

2016

Common data elements and data management: Remedy to cure underpowered preclinical studies

Lapinlampi N

Elsevier BV

<info:eu-repo/semantics/article>

© Elsevier B.V.

CC BY-NC-ND <https://creativecommons.org/licenses/by-nc-nd/4.0/>

<http://dx.doi.org/10.1016/j.eplepsyres.2016.11.010>

<https://erepo.uef.fi/handle/123456789/705>

Downloaded from University of Eastern Finland's eRepository

Accepted Manuscript

Title: Common Data Elements and Data Management:
Remedy to Cure Underpowered Preclinical Studies

Author: Niina Lapinlampi Esbjörn Melin Eleonora Aronica
Jens P. Bankstahl Albert Becker Cristophe Bernard Jan A.
Gorter Olli Gröhn Anu Lipsanen Katarzyna Lukasiuk
Wolfgang Löscher Jussi Paananen Teresa Ravizza Paolo
Roncon Michele Simonato Annamaria Vezzani Merab Kokaia
Asla Pitkänen



PII: S0920-1211(16)30210-8
DOI: <http://dx.doi.org/doi:10.1016/j.eplepsyres.2016.11.010>
Reference: EPIRES 5639

To appear in: *Epilepsy Research*

Received date: 16-10-2016

Accepted date: 19-11-2016

Please cite this article as: Lapinlampi, Niina, Melin, Esbjörn, Aronica, Eleonora, Bankstahl, Jens P., Becker, Albert, Bernard, Cristophe, Gorter, Jan A., Gröhn, Olli, Lipsanen, Anu, Lukasiuk, Katarzyna, Löscher, Wolfgang, Paananen, Jussi, Ravizza, Teresa, Roncon, Paolo, Simonato, Michele, Vezzani, Annamaria, Kokaia, Merab, Pitkänen, Asla, Common Data Elements and Data Management: Remedy to Cure Underpowered Preclinical Studies. *Epilepsy Research* <http://dx.doi.org/10.1016/j.eplepsyres.2016.11.010>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

*Short Communications***Common Data Elements and Data Management:****Remedy to Cure Underpowered Preclinical Studies**

Niina Lapinlampi¹, Esbjörn Melin², Eleonora Aronica³, Jens P. Bankstahl⁴, Albert Becker⁵, Cristophe Bernard⁶, Jan A. Gorter³, Olli Gröhn¹, Anu Lipsanen¹, Katarzyna Lukasiuk⁷, Wolfgang Löscher⁸, Jussi Paananen⁹, Teresa Ravizza¹⁰, Paolo Roncon¹¹, Michele Simonato¹¹, Annamaria Vezzani¹⁰, Merab Kokaia², Asla Pitkänen¹

¹*Department of Neurobiology, A. I. Virtanen Institute for Molecular Sciences, University of Eastern Finland, PO Box 1627, FI-70211 Kuopio, Finland*

²*Epilepsy Center, Wallenberg Neuroscience Center, Lund University, Lund, Sweden*

³*Academic Medical Center, Dept (Neuro)Pathology and Center for Neuroscience, Swammerdam Institute for Life Sciences, University of Amsterdam, Amsterdam, Stichting Epilepsie Instellingen Nederland, Heemstede, The Netherlands.*

⁴*Department of Nuclear Medicine, Hannover Medical School, Hannover, Germany*

⁵*Translational Epilepsy Research Section, University of Bonn Medical Center, Bonn, Germany*

⁶*Université d'Aix Marseille, Marseille, France*

⁷*Laboratory of Epileptogenesis, Nencki Institute of Experimental Biology of Polish Academy of Sciences, Warsaw, Poland*

⁸*Department of Pharmacology, Toxicology, and Pharmacy, University of Veterinary Medicine, Hannover, Germany*

⁹*Institute of Biomedicine, University of Eastern Finland, Kuopio, Finland*

¹⁰*Department of Neuroscience, Experimental Neurology, Mario Negri Institute for Pharmacological Research, Milan, Italy*

¹¹*University of Ferrara and University Vita-Salute San Raffaele, Milan, Italy*

Corresponding author: Asla Pitkänen, MD, PhD, Department of Neurobiology, A. I. Virtanen Institute for Molecular Sciences, University of Eastern Finland, PO Box 1627, FI-70211 Kuopio, Finland, Tel: +358-50-517 2091, Fax: +358-17-16 3030, E-mail: asla.pitkanen@uef.fi

Highlights

- Common Data Elements (CDEs) aid with systematic data collection and standardization.
- EPITARGET has published the first set of CDEs for preclinical studies of epilepsy.
- Good data management and handling supports scientific discovery.

Abstract

Lack of translation of data obtained in preclinical trials to clinic has kindled researchers to develop new methodologies to increase the power and reproducibility of preclinical studies. One approach relates to harmonization of data collection and analysis, and has been used for a long time in clinical studies testing anti-seizure drugs. EPITARGET is a European Union FP7-funded research consortium composed of 18 partners from 9 countries. Its main research objective is to identify biomarkers and develop treatments for epileptogenesis. As the first step of harmonization of procedures between laboratories, EPITARGET established working groups for designing project-tailored common data elements (CDEs) and case report forms (CRFs) to be used in data collection and analysis. Eight major modules of CRFs were developed, presenting >1000 data points for each animal. EPITARGET presents the first single-project effort for harmonization of preclinical data

collection and analysis in epilepsy research. EPITARGET is also anticipating the future challenges and requirements in a larger-scale preclinical harmonization of epilepsy studies, including training, data management expertise, cost, location, data safety and continuity of data repositories during and after funding period, and incentives motivating for the use of CDEs.

Key words: common data element; database; data management; epileptogenesis; epilepsy

1. Introduction

Due to many failures in translating promising preclinical treatments into clinic, there is increasing concern that the pharmaceutical industries' interest in brain-related diseases will vanish. Consequently, there will be no novel, more efficient, and better tolerated treatments for neurological and psychiatric diseases, including epilepsy. The problems in translation have been related to models used, differences in pathophysiology of the disease between experimental models and humans, and importantly, the lack of statistical power and reproducibility of preclinical studies (Simonato et al., 2014; Steward et al., 2012). Small sample sizes have led to low statistical power, and consequently, overestimation of effect size and poor reproducibility (Button et al., 2013).

Part of the problem is that the data between laboratories are incomparable because of miscommunication, bias in reporting, lack of standardized data collection guidelines (Landis et al., 2012) and diverse experimental procedures. Furthermore, continuously increasing amount of data requires commonly shared scientific practices and good data management to ensure data integrity, findability, accessibility, interoperability, and reusability between researchers and research groups.

Until recently, there have been very few attempts for harmonization of practices in preclinical studies (Lemmon et al., 2014; Smith et al., 2015). In 2010, however, the National Institutes of Health (NIH) launched an initiative called National Institute of Neurological Disorders and Stroke (NINDS) Common Data Element project (Stone, 2010) that has led to the generation of common data elements (CDEs) and case report forms (CRFs) for more than 10 neurological diseases; these

CDEs and CRFs can be used to harmonize clinical trials (<https://www.commondataelements.ninds.nih.gov>).

The need for standardization of preclinical studies between laboratories was recognized in the European Union (EU) 7th Framework (FP7) funded project “Targets and biomarkers for antiepileptogenesis” (EPITARGET), a consortium of 18 partners in 9 European Countries, 12 of which conduct preclinical studies. This led to the design of the first available CDEs for preclinical studies on epilepsy to help investigators to systematically collect, analyze, standardize and share preclinical epilepsy data. These CDEs and CRFs can now be downloaded from the EPITARGET web page (www.epitarget.eu). Here, we briefly summarize the procedures employed for generation of CDEs, the lessons learned, and the anticipated challenges ahead. We will also briefly discuss the good data management practices.

2. Methods

2.1. Terminology and generation of EPITARGET CDEs

A CDE can be defined as a basic unit which is common across all the study subjects. Examples include animal species, background strain, vendor information, and sex of an animal. CDEs can be divided into general core CDEs, disease-specific core CDEs, supplemental – highly recommended CDEs, supplemental CDEs, and exploratory CDEs (“Glossary,” 2016).

Core CDE is a data element that collects essential information applicable to any study, including either those which span across all disease and therapeutic areas or those that are specific to one disease area. **Supplemental-highly recommended CDEs** are data elements which are essential based on certain conditions or study types. **Supplemental CDEs** are data elements which are commonly collected but whose relevance depends upon the study design or type of research

involved. **Exploratory CDEs** are data elements that require further validation, but may fill current gaps in the CDEs and/or substitute for an existing CDE once validation is complete.

The CDEs describing the elements belonging to the same procedure (*e.g.*, CDEs for a given behavioral test such as Morris water-maze) are logically organized into a CRF. Next, CRFs are organized in modules, collating the CRFs related to the same entity (*e.g.*, “Imaging”) (**Fig. 1**).

To tailor the EPITARGET CDEs according to the project needs, EPITARGET partners formed working groups to generate the CDEs, CRFs, and Guidelines in their areas of expertise, for example, in modelling of epileptogenesis, behavioral testing, blood analysis, or imaging. Documents underwent several iterations over a 1-y period, during which the working groups communicated via teleconferences and workshops.

EPITARGET CRFs were organized into eight main modules: core animal characteristics, injury-related monitoring and procedures, post-injury monitoring, antiepileptogenesis treatment, laboratory tests, pathology, imaging, and assessment of functional outcome. The main modules were further divided into multiple sub-modules, describing the variables tailored to represent the experimental designs of EPITARGET (**Fig. 1**).

2.2. Implementation

The preclinical EPITARGET CDEs are currently available at the EPITARGET webpage (<http://www.epitarget.eu/cdes/>). Based on the CDE collection, a data dictionary was built. A data dictionary is a metadata repository, defining every variable and their relationship to other variables. The metadata contained in the data dictionary was used to structure the data collection instruments for the EPITARGET database created in Research Electronic Data Capture (REDCap) (Harris et al., 2009).

2.3. Data management

After acquisition, data is collected and entered into the REDCap database located at the Lund University (Sweden) data server. The data can be accessed and analyzed using investigators' own personal computers, computational servers, or cloud-based systems connecting to the data repository. Raw, analyzed and backup data will be stored in data repository for long-term use and reanalysis (Fig. 2).

3. Discussion

Lack of translation from preclinical discoveries to clinical practice in the treatment of neurological and psychiatric diseases has been well-recognized in the research community. The first actions to improve the translational potential of preclinical studies were taken by investigators in the stroke field by the STAIR initiative (Stroke Therapy Academic Industry Roundtable, 1999), and more recently in spinal cord injury (SCI) field by publishing minimum information standards for reporting SCI experiments (Lemmon et al., 2014). In parallel to development of EPITARGET CDEs, Smith et al. (2015) published the Working Group document, listing the CDEs for the needs of preclinical traumatic brain injury (TBI) research. Further steps were recently taken by the International League Against Epilepsy, American Epilepsy Society, and NIH by establishing a Task force for generation of preclinical CDEs for epilepsy research (Simonato et al., 2014). The CDEs generated by EPITARGET serve as useful templates for this Task force.

Even though the harmonization of data collection according to approved protocols can be expected to increase the experimental reproducibility between the study sites, for example, in preclinical multicentre studies designed to find biomarkers and treatments for epileptogenesis, several challenges remain. These include data management, continuity of data repositories,

updating of CDEs, training of investigators in use of CDEs, cost, data security, and last but not least, motivating investigators to use and develop CDEs.

The FAIR Guiding Principles, which have been proposed for proper data management practices and data stewardship, consist of four foundational principles: **F**indability, **A**ccessibility, **I**nteroperability and **R**eusability (Wilkinson et al., 2016). EPITARGET has tackled the challenge by developing and using CDEs, data dictionaries, and applying an easily implemented database platform (REDCap). These actions are expected to reduce miscommunication across EPITARGET research groups and increase data comparability. Well-managed data repository will facilitate *findability* and help to produce *accessible* and *interoperable* metadata and data itself. The metadata-driven web-based platform used for data collection will ensure easy and intuitive secure data capture across EPITARGET research groups. Furthermore, long-term storage will enable the *reusability* of the data in the future.

Data security was one of the issues that emerged during the development of the preclinical CDEs and the construction of a database. Although preclinical animal datasets do not need to be treated with the same anonymity and caution as patient-related data, there are still reasons to protect the data. For example, unauthorized use of the shared database would increase the risk of data misinterpretation. Also, it could lead to the failure to direct the credit to the centres producing the data. Therefore, there is a need to develop procedures which regulate the access to the preclinical database and its use.

Application of CDEs in daily experimental work will require not only to standardize laboratory practices according to the guidelines provided, but also *training* of investigators and technicians to use and manage the database. Efficient user-friendly software tools and researcher training should be available to help the first-time users in data management practises.

Even though CDEs and data management are anticipated to provide a remedy to cure underpowered preclinical studies, an important question is: at what cost? Database design and maintenance during and after the funding period, as well as development of user-friendly software tools, require special technical expertise. In addition to the extra time and effort invested by data managers, researchers and IT-personnel, storage space, backups, hardware, software licenses add to the **costs**. Thus, the data management expenses should be included in grant applications (Rocca-Serra et al., 2016).

Some funding organizations like the United States Department of Defense and the NIH already require the use of CDEs and upload of data into the data repository in some of their grant instruments, providing a strong incentive for investigators to use CDEs. The spinal cord injury research community compiled guidelines for publishing data which is another way of motivating the use of CDEs (Lemmon et al., 2014). However, we are currently in the very beginning of the road that will take us from harmonization of preclinical epilepsy studies to big data analysis of large animal datasets, for which we already have some examples available (Ferguson et al., 2013; Nielson et al., 2014). Challenges such as training, costs, and policies related to use of data repositories during and after the funding period are important issues to be solved, not only to benefit a single project such as EPITARGET but to help in establishing the CDEs in the epilepsy research community in general. Ultimately, we believe that the use of CDEs will facilitate the harmonization of (multicentre) preclinical studies, translation of preclinical findings to the clinic, and promote the development of novel treatments for patients at risk of epilepsy or its progression.

Acknowledgements

This study was supported by the FP7-HEALTH project 602102 (EPITARGET).

References

- Button, K.S., Ioannidis, J.P. a, Mokrysz, C., Nosek, B. a, Flint, J., Robinson, E.S.J., Munafò, M.R.,
2013. Power failure: why small sample size undermines the reliability of neuroscience. *Nat. Rev. Neurosci.* 14, 365–76. doi:10.1038/nrn3475
- Ferguson, A.R., Irvine, K.A., Gensel, J.C., Nielson, J.L., Lin, A., Ly, J., Segal, M.R., Ratan, R.R.,
Bresnahan, J.C., Beattie, M.S., 2013. Derivation of Multivariate Syndromic Outcome Metrics
for Consistent Testing across Multiple Models of Cervical Spinal Cord Injury in Rats. *PLoS One*
8. doi:10.1371/journal.pone.0059712
- Glossary [WWW Document], 2016. . *Natl. Inst. Neurol. Disord. Stroke*. URL
<https://www.commondataelements.ninds.nih.gov/glossary.aspx> (accessed 10.3.16).
- Harris, P.A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., Conde, J.G., 2009. Research electronic
data capture (REDCap)-A metadata-driven methodology and workflow process for providing
translational research informatics support. *J. Biomed. Inform.* 42, 377–381.
doi:10.1016/j.jbi.2008.08.010
- Landis, S.C., Amara, S.G., Asadullah, K., Austin, C.P., Blumenstein, R., Bradley, E.W., Crystal, R.G.,
Darnell, R.B., Ferrante, R.J., Fillit, H., Finkelstein, R., Fisher, M., Gendelman, H.E., Golub, R.M.,
Goudreau, J.L., Gross, R.A., Gubitza, A.K., Hesterlee, S.E., Howells, D.W., Huguenard, J., Kelner,
K., Koroshetz, W., Krainc, D., Lazic, S.E., Levine, M.S., Macleod, M.R., McCall, J.M., Moxley,
R.T., Narasimhan, K., Noble, L.J., Perrin, S., Porter, J.D., Steward, O., Unger, E., Utz, U.,
Silberberg, S.D., 2012. A call for transparent reporting to optimize the predictive value of
preclinical research. *Nature* 490, 187–91. doi:10.1038/nature11556
- Lemmon, V.P., Ferguson, A.R., Popovich, P.G., Xu, X.-M., Snow, D.M., Igarashi, M., Beattie, C.E.,
Bixby, J.L., 2014. Minimum information about a spinal cord injury experiment: a proposed

reporting standard for spinal cord injury experiments. *J. Neurotrauma* 31, 1354–61.

doi:10.1089/neu.2014.3400

Nielson, J.L., Guandique, C.F., Liu, A.W., Burke, D.A., Lash, A.T., Moseanko, R., Hawbecker, S., Strand, S.C., Zdunowski, S., Irvine, K.-A., Brock, J.H., Nout-Lomas, Y.S., Gensel, J.C., Anderson, K.D., Segal, M.R., Rosenzweig, E.S., Magnuson, D.S.K., Whittemore, S.R., McTigue, D.M., Popovich, P.G., Rabchevsky, A.G., Scheff, S.W., Steward, O., Courtine, G., Edgerton, V.R., Tuszynski, M.H., Beattie, M.S., Bresnahan, J.C., Ferguson, A.R., 2014. Development of a Database for Translational Spinal Cord Injury Research. *J. Neurotrauma* 31, 1789–1799. doi:10.1089/neu.2014.3399

Rocca-Serra, P., Salek, R.M., Arita, M., Correa, E., Dayalan, S., Gonzalez-Beltran, A., Ebbels, T., Goodacre, R., Hastings, J., Haug, K., Koulman, A., Nikolski, M., Oresic, M., Sansone, S.-A., Schober, D., Smith, J., Steinbeck, C., Viant, M.R., Neumann, S., 2016. Data standards can boost metabolomics research, and if there is a will, there is a way. *Metabolomics* 12, 14. doi:10.1007/s11306-015-0879-3

Simonato, M., Brooks-Kayal, A.R., Engel, J., Galanopoulou, A.S., Jensen, F.E., Moshé, S.L., O'Brien, T.J., Pitkanen, A., Wilcox, K.S., French, J.A., 2014. The challenge and promise of anti-epileptic therapy development in animal models. *Lancet Neurol.* 13, 949–960. doi:10.1016/S1474-4422(14)70076-6

Smith, D.H., Hicks, R.R., Johnson, V.E., Bergstrom, D.A., Cummings, D.M., Noble, L.J., Hovda, D., Whalen, M., Ahlers, S.T., LaPlaca, M., Tortella, F.C., Duhaime, A.-C., Dixon, C.E., 2015. Pre-Clinical Traumatic Brain Injury Common Data Elements: Toward a Common Language Across Laboratories. *J. Neurotrauma* 32, 1725–1735. doi:10.1089/neu.2014.3861

Steward, O., Popovich, P.G., Dietrich, W.D., Kleitman, N., 2012. Replication and reproducibility in spinal cord injury research. *Exp. Neurol.* 233, 597–605. doi:10.1016/j.expneurol.2011.06.017

Stone, K., 2010. Comparative effectiveness research in neurology: healthcare reform will increase the focus on finding the most effective--and affordable--treatment. *Ann. Neurol.* 68, A10–A11. doi:10.1002/ana.22113

Stroke Therapy Academic Industry Roundtable, 1999. Recommendations for Standards Regarding Preclinical Neuroprotective and Restorative Drug Development. *Stroke* 30, 2752–2759.

Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L.B., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Gonzalez-Beltran, A., Gray, A.J.G., Groth, P., Goble, C., Grethe, J.S., Heringa, J., 't Hoen, P.A., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M.A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., Mons, B., Roche, D.G., Kruuk, L.E.B., Lanfear, R., Binning, S.A., Bechhofer, S., Benson, D.A., Berman, H., Henrick, K., Nakamura, H., Wenger, M., Crosas, M., White, H.C., Carrier, S., Thompson, A., Greenberg, J., Scherle, R., Lecarpentier, D., Martone, M.E., White, E., Sandve, G.K., Nekrutenko, A., Taylor, J., Hovig, E., Wolstencroft, K., Bauch, A., Sansone, S.-A., González-Beltrán, A., Maguire, E., Sansone, S.-A., Rocca-Serra, P., González-Beltrán, A., Harland, L., Groth, P., Berman, H.M., Bourne, P.E., Berman, H.M., Watenpaugh, K., Westbrook, J.D., Fitzgerald, P.M.D., Rose, P.W., Kinjo, A.R., Gutmanas, A., Starr, J., Musen, M.A., 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* 3, 160018. doi:10.1038/sdata.2016.18

Figure 1. EPITARGET Common Data Elements (CDEs) are constructed into eight main modules: Core animal characteristics, injury-related monitoring and procedures, post-injury monitoring, antiepileptogenesis treatment, laboratory tests, pathology, imaging, and assessment of functional outcome. The 8 main modules include multiple sub-modules. For example, behavioral and cognitive outcome/spontaneous seizures/seizure susceptibility sub-modules are under “assessment of functional outcome” module. Each sub-module has a specific case report form (CRF) which contains the CDEs related in a logical order.

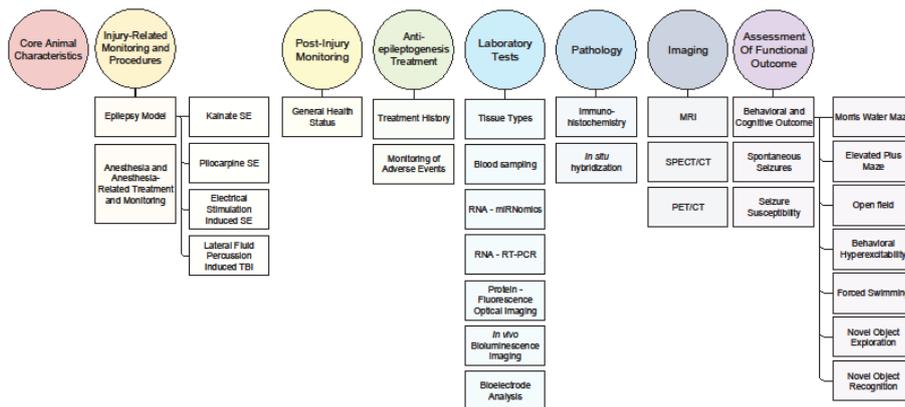


Figure 2. Data management and handling of the EPITARGET Database. Data is being collected by research groups. After data acquisition, relevant standardized case report forms (CRFs) are filled for each animal in the REDCap database located in data server. Raw, analyzed, and backup data is stored in data repositories for long-term use and reanalysis. Standardized data collection, as well as well-managed long-term data storage and management ensure data integrity, consistency, and availability for further use.

