A novel knowledge database of published pre-clinical data on spinal cord injury

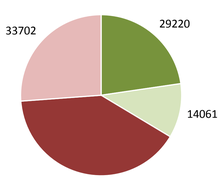
Nicole Brazda[1,2], Fabian Kruse[1], Michael Kruse[1], Tarek Kirchhoffer[1], Roman Klinger[3], Philipp Cimiano[3], Hans Werner Müller[1,2]

[1]Center for Neuronal Regeneration, Life Science Center Düsseldorf, [2]Molecular Neurobiology Laboratory, Heinrich Heine University Düsseldorf, [3]Cognitive Interaction Technology Excellence Cluster (CITEC), University Bielefeld, Germany

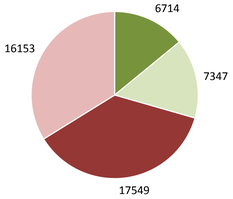
Introduction

Research in the field of central nervous system injury generates over 8,000 new publications each year and the output rises exponentially. For successfull clinical translation of basic research efforts it is mandatory to gain an objective overview of the existing knowledge. At the non-profit Center for Neuronal Regeneration, CNR e.V. (www.cnr.de) we make use of current information technology tools to collect and sort world-wide published basic and pre-clinical data on spinal cord injury (SCI) and therapy development by interdisciplinary cooperation of basic and clinical neuroscientists with data management and natural language processing experts. We aim to build up a comprehensive pre-clinical SCI database which will be made public for scientists and clinicians, for foundations and public funding organizations but also for decision makers in pharmaceutical companies planning clinical trials in this field. The aim of the CNR is to speed up translation of the most promising pre-clinical therapies into clinical trials. Today, the number of publications in the field of SCI and traumatic brain injury (TBI) adds up to around 140,000 PubMed listed original research papers. To build up a comprehensive database and develop tools for automated data extraction from scientific publications, the CNR focuses on SCI pre-clinical animal studies which comprise around 7,300 publications according to PubMed queries.

All "non-review" publications



Publications on „animal studies"



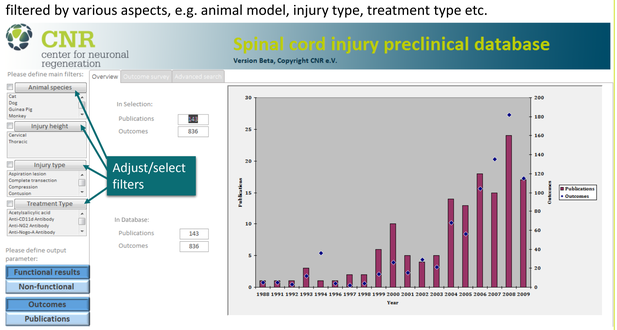
Methods

The principle concept of a pre-clinical SCI database was tested in a pilot software version programmed in Microsoft Access 2013® using visual basic elements. As a test-dataset 143 publications in the field of SCI therapy development were chosen which include those reviewed in Kwon et al. (J Neurotrauma 28, 2011,1525-43). Manual collection of data concerning the experimental setup (animal model, injury detail, treatment paradigm) and the results of the studies (histological outcomes, behavioral testings, etc.) was performed in a first phase. The depth of information gathered was iteratively re-adjusted to keep a reasonable time-frame for manual data extraction that would allow meta-analysis of data. Parallel to database construction, a SCI ontology is currently developed for a data extraction software tool which is designed by the Cognitive Interaction Technology Excellence Cluster (CITEC) of the University Bielefeld with the aim to speed up semi- or full-automatic collection of information from full-text articles to fill-up the CNR SCI database.

Results

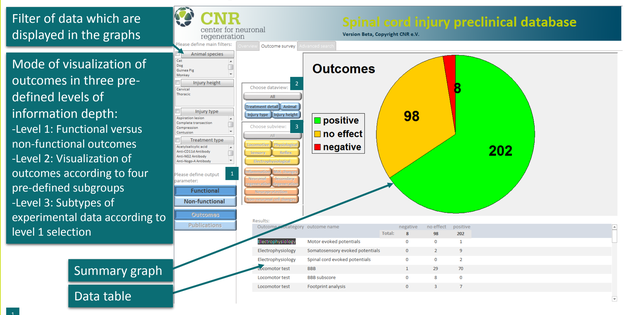
1. Dataset

The total number of results (termed „outcomes") represented in the CNR SCI database mounts up to 836 in a set of 143 publications comprising experiments on non-invasive neuroprotective strategies as reviewed in Kwon et al (2011). Publications and outcomes in the database can be filtered by various aspects, e.g. animal model, injury type, treatment type etc.



The filters are adjusted on the left side of the screen. Four filters are pre-defined, but can be changed according to the specific demands of the user by clicking on the name of the filter, e.g. „Animal species".

2. Outcome Survey

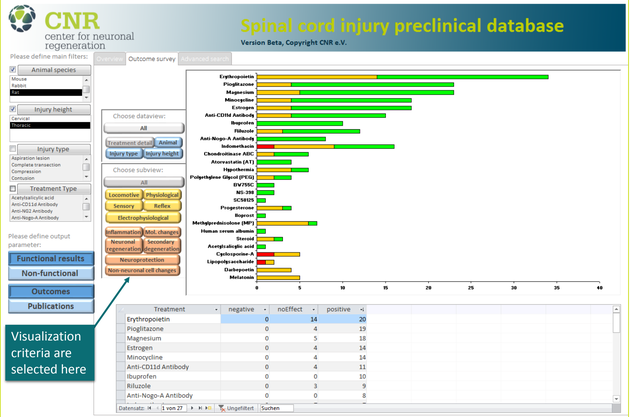


1

The **first level** of data-view summarizes the number of outcomes which correspond to the filters on the left side. The number of publications will be computed in later versions of the software as follows: as long as one positive outcome in a functional or non-functional test is present in a publication, it is counted as "positive" publication, even if negative results also occur in the publication. This seems reasonable since, e.g., dose-response curves often contain negative or neutral results which do not reflect the conclusion of the experiment.

2

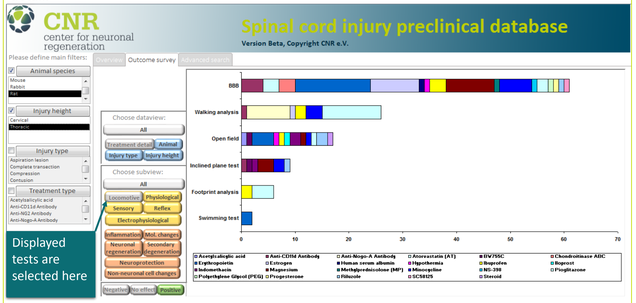
The **second level** of data-view modes shows outcomes according to the selected visualization type, e.g. according to the treatment applied. In this example, treatment substances are listed and data are sorted according to number of positive outcomes (green) from top to bottom (see below).



For all visualizations bar colors are defined as: green = positive outcome (e.g. enhanced axonal regeneration or decreased number of cell loss in treated animals *versus* the corresponding injured control animals), yellow = no effect (i.e. control and treated animals are not significantly different in the respective test), red = negative result (i.e. negative effect of the treatment, e.g. increased apoptosis in treated versus control injured animals).

3

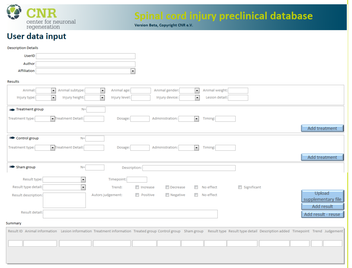
The **third level** of data-view (see below) shows a detailed list of experiments grouped according to the top level selection, e.g. first level selection "Functional", second level "Treatment detail", third level "Locomotive" displays a number of locomotor tests like BBB, walking analysis etc. and the number of positive, neutral or negative outcomes are displayed according to the treatment detail.



The visualization tools allow a fast overview and direct comparison of studies performed in SCI. Using the variable filter sets, the user can browse the different experimental fields (e.g. locomotor tests, sensory tests etc.) of therapy evaluation. On the other hand, comparison of different animal models, e.g. rat *versus* mouse or injury paradigms like thoracic *versus* cervical can be compared for different treatments.

3. Negative Data

The CNR SCI database will further allow **direct data input** by users via an input form sheet where unpublished data can be stored and made accessible for other researchers in this field.

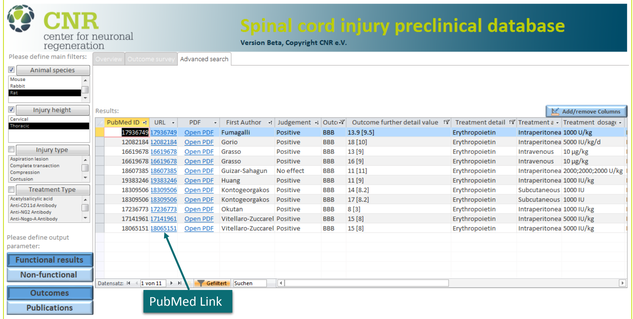


Until now reports of negative data in the field of SCI are very rare. In consequence, treatments showing no beneficial effect are usually not published. Therefore, it is assumed that a large number of unsuccessful experiments are repeated in multiple laboratories wasting precious human and financial resources.

The depth of information in the CNR SCI database is chosen in a way to approach this problem. Not only positive outcomes are documented, but also results of treatment groups which were "only" included to find the optimal treatment application, e.g. drug concentration or route of administration. Usually such results are not documented in the title or abstract of a publication and, therefore, can easily be overlooked.

4. Detailed Data View

The complete list of data can be accessed via the "Advanced search" function and can be customized according to the users needs.



In this view original publications and/or their PubMed links can be accessed. A large number of publications is already freely accessible *via* the PubMedCentral database. However, the largest amount of publications is still under copyright protection and thus only referred to by a link to its abstract. Nevertheless, all data in the CNR SCI database are drawn from full text articles if processed manually by expert curators.

5. Therapy Grading

The level-of-evidence of a therapy can be assessed by a scoring system as developed by Kwon et al, (2011). The scoring system was implemented in the CNR SCI database but, in addition, customized scores can be defined by the user.

Scores as defined by Kwon et al. can be customized by the user and applied to the dataset.

Therapy grading shows computed scores over a dataset of 117 publications on systemic neuroprotective therapies.

Therapy Grading

Therapy	Animal Species	Injury Model	Maximum Time Window of Efficacy	Clinically Meaningful Efficacy	Reproducibility/Replication	Total
Anti-CD11d Antibody	8	3	9	8	0	24
Atorvastatin (AT)	4	3	10	4	3	24
Erythropoietin	6	6	5	12	5	34
Estrogen	6	6	3	0	9	24
Hypothermia	4	18	3	12	5	42
Ibuprofen	4	4	10	8	7	33
Indomethacin	10	3	2	8	0	23
Inosine	6	1	8	4	7	26
Magnesium	4	6	11	12	12	45
Minocycline	6	8	13	16	5	48
Pioglitazone	4	3	5	12	3	27
Polyethylene Glycol (PEG)	10	9	11	4	4	38
Progesterone	4	3	2	4	0	13
Riluzole	4	12	5	4	7	32

For the pilot software version, scoring was only computed for the original dataset in Kwon et al (2011), comprising 117 publications.

The complete list of scoring criteria is available in the original publication by Kwon et al (PubMed ID: 20507235) or in the "Edit" window of the therapy grading score in the CNR SCI database software.

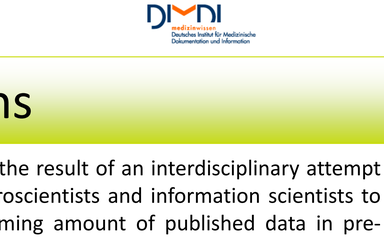
6. Future Perspectives

Meta-analysis of Data

Meta-analysis of data can be applied at two levels:

i. analysis of numeric experimental data as stored in the database, e.g. BBB scores of untreated injured animals, as shown in the example below (control groups in SCI experiments with contusion injuries using the NYU impactor). The legend lists PubMedIDs of the corresponding publications. As an example, animal gender is marked by different line colors. Accordingly, the outcome for different rat strains or spinal levels (height) of the lesion could be evaluated and compared;

ii. analysis of unexpected influences of experimental conditions on the number of positive outcomes, e.g. influence of rat strain, type of anaesthesia during surgery, animal weight etc.

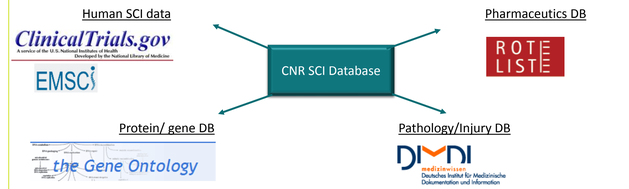


Ontology Development for SCI

An ontology is currently developed for pre-clinical therapies in SCI in order to customize information retrieval software tools for semi-automated knowledge extraction from full-text articles. So far, data have been manually filled into the pilot version of the database. For a set of over 7,000 publications in the field of SCI animal experiments involving therapy evaluation, a total of 21,000 person-hours can be calculated for manual extraction of data. For an estimated 2,000 hours annual work time, at least 10-12 domain experts in the field of SCI would spend one year exclusively on data extraction. This process is error-prone, therefore additional expert curators should monitor the process. With the help of semi-automated data-extraction, it is estimated that the amount of person-hours could be reduced by about two-thirds while the quality of data should be improved.

Links to Other Databases

The web-accessible future CNR SCI database will be linked to clinical databases like "European Multicenter Study about Spinal Cord Injury" (EMSCI) or clinicaltrials.gov, where patients´ data can be compared to animal model outcomes, e.g. the rate of spontaneous recovery (EMSCI) or the outcomes of clinical trials in comparison to the respective animal experiments if both were reported. In addition, gene and protein databases will be linked to the CNR SCI database in order to provide comprehensive information on molecular properties or side effects of drugs used for treatment, or on identified changes in gene regulation or signaling pathways after treatment.



Conclusions

The CNR SCI database is the result of an interdisciplinary attempt of basic and clinical neuroscientists and information scientists to cope with the overwhelming amount of published data in pre-clinical research on SCI and therapy development. It will be a public web-accessible online tool to support research in this field and offer an objective survey of the current knowledge available. This information might serve basic scientists, clinicians, governmental and private funding organizations as well as pharmaceutical companies for decision-making regarding their research strategy, financial support or selection of promising clinical trial strategies, respectively.

Acknowledgement

This work was supported by Allianz der Hoffnung-Foundation and Medical Faculty of the University of Düsseldorf

Corresponding authors email address: Nicole.brazda@cnr.de