods
- reduces the cost of duplicating data collection
- increases the impact and visibility of research
- promotes the research that created the data and its outcomes
- can provide a direct credit to the researcher as a research output in its own right
- provides important resources for education and training

MANAGING AND SHARING DATA
UK Data archive

BENEFITS OF DATA SHARING – CASE STUDY

- IPD collected from randomised controlled trials of anti-epileptic drugs (AEDs) as monotherapy
- Data obtained for 6418 patients from 20 trials comparing 8 AEDs
- Updated to 12745 patients from 36 trials comparing 10 AEDs

1. IMPROVE VALIDITY AND RELIABILITY OF META-ANALYSIS

- Overcome poor reporting and selective reporting biases
- Standardise outcome definition across studies
- More appropriate analysis
- Incorporate additional follow-up data
- Re-instate excluded patients

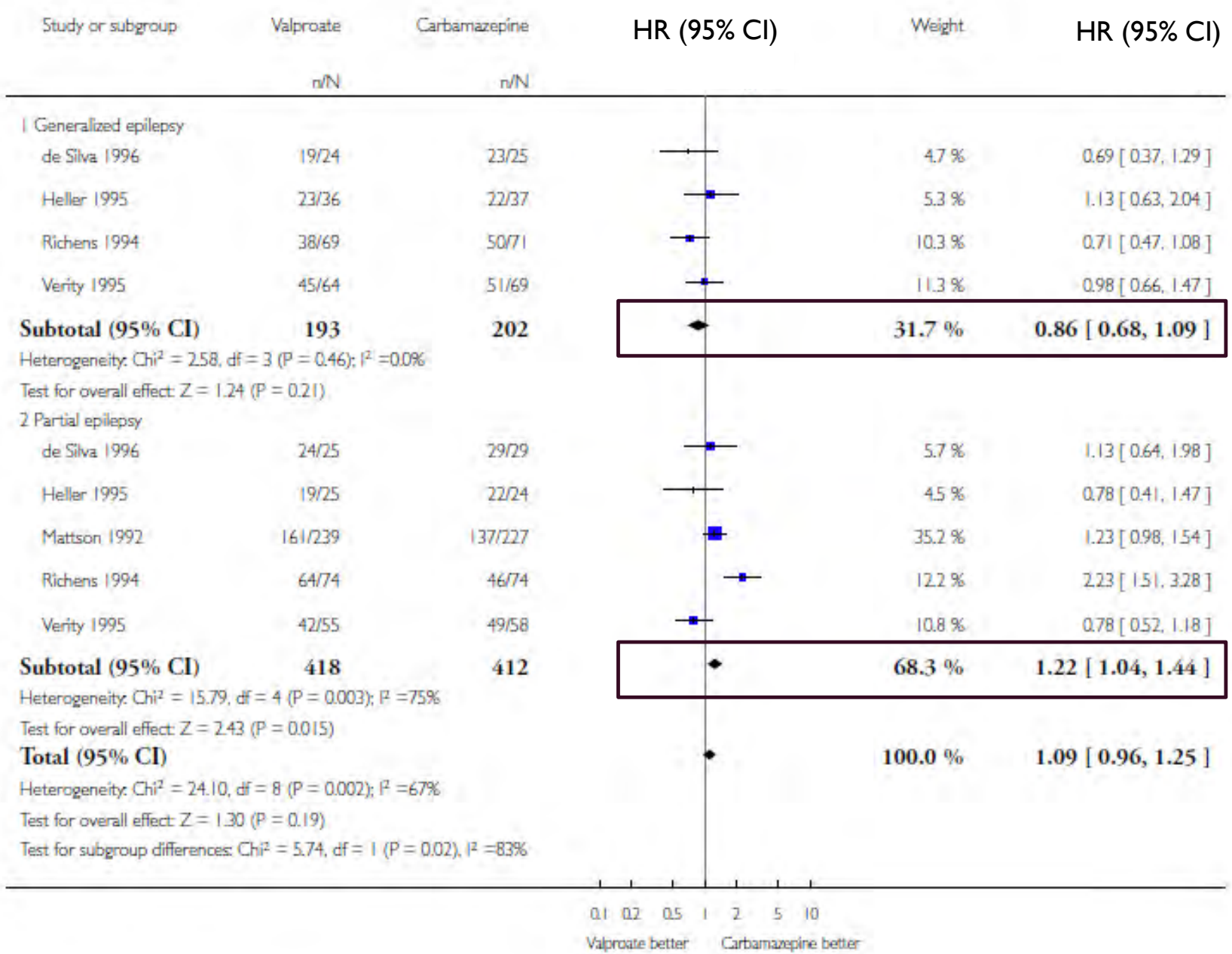
1. IMPROVE VALIDITY AND RELIABILITY OF META-ANALYSIS

- Nolan SJ, Marson AG, Weston J, Tudur Smith C. Carbamazepine versus phenobarbitone monotherapy for epilepsy: an individual participant data review. ***The Cochrane database of systematic reviews***. 2015 Jul 23;7:CD001904. **[1]**
- Nolan SJ, Marson AG, Weston J, Tudur Smith C. Carbamazepine versus phenytoin monotherapy for epilepsy: an individual participant data review. ***The Cochrane database of systematic reviews***. 2015 Aug 14;8:CD001911. **[3]**
- Nolan SJ, Marson AG, Weston J, Tudur Smith C. Phenytoin versus valproate monotherapy for partial onset seizures and generalised onset tonic-clonic seizures: an individual participant data review. ***The Cochrane database of systematic reviews***. 2016;4:CD001769.
- Nolan, SJ, Muller, M, Tudur Smith, C and Marson, AG (2013). Oxcarbazepine versus phenytoin monotherapy for epilepsy. ***The Cochrane database of systematic reviews*** **5**: CD003615. **[4]**
- Nolan, SJ, Tudur Smith, C, Pulman, J and Marson, AG (2013). Phenobarbitone versus phenytoin monotherapy for partial onset seizures and generalised onset tonic-clonic seizures. ***The Cochrane database of systematic reviews*** **1**: CD002217. **[10]**
- Tudur Smith, C, Marson, AG and Williamson, PR (2003). Carbamazepine versus phenobarbitone monotherapy for epilepsy. ***The Cochrane database of systematic reviews*** (1): CD001904. **[71]**
- Tudur Smith, C, Marson, AG, Clough, HE and Williamson, PR (2002). Carbamazepine versus phenytoin monotherapy for epilepsy. ***The Cochrane database of systematic reviews*** (2): CD001911. **[64]**
- Taylor, S, Tudur Smith, C, Williamson, PR and Marson, AG (2001). Phenobarbitone versus phenytoin monotherapy for partial onset seizures and generalized onset tonic-clonic seizures. ***The Cochrane database of systematic reviews*** (4): CD002217. **[67]**
- Tudur Smith, C, Marson, AG and Williamson, PR (2001). Phenytoin versus valproate monotherapy for partial onset seizures and generalized onset tonic-clonic seizures. ***The Cochrane database of systematic reviews*** (4): CD001769. **[65]**

2. TREATMENT EFFECT MODIFICATION

- Explore participant level covariates' influence on treatment effect
- Increased power to identify patients that may benefit most

Outcome: 3 Time to first seizure post randomization



3. ADDRESS NEW CLINICAL QUESTIONS

BMJ

RESEARCH

BMJ 2010;341:c6477

Risk of recurrence after a first seizure and implications for driving: further analysis of the Multicentre study of early Epilepsy and Single Seizures

Bonnett LJ, Tudur-Smith C, Williamson PR, Marson AG

Treatment outcome after failure of a first antiepileptic drug

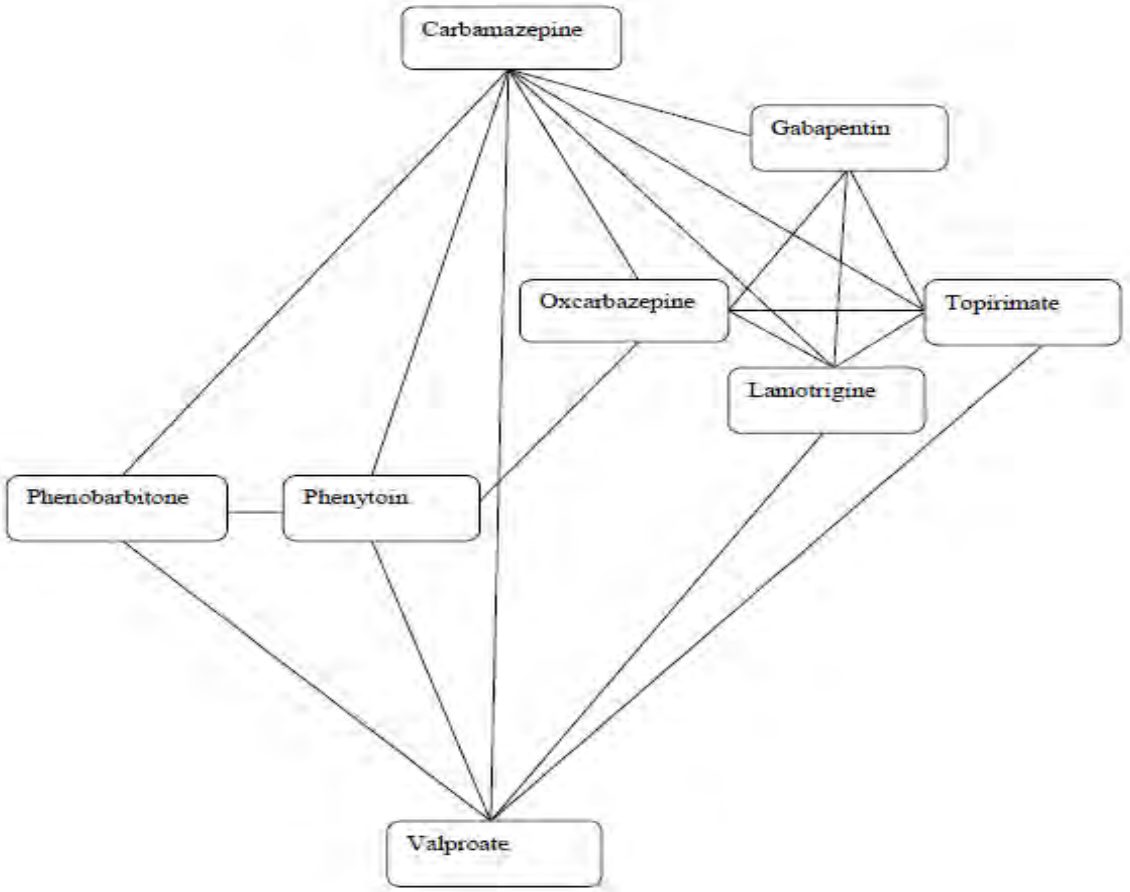
Neurology® 2014;83:552-560

Seizure recurrence after antiepileptic drug withdrawal and the implications for driving: further results from the MRC Antiepileptic Drug Withdrawal Study and a systematic review

Bonnett LJ, Shukralla A, Tudur-Smith C, *et al*. *J Neurol Neurosurg Psychiatry* (2010).

3. ADDRESS NEW CLINICAL QUESTIONS

28 possible pair-wise comparisons



20 trials
6418 patients
8 AEDS

Trials

BioMed Central

Research

Open Access

Multiple treatment comparisons in epilepsy monotherapy trials

Catrin Tudur Smith*[1], Anthony G Marson[2], David W Chadwick[2] and Paula R Williamson1[1]

Trials 2007, **8**:34

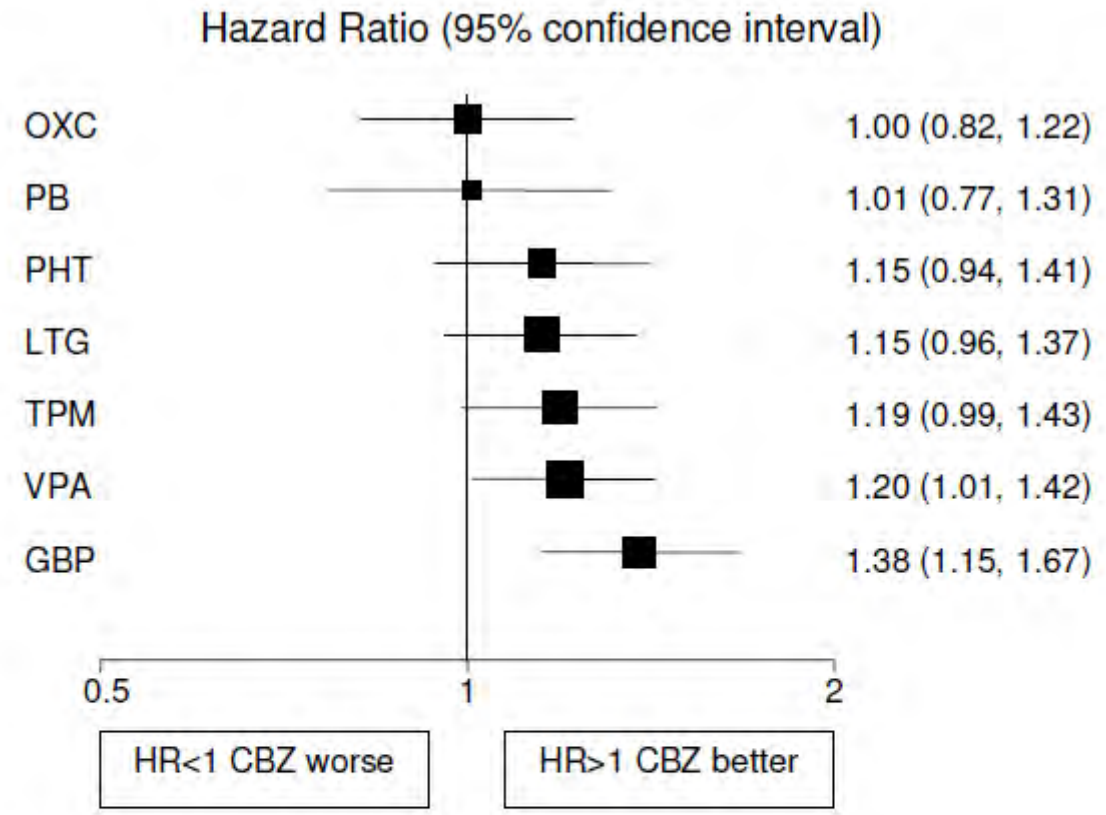


Figure 2
Time to 12 month remission for partial onset seizures (Hazard Ratio for each AED compared to standard CBZ). CBZ: Carbamazepine, **VPA**: Sodium Valproate, **PHT**: Phenytoin, **PB**: Phenobarbitone, **LTG**: Lamotrigine, **OXC**: Oxcarbazepine, **GBP**: Gabapentine, **TPM**: Topirimate

4. INFORM DESIGN OF NEW TRIAL

The SANAD study of effectiveness of carbamazepine, gabapentin, lamotrigine, oxcarbazepine, or topiramate for treatment of partial epilepsy: an unblinded randomised controlled trial

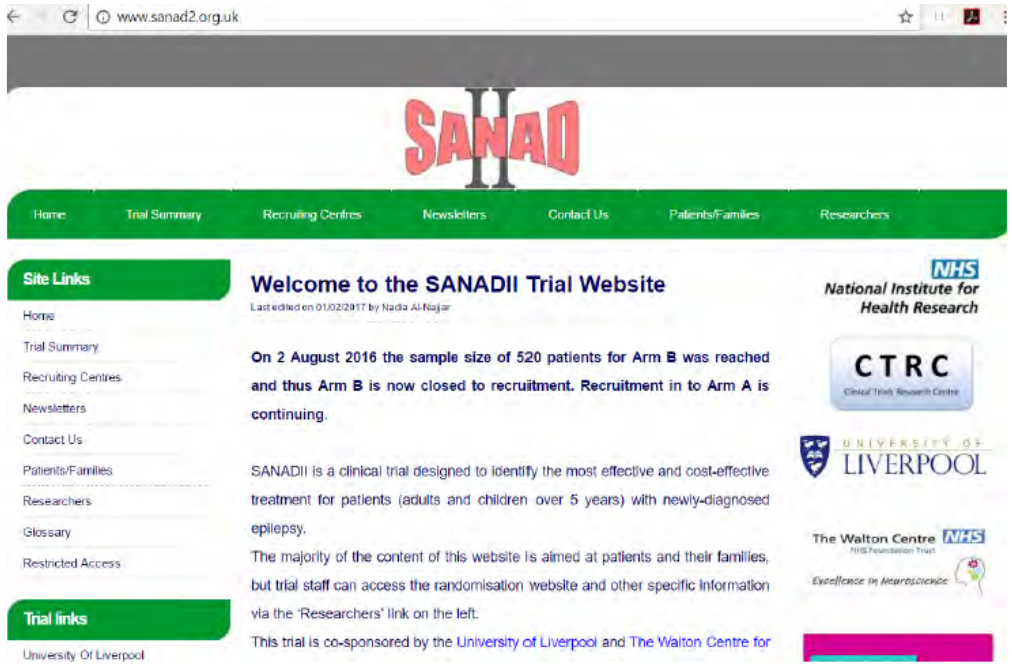
Anthony G Marson, Asya M Al-Kharusi, Muna Alwaidh, Richard Appleton, Gus A Baker, David W Chadwick, Celia Cramp, Oliver C Cockerell, Paul N Cooper, Julie Doughty, Barbara Eaton, Carrol Gamble, Peter J Goulding, Stephen J L Howell, Adrian Hughes, Margaret Jackson, Ann Jacoby, Mark Kellett, Geoffrey R Lawson, John Paul Leach, Paola Nicolaides, Richard Roberts, Phil Shackley, Jing Shen, David F Smith, Philip E M Smith, Catrin Tudur Smith, Alessandra Vanoli, Paula R Williamson, on behalf of the SANAD Study group.

Lancet 2007;369:1000-1015

The SANAD study of effectiveness of valproate, lamotrigine, or topiramate for generalised and unclassifiable epilepsy: an unblinded randomised controlled trial

Anthony G Marson, Asya M Al-Kharusi, Muna Alwaidh, Richard Appleton, Gus A Baker, David W Chadwick, Celia Cramp, Oliver C Cockerell, Paul N Cooper, Julie Doughty, Barbara Eaton, Carrol Gamble, Peter J Goulding, Stephen J L Howell, Adrian Hughes, Margaret Jackson, Ann Jacoby, Mark Kellett, Geoffrey R Lawson, John Paul Leach, Paola Nicolaides, Richard Roberts, Phil Shackley, Jing Shen, David F Smith, Philip E M Smith, Catrin Tudur Smith, Alessandra Vanoli, Paula R Williamson, on behalf of the SANAD Study group

Lancet 2007;369:1016-1026



5. DEVELOP AND APPLY NEW METHODOLOGY

- Better approach to analysis
- Improve future methods

STATISTICS IN MEDICINE
Statist. Med. 2005; **24**:1307–1319
Published online 31 January 2005 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/sim.2050

Investigating heterogeneity in an individual patient data meta-analysis of time to event outcomes

Catrin Tudur Smith[1,*,†], Paula R. Williamson[1] and Anthony G. Marson[2]

ENCOURAGING DATA SHARING

INSTITUTE OF MEDICINE
OF THE NATIONAL ACADEMIES

Advising the Nation. Improving Health.

EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

The BMJ **requires data sharing on request for all trials**

ACCESSING IPD FROM CLINICAL TRIALS

13 sponsors
>2100 studies

2 sponsors
> 100 studies

SHARING IPD FROM PUBLICLY FUNDED TRIALS

- 60% of trials registered on clinicaltrials.gov are non-industry sponsored
- Some examples of proactive sharing


Sandercock et al. Trials 2011, 12:101
http://www.trialsjournal.com/content/12/1/101

TRIALS

RESEARCH Open Access

The International Stroke Trial database

Peter AG Sandercock[1*], Maciej Niewada[2,3], Anna Członkowska[2,3] and for the International Stroke Trial Collaborative Group

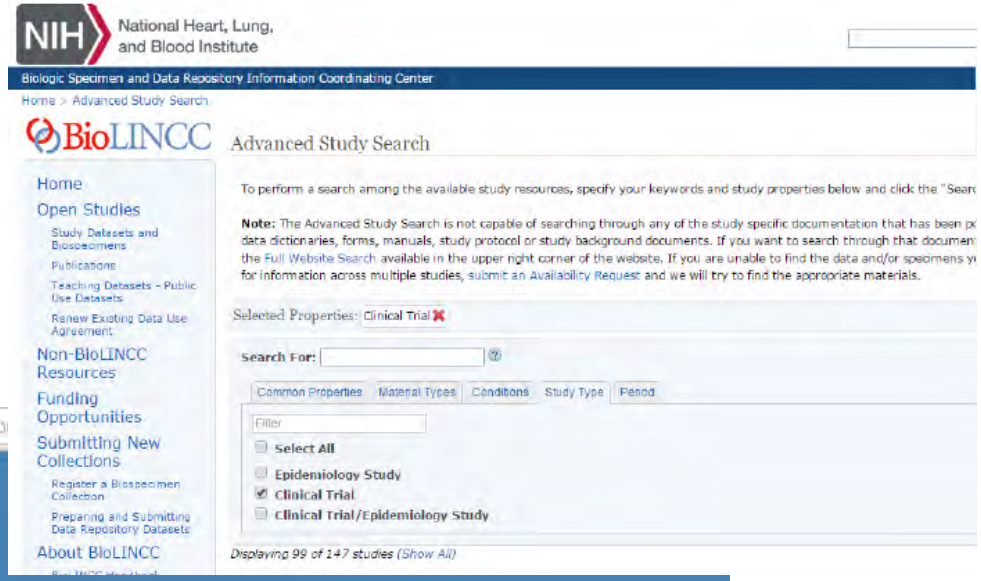
Abstract

Background: We aimed to make individual patient data from the International Stroke Trial (IST), o randomised trials ever conducted in acute stroke, available for public use, to facilitate the planning and to permit additional secondary analyses.

Methods: For each randomised patient, we have extracted data on the variables assessed at rand early outcome point (14-days after randomisation or prior discharge) and at 6-months and provid analysable database.

Results: The IST dataset includes data on 19 435 patients with acute stroke, with 99% complete f 26.4% patients were aged over 80 years at study entry. Background stroke care was limited and n patients received thrombolytic therapy.

Conclusions: The IST dataset provides a source of primary data which could be used for planning sample size calculations and for novel secondary analyses. Given the age distribution and nature background treatment given, the data may be of value in planning trials in older patients and in settings.


NIH National Heart, Lung, and Blood Institute

Biologic Specimen and Data Repository Information Coordinating Center

Home > Advanced Study Search

BioLINCC Advanced Study Search

Home
Open Studies
Study Datasets and Biospecimens
Publications
Teaching Datasets - Public Use Datasets
Renew Existing Data Use Agreement
Non-BioLINCC Resources
Funding Opportunities
Submitting New Collections
Register a Biospecimen Collection
Preparing and Submitting Data Repository Datasets
About BioLINCC

To perform a search among the available study resources, specify your keywords and study properties below and click the "Searc

Note: The Advanced Study Search is not capable of searching through any of the study specific documentation that has been p data dictionaries, forms, manuals, study protocol or study background documents. If you want to search through that documen the Full Website Search available in the upper right corner of the website. If you are unable to find the data and/or specimens y for information across multiple studies, submit an Availability Request and we will try to find the appropriate materials.

Selected Properties: Clinical Trial

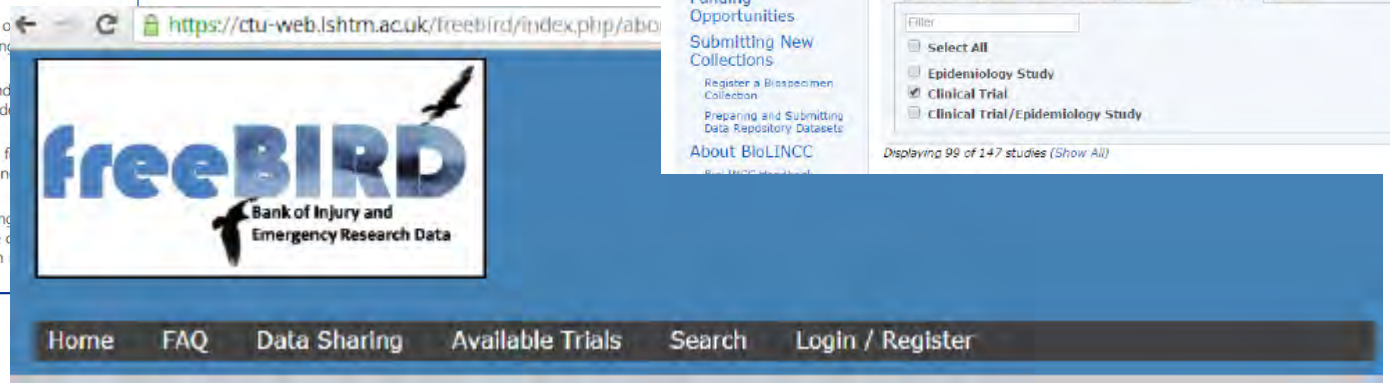
Search For:

Common Properties | Material Types | Conditions | Study Type | Period

Filter

Select All
Epidemiology Study
Clinical Trial
Clinical Trial/Epidemiology Study

Displaying 99 of 147 studies (Show All)


https://ctu-web.lshtm.ac.uk/freebird/index.php/abo

freeBIRD
Bank of Injury and Emergency Research Data

Home FAQ Data Sharing Available Trials Search Login / Register

More needs to be done to share data, particularly from publicly funded trials

BMJ 2013;347:f6927 doi: 10.1136/bmj.f6927 (Published 2 December 2013)

VIEWS & REVIEWS

OPEN DATA CAMPAIGN

Why did it take 19 months to retrieve clinical trial data from a non-profit organisation?

Asbjørn Hróbjartsson *The Nordic Cochrane Centre, Copenhagen, Denm*

Jaspers and Degraeuwe *Systematic Reviews* 2014, 3:97
http://www.systematicreviewsjournal.com/content/3/1/97

SYSTEMATIC REVIEWS

LETTER Open Access

A failed attempt to conduct an individual patient data meta-analysis

Gerald J Jaspers and Pieter LJ Degraeuwe*

CTU SURVEY

ELSEVIER

Journal of Clinical Epidemiology

Journal of Clinical Epidemiology ■ (2015) ■

ORIGINAL ARTICLE

UK publicly funded Clinical Trials Units supported a controlled access approach to share individual participant data but highlighted concerns

Carolyn Hopkins[a], Matthew Sydes[b], Gordon Murray[c], Kerry Woolfall[d], Mike Clarke[e], Paula Williamson[a], Catrin Tudur Smith[a,*]

UK Clinical Research Collaboration network of 45 CTUs

MRC | Hubs for Trials Methodology Research

23 (51%) CTUs responded to the survey

CTU SURVEY

- 30% had made a request for IPD in the last 12 months
- 65% had received a request for IPD in the last 12 months
- Most common reason was for meta-analysis

- 8 (35%) CTUs indicated that consent is generally sought for data to be used outside the original scope of trials
- A standard phrase was not being used across the CTUs

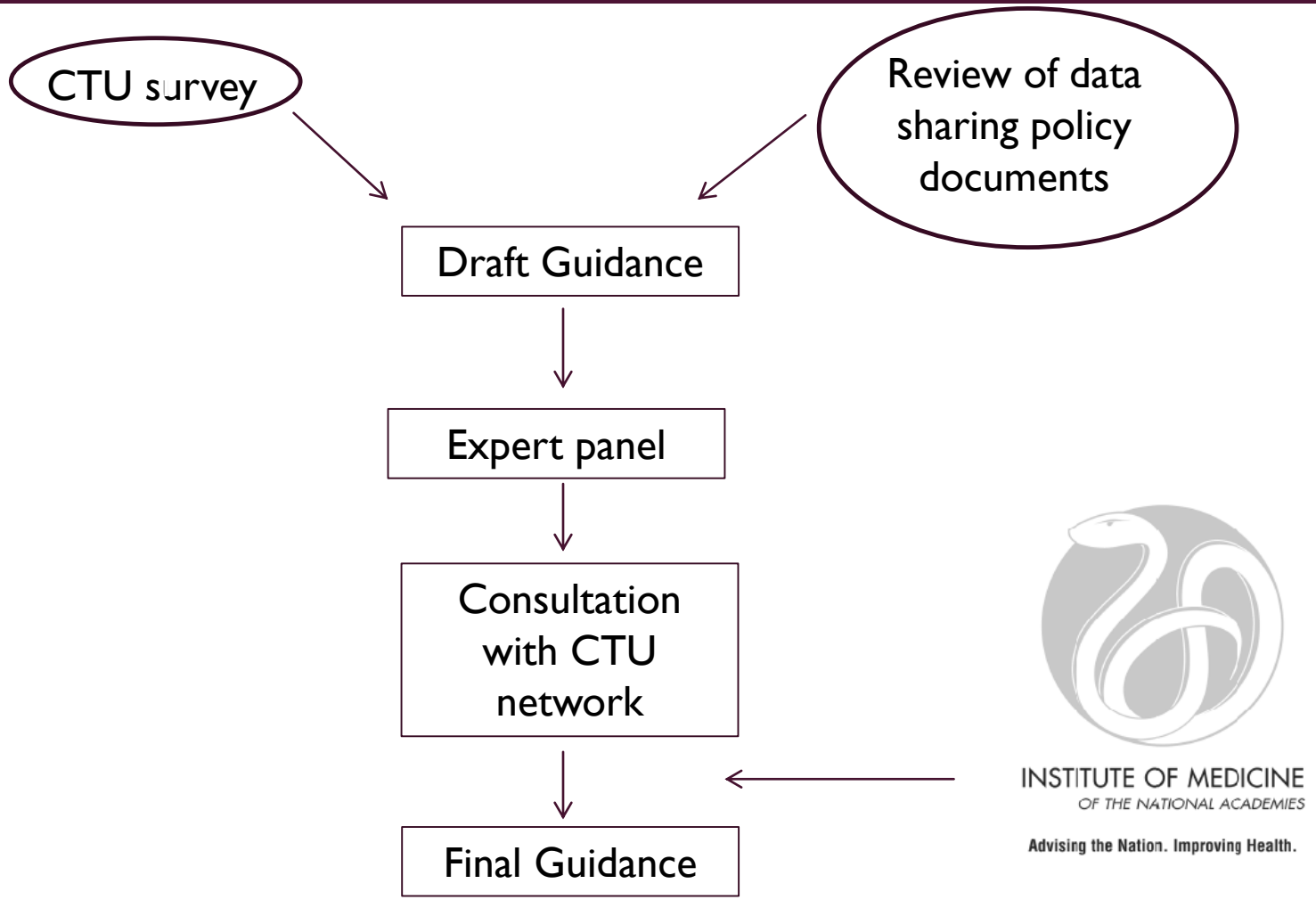
FUTURE DATA SHARING

- 70% would transfer their clinical trial data to a central facility (with restrictions/conditions)

- Approval process preference
  - Open access 0
  - Learned intermediary 25%
  - Internal Review process 75%

- Concerns: Misuse of data, resources, loss of IP

GUIDANCE


CTU survey
Review of data sharing policy documents
Draft Guidance
Expert panel
Consultation with CTU network
Final Guidance
INSTITUTE OF MEDICINE
OF THE NATIONAL ACADEMIES
Advising the Nation. Improving Health.

GUIDANCE

New Data Sharing Guidance document published

15 May 2015

The NIHR is pleased to support a new data sharing guide for publicly funded clinical trials.

Data sharing has many advantages that benefit patients and advance clinical research. In addition, it ensures that all research data can be utilised to the maximum, and helps prevent duplication and waste.

The guidance document summarises good clinical practice principles for publicly funded CTUs to follow, when sharing individual participant data from clinical trials.

Tudur Smith C, Hopkins C, Sydes MR, Woolfall K, Clarke M, Murray G, Williamson PR. **How should individual participant data (IPD) from publicly funded clinical trials be shared?** BMC Medicine. 2015;13(1):1-7

Group of the UK CRC Registered CTUs Network. The National Institute for Health Research (NIHR) has confirmed it is supportive of the application of this guidance.

http://www.network-hubs.org.uk/files/7114/3682/3831/Datasharingguidance2015.pdf

EXAMPLES OF GOOD PRACTICES

- A CTU data sharing policy (scope of trials, data request process, data release process)
- Data use agreement important
- Identify roles and responsibilities for data sharing
- Include plans for data sharing in the protocol and data management plan
- Include a data sharing statement in the consent form and patient information leaflet

  I understand that the information collected about me will be used to support other research in the future, and may be shared anonymously with other researchers (HRA consent form template)

- Prepare 'data pack' ready for sharing (De-identified datasets, protocol with amendments, blank CRFs, dataset specifications (or annotated CRFs) including data variable amendments

Tudur Smith C, Hopkins C, et al. BMC Med. 2015 Dec 17;13:298.

DE-IDENTIFICATION

- Process of turning data into a form which does not identify individuals and where identification is not likely to take place

DE-IDENTIFICATION METHODS

- **Remove direct identifiers from a dataset**
  - eg *Remove respondents' names or replace with a code. Remove addresses, postcode information, institution and telephone numbers.*

- **Aggregate or reduce the precision of a variable**
  - *eg Record the year of birth rather than the day, month and year; record postcode sectors (first 3 or 4 digits) rather than full postcodes*

- **Restrict the upper or lower ranges of a continuous variable to hide outliers** if the values for certain individuals are unusual or atypical within the wider group researched.
  - eg *Annual salary could be 'top-coded' to avoid identifying highly paid individuals*

Taken from the UK Data Archive
http://www.data-archive.ac.uk/create-manage/consent-ethics/anonymisation?index=1

BEST PRACTICES FOR ANONYMISATION OF QUALITATIVE DATA

- do not collect disclosive data unless this is necessary, e.g. do not ask for full names if they cannot be used in the data
- plan anonymisation at the time of transcription or initial write up
- use pseudonyms or replacements that are consistent within the research team and throughout the project, e.g. use the same pseudonyms in publications or follow-up research
- identify replacements in text clearly, e.g. with [brackets]
- retain unedited versions of data for use within the research team and for preservation
- create an anonymisation log of all replacements, aggregations or removals made - store such a log separately from the anonymised data files

Taken from the UK Data Archive
http://www.data-archive.ac.uk/create-manage/consent-ethics/anonymisation?index=2

EXAMPLE ANONYMISATION LOG

Interview and page number	Original	Changed to
Int1		
p1	Age 27	Age range 20-30
p1	Spain	European country
p3	Manchester	Northern metropolitan city or English provincial city
p2	20th June	June
p2	Amy (real name)	Moira (pseudonym)
Int2		
p1	Francis	my friend
p8	Station Road primary school	a primary school
p10	Head Buyer, Produce, Sainsburys	Senior Executive with leading supermarket chain

Taken from the UK Data Archive
http://www.data-archive.ac.uk/create-manage/consent-ethics/anonymisation?index=2

FINAL REMARKS

- Sharing data from clinical research studies can be very valuable
- But this must be done responsibly, protecting participants confidentiality whilst maximising the utility of data
  - Consent for sharing should be obtained from participants
  - Suitable de-identification methods
  - Data use agreement
- Planning data sharing during the design stage can save time
- When designing new studies and identifying new research questions consider whether existing data could be used

ACKNOWLEDGEMENTS

Carolyn Hopkins, Matt Sydes, Kerry Woolfall,
Mike Clarke, Gordon Murray, Paula Williamson, Sarah Nolan, Liz Tremain, Tony Marson, Duncan Appelbe, Richard Appleton

Guidance project was funded by the MRC HTMR (MR/L004933/1-R39)