Publication and access to research datasets

UCL & Yale workshop, 9th October 2015

Iain Hrynaszkiewicz
Head of Data and HSS Publishing, Open Research
Nature Publishing Group & Palgrave Macmillan

iain.hrynaszkiewicz@nature.com
@iainh_z
Publishers/journals and data access

- More reliable evidence – and papers
- Journal mission/goals
- Content innovation (and more use and reuse)
- Reliability (peer review)
- Discoverability and visibility (bibliographic databases)
- Permanence (content and links)
- Credit/incentives (article types and citations)
- Encouraging and implementing good practice and policies
Journal data policies

- Willingness to share stated (Annals Internal Medicine)
- Data sharing implied by submission (BioMed Central*)
- Data sharing implied as a condition of publication (Nature*)
- Mandated data sharing with statement in paper (PLOS, BMJ)
- Mandated data sharing with statement and link to data (non-medical journals e.g. ecology, animal genomics)
- Mandated open data as a condition of submission (Scientific Data, GigaScience, F1000Research)

*Minimum publisher requirement – some disciplines/journals may mandate open access to data

Data sharing via supplementary files

doi:10.1186/1745-6215-12-101
Data sharing via repository links

Research

Restoring Study 329: efficacy and harms of paroxetine and imipramine in treatment of major depression in adolescence

*BMJ* 2015;351 doi: http://dx.doi.org/10.1136/bmj.h4320 (Published 16 September 2015)

Cite this as: *BMJ* 2015;351:h4320

Joanna Le Noury, research psychologist¹, John M Nardo, retired clinical assistant professor², David Healy, professor¹, Jon Jureidini, clinical professor³, Melissa Raven, postdoctoral fellow³, Catalin Tufanaru, research associate⁴, Elia Abi-Jaoude, staff psychiatrist⁵

Author affiliations ▼

Correspondence to: J Jureidini jon.jureidini@adelaide.edu.au

Accepted 3 August 2015
Data sharing via repository links

Restoring Study 329: efficacy and harms of paroxetine and imipramine in treatment of major depression in adolescence

planned (and, if relevant, registered) have been explained.

Data sharing: Clinical study reports, detailed data tables, and programming code are available on the Dryad Digital Repository (http://dx.doi.org/10.5061/dryad.bv8j6) and at www.Study329.org/.

This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

References


Cite this as: BMJ 2015;351:h6509.
Data sharing via repository links

Research

Restoring Study 329: efficacy and harms of paroxetine and imipramine in treatment of major depression in adolescence


Article

Joanna Le Nee Mansell, professor, chair, and associate, BSc, MD, FRCP, professor, chair, and associate, BSc, MD, FRCP
Author affiliations

 Correspondence to: 
Accepted 31 March 2015

Ref


Data from: Restoring Study 329: efficacy and harms of paroxetine and imipramine in treatment of major depression in adolescence

Files in this package

Content in the Dryad Digital Repository is offered “as is.” By downloading files, you agree to the Dryad Terms of Service. To the extent possible under law, the authors have waived all copyright and related or neighboring rights to this data. [CC BY-NC-ND] [Open Data]

<table>
<thead>
<tr>
<th>Title</th>
<th>RIAT Restoring Study 329 data files</th>
</tr>
</thead>
<tbody>
<tr>
<td>Downloaded</td>
<td>28 times</td>
</tr>
<tr>
<td>Description</td>
<td>Zip file containing: Original SKB CSR (both Appendix D &amp; Appendix G); ii) RIAT Excel spreadsheet for Harms/Adverse Event data; RIAT Excel spreadsheet for Withdrawals/dropouts from the study. R code used to analyse Efficacy data: Original SKB Study 329 Trial protocol which was followed in Restoring Study 329 (RIAT).</td>
</tr>
<tr>
<td>Download</td>
<td>README txt (767 bytes)</td>
</tr>
<tr>
<td>Download</td>
<td>Dryad zip (60.44Mb)</td>
</tr>
<tr>
<td>Details</td>
<td>View File Details</td>
</tr>
</tbody>
</table>

nature publishing group npg
Anonymisation

"...datasets that contain three or more indirect identifiers, such as age or sex, should be reviewed by an independent researcher or ethics committee"

Hrynaszkiewicz et al., *BMJ* 2010;340:c181

<table>
<thead>
<tr>
<th>Identifier (information sources)</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct</td>
<td></td>
</tr>
<tr>
<td>Name</td>
<td></td>
</tr>
<tr>
<td>Initials</td>
<td></td>
</tr>
<tr>
<td>Address, including full or partial postal code</td>
<td></td>
</tr>
<tr>
<td>Telephone or fax numbers or contact information</td>
<td></td>
</tr>
<tr>
<td>Electronic mail addresses</td>
<td></td>
</tr>
<tr>
<td>Unique identifying numbers</td>
<td></td>
</tr>
<tr>
<td>Vehicle identifiers</td>
<td></td>
</tr>
<tr>
<td>Medical device identifiers</td>
<td></td>
</tr>
<tr>
<td>Web or internet protocol addresses</td>
<td></td>
</tr>
<tr>
<td>Biometric data</td>
<td></td>
</tr>
<tr>
<td>Facial photograph or comparable image</td>
<td></td>
</tr>
<tr>
<td>Audiotapes</td>
<td></td>
</tr>
<tr>
<td>Names of relatives</td>
<td></td>
</tr>
<tr>
<td>Dates related to an individual (including date of birth)</td>
<td></td>
</tr>
<tr>
<td>Indirect—may present a risk if present in combination with others in the list</td>
<td></td>
</tr>
<tr>
<td>Place of treatment or health professional responsible for care</td>
<td>Could be inferred from investigator affiliations</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
</tr>
<tr>
<td>Rare disease or treatment</td>
<td></td>
</tr>
<tr>
<td>Sensitive data, such as illicit drug use or &quot;risky behaviour&quot;</td>
<td></td>
</tr>
<tr>
<td>Place of birth</td>
<td></td>
</tr>
<tr>
<td>Socioeconomic data, such as occupation or place of work, income, or education</td>
<td>MRC requirement is for &quot;rare&quot; occupations only</td>
</tr>
<tr>
<td>Household and family composition</td>
<td></td>
</tr>
<tr>
<td>Anthropometry measures</td>
<td></td>
</tr>
<tr>
<td>Multiple pregnancies</td>
<td></td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
</tr>
<tr>
<td>Small denominators—population size of &lt;100</td>
<td></td>
</tr>
<tr>
<td>Very small numerators—event counts of &lt;3</td>
<td></td>
</tr>
<tr>
<td>Year of birth or age (this article)</td>
<td>Age is potentially identifying if the recruitment period is short and is fully described</td>
</tr>
<tr>
<td>Verbatim responses or transcripts</td>
<td></td>
</tr>
</tbody>
</table>
Data on (reasonable) request - issues

• Meta-analysis fails to launch when <40% IPD available – unanswered requests and refusal to share


• Poor availability of psychological research data (only 64/249 datasets available)


• Data received from 1/10 authors publishing in *PLOS Medicine* and *PLOS Clinical Trials*

  *PLoS ONE 4(9): e7078. doi:10.1371/journal.pone.0007078*
• Sensitive data repositories (e.g. UKDA)
  Permanence, curation, persistent identifiers, versioning

• Data-on-request services (e.g. YODA)
  Independent governance, scientific review and transparency of access requests, DUAs

• Journals/publishers
  Peer review, visibility, credit/citations, robust links

= Better way to publish data on request

Open access Data Descriptor

http://www.nature.com/articles/sdata201531
Linked to open access data

http://dx.doi.org/10.7910/DVN/25833
All approved repositories:
http://www.nature.com/sdata/data-policies/repositories
And restricted access data

http://www.nature.com/articles/sdata201531
Appendix: Repository approval criteria

- Supported and recognized by scientific community
- Ensure long-term persistence and preservation of datasets
- Provide data curation
- Implement community-endorsed reporting requirements
- Provide for confidential review of datasets
- Provide stable identifiers
- Allow public access to data without unnecessary restrictions

http://www.nature.com/sdata/about/faq#q21
http://www.nature.com/sdata/data-policies/repositories