# A Unified Framework for Multimodal Animal Studies: Data Management, MRI Quality Control, and Proportional Recovery in Mice

Inaugural Dissertation

zur

Erlangung des Doktorgrades

*philosophiae doctor* (PhD) in Health Sciences

der Medizinischen Fakultät

der Universität zu Köln

vorgelegt von

Aref Kalantari
aus Yazd

Köln

2026

Betreuer:

PD Dr. Markus Aswendt

Gutachter:

Prof. Dr. Thilo van Eimeren

Prof. Dr. Joseph Kambeitz

Datum der mündlichen Prüfung: 11.12.2025

*"The important thing is not to stop questioning. Curiosity has its own reason for existing. One cannot help but be in awe when contemplating the mysteries of eternity, of life, of the marvelous structure of reality. It is enough if one tries merely to comprehend a little of this mystery every day. Never lose a holy curiosity."*

*– Albert Einstein*

# List of Publications

1. Mahani FSN, Mahani F, **Kalantari A**, Diedenhofen M, Green C, Wiedermann D, et al. Dynamic structural connectivity changes in cortical and cortico-striatal strokes in mice. bioRxiv. 2025 Mar 27. doi:10.1101/2025.03.27.644898.

2. **Kalantari A**, Hambrock C, Grefkes C, Fink GR, Aswendt M. Proportional recovery in mice with cortical stroke. Exp Neurol. 2025 Apr;386:115180. doi:10.1016/j.expneurol.2025.115180. Epub 2025 Feb 4. PMID: 39914643.[1]

3. **Kalantari A**, Shahbazi M, Schneider M, Raikes AC, Frazão VV, Bhattrai A, et al. Automated quality control of small animal MR neuroimaging data. Imaging Neurosci. 2024 Oct 17;2:1-23. doi:10.1162/imag_a_00317. PMID: 40212822; PMCID: PMC11980538.[1]

4. Mahani FSN, **Kalantari A**, Fink GR, Hoehn M, Aswendt M. A systematic review of the relationship between magnetic resonance imaging based resting-state and structural networks in the rodent brain. Front Neurosci. 2023 Jul 24;17:1194630. doi:10.3389/fnins.2023.1194630. PMID: 37554291; PMCID: PMC10405456.

5. **Kalantari A**, Szczepanik M, Heunis S, Mönch C, Hanke M, Wachtler T, Aswendt M. How to establish and maintain a multimodal animal research dataset using DataLad. Sci Data. 2023 Jun 5;10(1):357. doi:10.1038/s41597-023-02242-8. PMID: 37277500; PMCID: PMC10241774.[1]

6. Götz J, Wieters F, Fritz VJ, Käsgen O, **Kalantari A**, Fink GR, et al. Temporal and spatial gene expression profile of stroke recovery genes in mice. Genes (Basel). 2023 Feb 9;14(2):454. doi:10.3390/genes14020454. PMID: 36833381; PMCID: PMC9956317.

7. Wilke JBH, Hindermann M, Moussavi A, Butt UJ, Dadarwal R, Berghoff SA, **Kalantari A**, et al. Inducing sterile pyramidal neuronal death in mice to model distinct aspects of gray matter encephalitis. Acta Neuropathol Commun. 2021 Jun 30;9(1):121. doi:10.1186/s40478-021-01214-6. PMID: 34215338; PMCID: PMC8253243.

8. Jones WD, **Kalantari A**, Otte A. [Neurostimulation by music – an overview]. MMW Fortschr Med. 2020 Mar;162(Suppl 4):3-8. doi:10.1007/s15006-020-0227-z. Epub 2020 Mar 19. PMID: 32189317. German.

[1]This publication is part of the present cumulative dissertation.

# Abbreviations

| | |
|---|---|
| AIDAqc | Atlas-based Imaging Data Analysis – quality control |
| AIDAmri | Atlas-based Imaging Data Analysis – MRI processing |
| ANOVA | Analysis of Variance |
| ARA | Allen Brain Reference Atlas |
| ARRIVE | Animal Research: Reporting of In Vivo Experiments |
| BL | Baseline |
| BIDS | Brain Imaging Data Structure |
| b0 | baseline (diffusion MRI reference image) |
| CKAN | Comprehensive Knowledge Archive Network |
| COI | Center of Intensity |
| CON | Control group |
| CSV | Comma-Separated Values |
| DataLad | Data management software for version-controlled data |
| DFG | Deutsche Forschungsgemeinschaft |
| DOI | Digital Object Identifier |
| DTI | Diffusion Tensor Imaging |
| EE | Elliptic Envelope |
| EPI | Echo Planar Imaging |
| EBRAINS | European Brain Research Infrastructures |
| FAIR | Findable, Accessible, Interoperable, and Reusable |
| FMA-UE | Fugl-Meyer Assessment of Upper Extremity |
| FOV | Field of View |
| FP | False Positive |
| FN | False Negative |
| FLASH | Fast Low Angle Shot (MRI sequence) |
| FM-UE | Fugl-Meyer Upper Extremity |
| GIN | G-Node Infrastructure |
| GLP | Good Laboratory Practice |
| Git | Version control system |
| Git-annex | Git extension for managing large files |
| GUI | Graphical User Interface |
| HDD | Hard Disk Drive |
| HTTPS | Hypertext Transfer Protocol Secure |
| IF | Isolation Forest |
| IQR | Interquartile Range |
| IT | Information Technology |
| JSON | JavaScript Object Notation |

| | |
|---|---|
| LOF | Local Outlier Factor |
| MCAO | Middle Cerebral Artery Occlusion |
| MI | Mutual Information |
| MRI | Magnetic Resonance Imaging |
| MRR | Mouse Recovery Rule |
| NIfTI | Neuroimaging Informatics Technology Initiative |
| N.A. | Not Available |
| OCSVM | one-class Support Vector Machine |
| OMERO | Open Microscopy Environment Remote Objects |
| OSF | Open Science Framework |
| PACS | Picture Archiving and Communication System |
| PRR | Proportional Recovery Rule |
| QC | Quality Control |
| QA | Quality Assurance |
| RDM | Research Data Management |
| REDcap | Research Electronic Data Capture |
| ROI | Region of Interest |
| rs-fMRI | resting-state functional MRI |
| SMA | Supplementary Motor Area |
| SNR | Signal-to-Noise Ratio |
| SSD | Solid State Drive |
| SSH | Secure Shell |
| SPs1c4m1 | Internal animal ID pattern (project, subproject, cage, mouse number) |
| SEM | Standard Error of the Mean |
| TE | Echo Time |
| TP | True Positive |
| TR | Repetition Time |
| Tx | Transmit coil |
| Rx | Receive coil |
| TVA | Tierversuchsantrag (German: animal experiment protocol) |
| tSNR | temporal Signal-to-Noise Ratio |
| URL | Uniform Resource Locator |
| XNAT | Extensible Neuroimaging Archive Toolkit |
| YAML | Markup Language |
| YODA | DataLad structure principle |
| 3R | Replace, Reduce, Refine principle |

# Zusammenfassung

Die präklinische Schlaganfallforschung untersucht die akute Hirnverletzung, die zugrunde liegende Pathophysiologie und die Erholungsmechanismen, um Therapien zu entwickeln, die in die klinische Praxis übertragbar sind. Die Integration multimodaler Daten zu Verhaltensparametern und /ci magnetresonanztomographischen (MRT-)Biomarkern birgt großes Potenzial für die Entdeckung neuartiger Mechanismen und die Überwachung therapeutischer Effekte. Dieser Ansatz erzeugt jedoch eine enorme Menge hochdiverser Daten in Format und Umfang, angefangen bei Datentabellen zur Dokumentation von Verhaltensmetriken bis hin zu MR-Bilddatensätzen. Bis jetzt wurden die Daten in uneinheitlichen Formaten gespeichert, was die Reproduzierbarkeit erschwert hat. Außerdem ist die Gewährleistung einer hohen Qualität der MRT-Daten bei großen Datensätzen mit erheblichem manuellem Aufwand verbunden. Zur Bewältigung dieser Herausforderungen etablierte diese Arbeit ein einheitliches Rahmenwerk für multimodale Tierstudien, basierend auf den FAIR-Prinzipien (Findable, Accessible, Interoperable, Reusable) des Datenmanagements sowie den 3R-Prinzipien (Replacement, Reduction, Refinement) der Tierversuchsforschung. Es wurde eine Infrastruktur implementiert, die *DataLad* für versionierte Workflows und *GIN* für die langfristige öffentliche Archivierung nutzt, wodurch eine transparente, effiziente und nachhaltige Forschungsdatenpipeline entsteht. Darauf aufbauend wurde *AIDAqc* entwickelt, eine automatisierte Qualitätssicherungs-Pipeline, die quantitative Bildgebungsmetriken extrahiert und in einem Mehrheitsentscheidungs-Framework mehrere Machine-Learning-Modelle einsetzt, um Ausreißer Scans effizient und intelligent zu detektieren. Validiert an 23 realen Datensätzen aus 19 internationalen Einrichtungen, gewährleistet *AIDAqc* eine objektive Erkennung minderwertiger Daten vor der weiteren Verarbeitung und demonstriert Robustheit sowie Generalisierbarkeit über unterschiedliche Settings hinweg. Abschließend wurde die Plattform in einem Maus-Schlaganfallmodell eingesetzt, um die *Proportional Recovery Rule* zu testen, ein etabliertes klinisches Modell, das eine lineare Beziehung zwischen initialer motorischer Beeinträchtigung und anschließender funktioneller Verbesserung beschreibt. Hierfür wurde die *Mouse Recovery Rule* vorgeschlagen, die zeigt, dass etwa 90 % der verlorenen motorischen Funk-

tion wiedererlangt werden, vergleichbar mit den rund 70 % Erholung beim Menschen. Der entwickelte integrierte Ansatz stellt eine skalierbare Vorlage für das multimodale Datenmanagement in der translationalen Tierforschung dar und dient als Beispiel für eine neue Form datengetriebener experimenteller Schlaganfallstudien.

# Abstract

Preclinical stroke research investigates the acute brain injury, the related pathophysiology, and recovery mechanisms to develop therapies translatable to clinical practice. Integrating multimodal data of behavioral outcomes and magnetic resonance imaging (MRI biomarker), holds great promise for the discovery of novel mechanisms and monitoring of therapeutic effects. But that approach generates a vast amount of highly diverse data in format and scale, ranging from data tables documenting behavioral metrics to MR image datasets. Previously, data was stored in disparate formats, hindering reproducibility. Moreover, ensuring high-quality MRI data from large datasets requires time-intensive manual labor. To address these challenges, this thesis established a unified framework for multimodal animal studies, grounded in Findable, Accessible, Interoperable, and Reusable data management and aligned with the Replacement, Reduction, and Refinement principles of animal research. An infrastructure was implemented that leverages *DataLad* for version-controlled workflows and *GIN* for long-term public archiving, resulting in a transparent, efficient, and sustainable research data pipeline. Building on this foundation, *AIDAqc* was developed: an automated quality-control pipeline that extracts quantitative imaging metrics and employs multiple machine-learning models in a majority-vote framework, detecting outlier scans efficiently and intelligently. Validated on 23 real-world datasets from 19 international institutes, AIDAqc ensures objective detection of poor-quality data before further processing, demonstrating robustness and generalizability across diverse settings. Finally, the platform was employed to test the Proportional Recovery Rule, a well-established clinical model describing a linear relationship between initial motor impairment and subsequent functional improvement, in a mouse stroke model. To this end, a *Mouse Recovery Rule* was proposed, demonstrating that approximately 90% of lost motor function is regained, comparable to the 70% recovery observed in humans. The integrated approach developed in this thesis represents a scalable blueprint for multimodal data management in translational animal research and serves as an example for a novel type of large data-driven experimental stroke studies.

# Contents

**List of Publications** **V**

**Abbreviations** **VI**

**Zusammenfassung** **VIII**

**Abstract** **X**

**1 Introduction** **1**
Background . . . . . 1
Objectives . . . . . 4
Use of Artificial Intelligence . . . . . 5

**2 How to establish and maintain a multimodal animal research dataset using DataLad** **6**

**3 Automated quality control of small animal MR neuroimaging data** **19**

**4 Proportional recovery in mice with cortical stroke** **43**

**5 Discussion** **53**
Examples of Usage . . . . . 55
Limitations . . . . . 56
Conclusion & Future Prospects . . . . . 57

**Bibliography** **58**

**Appendix: Supplements** **61**

**Acknowledgments** **84**

**Erklärung** **86**

# Chapter 1

# Introduction

## Background

Preclinical research using standardized animal models to study disease mechanisms and test new therapies before clinical trials plays a vital role in the biomedical workflow. Most often conducted in rodents, these studies form the essential link between discoveries in the laboratory and clinical applications. As an example, in the late 1950s, thalidomide was marketed for morning sickness without proper animal testing in pregnant models, resulting in more than 10,000 infants worldwide born with severe limb and organ malformations [1]. That tragedy led regulators everywhere to require thorough animal studies in rodents and at least one other species before any new treatment can be tried in people, underscoring exactly why rigorous preclinical research is so essential.

The 3R principle, first introduced by Russell and Burch in 1959, stands for Replacement, Reduction, and Refinement of animal use in research. It urges scientists to replace animals with alternative methods whenever possible, reduce the number of animals used to a minimum, and refine experimental procedures to minimize suffering and improve welfare [2]. Complementary to this, the FAIR guidelines, introduced by Wilkinson et al. in 2016, ensure that research data are Findable, Accessible, Interoperable, and Reusable, aiming to maximize data transparency, reproducibility, and reuse across the scientific community [3]. These two frameworks remain the most important regulations governing preclinical research today. In preclinical research, which is inherently smaller in scale than general research, many hurdles must be overcome by researchers. When we examine it alongside today's data landscape, it becomes clear that there is substantial room for improvement. The OpenNeuro platform, a renowned neuroimaging repository [4], reports that only 3%

of all datasets originate from mice and rats. The goal is not to compete with human studies but to foster a research culture in which every investigator publishes and shares their data. Only then would this percentage grow substantially. The other problem is that vast amounts of valuable preclinical data are forgotten or even lost over time due to poor organization, incomplete documentation, institutional policies, or technical limitations. Such data often contributes to only a handful of studies when they could support many more. Vines et al. (2014) demonstrated that data availability declines by 17% per year after publication, simply due to aging infrastructure and broken links [5]. Moreover, when older data do reach an external researcher, non-standardized storage practices, especially for preclinical datasets, and the time and effort needed to interpret poorly documented files, often stored by hand or in ad-hoc spreadsheets, make the reuse impractical, despite the data's intrinsic value. Gabelica et al. (2022) found that of 1792 contacted biomedical Open Access manuscripts only 6.8% provided the requested data, underscoring a near scandalous gap in data sharing [6]. Crucially, motivation plays a central role. Giehl et al. (2024) report that 70% of researchers find data publication and sharing too time consuming without dedicated funding and fear insufficient recognition [7]. These challenges underscore the urgent need for an integrated framework that brings together robust data management practices and standardized sharing protocols. Without such a framework, preclinical research datasets remain vulnerable to loss, fragmentation, and inconsistent documentation.

Preclinical models hold a unique niche in stroke research by allowing the study of brain processes under controlled, homogeneous conditions, something often impossible in human cohorts due to heterogeneity in stroke size and location. The evolution of stroke research into a modern discipline owes much to neuroimaging advances. In the 1970s, Computed Tomography (CT) introduced by Hounsfield and Cormack enabled rapid, noninvasive detection of cerebral hemorrhages, improving emergency care [8]. CT, however, had limited sensitivity for acute ischemia. The 1990s witnessed Magnetic Resonance Imaging (MRI) innovations that could detect ischemic strokes within minutes. Beyond diagnosis, neuroimaging has revealed how brain networks adapt and reorganize after stroke, shifting research from identifying damage to exploring recovery mechanisms and plasticity. This imaging focus in the neuroimaging research community introduces its own challenges. MRI is highly data-intensive, amplifying the need for robust data strategies, precisely

the focus of this work. Another central issue in preclinical studies is quality control (QC). Many reproducibility failures stem from inadequate QC and a lack of transparency. QC must span all aspects and levels of a study. In preclinical MRI specifically, QC has traditionally emphasized phantom testing and hardware performance evaluation to verify device standards and sequence consistency [9, 10], with comparatively little focus on data-centric QC measures such as artifact detection, motion assessment, and automated outlier screening. The absence of automated QC pipelines forces manual checks that are subjective and time-consuming, discouraging routine use [11]. Guidelines around 3R and FAIR encourage merging datasets from multiple studies to address new questions, yet manual QC of large, combined cohorts is not scalable. Although merging can be beneficial, it brings challenges for QC. Manual outlier detection is not always practical, and rigid quality thresholds risk discarding entire datasets. Therefore, scalable automated QC approaches that incorporate data-centric metrics, such as artifact detection, motion assessment, and efficient outlier screening, are essential to maintain data integrity across heterogeneous preclinical imaging cohorts.

Building on these components, the multimodal data management and automated QC were applied in a preclinical stroke study. A stroke occurs when blood flow to the brain is suddenly disrupted, causing rapid damage to brain tissue that is often permanent. There are two main types of stroke: ischemic stroke, which makes up about 85% of cases and happens when an artery is blocked, and hemorrhagic stroke, which occurs when a blood vessel bursts, causing bleeding in or around the brain. The clinical consequences of stroke depend on the location and extent of brain damage, but commonly include motor deficits, such as hemiparesis or spasticity, as well as sensory loss, language impairment, cognitive dysfunction, and emotional disturbances. Motor impairments are among the most frequent and disabling outcomes contributing significantly to loss of independence and quality of life. Stroke remains one of the leading causes of death and the leading cause of long-term adult disability worldwide [12]. According to European data pooled from 22 countries, in-hospital stroke mortality was 12% at 30 days and 16% at 3 months, with only 36% of survivors receiving a 3-month follow-up [13]. In Germany, with a population of about 84 million in 2022, approximately 0.3% of the population experienced a stroke. In preclinical research, strokes are modeled in rodents, using established techniques to mimic human ischemic brain injury. The most widely used approach is middle cerebral artery occlusion

(MCAO), in which the middle cerebral artery is temporarily or permanently occluded to induce focal ischemia, typically damaging the striatum and overlying sensorimotor cortex [14]. Another common method is photothrombotic stroke, which uses a photosensitive dye injected into the bloodstream, followed by shining light with a laser on a specific area of the skull or even inside the brain. This causes localized blood vessel damage and clot formation, creating precise and consistent brain lesions in the cortex [15]. Both models result in quantifiable motor deficits, such as impaired forelimb use, abnormal gait, and poor performance in coordination tasks, enabling behavioral assessments of functional recovery [16]. However, recovery mechanisms in rodents often differ from those observed in human patients, which complicates the direct translation of findings. Clinically, the Proportional Recovery Rule (PRR), introduced by Prabhakaran et al. (2008), posits that most stroke patients recover approximately 70% of their lost motor function within the first three to six months [17]. What occurs after this period, or in patients who deviate from the proportional recovery pattern, remains an open question. In rodent models, the PRR has only rarely been explored. Determining whether animals exhibit similar recovery trajectories is critical for improving the translational value of preclinical stroke studies and guiding the development of more predictive models for human recovery.

# Objectives

Working with complex experimental data containing imaging (MRI), video (sensorimotor tests), and meta information (group, age, time point, etc.) culminates in two foundational challenges: mastering version control for both code and data to create a clear, shareable workflow blueprint, and automated quality control to prevent artifacts from undermining downstream analyses. The work was structured into the following objectives:

1. Establish a FAIR-compliant, version-controlled framework for curating, documenting, and sharing multimodal animal datasets, providing a clear structure and enabling reproducible collaboration between researchers and across remote and on-site locations.

2. Develop and validate an automated and efficient QC pipeline that detects poor quality MRI scans, ensuring reliable quantitative analysis in mouse stroke studies.

3. Apply the data management workflow to a large (N>100 animals) dataset with the aim of applying the proportional recovery rule in mice, building a translational bridge between preclinical findings and clinical insight.

Together, these objectives aim to improve the transparency, reliability, and relevance of preclinical stroke research by using strong data management and quality control practices.

## Use of Artificial Intelligence

During the preparation of this cumulative dissertation, I utilized OpenAI's ChatGPT [18] as a writing assistant. Its use included aiding in improving my own formulations and serving as a translation tool to enhance my understanding. Throughout the process, I maintained intellectual control by carefully reviewing and modifying all AI-generated content. ChatGPT also assisted me with LaTeX formatting and iterative refinement of language in the introductory and discussion chapters. The use of ChatGPT was limited to linguistic and technical assistance; all scientific content, analyses, and conclusions are entirely my own work. This use complies with the University of Cologne's guidelines on AI-supported academic work [19].

# Chapter 2

# How to establish and maintain a multimodal animal research dataset using DataLad

*Aref Kalantari, Michał Szczepanik, Stephan Heunis, Christian Mönch, Michael Hanke, Thomas Wachtler & Markus Aswendt*

ORIGINAL ARTICLE

Scientific Data (2023)

The complexity of translational studies, encompassing diverse data types such as images, text documents, videos, behavioral measurements, and microscopy across multiple experimental groups and time points, motivated the development of a unified, cloud-based framework for data management. This system leverages DataLad to ensure comprehensive tracking of all files and changes, minimizing data loss and facilitating reproducibility. The workflow comprises four key steps: (1) Set up your project by creating a DataLad repository to store data and, (2) Keep track of changes by recording every addition or modification of data and analysis scripts, which creates a complete version history, (3) Share with your team by connecting the local repository to a secure online platform (GIN) for seamless collaboration and (4) Publish your dataset by requesting a permanent DOI from GIN for the dataset, ensuring citability.

Within the project, data are organized into distinct folders for input files, output results, code scripts, and analyses, allowing independent updates to each component. By adopting this workflow, researchers can maintain data integrity, transparency, and accessibility, thereby enhancing the reliability and reproducibility of preclinical studies and promoting the FAIR principles.

www.nature.com/scientificdata

scientific data

Check for updates

OPEN

ARTICLE

# How to establish and maintain a multimodal animal research dataset using DataLad

Aref Kalantari[1], Michał Szczepanik[2], Stephan Heunis[2], Christian Mönch[2], Michael Hanke[2,3], Thomas Wachtler[4] & Markus Aswendt[1,5]✉

**Sharing of data, processing tools, and workflows require open data hosting services and management tools. Despite FAIR guidelines and the increasing demand from funding agencies and publishers, only a few animal studies share all experimental data and processing tools. We present a step-by-step protocol to perform version control and remote collaboration for large multimodal datasets. A data management plan was introduced to ensure data security in addition to a homogeneous file and folder structure. Changes to the data were automatically tracked using DataLad and all data was shared on the research data platform GIN. This simple and cost-effective workflow facilitates the adoption of FAIR data logistics and processing workflows by making the raw and processed data available and providing the technical infrastructure to independently reproduce the data processing steps. It enables the community to collect heterogeneously acquired and stored datasets not limited to a specific category of data and serves as a technical infrastructure blueprint with rich potential to improve data handling at other sites and extend to other research areas.**

## Introduction

Data management and sharing require best practices as recently introduced for human MRI[1,2]. In our experience, most laboratories rely on non-standardized data storage on local hard drives or network drives with insufficient user management and backup capacity. Despite the fact that only a minority of MRI studies are using small animals, it is alarming that on OpenNeuro, a widely used neuroimaging data-sharing platform[3], only 3% of datasets contain data from mice or rats. Similarly, on another popular data-sharing platform, not specific for neuroimaging, Zenodo[4], only about 30% of MRI datasets are from mice or rats. In addition, it is surprising and contrary to FAIR principles[5], if in the majority of these neuroimaging datasets, only the imaging data are provided. This excludes a large part of the accompanying data, e.g., the microscopy files used for *in vivo* cross-validation. We also identified a clear lack of step-by-step guides or automated routines required to reproduce the processed data. These examples underline previous reports[6] that small animal data sharing is far from common and that there is no standardization in terms of data acquisition, storage, and sharing. If data are not shared and thus not available for re-use as it is the case for 93% of biomedical open-access publications[7], this also contrasts strongly with the 3 R principle of minimizing the number of animal experiments[8]. Therefore, it remains very difficult to compare studies between different laboratories, which contributes to the reproducibility crisis[9], and small animal (neuroimaging) studies are no exception[10].

We envision a change toward the conditions of good scientific practice and the principles of FAIR - Findable, Accessible, Interoperable, Reusable[5] and Open Science[2] to improve the reliability and recognition of animal studies. Our goal was to create an easily applicable approach for setting up a multimodal dataset that provides access to raw and processed data, methods, results, and their provenance. Proper research data management (RDM), as it is also increasingly required by funding agencies and publishers, is key to meeting these standards[2,11,12].

[1]University of Cologne, Faculty of Medicine and University Hospital Cologne, Department of Neurology, Cologne, Germany. [2]Psychoinformatics Lab, Institute of Neuroscience and Medicine, Brain & Behaviour (INM-7), Research Centre Jülich, Jülich, Germany. [3]Institute of Systems Neuroscience, Medical Faculty, Heinrich Heine University, Düsseldorf, Germany. [4]Faculty of Biology, Ludwig-Maximilians-Universität München, Planegg-Martinsried, München, Germany. [5]Cognitive Neuroscience, Institute of Neuroscience and Medicine (INM-3), Research Centre Jülich, Jülich, Germany. ✉e-mail: markus.aswendt@uk-koeln.de

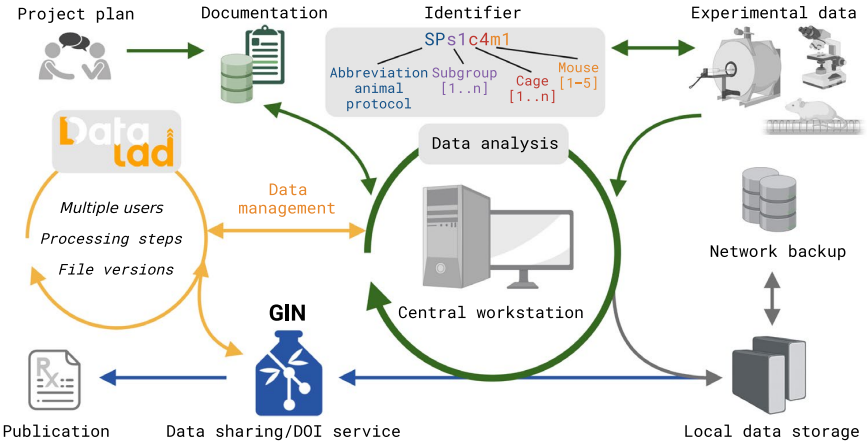



**Fig. 1** Green arrows: workflow for project planning, data acquisition, processing, storage; gray arrows: backup plan on local and network storages; orange arrows: integration of DataLad for version control; blue arrows: publication process using GIN as the online hosting service. Figure created with Biorender.com.

Here, we describe our strategy for data organization, metadata collection, and data/analysis tracking using three established tools: our relational database[13], the data platform GIN (G-Node Infrastructure services, https://gin.g-node.org), and the research data management software DataLad[14]. The database is used to collect all experimental metadata about the complete timeline of longitudinal and multimodal animal experiments, including MRI, histology, electrophysiology, and behavior. GIN and DataLad are both based on Git, a popular version control system, and git-annex, which extends Git's capabilities, especially with respect to managing large files. GIN is an open-source, web-based data management service with various features for collaborative data handling, e.g., built-in versioning, secure access, persistent data identifiers for publication (DOI), automatic indexing, and data validation. DataLad is a data management software designed to support the various stages of the development of digital objects. Importantly, DataLad can be seen as an overlay on top of existing data structures and services: Tracking files does not change the files themselves or the location from which they can be retrieved by data processing tools.

## Results

**Workflow.** Over the last 5 years, we have established a strategy for 1) project planning, 2) documentation, 3) data acquisition and storage, and 4) data sharing (Fig. 1). Project planning and experimental details are recorded in an in-house relational cloud-based database[13]. A key element for both the database and the data storage is the identifier, the study ID for each animal, used in a standardized file name structure to make the data findable. The directory structure for the raw data follows the permit of performing animal experiments. The data for a specific project is organized following the YODA principles (https://handbook.datalad.org/en/latest/basics/101-127-yoda.html), which is compatible with existing standards, e.g., the BIDS structure[15] (Fig. 2A). An automatic incremental backup routine was installed, which transfers the data from an external diskstation linked to the main analysis workstation to a centrally managed network drive. In preparation for publication and to facilitate data reproducibility, the experimental raw and processed data is made publicly available on GIN, and post-processing details and pipelines are specified - either in the publication or on a GitHub page (https://github.com/aswendtlab/Project_C3a_peri-infarct).

DataLad is used as the central data management tool (Fig. 1) and for version control: It keeps track of which files were modified, when, and by whom, and provides the ability to restore previous states. To this end, DataLad is agnostic of the data type and provides a unified interface for managing code and data files (typically handled by Git and git-annex, respectively).

**Dataset nesting.** Within DataLad datasets, it is possible to nest (an unlimited number of) other DataLad datasets, each sub-dataset remaining a standalone component with its own history and siblings. To achieve this structure the following stages 1 through 4 can be performed independently within an established dataset. In this way, the user can subdivide the project e.g. by publication, data type, or storage location. To clarify this, it is possible to place a top-level dataset together with the sub-datasets in the online repository service. Here, the data tracked by a dataset directly is stored as usual, but the sub-datasets are available only as links to their own repository, which can, but do not have to be hosted by the same online repository provider. In our case, all of the project code is stored and maintained in GitHub repositories (https://github.com), and installed as multiple sub-datasets within the main top-level dataset **Project1** on GIN (Fig. 2A). The structure introduced by the Yoda principles makes the entire project self-contained. No further element outside the project is needed to get from inputs to results. In contrast, no self-containment is possible for the structure in Fig. 2C.

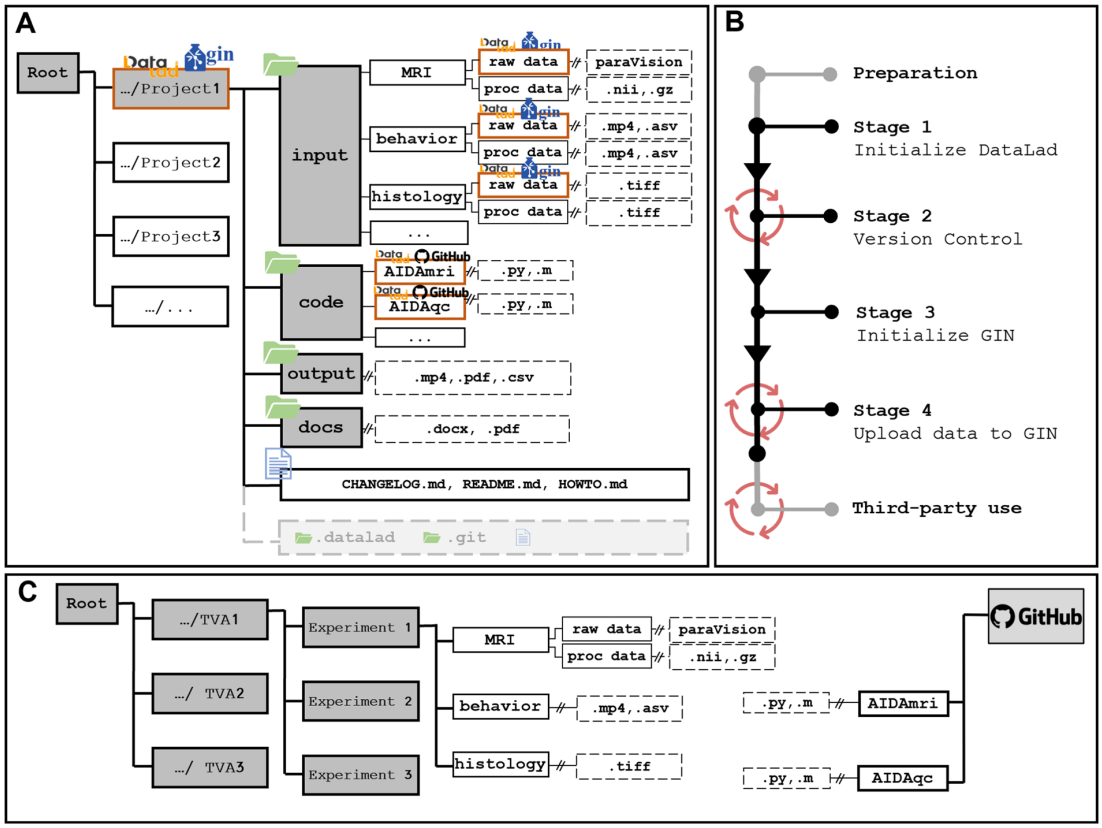



**Fig. 2** (**A**) YODA-directory structure and integration of DataLad. The root directory of the operating system (OS), the list of projects, and project-related content (folders + file types). The dashed and "//" separated connections indicate that there can be additional levels, e.g., the time of measurements. During the conversion of the Project1 folder into a DataLad dataset, the corresponding DataLad files, (gray boxes), are created as additional information in the folder, without affecting the rest of the data. "raw_data" folders and all folders in "code" are independent subdatasets in the context of nesting and establishing decentralization with "code" being located in Github instead of GIN. (**B**) Step-by-step guide for creating the DataLad dataset is composed of 4 mandatory subsequent stages and framed by an initial preparation phase and an optional third-party use scenario (red circles highlight recurring stages). (**C**) Folder structure based on the permit of performing animal experiments without DataLad (TVA = Tierversuchsantrag (German: animal protocol). This is the old structure before the YODA structure was implemented, all processing was also done within this structure, using the pipelines parallelly from another storage separated from this structure. The original raw data is still kept in this structure as an archive of raw data, but the rest is abolished.

INFO: If a dataset or a subdataset is to be implemented for a folder that contains only coding scripts, stages 1 to 4 should be slightly optimized. Briefly, git is much better suited for version control of scripts that contain code. When step 1.2 is run, everything is handled by git-annex by default, the alternative is to use `datalad create --force -c text2git`. The additional configuration puts git in charge of the contents of the datasets.

**Step-by-step data management protocol.** Based on our workflow development (Fig. 1), this use case description provides a step-by-step guide for *non-expert users* on how to manage large experimental datasets containing complex, multimodal files, i.e. the raw and processed data as well as the code and output files.

For the hard- and software requirements and installation routines, see the Methods section and Supplementary Material. The workflow consists of 4 mandatory stages (Fig. 2B): 1. Initialize a DataLad dataset, 2. Version controlling, 3. Initializing the remote (online) repository, and 4. Uploading to the remote repository. We also describe an optional use-case scenario for third-party use, i.e., collaborating on the same dataset with other researchers, and dataset publication.

*Preparation.* In order to create the data structure, start with the main **Project** folder and sort the raw data in the related **input** folder (Fig. 2A).

When converting the **Project** folder to a DataLad dataset, the corresponding DataLad files are created automatically without changing the files and structure. DataLad files contain the version and the history of all files in the project folder. **Project1** serves as a super-dataset with the associated **raw_data** and **proc_data** folders as sub-datasets with their own repositories on the repository services GIN and GitHub, respectively. With this structure, DataLad can register actions such as processing **raw_data** from a specific pipeline in **code** and store the output of the pipeline in **proc_data**. These actions are saved in the history information of the super-dataset **Project1**.

In our case, the data from multiple animal protocol folders are copied to the project folder and the original raw data (in the archive) remains untouched. This process can be tracked if Datalad is already effective for all storages. In case of other data structures, e.g. using PACS server, the transfer of the data into the main **Project** folder would be similar. The idea is to have all project-relevant data available for analysis and later on sharing. Each sub-folder of **input** is named according to the methodology (e.g., MRI, behavior, histology, etc.) and can contain various file types (Table 2). The other folders include the **code, results**, and **docs**. In this example, **code** contains the Atlas-based Imaging Data Analysis tools AIDAqc and AIDAmri[16] for automated quality control, processing of multimodal MRI[16], and registration with the Allen Mouse Brain Atlas[17]. The **docs** folder contains all necessary (metadata) information and documentation to reproduce the MRI dataset (e.g. which study ID belongs to which experimental group, time points, MRI scan type, etc.).

The following stages are executed to start the conversion of the **Project1** folder into a DataLad dataset (Fig. 2B). A video guide is available (https://gin.g-node.org/Aswendt_Lab/Example_Dataset_Kalantari_Paper/src/master/doc).

*Stage 1: Initialize a datalad dataset.* 1.1 To begin, open the terminal and change the directory to the target folder. In this example, it is **Project1**.

```
»cd Project1
```

1.2 Once in the target folder, run the following command to create the DataLad dataset-specific folders and files in the **Project1** folder (Fig. 2A).

```
» datalad create-force
```

**INFO:** DataLad provides numerous **commands** and each command has different **options**. You can create an empty DataLad *dataset* folder by `datalad create folder_name`. Here, in step 2, the `create` command was used to initialize a DataLad *dataset* in the current working directory, and the `--force` option allowed dataset creation in a folder already containing data. For more information about other options you can use`--help`.

*Stage 2: Version control (local data).* 2.1 After the initialization stage 1, continue with the following step. Note that this step may take some time depending on the hardware[1].

```
» datalad status
»datalad save -m "user message"
```

Once the dataset is created, steps that modify its content are recorded by executing the `datalad save` command. Stage 2 shows the first practical use of DataLad datasets: recording changes. In our case, the **raw_data** folder contains, e.g. MRI raw data. After processing using AIDAmri[16], the results are stored in the **proc_data** folder (Fig. 2A).

**INFO:** The `datalad status` command can be considered as an inspection tool that prints the current status of each file, whether it is tracked or untracked by DataLad, deleted or modified. After initializing in step 2.1, all status printouts would show "untracked". Any newly added files will not be tracked until `datalad save` is explicitly executed. After a file has been saved and the contents subsequently changed, the status will be printed differently. In general, `datalad status` is used for informational purposes only and may be helpful in identifying recent (unsaved) changes.

The `datalad save` command, on the other hand, records new changes in the.git folder. The save command also has several options, where `-m` associates the save with a user message which can describe the purpose of the change. These messages can be beneficial for the user to find a particular version or to recall changes in the dataset more easily. If you are interested in more options for a command, use the `--help` option. For example, `datalad save --help` displays all available options for the command `save`.

2.2 (Optional) If changes are done programmatically (e.g. by running a custom script), this can be done with a command like the following:

```
» datalad run -m "user message" --input … python<code.py>
```

The run command registers information about the input and output data of the executed code (here a Python script) and registers it in the history information of the dataset. The `datalad save` command is automatically

([1]Refers to features such as internally or externally connected storage drives, network drives, storage types such as hard disk drive (HDD) or solid state drives (SSD)).

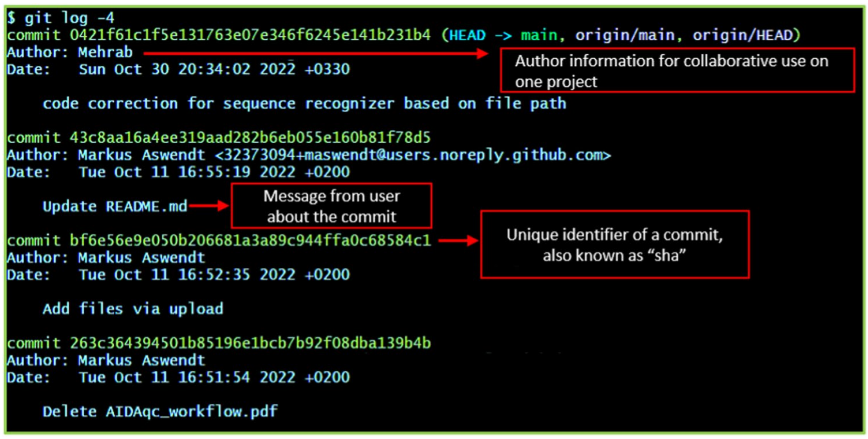



**Fig. 3** Sample `git log` output of a dataset showing information about the author, date, changes, and the corresponding commit identifier, also called "sha".

executed when using `datalad run`, but in addition to the user-provided message, a re-executable run record is also stored to capture provenance.

It is important to emphasize that Stage 2 is a repetitive Stage (Fig. 2B), i.e., it can be repeated to record consecutive changes or new states. Over time, this process creates a logbook, in which all the actions that a user has performed on the project over time are recorded. How to access and use this logbook is explained in the following step.

2.3 All registered information about the recorded changes can be accessed by:

```
» git log
```

This will display all the changes that have been made since version control began. Optionally by typing:

```
» git log -2
```

Only the last two changes or **commits** will be printed (Fig. 3).

**INFO:** A **commit** is a snapshot of the current state of the project. The difference between two commits in one datalad dataset means the changes that have happened between those two states of the project. Some data might be produced, changed, moved, or deleted. There are other ways to access the history such as **tig** (https://jonas.github.io/tig/) or some with their own graphical user interface (GUI) like **gitgui, gitk**, etc (https://git-scm.com/downloads/guis).

*Stage 3: Initialize a GIN repository as a DataLad dataset sibling.* In this scenario, we are using the data platform GIN (other options are available, see Table 1). A prerequisite for the following steps is an operational GIN account (https://gin.g-node.org/user/sign_up) with an associated ssh key pair (http://handbook.datalad.org/en/latest/basics/101-139-gin.html) between the GIN account and the account on the local computer. Note: This step may take some time depending on the network speed.

3.1. Create a repository with a project-specific dataset name under the GIN account and copy the project-specific ssh link generated by GIN.
3.2. Execute the following command with your own credentials in the Project1 folder.
`datalad siblings add-dataset.-name gin-url git@gin.g-node.org:/username/dataset-name.git`

**INFO:** `datalad siblings` is used for all actions that affect other copies of the dataset. Here the `add` option is used to link the dataset to a new location where the siblings or the copy of the dataset should go. Next, `--dataset` defines the path of the dataset being configured, here it is defined by the "`.`", meaning the current folder location. `--name` is the name of the sibling which can be defined by the user. However, it is logical to name it according to the online repository service for ease of use and to avoid confusion. Here we use `gin`. The `--url` option is for the link available on the project page acquired in step 3.1, which is generated explicitly for the project created there. For some repositories, like GIN, dedicated commands are provided that automate the creation and registration of remote siblings in a single step. For GIN this command is `create-sibling-gin`.

| Type | Name | Storage space limit/ File size limit | Open source/ Costs | Data types | Versio-ning | Meta- data | DOI Service | Reference/Link |
|---|---|---|---|---|---|---|---|---|
| Research data platform | GIN | No/No | Yes/No | No restrictions | Yes | Yes | Yes | https://gin.g-node.org |
| Research data platform | EBRAINS | No/No | No/No | No restrictions | Yes | Yes | Yes | https://ebrains.eu/service/share-data/ |
| Data hosting | Harvard Dataverse | 1TB/2.5GB | No/No | No restrictions | Yes | Yes | Yes | https://dataverse.harvard.edu |
| Research data platform | OpenNeuro | No/No | No/No | MRI, PET, MEG, EEG (BIDS format only) | Yes | Yes | Yes | https://openneuro.org |
| Research data platform | OSF | 50GB/No | Yes/No | No restrictions | Yes | Yes | Yes | https://osf.io/ |
| Data hosting | Zenodo | No/50GB | No/No | No restriction | Yes | Yes | Yes | https://www.zenodo.org/ |
| Data hosting | Dryad | No/No | No/No | No restrictions | Yes | Yes | Yes | https://datadryad.org/ |
| Data hosting | figshare | 1TB/20GB | No/Yes | No restrictions | Yes | Yes | Yes | https://figshare.com |
| Cloud storage | Amazon S3 | No/5TB | No/Yes | No restrictions | Yes | No | No | https://aws.amazon.com/ |
| Cloud storage | Google Drive | 20TB/5TB | No/Yes | No restrictions | No | No | No | https://www.google.com/intl/eng/drive/ |

**Table 1.** Overview of currently available research data platforms, data hosting services, and cloud storage serving different needs for data storage, version-control, metadata collection, and or sharing.

*Stage 4: Upload data to GIN.* After the remote repository is set up and a connection between the two is established, the project can be uploaded (pushed) to GIN.

4.1. Start by entering the following command in the terminal:

```
datalad push --to gin
```

Just like stage 2, this step can be repetitive (Fig. 2B). As a dataset and its history information progress, new commits are added, and updates can be pushed to the remote repository. This is done efficiently, meaning that only changed files are transferred. Versioning makes the state of the uploaded dataset unique. In addition, GIN is fully compatible with git and git-annex, and its web interface can, for example, display the history of changes and associated commit messages.

Depending on the size of the project and the quality of the Internet connection, the execution of this step may take some time. Therefore, it may be wise not to try to push all the data in one batch, i.e., split it up and do the push process in multiple batches. This is possible by adding the path of a specified smaller part of the project to the command:

4.2 (Optional) `datalad push --to gin <path_to_smaller_project_part>`

**Working with published data.** Anyone interested in the project can now access and download not only the data but also the history of the project as it has evolved over time, through what is called "cloning". From a third-party perspective, the first step is to visit the respective repository on the GIN website, where the SSH link or the HTTPS link can be found respectively, depending on whether the dataset is to be modified or only downloaded.

**INFO:** On the projects page of a GIN account, there are three links for accessing the repository, SSH, HTTPS, and GIN. In simple terms, SSH and HTTPS are different ways of communicating between the user's operating system and the online repository service. The main difference in our use case is that an SSH connection is required if we want to upload or "push" the data to the remote repository. For downloading or "cloning" the data, an HTTPS connection is sufficient. As briefly mentioned in stage 3, setting up an SSH link requires additional steps, but HTTPS links can be used without further concern.

1. Copy the SSH/HTTPS link from the GIN project web page.
2. Open the terminal and navigate to a folder where the project content should be located and run the following command, replacing the example URL with the copied link.
   For SSH links:
   `datalad clone git@gin.g-node.org :/username/dataset-name.git`
   For HTTPS links:
   `datalad clone https://gin.g-node.org/username/dataset-name`
   If the HTTPS link is chosen, note that the link does not have a `.git` extension.

**INFO:** A useful feature of the `datalad clone` command is that it does not download the entire dataset at once. It downloads only the folder structure, the small files, and the filenames of large files, i.e., files that are handled by git-annex. The `datalad get` command can then be used to selectively download the contents of large files so that the entire content becomes available locally. This can be useful from two points of view: first, if the entire dataset is very large and only some parts are of interest, only those parts can be selectively downloaded. Second, if only the project structure and file name are of interest, in this case, `datalad get` would not be called at all.

| Data types | Description | DataLad tracking |
|---|---|---|
| .txt | Text format | Holistic |
| .xlsx | Microsoft Excel spreadsheets | Difference |
| .csv | Delimited text file | Holistic |
| .docx | Microsoft Word documents | Difference |
| .nii | File format for neuroimaging | Holistic |
| .gz | compressed archive | Difference |
| .rar | compressed archive | Difference |
| .py | file containing code written in Python | Holistic |
| .mat | binary MATLAB® files that store workspace variables | Difference |
| .m | file containing code written in MATLAB | Holistic |
| .png | image format with lossless compression | Difference |
| .jpg | image format with lossy compression | Difference |
| .lif,.vsi,.zvi | vendor-specific microscopy file format (Leica, Olympus, Zeiss) | Difference |
| .yml | configuration files | Holistic |
| .json | open standard file format and data interchange format | Holistic |

**Table 2.** Typical file formats in our project.

If you want to download the entire project, continue with the following command in the DataLad dataset folder created after the last step.

3. Type in the following:
   `datalad get`
   The command can be restricted to a specific path (directory or file) to selectively retrieve the contents.

As introduced in the previous section, dataset nesting can be very useful from a third-party perspective. In our example, the top-level dataset **Project1** contains multiple datasets for the "raw_data" folders and different code pipelines (Fig. 2A). The intention for this setup was, first, to make the raw data and codes separately available to third-parties without having to download all the peripheral project structures, and secondly, to preserve updatability of the various pipelines in the **code** folder, i.e., as the pipelines are updated over time, cloning the top-level dataset automatically associates the most up-to-date pipelines with it.

**Metadata collection and dataset publication.** Since our dataset is intended for publication and reuse, it is critical to annotate it with relevant information. On creation, GIN repositories are by default private, i.e., access restricted, but can be set to public through a tick box in the settings menu. Public datasets on GIN can receive a unique persistent Digital Object Identifier (DOI), which makes them citable. Next to the publication metadata, we include documentation about the dataset, such as experimental groups and time points, as well as modality-specific information, e.g., about MRI sequences, which is retrieved from our relational database.

## Discussion

Here, we provide a step-by-step guide for non-expert users to implement a FAIR data workflow applicable to small animal data. In this workflow, we are using a simple but efficient local backup scheme and standardized data structure in combination with GIN as a solution for public data availability. DataLad was utilized as a data management tool to provide a foundation for transparent and reliable collaboration between researchers, with the intention of encapsulating all elements of a project, such as input data, codes, and results, along with information about how they are connected[18].

Our motivation to implement this workflow was to ensure data preservation, efficient collaboration, data sharing, and increased reproducibility. Such practical needs related to research data management are widespread in the field and so far only single solutions have been proposed (Kuhn Cuellar *et al.*[19]). A recent survey on data management and sharing in (human) neuroimaging shows the significant challenges associated with properly managing and sharing neuroimaging data, with the biggest limitations being time and lack of best practices[20]. Researchers consider themselves less mature in data sharing than in data analysis and collection practices. To overcome this limitation, our focus was on an easy-to-implement workflow that uses only freely available software and does not require any special prior knowledge.

A standardized subject identifier and file/folder structure (see Methods section) are the foundation of efficient research data management. The created identifier for each animal and standardized file naming helps to make the data traceable even in the absence of other metadata. In our case, the animal protocol determines which and how often a method can be applied. In order to provide full transparency according to the guidelines of good scientific practice using animals[21] and to simplify documentation for the audits by the local authorities, we collect all experimental data in an electronic relational database[13] and store the raw data in the structure predefined by the animal protocol. This raw data directory serves later on as an archive, which remains untouched. For the project/publication-related processing, all raw data - which can come from different animal licenses - is copied into the input folder to leave the raw data untouched for security while providing maximum

flexibility to work with the data. The YODA structure contains not only the **input folder** but also **output, docs**, and **code**. Here, our approach differs from previously suggested folder structures with a 3-level hierarchy, i.e., laboratory level (organization holding several projects), projects level, and subsequent experimental level[22]. For publication or collaboration, the whole YODA structure is uploaded to the online data hosting service, in our case GIN. Importantly, DataLad is flexible in where the data is stored. If users have their complete data in alternative storage solutions such as on XNAT or PACS server for radiological data[23] or OMERO[24] as for microscopy data, there are dedicated DataLad extensions (https://docs.datalad.org/), or additional extensions can be developed with minimal effort to support additional services. Our workflow is not excluding other software and data hosting services, for example, CKAN (https://ckan.org), Harvard Dataverse (https://dataverse.harvard.edu), and BarcelonaBeta Brain[25], with their own strategy to handle complex multimodal research data. In contrast, it provides a versatile and interoperable RDM, with the necessary flexibility and simplicity to be adapted to local experimental paradigms and existing IT infrastructures in small laboratories to large multisite consortia[12,18].

**Metadata.** Metadata is a very important ingredient in current research practices, especially when it comes to transparent collaboration between researchers according to FAIR and neuroimaging community-specific guidelines[5,26], For MRI saved as DICOM or NiFTI file formats, there have been several attempts to make the header file more metadata rich, with the BIDS standard being the most advanced and widely used option for structuring files according to MRI sequences and including standardized metadata[15]. Our workflow is fully compatible with the BIDS format (see example in https://doi.org/10.12751/g-node.3yl5qi[27]) but we decided to make BIDS no requirement to use the workflow as for example is the case for the data platform OpenNeuro, to maximize the flexibility for researchers working with different file formats. Although we fully support the efforts to maximize standardization at the level of metadata, in our own experience working with multimodal data (from MRI, behavior, electrophysiology, and microscopy), it is necessary as an intermediate step, to prioritize the existence of as much metadata as possible in machine-readable files (txt, csv, json), which do not necessarily have to be the current community-based structure. We, therefore, recommend using the workflow in combination with a relational database such as our own development[13], REDcap[28], or others, which allows the generation of such machine-readable files. With basic programming expertise, it will be easy for third-party users to search and filter the shared data based on these files.

If the processing requires multiple steps or processing tools, we further suggest to include a more detailed step-by-step guide (e.g. https://github.com/aswendtlab/ Project_C3a_peri-infarct), which goes beyond the requirements of metadata reporting in existing standards, i.e. BIDS. Such step-by-step guides might be necessary for data replication in case automated routines, i.e., datalad run, cannot be used.

**Limitations and Outlook.** It should be mentioned that DataLad requires some upfront investment of time and resources, but the efficiency-, quality-, and reliability gains down the line make those investments worthwhile in our experience.

The inability to reproduce the results of a particular study is not unique to neuroimaging in small animals. This is mainly due to the lack of transparency and methodological details, such as a step-by-step protocol in scientific papers. Reproducing a study requires access to a variety of other documents, such as scanning parameters, pre-and post-processing procedures, and individual subject characteristics. To this end, in addition to the workflow presented here, we take a pragmatic approach, i.e., we make the raw and processed data publicly available on GIN and specify the post-processing details on a project-specific GitHub page.

Converting the data folders into a DataLad dataset not only provides the benefits described above, i.e., track modifications, easy sharing of datasets by publishing them to siblings, and efficient management of large files by reducing transport and storage needs. As stated before DataLad opens up your datasets for much broader use and provides easy reuse of published datasets in other DataLad datasets through the DataLad subdataset mechanism. DataLad also offers provenance tracking (re-executable annotation of changes), and metadata management. Together with DataLad's run-command and rerun-command, which allow "tracked execution" of operations on a dataset, DataLad enables truly reproducible research[29].

**Conclusion.** Open data sharing embedded in a proper RDM protocol will not resolve shortcomings in study design or analysis strategy, but most importantly it provides other researchers to identify potential shortcomings and address them in future work, which prevents the repetition of errors[20]. Our efforts serve as a blueprint for other areas of preclinical imaging with a high impact on the research practice in the animal neuroimaging community and beyond. In line with the 3 R principle to reduce the number of animals, this workflow enables the community to collect heterogeneously acquired and stored datasets to foster mouse brain simulation experiments and implement a cross-species modeling approach. Moreover, the close collaboration with efforts for data and metadata standardization, provenance tracking and workflow management, integration into an interoperable research data platform ecosystem, and specification of the computational model using standardized means will help to strengthen the interaction of all participants and will ameliorate the current translational gap between basic and clinical researchers.

## Methods

In this section, we describe the details of the material necessary to reproduce the workflow.

**Hardware.** The workflow described here (Fig. 1A) was developed using a Mac Pro (Late 2013) as the main workstation run in server mode. In this way, multiple users can access the Mac via the built-in screen or file sharing (VNC/SMB protocols) and retrieve data and run programs in parallel. A similar way of working is possible

with workstations running Linux or Windows 10/11 installations via remote desktop connections. The data storage strategy was developed according to best practices[30] including a backup plan, different backup locations, backup validity checks, and the possibility to extend data storage. The main local storage - directly connected to the workstation - is a 20TB LaCie Thunderbolt drive operating in Raid-5 mode (one hard drive can be damaged while the file system remains intact). All data are first manually copied to this local storage by the responsible experimenter. The path to the raw data is documented in the electronic database[13]. Automated and incremental backups are performed using Carbon Copy Cloner (Bombich Software, Inc., USA) weekly from the local storage to network storage (managed by the IT Department of the University Hospital).

**Software.** The workflow (Fig. 1A) builds upon free and open-source software: Python (https://www.python.org/, RRID:SCR_008394), GIN (https://gin.g-node.org, RRID:SCR_015864), and DataLad (https://www.datalad.org, RRID:SCR_003931). Other data hosting options exist (Table 1), which can be easily integrated in the provided workflow. Representative installation guides are provided in the Supplementary Material. For the most up-to-date installation instructions, see the related websites.

*DataLad.* DataLad is used for distributed version control: creation, synchronization, and tracking of linked dataset copies (called siblings). A DataLad dataset can have one or several siblings, stored either locally (backup drives, local servers, different workstations) or online (e.g. GIN). Changes made to one dataset copy can be synchronized with other copies, and synchronization is always explicit (it is easy to know whether versions diverge or not). File content availability is tracked, and content available remotely can be retrieved on demand to save space locally. This also means that file exchange, backup, and publishing are all done through the same software interface, even if the locations are different.

*Definitions.*

1. Repository**: Folder with files and subfolders as a structural unit for code or data management;** can be any set or combination of files with different folder structures and different data types, sometimes it can even be empty and contain no files.
2. **DataLad dataset:** repository on which DataLad is executed and version control is running.
3. **Initializing a DataLad dataset:** to create an empty dataset that is subject to DataLad's version control. If files are added to this dataset after that, they will be tracked by DataLad.
4. Initializing **DataLad in an already existing repository:** the transition of a dataset in a file/folder structure to a DataLad-managed dataset.
5. **Tracking/version control:** if a file changes because it has been replaced or edited by a user, DataLad logs those changes over time accordingly.
6. **A DataLad sibling/clone:** can be defined as a copy of the dataset. This does not necessarily mean that it is an exact copy, or that all the data is fully available. A placeholder is a file with the same name as the original file but without its content.
7. **Git:** a free and open-source version control system used to handle small to very large projects efficiently
8. **Git-annex:** git-annex is a distributed file synchronization system that aims to solve the problem of sharing and synchronizing collections of large files.
9. **Dataset nesting:** datasets can contain other datasets (subdataset), which can in turn contain subdatasets, and so on. Every dataset that contains another subdataset can be called a **superdataset**. The **top-level dataset** is the superdataset which has the highest level in the hierarchy of datasets.

*Data storage solutions.* Online data hosting providers offer the infrastructurefor researchers to upload their data and share it with others. These services differ in their primary focus: cloud storage (file storage and limited sharing, no metadata indexing), simple data hosting services (long term hosting, automated metadata indexing), and research data platforms (long term hosting, sharing, collaboration and data management) (Table 1).

Services and platforms that provide persistent identifers (DOI) are key to adopting practices of FAIR and open science in line with aspects of reproducibility and data reuse. However, only a minority of studies share their data, which might change with the increasing demand from publishers and funding institutions. We have chosen the research data platform GIN (https://gin.g-node.org/G-Node/Info/wiki/), which is supported by the German government (BMBF Grant 01GQ1302) and LMU Munich, and developed open-source by the German Neuroinformatics Node (G-Node). GIN is a registered resource (https://doi.org/10.17616/R3SX9N) and fulfills the criteria for repositories and scientific gateways endorsed by the INCF[31].

To make a dataset eligible for the GIN-DOI service, specific metadata, including a GIN-specific file called "datacite.yml" needs to be created, which contains information about the authors, title, description, keywords, and license according to the DataCite schema (https://schema.datacite.org/) need to be provided. On GIN, this metadata needs to be in a file called "datacite.yml" in the root of the repository. A license should be chosen (e.g., https://creativecommons.org/choose/) to specify attribution, derivatives, and sharing requirements, with the license text to be included in a text file called LICENSE. The GIN DOI service ensures permanent access to the dataset using a persistent identifier (https://gin.g-node.org/G-Node/Info/wiki/DOI).

INFO: The choice of a data hosting service should be made carefully, as various factors play a role, e.g. the availability, anonymity encryption, the general target group of external or internal users, the scope of the data, frequency of usage, etc. Importantly, DataLad will work with all of the listed services. DataLad version control is applied regardless of where the data resides, i.e. locally or in an online data store. Nevertheless,

an adequate backup is necessary in order not to lose access to the data, the DataLad records and the associated history. Therefore, it is common to store a replica of the dataset and its version-controlled history somewhere else with sufficient storage space.

**Data structure/data types.** *Folder structure.* We are working with two data structures, one in which the (raw) data is stored according to the experiments described in the approved animal protocol (Fig. 2C), and the second one in which the project-specific data is stored (Fig. 2A). First, we collect all raw data in a uniform folder structure according to the main projects and subprojects in our animal experiment licenses and as approved by the local authorities.A typical raw data folder (MRI/raw_data/P1/SPs1c4m1_1_1_20180105_094610) specifies the experiment (MRI), the type of data (raw data), the time point (P1, i.e., post-stroke day 1), the studyID (SPs1c4m1), and a MRI hardware (Bruker) specific code (i.e., study number, reconstruction number, date as YYYY-MM-DD, and time as hh-mm-ss). We recommend saving the raw data in the folder structure related to the animal protocol, which simplifies documentation for the authorities. In a second step, before initializing DataLad, the raw data is copied from one or multiple raw data folders in the YODA structure for each project/publication. The YODA structure ensembles data input, output, docs, and code in a project-specific structure (Fig. 2A).

*Identifier.* In order to keep track of the large number of experiments, we created unique identifiers for each mouse/scan (Fig. 1). The identifier (Study ID) combines elements of the animal protocol, the (sub)project, the cage, and the animal number per cage. For example, **SPs1c4m1** relates to project SPasticity, subproject 1, cage 4, mouse **1**. In principle, other information, such as sex and genotype can be added according to common nomenclature, e.g., **SPs1c4m1mRbp4**, relating to a male (m) Tg(Rbp4-cre)KL100Gsat (short Rbp4) mouse. Importantly for translational studies, the study IDs should not be modified to reveal the experimental group to the user, i.e., the user remains blinded during data collection and analysis. In any case, the main functions of the identifier should be preserved, i.e., to anonymize the subject (in terms of experimental details) and to remain identifiable by keyword-style search/sort terms.

**INFO:** The IDs refer to the animal protocol and are stored in a fixed folder structure similar to that used in the BIDS format. Our file convention for naming and structuring files was specified before BIDS became popular for animal MRI data. Therefore, we have datasets with underscores and hyphens as separators, e.g. TP_T1_4_1. If users plan to use BIDS format, the workflow can still be applied as DataLad is BIDS compatible. However, as BIDS is sensitive to specific characters, e.g. hyphens and underscores, the IDs should contain only numbers and letters. A possible alternative for **SP_T1_4_1** would read **SPsT1c4m1**, which replaces the underscores with the initial letters of subproject (s), cage (c), and mouse (m). Note: if special characters were included in the study IDs in the past, making the files BIDS-compatible requires much more attention, e.g. writing the correct information also in the NIfTy header and in all related metadata files.

*File name.* Whenever possible, the elements of the file names are either generated automatically or should be assigned according to a standardized scheme, which includes the StudyID, test/experiment, and the time point (if applicable). This way, the file name will be unique and contains already essential metadata. As an example, SPs1c4m1CytD4, would relate to the Cylinder Test (Cyt), a mouse behavior test, which was performed on day 4 (D4) with study ID SPs1c4m1. The file/folder paths are stored in the electronic database[13] along with important metadata information about the experiment. In the case of MRI, this includes the anesthesia protocol (including the exact timing and dosage), configuration details (e.g. coil, gradient), and the list of MRI scans.

According to the basic rules of data storage[30], open data formats are used whenever possible. Our datasets contain a wide range of files, ranging from small text files to larger MRI files and video recordings. There is no general restriction in the file format to be compatible with the workflow (Table 2).

INFO: For any text file format (e.g. txt, csv, json) DataLad tracks the changes in the file on a line-by-line basis (*holistic*). As a result, each line in e.g. Python code can be attributed to a commit (and author) that changed it last (using Git functionality). For all other data formats, DataLad tracks the *difference* on a per-file basis using file checksums, i.e., information on when and who changed the document is stored but not which part of the file changed. In terms of Open Science and long-term usability, it is recommended to use line-by-line file types (machine-readable, e.g. csv) whenever possible.

## Data availability

We recommend using the test dataset[27] (https://doi.org/10.12751/g-node.3yl5qi). It contains a basic yoda structure (Fig. 2A) and is sufficiently small to allow rapid processing.

## Code availability

DataLad and GIN are freely available. The manuscript contains all code to reproduce the workflow.

Received: 19 December 2022; Accepted: 15 May 2023;
Published online: 05 June 2023

## References

1. Nichols, T. E. *et al.* Best Practices in Data Analysis and Sharing in Neuroimaging Using MRI. *Nature Neuroscience* **20**(3), 299–303 (2017).
2. Niso, G. *et al.* Open and Reproducible Neuroimaging: From Study Inception to Publication. https://doi.org/10.31219/osf.io/pu5vb (2022).

3. Markiewicz, C. J. *et al*. The OpenNeuro Resource for Sharing of Neuroscience Data. *eLife* **10** (October). https://doi.org/10.7554/eLife.71774 (2021).
4. European Organization for Nuclear Research, and OpenAIRE. Zenodo. CERN. https://doi.org/10.25495/7GXK-RD71 (2013).
5. Wilkinson, M. D. *et al*. The FAIR Guiding Principles for Scientific Data Management and Stewardship. *Scientific Data* **3**(March), 160018 (2016).
6. Mandino, F. *et al*. Animal Functional Magnetic Resonance Imaging: Trends and Path Toward Standardization. *Frontiers in Neuroinformatics* **13**, 78 (2019).
7. Gabelica, M., Bojčić, R. & Puljak, L. Many Researchers Were Not Compliant with Their Published Data Sharing Statement: A Mixed-Methods Study. *Journal of Clinical Epidemiology* **150**(October), 33–41 (2022).
8. The Principles of Humane Experimental Technique. *The Medical Journal of Australia* **1**(13), 500–500 (1960).
9. Begley, C. G. & Ioannidis, J. P. A. Reproducibility in Science: Improving the Standard for Basic and Preclinical Research. *Circulation Research* **116**(1), 116–26 (2015).
10. Poldrack, R. A. *et al*. Scanning the Horizon: Towards Transparent and Reproducible Neuroimaging Research *Nature Reviews. Neuroscience* **18**(2), 115–26. (2017).
11. Couture, J. L., Blake, R. E., McDonald, G. & Ward, C. L. A Funder-Imposed Data Publication Requirement Seldom Inspired Data Sharing. *PloS One* **13**(7), e0199789 (2018).
12. Hanke, M. *et al*. In Defense of Decentralized Research Data Management. *Neuroforum* 0 (0). https://doi.org/10.1515/nf-2020-0037 (2021).
13. Pallast, N., Wieters, F., Nill, M., Fink, G. R. & Aswendt, M. 2018. Cloud-Based Relational Database for Multimodal Animal Data. *Database: The Journal of Biological Databases and Curation* https://doi.org/10.1093/database/bay124 (January 2018).
14. Halchenko, Y. *et al*. DataLad: Distributed System for Joint Management of Code, Data, and Their Relationship. *Journal of Open Source Software* **6**(63), 3262 (2021).
15. Gorgolewski, K. J. *et al*. The Brain Imaging Data Structure, a Format for Organizing and Describing Outputs of Neuroimaging Experiments. *Scientific Data* **3**(June), 160044 (2016).
16. Pallast, N. *et al*. Processing Pipeline for Atlas-Based Imaging Data Analysis of Structural and Functional Mouse Brain MRI (AIDAmri). *Frontiers in Neuroinformatics* **13**(June), 42 (2019).
17. Wang, Q. *et al*. The Allen Mouse Brain Common Coordinate Framework: A 3D Reference Atlas. *Cell* **181**(4), 936–53.e20 (2020).
18. Wachtler, T. *et al*. NFDI-Neuro: Building a Community for Neuroscience Research Data Management in Germany. *Neuroforum* 0 (0). https://doi.org/10.1515/nf-2020-0036 (2021).
19. Kuhn, L. *et al*. A Data Management Infrastructure for the Integration of Imaging and Omics Data in Life Sciences. *BMC Bioinformatics* **23**(1), 61 (2022).
20. Borghi, J. A. & Van Gulick, A. E. Data Management and Sharing in Neuroimaging: Practices and Perceptions of MRI Researchers. *PloS One* **13**(7), e0200562 (2018).
21. Percie du Sert, N. *et al*. The ARRIVE Guidelines 2.0: Updated Guidelines for Reporting Animal Research. *Journal of Cerebral Blood Flow and Metabolism: Official Journal of the International Society of Cerebral Blood Flow and Metabolism* **40**(9), 1769–77. (2020).
22. Colomb, J., Arendt, T. & Sehara, K. The Gin-Tonic team. *Towards a Standardized Research Folder Structure*. https://doi.org/10.25815/WCY6-M233 (2021).
23. Marcus, D. S. *et al*. The Extensible Neuroimaging Archive Toolkit: An Informatics Platform for Managing, Exploring, and Sharing Neuroimaging Data. *Neuroinformatics* **5**(1), 11–34 (2007).
24. Swedlow, J. R. 2007. The Open Microscopy Environment: A Collaborative Data Modeling and Software Development Project for Biological Image Informatics. In *Imaging Cellular and Molecular Biological Functions*, 71–92. Berlin, Heidelberg: Springer Berlin Heidelberg.
25. Huguet, J. *et al*. Management and Quality Control of Large Neuroimaging Datasets: Developments From the Barcelonaβeta Brain Research Center. *Frontiers in Neuroscience* **15**(April), 633438 (2021).
26. Poline, J. B. *et al*. Data Sharing in Neuroimaging Research. *Frontiers in Neuroinformatics* **6**(April), 9 (2012).
27. Aswendt, M. & Kalantari, A. A DataLad dataset for an exemplary structure of a multimodal animal data repository, *G-Node*, https://doi.org/10.12751/g-node.3yl5qi (2023).
28. Harris, P. A. *et al*. Research Electronic Data Capture (REDCap)–a Metadata-Driven Methodology and Workflow Process for Providing Translational Research Informatics Support. *Journal of Biomedical Informatics* **42**(2), 377–81 (2009).
29. Wagner, A. S. *et al*. FAIRly Big: A Framework for Computationally Reproducible Processing of Large-Scale Data. *Scientific Data* **9**(1), 80 (2022).
30. Hart, E. M. *et al*. Ten Simple Rules for Digital Data Storage. *PLoS Computational Biology* **12**(10), e1005097 (2016).
31. Sandström, M. *et al*. Recommendations for Repositories and Scientific Gateways from a Neuroscience Perspective. *Scientific Data* **9**(1), 212 (2022).

## Acknowledgements

This work was funded by the Friebe Foundation (T0498/28960/16) and the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 431549029 – SFB 1451. We acknowledge support for the Article Processing Charge from the DFG (German Research Foundation, 491454339).

## Author contributions

Conceptualization: A.K., M.A., M.S., Data Curation: A.K., Formal Analysis: A.K., M.A., Funding acquisition: M.A., M.H., Investigation: A.K., M.A., Methodology: A.K., M.S., T.W., M.H., M.A., Project administration: M.A., Resources: M.A., Software: A.K., M.A., Supervision: M.A., M.H., Visualization: M.A., A.K., Writing – original draft: M.A., A.K., M.S., S.H., T.W., Writing – review & editing: M.A., A.K., C.M., S.H., M.H., T.W.,All authors read, edited, and approved the final manuscript.

## Funding

Open Access funding enabled and organized by Projekt DEAL.

## Competing interests

MS, SH, CM, & MH are developers of the free and open source software DataLad. TW is involved in the development of the free and open-source software GIN and provision of the GIN platform. The other authors report no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at https://doi.org/10.1038/s41597-023-02242-8.

**Correspondence** and requests for materials should be addressed to M.A.

**Reprints and permissions information** is available at www.nature.com/reprints.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2023

# Chapter 3

# Automated quality control of small animal MR neuroimaging data

*Aref Kalantari*, *Mehrab Shahbazi*, *Marc Schneider*, *Adam C. Raikes*, *Victor Vera Frazão*, ..., *Giovanna D. Ielacqua* & *Markus Aswendt*

ORIGINAL ARTICLE

Imaging Neuroscience (2024)

The second paper presents AIDAqc, a Python-based command-line tool for automated quality assessment of small animal MRI data. The tool was developed to harmonize large, heterogeneous MRI datasets and to establish objective criteria for image quality across studies employing different scanners, coils, and protocols. AIDAqc computes key image quality metrics like signal-to-noise ratio (SNR), temporal SNR, motion severity, and ghosting artifacts from raw scans without manual region-of-interest specification. It employs interquartile range thresholds alongside four outlier-detection algorithms (one-class SVM, isolation forest, local outlier factor, and elliptic envelope) in a majority voting scheme to identify outlier images. Validation involved synthetic noise (Gaussian, salt-and-pepper, speckle) and motion artifacts, demonstrating complete detection of induced errors. Evaluation of approximately 2600 anatomical, 1200 diffusion and 700 functional scans of mice, rats, rabbits, hamsters and gerbils in 19 international laboratories (7 T–16.4 T) yielded precision exceeding 70% emphasizing the need to overcome limited manual quality control, which showed low inter-rater agreement (Fleiss $\tilde{\kappa}$ = 0.17).

The tool accepts Bruker raw, BIDS/NIfTI formats and produces machine-readable CSV summaries, figures, and image sequences. AIDAqc is publicly available to support integration into existing workflows and to enhance reproducibility, standardization, and transparency in preclinical MRI research.

Imaging Neuroscience
an open access journal

Software Toolbox

# Automated quality control of small animal MR neuroimaging data

**Aref Kalantari[a], Mehrab Shahbazi[b], Marc Schneider[a], Adam C. Raikes[c], Victor Vera Frazão[a], Avnish Bhattrai[c], Lorenzo Carnevale[d], Yujian Diao[e,f], Bart A. A. Franx[g], Francesco Gammaraccio[w], Lisa-Marie Goncalves[i], Susan Lee[j], Esther M. van Leeuwen[g], Annika Michalek[i], Susanne Mueller[k,l,m], Alejandro Rivera Olvera[n], Daniel Padro[o], Mohamed Kotb Selim[p], Annette van der Toorn[g], Federico Varriano[ab], Roël Vrooman[n], Patricia Wenk[i], H. Elliott Albers[j], Philipp Boehm-Sturm[k,l,m], Eike Budinger[i,ad], Santiago Canals[p], Silvia De Santis[p], Roberta Diaz Brinton[c], Rick M. Dijkhuizen[g], Elisenda Eixarch[q], Gianluigi Forloni[r], Joanes Grandjean[n,s], Khan Hekmatyar[j], Russell E. Jacobs[t], Ileana Jelescu[e,f], Nyoman D. Kurniawan[u], Giuseppe Lembo[d,v], Dario Livio Longo[h], Naomi S. Sta Maria[t], Edoardo Micotti[r], Emma Muñoz-Moreno[x], Pedro Ramos-Cabrer[o,y], Wilfried Reichardt[z], Guadalupe Soria[aa,ab], Giovanna D. Ielacqua[ac], Markus Aswendt[a]**

[a]University of Cologne, Faculty of Medicine and University Hospital Cologne, Department of Neurology, Cologne, Germany
[b]Hamedan University of Technology, Faculty of Medical Engineering, Hamedan, Iran
[c]Center for Innovation in Brain Science, University of Arizona, Tucson, AZ, USA
[d]IRCCS INM Neuromed, Department of AngioCardioNeurology and Translational Medicine, Pozzilli, Italy
[e]Department of Radiology, Lausanne University Hospital (CHUV), Lausanne, Switzerland
[f]CIBM Center for Biomedical Imaging, Lausanne, Switzerland
[g]Translational Neuroimaging group, Center for Image Sciences, University Medical Center Utrecht, Utrecht, The Netherlands
[h]Institute of Biostructures and Bioimaging (IBB), National Research Council of Italy (CNR), Turin, Italy
[i]Leibniz Institute for Neurobiology (LIN), Combinatorial Neuroimaging Core Facility (CNI), Magdeburg, Germany
[j]Center for Behavioral Neuroscience, Neuroscience Institute, Advanced Translational Imaging Facility, Georgia State University, Atlanta, Georgia, USA
[k]Charité 3R | Replace, Reduce, Refine, Charité-Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany
[l]Department of Experimental Neurology and Center for Stroke Research Berlin, Charité-Universitätsmedizin Berlin, Berlin, Germany
[m]NeuroCure Cluster of Excellence and Charité Core Facility 7T Experimental MRIs, Charité-Universitätsmedizin Berlin, Berlin, Germany
[n]Donders Institute for Brain, Cognition, and Behaviour, Radboud University Medical Centre, Nijmegen, The Netherlands
[o]Center for Cooperative Research in Biomaterials (CIC biomaGUNE), Basque Research and Technology Alliance (BRTA), Donostia-San Sebastián, Spain
[p]Instituto de Neurociencias, CSIC/UMH, San Juan de Alicante, 03550 Alicante, Spain
[q]BCNatal Fetal Medicine Research Center (Hospital Clínic and Hospital Sant Joan de Déu), Universitat de Barcelona; Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS); and Centre for Biomedical Research on Rare Diseases (CIBER-ER), Spain
[r]Istituto di Ricerche Farmacologiche Mario Negri IRCCS, Department of Neuroscience, Milan, Italy
[s]Department of Medical Imaging, Radboud University Medical Center, Nijmegen, The Netherlands
[t]Department of Physiology and Neuroscience, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA
[u]Centre for Advanced Imaging, Australian Institute for Bioengineering and Nanotechnology, The University of Queensland, St. Lucia, Australia
[v]Sapienza University of Rome, Department of Molecular Medicine, Rome, Italy
[w]Molecular Imaging Center, Department of Molecular Biotechnology and Health Sciences, University of Turin, Turin, Italy
[x]MRI Core Facility, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Barcelona, Spain
[y]Ikerbasque, Basque Foundation for Science, Bilbao, Spain
[z]Division of Medical Physics, Department of Diagnostic and Interventional Radiology, University Medical Center Freiburg, Faculty of Medicine, University of Freiburg, Freiburg, Germany
[aa]CIBER de Bioingeniería, Biomateriales y Nanomedicina, Instituto de Salud Carlos III
[ab]Laboratory of Surgical and Experimental Neuroanatomy, Faculty of Medicine and Health Sciences, Institute of Neurosciences, University of Barcelona, Barcelona, Spain
[ac]Preclinical Research Center (PRC), Max-Delbrück Center for Molecular Medicine in the Helmholtz Association, Berlin, Germany
[ad]Center for Behavioral Brain Sciences, Magdeburg, Germany

Corresponding Author: Markus Aswendt (markus.aswendt@uk-koeln.de; m.aswendt@gmail.com)

Received: 29 December 2023 Revision: 28 August 2024 Accepted: 4 September 2024 Available Online: 27 September 2024

The MIT Press

© 2024 The Authors. Published under a Creative Commons Attribution 4.0 International (CC BY 4.0) license.

**Imaging Neuroscience**, Volume 2, 2024
https://doi.org/10.1162/imag_a_00317

**A. Kalantari, M. Shahbazi, M. Schneider et al.** **Imaging Neuroscience, Volume 2, 2024**

**ABSTRACT**

Magnetic resonance imaging (MRI) is a valuable tool for studying brain structure and function in animal and clinical studies. With the growth of public MRI repositories, access to data has finally become easier. However, filtering large datasets for potential poor-quality outliers can be a challenge. We present AIDAqc, a machine-learning-assisted automated Python-based command-line tool for small animal MRI quality assessment. Quality control features include signal-to-noise ratio (SNR), temporal SNR, and motion. All features are automatically calculated and no regions of interest are needed. Automated outlier detection for a given dataset combines the interquartile range and the machine-learning methods one-class support vector machine, isolation forest, local outlier factor, and elliptic envelope. To evaluate the reliability of individual quality control metrics, a simulation of noise (Gaussian, salt and pepper, speckle) and motion was performed. In outlier detection, single scans with induced artifacts were successfully identified by AIDAqc. AIDAqc was challenged in a large heterogeneous dataset collected from 19 international laboratories, including data from mice, rats, rabbits, hamsters, and gerbils, obtained with different hardware and at different field strengths. The results show that the manual inter-rater agreement (mean Fleiss Kappa score 0.17) is low when identifying poor-quality data. A direct comparison of AIDAqc results, therefore, showed only low-to-moderate concordance. In a manual post hoc validation of AIDAqc output, precision was high (>70%). The outlier data can have a significant impact on further postprocessing, as shown in representative functional and structural connectivity analysis. In summary, this pipeline optimized for small animal MRI provides researchers with a valuable tool to efficiently and effectively assess the quality of their MRI data, which is essential for improved reliability and reproducibility.

**Keywords:** standardization, reproducibility, machine learning, motion detection, image artifacts, majority voting

## 1. INTRODUCTION

Quality control (QC) is an integral part of overcoming reproducibility issues and lack of transparency toward open data sharing for magnetic resonance imaging (MRI) (Niso et al., 2022). Numerous software developments are available for human MRI, for example, MRIQC (Esteban et al., 2017), pyfMRIqc (Williams et al., 2023), VisualQC (Raamana, 2018), and FreeSurfer (Backhausen et al., 2016) as well as standardized protocols, for example, in the Human Connectome Project (Marcus et al., 2013), ENIGMA (Thompson et al., 2020), and INDI initiative (Zarrar et al., 2015). In contrast, the focus of previous small animal approaches was mainly on the use of standardized phantoms but not software (Mannheim et al., 2018; Osborne et al., 2017). A recent survey highlighted the need for more efficient animal MRI QC, as the majority of users (63%) do not follow any QC protocols and 19% declare that no regular maintenance for quality assurance (QA) of the machine is performed (Tavares et al., 2023). This is in stark contrast to industry and clinical guidelines, for example, the AAPM acceptance testing and QA procedures for MRI facilities (Jackson et al., 2010) or the ACR MRI QC manual (Price et al., 2015), which have been shown to improve further MRI analysis (Bedford et al., 2023; Gilmore et al., 2021).

Both QA and QC standards aim to facilitate stable, high-quality MRI data. In addition to the desired homogeneous image quality in the in-house laboratory, QC/QA procedures are necessary for longitudinal imaging sessions between different laboratories (Sreedher et al., 2021). MR image quality is mainly determined by noise and image artifacts. Noise arises from multiple sources including intrinsic scanner and thermal noise as well as physiological noise, which originates from motion, cardiac cycle, and respiration (Bianciardi et al., 2009). Typical image artifacts include occasional noise, differences in susceptibility, that is, differences in the magnetic field, mostly at the border between different tissue compartments (Budrys et al., 2018), and ghosting, that is, a blurry copy of the sample that appears in the field-of-view displaced from its true location. Ghosting can be induced by periodic movements, including involuntary movements, cardiac and respiratory motion, vessel pulsation, and blood and CSF flow (*discrete ghosts*) (Zaitsev et al., 2015) or hardware- and software-related issues, such as improper shimming, strong heating of the gradients, and induction of eddy currents (*Nyquist ghosts*) (Chen & Wyrwicz, 2004). In small animal imaging, movement is restricted by anesthesia and animal fixation in the MRI bed, which can be combined with muscle relaxants and intubation to achieve a stable respiratory and cardiac cycle (Grandjean et al., 2014). Nevertheless, physiological factors and movement of parts of the body can induce motion in the image. This leads to an increased level of noise, blurring, and ghosting artifacts (Zaitsev et al., 2015), which are difficult to distinguish and eliminate via postprocessing. Due to the large variety of hardware/software combinations, the most prominent components influencing image quality may

vary. For functional MRI (fMRI) in rodents, a previous study has shown that physiological fluctuations, mainly originating from the respiratory cycle, introduce severe artifacts and lead to misinterpretation of corresponding connectivity measures (Kalthoff et al., 2011). Different from faster MRI sequences with easier visual interpretation of motion artifacts, for example, as ghosting or blurring, in dynamic MRI scans, motion artifacts can cause signal changes that may severely confound statistical analysis rendering results unreliable (Havsteen et al., 2017).

The manual identification of poor image quality, which is still the most frequently used method, can only be practically applied to a rather small number of datasets (according to our expertise this relates to <50 datasets) and remains a very subjective method with high interassessor variability (Williams et al., 2023). This is mainly due to unavoidable human error and a missing standard of how to draw and where to place regions of interest (ROI) to quantify signal-to-noise ratio (SNR) or temporal SNR (tSNR). Predefined standardized regions as implemented in the image artifact analysis of the American College of Radiology are only available for phantoms (Etman et al., 2017). As an alternative, machine-learning algorithms for specific quality metrics were tested, for example, to automatically detect motion artifacts (Fantini et al., 2021; Lorch et al., 2017). However, such deep learning approaches require large training datasets, which is difficult to achieve with small animal MRI and in addition would require expert knowledge to generate a ground truth. Automation without the need to train the model would provide a unique opportunity to assess the quality of images for less experienced users or a scenario of data reuse. However, to date, user-independent and large-scale-applicable calculation of image quality metrics in small animal MRI is missing. To address this issue, we developed AIDAqc, a user-friendly command-line tool tailored to the needs of small animal MR neuroimaging, while maintaining high flexibility with respect to animal model, MR hardware, and sequences. The tool generates a detailed summary of the dataset parameters and performs automatic calculations of image quality metrics, which are assessed using a majority-voting approach to simplify and objectify the categorization of image quality. Here, we describe the AIDAqc pipeline, present the validation using a large heterogeneous dataset collected from 19 international MRI laboratories, and discuss the interpretation of quality metrics and their influence on further analysis.

## 2. METHODS

### 2.1. Python-based Workflow for automated QC

AIDAqc was developed as part of the AIDA (Atlas-based Imaging Data Analysis Pipeline) software family (https://github.com/aswendt-lab) with the suffix *qc* relating to *quality control*. The automated pipeline was tested in Python 3.12.0 using the libraries *Numpy*, *Pandas*, *Matplotlib*, *Sklearn*, *Nibabel*, and *Alive_progress*. The installation was tested successfully using a custom Anaconda software environment (conda 23.10.0), on Windows 10 & 11, macOS 12.6.8, and Linux Ubuntu 22.04. Detailed instructions on how to install and use AIDAqc can be found online (https://github.com/Aswendt-Lab/AIDAqc).

The following subsections explain each stage of the pipeline in detail (Fig. 1): (1). Parsing Input Data, (2). Feature Calculation, and (3).Outlier Detection.

#### *2.1.1. Stage I: Parsing input data*

To increase the flexibility of the pipeline, the Neuroimaging Informatics Technology Initiative (NIfTI) data format, Brain Imaging Data Structure (BIDS) data folder structure, as well as the raw Bruker (Bruker, BioSpin, Ettlingen, Germany) data format (i.e., the complete Bruker study folder including image and metadata from one subject and time point) can be used as input. Note: the image data need to be oriented as coronal brain sections. In the initial parsing stage, the input folder is automatically and recursively searched for these MRI files. As a result, the file paths are validated to contain a unique file and sorted into anatomical (T1- or T2-weighted), diffusion (diffusion-weighted, diffusion tensor imaging, etc.), and functional (fMRI and rs-fMRI) scans based on filenames and related metadata. Other scans, for example, localizers, and files with missing necessary metadata are excluded from further processing.

#### *2.1.2. Stage II: Feature calculations*

Automatically calculated features include (t)SNR, motion, and ghosting. At this stage, the quantitative measures are automatically obtained from the input data, that is, no user input is required. Specific measures, for example, tSNR, will be calculated only for fMRI.

**2.1.2.1. SNR.** The SNR is the most rudimentary but also versatile quantitative measure of MR image quality for most sequences. However, the calculation can get complex due to the manual definition of regions of interest (ROIs), that is, at least one ROI inside the *sample* and one ROI outside the sample representing the *noise area*. Here, two approaches were implemented: the “standard” method, which automatically defines ROIs based on the center of intensity (COI) of the image, and the Chang method (Chang et al., 2005), which is independent of ROI selection. For both methods, the SNR is reported in decibel (dB).

*2.1.2.1.1. Standard method*. SNR is calculated by dividing the mean signal intensity by the standard deviation

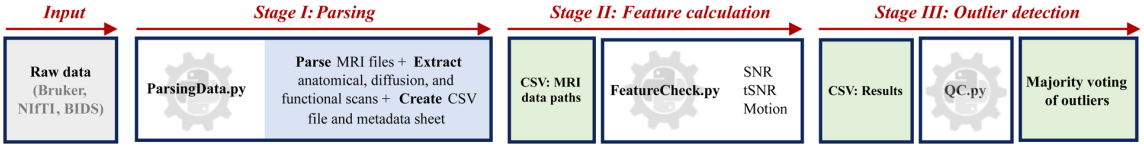



**Fig. 1.** Schematic overview of the AIDAqc pipeline. Starting with input data formats, three consecutive stages of automatic parsing, feature calculation, and multiple outlier detection techniques combined to a major vote for the outlier status. Note: naming convention—anatomical (all T1- or T2-weighted scans), diffusion (all diffusion-weighted, DWI, or diffusion tensor imaging, DTI, scans), and functional (fMRI, and rs-fMRI scans).

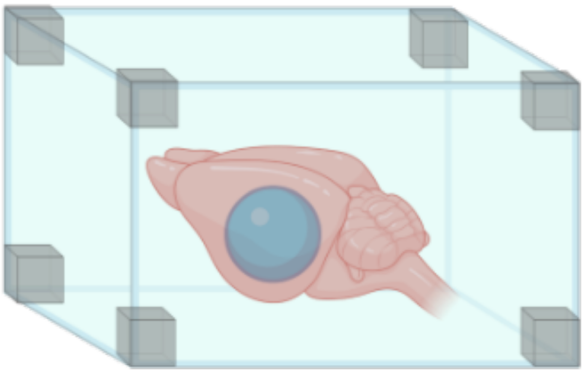


**Fig. 2.** SNR calculation without manually placed regions of interest (ROI). Graphical illustration of automated ROI selection to distinguish signal (blue sphere) from noise (gray cuboids), which are used to calculate the SNR.

of the noise (Henkelman, 1985; Kaufman et al., 1989). ROIs are placed automatically in a three-step procedure (Fig. 2).

1. **Center of intensity (COI):** To robustly determine the center of the object in the image, the coordinates of the COI are extracted as a starting point for step 2.
2. **COI sphere:** A sphere with the COI as its center is used to create a mask for averaging all the voxels within this sphere as a reference value for the "true signal." To ensure that the sphere does not extend beyond the signal volume of interest, a relative value for the sphere radius is set based on the image dimensions.
3. **Cuboids:** In all eight corners of the image volume with a size relative to the dimensions of the volume, cuboid ROIs are created. All voxels in these cuboids are averaged and used as a reference for the *noise*.

If *I* represents the three-dimensional image, then *S* and *C* are subsets of *I* representing the sphere and the cuboid space, respectively (Eq. 1a). *S(x,y,z)* and *C(x,y,z)* are image functions representing a specific voxel in *I*, defined by the coordinate *x,y,z* in the three-dimensional image space *I*. Upper limits of the subsets *S* and *C* are shown as $(i_s, j_s, k_s)$ and $(i_c, j_c, k_c)$, respectively (Eq. 1a). The SNR is calculated by dividing the mean intensity of the sphere by the standard deviation of the cuboids with *N* being the number of voxels in the image subset of the image *I* (Eq. 1b). The SNR value is commonly reported in decibel, dB (Eq. 1c). In the case of 4D diffusion images, the first b0 image volume is chosen to represent the three-dimensional image *I*.

$$(x,y,z),\ C(x,y,z) \subset I(x,y,z) \tag{1a}$$

$$\mu_S = \frac{\sum_{x=1}^{i_s}\sum_{y=1}^{j_s}\sum_{z=1}^{k_s} S(x,y,z)}{N_s}, \qquad \sigma_c = \sqrt{\frac{\sum_{x=1}^{i}\sum_{y=1}^{j}\sum_{z=1}^{k}\left(C(x,y,z)-\mu_c\right)^2}{N_c}} \tag{1b}$$

$$SNR_{standard} = \frac{\mu_s}{\sigma_c} = 20 \cdot \log\left(\frac{\mu_s}{\sigma_c}\right) \text{ dB}. \tag{1c}$$

*2.1.2.1.2. Chang method*. This approach uses a histogram analysis of the image signal intensity to calculate the distribution of noise (Chang et al., 2005). The strength of the method lies in the fact that it can be used for developing a processing workflow for large-scale MR data without the need for manual region selection. Here, this method is used to acquire a value for SNR based on Equation 2 with *E (z)* as the SNR calculated for each slice (z) with the Chang function (see details of the function in Chang et al. (2005)). *N* is the number of voxels in the image subset of the image *I*. In the case of 4D diffusion images, the Chang method (different to the standard method) creates an average SNR of all diffusion directions except the b0 image.

$$\mu_{slice}(z) = \frac{\sum_{x=1}^{N}\sum_{y=1}^{N} S(x,y,z)}{N_s}, \qquad \sigma_{chang}(z) = \text{Chang}\left(I(x,y,z)\right) \tag{2a}$$

$$E(z) = \frac{\mu_{\text{slice}}(z)}{\sigma_{\text{chang}}(z)} = 20 \cdot \log\left(\frac{\mu_{\text{slice}}(z)}{\sigma_{\text{chang}}(z)}\right) \text{dB} \quad (2b)$$

$$\text{SNR}_{\text{chang, Final}} = \frac{\sum_{z=1}^{Slices} E(z)}{N}. \quad (2c)$$

**2.1.2.2. tSNR.** The temporal signal-to-noise ratio (tSNR) is defined for sequences acquired over time (Welvaert & Rosseel, 2013). Therefore, there is an additional fourth dimension in a three-dimensional volume, here represented with the image function $I(x,y,z,t)$ in Equation 3. For this case, the center of intensity (COI) of the averaged volume over time is calculated and the sphere with the COI as its center is defined. The sphere image function as a subset of S is represented with $S(x,y,z,t)$. $\boldsymbol{N_s}$ is the number of voxels in the sphere. From a practical point of view, to calculate one quantitative value for the tSNR, first, a tSNR map is calculated (Eq. 3c) and the average of the tSNR over all voxels in the sphere is the final tSNR value.

$$S(x,y,z,t) \subset I(x,y,z,t) \quad (3a)$$

$$\mu_t(x,y,z) = \frac{\sum_{t=1}^{T} S(x,y,z,t)}{N_s}, \qquad \sigma_t(x,y,z) = \sqrt{\frac{\sum_{t=1}^{T}\left(S(x,y,z,t) - \mu_t(x,y,z)\right)^2}{N_s}} \quad (3b)$$

$$\text{tSNR}(x,y,z) = \frac{\mu_t(x,y,z)}{\sigma_t(x,y,z)} = 20\log\left(\frac{\mu_t(x,y,z)}{\sigma_t(x,y,z)}\right) \text{dB} \quad (3c)$$

$$\text{tSNR}_{\text{Final}} = \sum_{x=1}^{i_s}\sum_{y=1}^{j_s}\sum_{z=1}^{k_s} \text{tSNR}(x,y,z). \quad (3d)$$

**2.1.2.3. Ghosting and motion.** Mutual information (MI) was used as a metric for the identification of MRI ghosting artifacts (Fig. 3a, b). It is known that one of the sources of ghosting artifacts is the sensitivity of echo-planar imaging (EPI) scans to magnetic field inhomogeneities and an additive consequence of hardware-originated artifacts. The *Nyquist ghost* is one of the artifacts happening with EPI sequences where the ghost is located half of the field of view (FOV) away from the image (Buonocore & Gao, 1997; Reeder et al., 1997; Yang et al., 1996). Additionally, there are *discrete ghosts*, which occur in images where the periodic motion of the subject is present, like respiratory or cardiac motions. These regular motions often spread across the whole image in the form of ghosts emerging from the principles of the Fourier transform (Axel et al., 1986; Storey et al., 2002; Wood & Henkelman, 1985).

In practice, irrespective of the type of the ghost for a 3D image volume, the middle slice is automatically extracted. For 4D image data, that is, diffusion and functional scans, the average across the fourth dimension is used to create a representative 3D image. In the next step, the MI is calculated between the selected image and the shifted image by *n* voxels, with *n* ranging from 1 to N, with N being the size of the image in x or y directions. According to the MI theorem, the MI value decreases as the shifts increase because the pair of images fits less well. Considering a "ghost" has happened, along one of these shifted image versions, the main target of the image (in our case the brain) will match its ghost in the shifted image. This process will generate peaks in the MI as the shifts increase, which are automatically detected. If a peak exceeds a defined threshold, there is a high probability that it is due to a ghost.

Mutual information (MI) is also a sensitive measure of detecting translational motion between images (Godenschweger et al., 2016) and was here used to compare brain structures along the time dimension (Fig. 3c). To obtain a quantitative value for the severity of motion in the image, the Mutual Information (MI) was calculated between a reference image and subsequent images for the time dimension in functional scans and the direction dimension in diffusion scans. The choice of reference image depends on the number of repetitions in the fourth dimension. For smaller datasets with, for example, 10 repetitions, the 1st repetition (TR or b = 0, depending on the acquisition type) is taken as the reference. However, for larger datasets with more repetitions, the 10th repetition is used as the reference to ensure the signal has reached a steady state and is free from initial transient effects. This practice helps to provide a more reliable measure of motion by avoiding the instability often present in the first few repetitions. In the case of no or small amounts of motion (Fig. 3d), the MI is close to 1 and relatively stable over different time points or directions. In the case of severe motion (Fig. 3e), there is a characteristic drop in the MI (Fig. 3c). Motion severity as a single quantitative value for each scan reflects the standard deviation of the MI vector. To reduce the processing time for 4D image volumes, MI is only calculated for the slice with the highest average intensity over time.

### *2.1.3. Stage III: Outlier detection*

After the extraction of quality features, AIDAqc applies one univariate and four multivariate outlier detection methods using a selection of statistical and machine-learning (ML) algorithms: *interquartile range (IQR)*, *one-class support vector machine (ocSVM)*, *isolation forest (IF)*, *local outlier factor (LOF)*, and *elliptic envelope (EE)*.

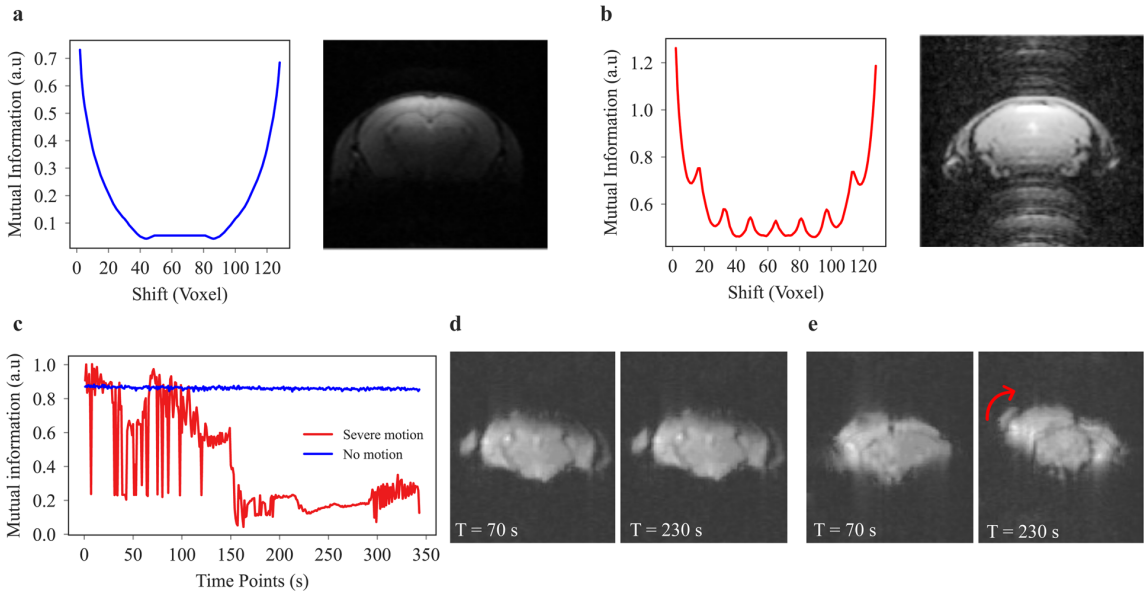


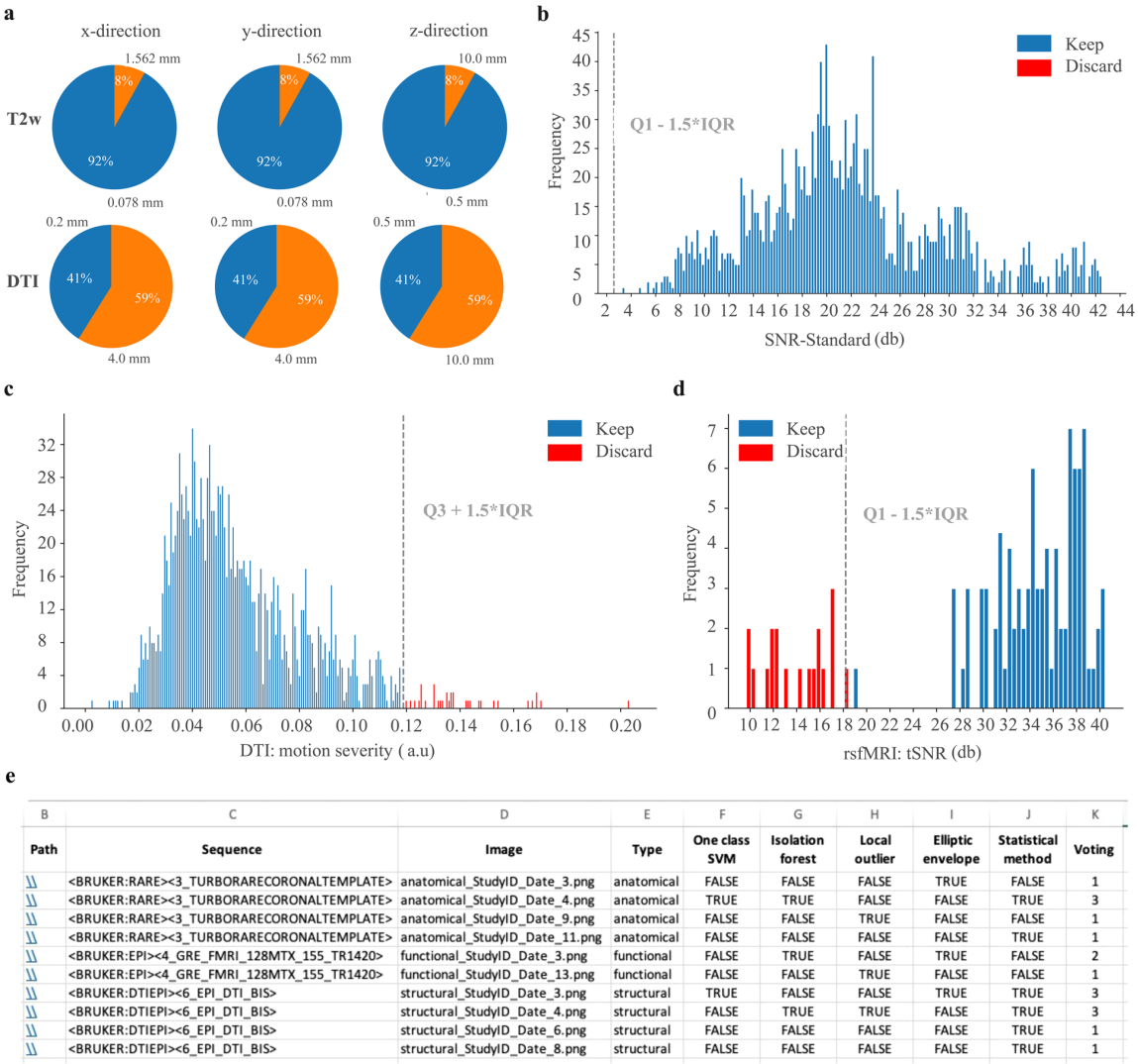

**Fig. 3.** Mutual information (MI) to detect ghosting artifacts and motion. (a) Image without ghosting: smooth and bowl-shaped MI curve indicating the whole cycle of image shifting. An initial decrease due to an increasing mismatch between the reference image and the shifted versions is followed by an increase to its initial state when the shifted image arrives at the reference. (b) Complementary MI curve for an image with ghosting. MI peaks at various shifts throughout the cycle. (c) Representative MI was plotted for consecutive time points of two rs-fMRI scans, one with strong motion artifacts and the other with no motion artifacts. (d) There is near to no motion visually detectable in the receptive images of time = 70 versus 230 s. (e) The motion can be detected visually in the receptive images of time = 70 versus 230 s. A mixture of translational and rotational motions shifts the image approximately 20 voxels.

This approach provides a diverse range of perspectives on the potential outlier status and overcomes the biased approach of only applying one algorithm, which is not suited to detect outliers from a broad range of datasets. The IQR identifies outliers by measuring data spread and is particularly effective for datasets with non-normal distributions (Vinutha et al., 2018). The ocSVM specializes in anomaly detection by defining hyperplanes, making it robust to high-dimensional datasets and capable of capturing complex patterns. The LOF detects anomalies based on local density deviations, offering sensitivity to local variations in data density and robustness to noise, complementing the limitations of ocSVM in handling local anomalies (Budiarto et al., 2019). In scenarios where the ocSVM might struggle with nonlinear separations, Isolation Forest (IF) excels in isolating anomalies by constructing decision trees, providing a computationally efficient alternative (Mohammed et al., 2021). The EE approach offers probabilistic outlier detection by fitting robust covariance estimations, making it suitable for multivariate datasets with elliptical distributions (Ashrafuzzaman et al., 2020).

Outcomes from this multialgorithm approach are aggregated using majority voting on a per-image basis. Each image is assigned a score ranging from 0 (not detected as an outlier in any algorithm) to 5 (reported as an outlier in all algorithms). This enables flexible thresholding on a per-study basis (rather than a singular, binary "outlier" assignment) and minimizes the potential for false positives and negatives in a single-algorithm approach.

#### *2.1.4. Scan reports and automatically created files*

AIDAqc provides detailed reports for the user to get an overview of the dataset and to verify the outlier detection in machine-readable CSV files. The reports include information on spatial resolution and are presented in pie charts and distribution plots (Fig. 4). Positive votes that an algorithm has detected an image as an outlier are included in the "*voting.csv*" table, together with the evaluations provided by each outlier approach. This way, the user can identify which algorithm was important to flag the data as an outlier and investigate the reasons further if necessary. Next to the scan report, the single-slice inspection function automatically saves the middle slice images in the respective output folder from each anatomical, diffusion, and functional scan, respectively. The images are named based on a unique scheme, that is, sequence type followed by a number that is incremented for each image and the subject name. Additionally, the




| Path | Sequence | Image | Type | One class SVM | Isolation forest | Local outlier | Elliptic envelope | Statistical method | Voting |
|---|---|---|---|---|---|---|---|---|---|
| \\ | <BRUKER:RARE><3_TURBORARECORONALTEMPLATE> | anatomical_StudyID_Date_3.png | anatomical | FALSE | FALSE | FALSE | TRUE | FALSE | 1 |
| \\ | <BRUKER:RARE><3_TURBORARECORONALTEMPLATE> | anatomical_StudyID_Date_4.png | anatomical | TRUE | TRUE | FALSE | FALSE | TRUE | 3 |
| \\ | <BRUKER:RARE><3_TURBORARECORONALTEMPLATE> | anatomical_StudyID_Date_9.png | anatomical | FALSE | FALSE | TRUE | FALSE | FALSE | 1 |
| \\ | <BRUKER:RARE><3_TURBORARECORONALTEMPLATE> | anatomical_StudyID_Date_11.png | anatomical | FALSE | FALSE | FALSE | FALSE | TRUE | 1 |
| \\ | <BRUKER:EPI><4_GRE_FMRI_128MTX_155_TR1420> | functional_StudyID_Date_3.png | functional | FALSE | TRUE | FALSE | TRUE | FALSE | 2 |
| \\ | <BRUKER:EPI><4_GRE_FMRI_128MTX_155_TR1420> | functional_StudyID_Date_13.png | functional | FALSE | FALSE | TRUE | FALSE | FALSE | 1 |
| \\ | <BRUKER:DTIEPI><6_EPI_DTI_BIS> | structural_StudyID_Date_3.png | structural | TRUE | FALSE | FALSE | TRUE | TRUE | 3 |
| \\ | <BRUKER:DTIEPI><6_EPI_DTI_BIS> | structural_StudyID_Date_4.png | structural | FALSE | TRUE | TRUE | FALSE | TRUE | 3 |
| \\ | <BRUKER:DTIEPI><6_EPI_DTI_BIS> | structural_StudyID_Date_6.png | structural | FALSE | FALSE | FALSE | FALSE | TRUE | 1 |
| \\ | <BRUKER:DTIEPI><6_EPI_DTI_BIS> | structural_StudyID_Date_8.png | structural | FALSE | FALSE | FALSE | FALSE | TRUE | 1 |

**Fig. 4.** Example of the automatically created outputs as part of the scan report for a representative dataset. Graphs and tables summarizing the (a) spatial resolution (in x-, y-, and z-direction), (b) distribution of anatomical SNR, (c) distribution of motion based on the standard deviation of the mutual information calculated for diffusion scans, (d) distribution of the tSNR for functional scans, and (e) file location, metadata details, and individual outlier votings as listed in the votings. csv file (representative dataset: 94_m_As). Note: in (b–d) red bars indicate outliers based on the statistical definition of quartiles (Q) added by the 1.5-fold interquartile range (IQR). The dashed gray line indicates the predefined outlier threshold. In (e) the file paths were shortened by the placeholder StudyID and date.

image file names are listed in the final feature calculated CSV files.

This information fulfills the purpose to (i) get a first overview of the dataset and to identify scans with different voxel sizes—especially when accidentally a wrong sequence or parameters were used. This feature facilitates quick browsing of images without the need to open each dataset separately in a different program. The output is restricted to the middle slice, which is particularly valuable as it most likely covers larger areas of the subject. While the pipeline is running, users can observe how these images are created one after the other. This allows for quick visual inspection of the subjects. If the orientation of the slice differs in one of these images, the users will promptly recognize that some settings for that subject were different from the usual pattern compared with the others. Furthermore, noisy images are easily discernible at first glance, further enhancing the efficiency of the inspection process. This method can be regarded as a basic but convenient complement to the pipeline. In addition to BIDS datasets, which come with specific metadata and unique file and sequence names, detailed

listing as *csv* files is implemented in AIDAqc for Bruker datasets as well. Here, the sequence names, which have been assigned manually by the user, are listed next to the sequence type (Fig. 4e and Supplementary Fig. 9). Further practical instructions on how to interpret AIDAqc results are documented online (https://github.com/Aswendt-Lab/AIDAqc).

### 2.2. Artifact simulation to evaluate the reliability of quality metrics and qutlier detection

#### *2.2.1. Quality metrics*

To evaluate the reliability of individual quality control (QC) metrics, a simulation of noise and motion was performed using custom-written code in Python (version 3.11) using the scikit-image library (van der Walt et al., 2014) and Matlab (Matlab Version R2023a, The MathWorks Inc., Natick, USA), respectively. Anatomical and functional scans from the combined datasets 94_m_As and 94c_m_As (Table 1) were used.

All of the subjects were induced with noise and then AIDAqc was applied on ~60 subjects (30 original, 30 noise induced). This was done to prove the reliability of each metric.

For noise simulation, a constant amount of noise (variance of 0.2) was added to every subject's anatomical scans. Additionally, a gamma function was applied to reduce the overall brightness of the images, mimicking the shadows present in MR images (gamma value of 0.6). For motion simulation, the spatial image dimensions of the functional scans were manipulated by adding variable translational motion along the time dimension. Three different types of noise were added: Gaussian, salt and pepper (S&P), and speckle noise. Each type of noise serves a specific purpose: Gaussian noise simulates general background noise, salt and pepper noise represents sudden, isolated artifacts, and speckle noise captures spatial variations in tissue texture or structure (Khan et al., 2019). Additionally, for the functional scans, motion was induced by shifting the image volume over the time course, adding random degrees of motion to each subject's image data. SNR and motion severity were computed both before and after adding noise and motion, and AIDAqc's results regarding SNR in both the Chang and standard methods, as well as motion severity, were calculated.

#### *2.2.2. Outlier detection*

To assess the reliability of the outlier detection, we applied the same as for quality metrics to introduce noise and motion, here only for one randomly selected scan from each sequence instead of all subjects, effectively creating single scans with simulated artifacts in the whole dataset (combined datasets 94_m_As and 94c_m_As). Thus, one scan at a time was induced with artifacts and then AIDAqc was applied 30 times on ~31 subjects (30 original, 1 artifact induced) each time with another scan as outlier.

### 2.3. Validation of outlier detection

#### *2.3.1. Initial validation approach*

The validation of the outlier detection was conducted in two phases, addressing different aspects. In the initial phase, five experienced users manually rated the image quality of all datasets (Table 1). The individual images from the "manual_slice_inspection" folder ("static" images as part of the automatically generated output derived from the 3D/4D images), encompassing anatomical, diffusion, and functional MRI sequences, were examined. Raters identified potential outliers, that is, "bad quality data," based on subjective evaluations of the image in comparison with all

**Table 1.** Summary of datasets used for developing AIDAqc.

| # | Repository | Datasets | Sequences | Data format | # | Repository | Datasets | Sequences | Data format |
|---|---|---|---|---|---|---|---|---|---|
| **1** | Aswendt | **94_m_As** | T2w, fMRI, DWI | Bruker | **13** | Rivera-Olvera | **117_m_Ri** | T2w, DWI, fMRI | Bruker |
| **2** | Aswendt | **94c_m_As** | T2w, fMRI, DWI | Bruker | **14** | Sta Maria | **7_m_St** | T2w, DWI | Nifti |
| **3** | Boehm-Sturm | **7_m_Bo** | T2w, DWI | Bruker | **15** | Selim | **7_m_Se** | T2w, T1w, DWI | Bruker |
| **4** | Carnevale | **7_m_Ca** | T2/DWI | Bruker | **16** | Selim | **7_r_Se** | T2w, T1w, DWI | Bruker |
| **5** | Franx | **94_r_Fr** | T2w, DWI, T1w | Nifti | **17** | Soria | **7_r_So** | T2w, fMRI | Nifti |
| **6** | Hekmatyar | **7_h_He** | T2w, fMRI | Nifti | **18** | Van Leeuwen | **94_m_Va** | T2w, DWI | Nifti |
| **7** | Kurniawan | **164_m_Ku** | T2w, DWI | Bruker | **19** | Vrooman | **117_m_Vr** | T2w, fMRI | Nifti |
| **8** | Micotti | **7_m_Mi** | T2w, DWI | Bruker | **20** | Wenk | **94_m_We** | T2w, fMRI, DWI | Bruker |
| **9** | Muñoz-Moreno | **7_rab_Mu** | T1w, DWI | Nifti | **21** | Wenk/Goncalves | **94_r_We** | T2w, fMRI, DWI | Bruker |
| **10** | Brinton | **7_m_Br** | T2w, DWI | Nifti | **22** | Wenk/Michalek | **94_g_We** | T2w, fMRI, fMRI, DWI | Bruker |
| **11** | Ramos-Cabrer | **117_m_Ra** | T2w, DWI, fMRI | Bruker | **23** | Longo | **7_m_Lo*** | T2w, DWI | Bruker |
| **12** | Reichardt | **94_m_Rei*** | T2w | Bruker | | | | | |

other images within one dataset. The level of agreement among raters when evaluating the same set of data for classifications, also known as the inter-rater agreement, was analyzed and compared with the results generated by AIDAqc. The classification features true/false positive (TP/FP) and true/false negative (TN, FN), sensitivity, specificity, and accuracy were calculated for quantitative comparison (Supplementary Table 3).

#### *2.3.2. Revised post hoc validation*

Because of significant inter-rater variability in the initial approach, a revised and more comprehensive validation strategy was implemented. In this second phase, six experienced users conducted a detailed manual screening of the complete 3D/4D image files that AIDAqc identified as high-threshold outliers (AIDAqc voting thresholds of 4 and 5). Thresholds were introduced to create cutoff scenarios, in which the majority voting results are grouped into very strict (high threshold) or not as strict (low threshold) outliers. In a systematic way, different thresholds for the number of manual raters and AIDAqc outlier algorithms were compared to find an optimum (Supplementary Material: "Role of the rating threshold for the agreement between manual raters and AIDAqc").

### 2.4. Statistics and visualization

The statistical analyses were conducted using GraphPad Prism version 9.5.1 (GraphPad Software, Boston, MA, United States, www.graphpad.com) and custom code written in Python version 3.11.4, with the libraries statsmodels (Seabold & Perktold, 2010) and scipy.stats (Virtanen et al., 2020).

Fleiss' kappa score, a statistical measure utilized to assess the inter-rater agreement between multiple raters when categorizing items into different groups, was used to evaluate the reliability of manual ratings of image quality as "good" or "bad." The interpretation of Fleiss' kappa scores indicates various levels of agreement: scores less than 0 indicate poor agreement, 0.01–0.20 suggest slight agreement, 0.21–0.40 imply fair agreement, 0.41–0.60 signify moderate agreement, 0.61–0.80 denote substantial agreement, and scores between 0.81 and 1.00 represent almost perfect agreement.

Additionally, various data analysis and visualization tasks were done using the Python libraries pandas, numpy, matplotlib, seaborn, and nibabel. DTI and rs-fMRI post-processing for selected datasets were conducted using an atlas-based approach with our in-house software AIDAmri (https://github.com/Aswendt-Lab/AIDAmri). Briefly, AIDAmri performs brain extraction and data correction (slice time correction, motion correction, spatial smoothing with full width at half maximum, and high-pass filtering for rs-fMRI), followed by a multistep registration with the Allen Mouse Brain Reference atlas (CCF v3). Functional connectivity is derived by a seed-based Pearson correlation among all atlas regions. Structural connectivity matrices are produced with deterministic fiber tracking using DSI Studio; for details see Pallast et al. (2019).

### 2.5. Ethics statement

The data corresponding to 94c_m_As and 94_m_As from Table 1 were acquired in strict adherence to the ARRIVE guidelines for reporting in vivo animal experiments and the IMPROVE guidelines for stroke animal models, as recommended by Kilkenny et al. (2010) and Percie du Sert et al. (2020). These datasets were processed following comprehensive ethical protocols approved by the Landesamt für Natur, Umwelt and Verbraucherschutz North Rhine-Westphalia, Germany, under animal protocol numbers 84–02.04.2016. A461 and 84–02.04.2014.A305. Further ethical clearance was obtained from the Gothenburg Ethics Committee, Sweden, under animal permit number 1551/2018. In the case of data sourced from other groups within the research, all animal experiments were approved by the respective local ethical authorities.

## 3. RESULTS

### 3.1. AIDAqc workflow

AIDAqc was designed as a three-stage open-source pipeline of Python scripts (https://github.com/Aswendt-Lab/AIDAqc), which requires minimal user input (Fig. 1). The installation and application of AIDAqc were independently tested by four scientists and optimized using the datasets (see next section). T1, T2, DTI, and fMRI scans can be used—independent of the underlying MR sequence (spin echo, gradient echo, FLASH, etc.)—as input. AIDAqc accepts a vendor-specific data format by Bruker Biospin and the widely used vendor-independent BIDS/NIfTI data format. No prior data sorting is necessary as the main input folder will be iteratively searched. Each of the three stages can be run individually but the result of the third stage depends on the sequential processing of stages 1 and 2, respectively. The results are reported automatically in machine-readable csv files as well as figures and image sequences.

### 3.2. Dataset statistics

To test AIDAqc on a wide range of MRI setups, sequences, and animal models, 65 international MRI experts were invited as part of an initiative from the STANDARD working

group of the European Society for Molecular Imaging (documented on GitHub: https://github.com/Aswendt-Lab/MRI_Standardization_AIDAqc).

Finally, n = 19 laboratories provided a total of n = 23 small animal datasets to provide datasets consisting of about 2,600 anatomical, 1,200 diffusion, and 700 functional scans. These datasets had not been prescreened for outliers and were used to test and validate AIDAqc. In applying AIDAqc to test its efficacy, each of the 23 datasets was processed independently. This separate analysis allowed us to maintain the integrity of the outlier detection process, ensuring that any identified quality issues were specific and relevant to each dataset's unique characteristics. All data are publicly available (see section 5: "Data and Code Availability"). The majority of datasets were obtained with mice (68.2%) and rats (18.2%), and single studies used hamsters, rabbits, and gerbils (Fig. 5a). MRI was acquired with a

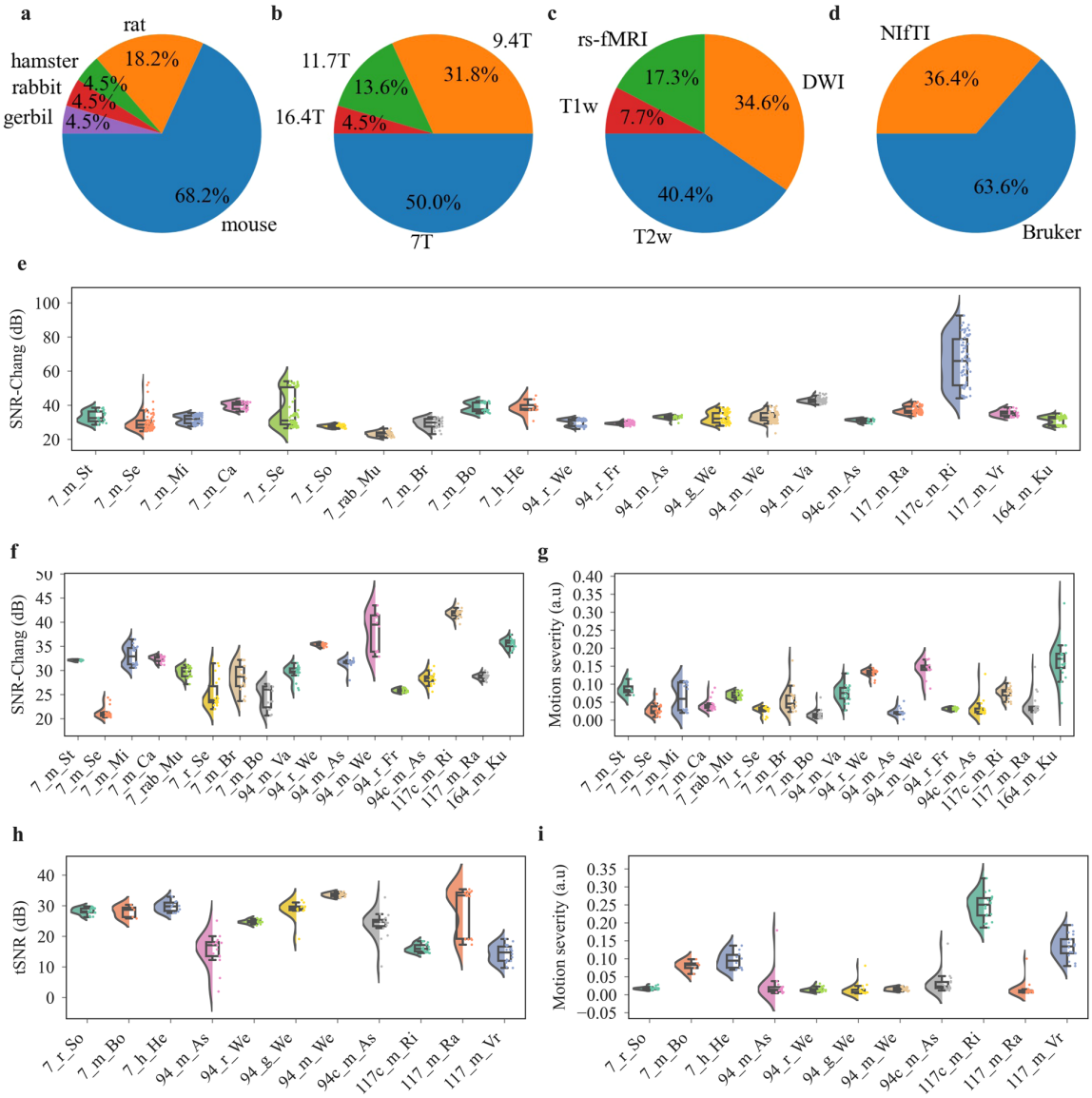



**Fig. 5.** Summary of dataset distribution regarding basic statistical information (a–d) and quality features (e–i). Statistical distribution between all datasets regarding the (a) species, (b) field strength, (c) sequence, and (d) data format. In (d) "Bruker" refers to the Bruker raw data format and NIfTI to the Neuroimaging Informatics Technology Initiative data format. Summary of quality features for all (e) anatomical scans, (f–g) diffusion scans, and (h–i) functional scans. Note that the distribution of datasets for (e–i) is different, as not all datasets contain diffusion and functional scans in addition to anatomical scans.

broad range of hardware in terms of field strength (7–16.4T) and receive/transmit coil arrangement (Fig. 5b). Most studies were conducted at 7T (50%) followed by 9.4T (31.8%) using anatomical (T1- or T2-weighted MRI) and diffusion-weighted imaging (DWI), as well as resting-state functional MRI (rs-fMRI) (Fig. 5c). Most data were provided in the Bruker raw data format (63.6%) next to the NIfTI format of which only six datasets were BIDS-compliant (Fig. 5d).

The datasets further varied substantially in terms of FOV, matrix size, and image resolution (see Supplementary Fig. 1 and extended version of Table 1 in Supplementary Table 1). In addition, we also applied AIDAqc in a proof-of-concept approach to two abdominal datasets (Supplementary Fig. 2) and previously published large repositories containing mouse and rat data (Supplementary Table 2). The AIDAqc output is summarized in the online repository (https://doi.gin.g-node.org/10.12751/g-node.q82cjj/), featuring detailed reports for SNR, tSNR, and motion severity.

The mean SNR in anatomical scans (neuroimaging datasets as listed in Table 1), as calculated using the Chang method, was 32.1 dB ± 4.6 dB (Fig. 5e). For the diffusion scans, the mean SNR was 28.4 dB ± 5.0 dB (Fig. 5f), and in functional scans, we measured 27.4 dB ± 5.9 dB tSNR (Fig. 5h). Temporal SNR across functional MR sequences varied with large differences in variance between specific studies, for example, *117_m_Ra*, which was acquired with two different sequences, visible as two tSNR populations (Fig. 5f). The motion severity was rather stable for 16 out of 18 diffusion and 9 out of 11 functional studies, respectively (Fig. 5g, i). The motion severity varied between 0.017 (7_m_Bo) and 0.217 (7_m_Lo) in diffusion scans and between 0.002 (117_m_Ra) and 0.323 (117c_m_Ri) in functional scans, respectively.

Initially, the intention was to also apply AIDAqc to a published large, multicenter dataset to demonstrate the tool's adoption and efficacy (see Section 4.3). However, it became apparent that these datasets had already undergone screening for poor-quality data, rendering them unsuitable for our intended purpose. As a result, we also applied the tool to all NIfTI datasets from Table 1 in a *collective manner* (Supplementary Fig. 8), and outliers in addition to the normal statistical outliers were registered in the *voting.csv* through the majority vote approach.

### 3.3. Automated SNR measurements

#### *3.3.1. Comparison of SNR across datasets*

Given the differences in MRI hardware (scanner, coils), sequences, shimming procedure, and animal experiment, here it was not meaningful to statistically compare all individual datasets. Importantly, the effects of differences in MRI hardware (gradients, coils) overrode the positive correlation of field strength and SNR (Fig. 5). In studies selected for a relatively homogeneous voxel volume of 9.5 nl ± 1.2 nl, there were significant differences in SNR values for different magnetic field strengths, also for selected studies with the same magnetic field strength and an increasing voxel volume, there were significant differences observed. As expected, SNR was scaled with magnetic field strength and voxel volume (Supplementary Fig. 3). Within studies including both anatomical and functional scans, SNR and tSNR correlated positively (Spearman $r = 0.57$, $p < 0.001$) (Fig. 6a). A similar positive correlation was found between anatomical SNR and diffusion SNR ($r = 0.30$, $p < 0.001$) (Fig. 6b).

We further quantitatively compared the two SNR methods (SNR-Standard vs. SNR-Chang) for all anatomical datasets (Supplementary Fig. 4). The SNR methods correlated positively (Spearman $r = 0.16$, $p < 0.001$). For a randomly chosen selection of five datasets, we further validated that the quantitative SNR values represent the qualitative "visual" appearance of the images, that is, higher SNR corresponds to more brightness and contrast in the image (Fig. 7a, b, image #3–4). The visual comparison (Fig. 7b) also suggests that SNR-Chang is less

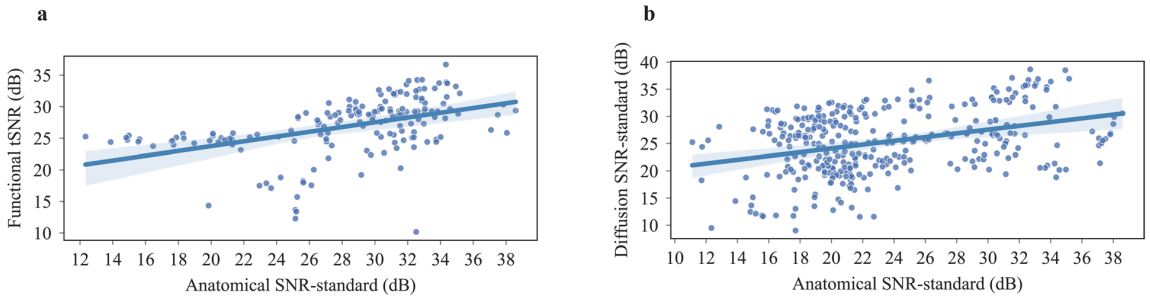



**Fig. 6.** Scatterplots (a) Spearman correlation between functional tSNR and anatomical standard SNR. (b) Spearman correlation between diffusion standard SNR and anatomical standard SNR of subjects who had both scans.

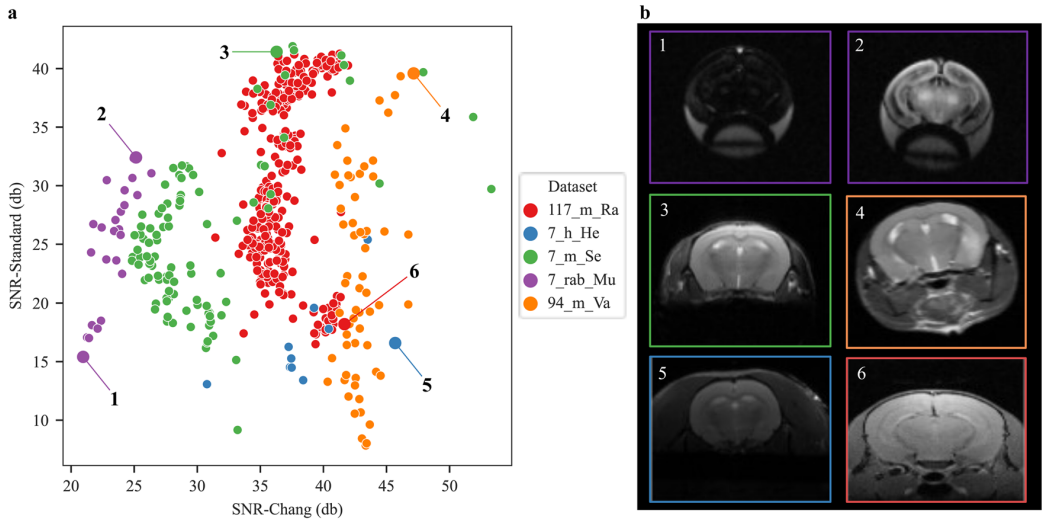



**Fig. 7.** Scatterplot of the relationship between SNR values calculated using the Standard and Chang method. (a) SNR-Standard and SNR-Chang comparison for five randomly selected datasets (7_h_He, 7_m_Se, 7_rab_Mu, 94_m_Va, and 117_m_Ra). (b) Selected images from the different datasets reflect similar SNR (examples 1 and 4) and different SNR (examples 2, 3, 5, and 6).

sensitive for images with brightness differences (Fig. 7b, #1–2). The SNR-Standard method, however, can be falsified when the ratio of subject to field of view is high (Fig. 7b, #5–6).

### 3.4. Artifact simulation to evaluate the reliability of quality metrics and outlier detection

#### *3.4.1. Quality metrics*

The analysis of data with simulated (added) Gaussian, salt and pepper, and speckle noise (Fig. 8a, b) revealed a significantly lower SNR compared with the original image using both SNR methods. Similarly, adding motion to the rs-fMRI scans significantly increased the motion severity index (Fig. 8c). Welch's *t*-test was applied for comparison, indicating significant differences ($p < 0.001$) between the modified and the original group data for each metric.

#### *3.4.2. Outlier detection*

In the approach to showcase the reliability of outlier detection and majority voting, a significant difference in majority votes between the artifact-induced and original images was detected ($p < 0.001$), with the artifact-induced images consistently receiving higher majority votes.

Representative examples show that the majority vote was consistently higher in the artifact-induced images than in the original (Fig. 8d). For the anatomical image, the number of ML algorithms flagging the original image as an outlier was 1, whereas in the contaminated counterpart image, it increased to 4 (Fig. 8e). Similarly, in the case of the diffusion sequence, the majority vote was 0 for the original image and increased to 4 for the contaminated image (Fig. 8f). In functional sequences, the contaminated image was flagged with a majority vote of 3 compared with 1 for its original counterpart (Fig. 8g).

### 3.5. Validation of outlier detection

#### *3.5.1. Initial validation approach*

To gain a ground truth for the comparison of classification, we worked with the votings for all datasets from n = 5 experts with 5–10 years of experience in small animal MRI. The experts were asked to identify bad-quality images in the 23 datasets including 4,452 images. The assessment was purely objective and referred to visible artifacts or noise in the images, without any further specific step-by-step evaluation. As the main result, the mean inter-rater agreement for all data (Fig. 9) was low across anatomical, functional, and diffusion scans (mean Fleiss Kappa score $0.2 \pm 0.2$, $0.2 \pm 0.2$, and $0.1 \pm 0.2$). For the datasets with a substantial inter-rater agreement, for example, 7_h_He and 117_m_Ra (Fleiss Kappa score >= 0.6), and with no inter-rater agreement, for example, 7_m_Br and 94_m_Rei (Fleiss Kappa score 0.0), respectively, there was visually no related identifiable image feature or pattern.

In addition, we conducted a comprehensive analysis using a confusion matrix (Supplementary Material Table 3 and related text). This analysis was based on calculations

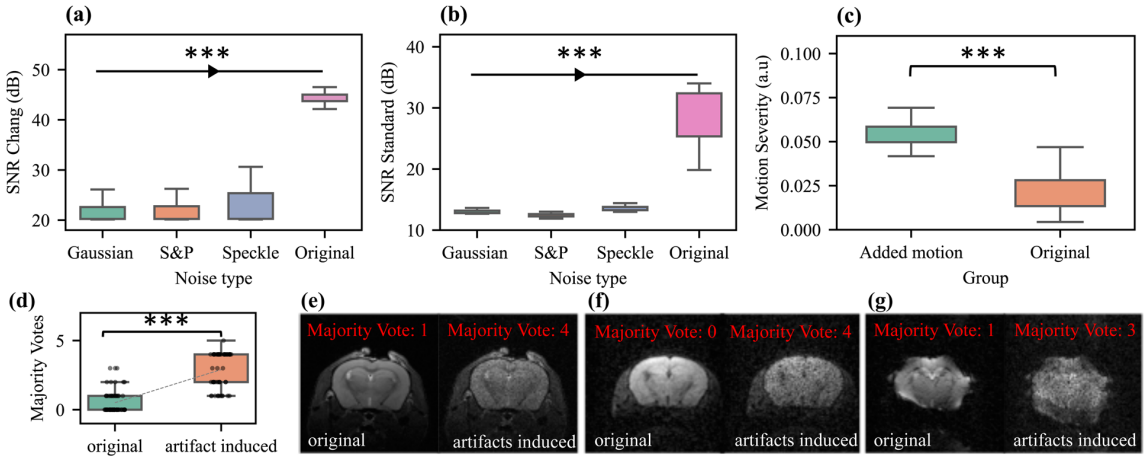



**Fig. 8.** Artifact simulation to evaluate the reliability of quality metrics and outlier detection. Comparison of signal-to-noise ratio (SNR) and motion severity metrics between the induction of different noise types and motion versus the original image in (a) SNR-Chang, (b) SNR Standard Method, and (c) Motion Severity. The boxplots illustrate the distribution before (original) and after noise addition and motion induction. Significant differences (*** indicates $p < 0.001$) were observed between each noise-induced group and the original group for using Welch's *t*-test. The evaluation of majority votes on artifact-induced images revealed significantly higher votes compared to the original images (*** indicates $p < 0.001$). (d) Consistent higher majority votes in the artifact-induced images compared to the original. Representative anatomical (e), diffusion (f), and functional (g) images before and after addition of artifacts with related majority votes.

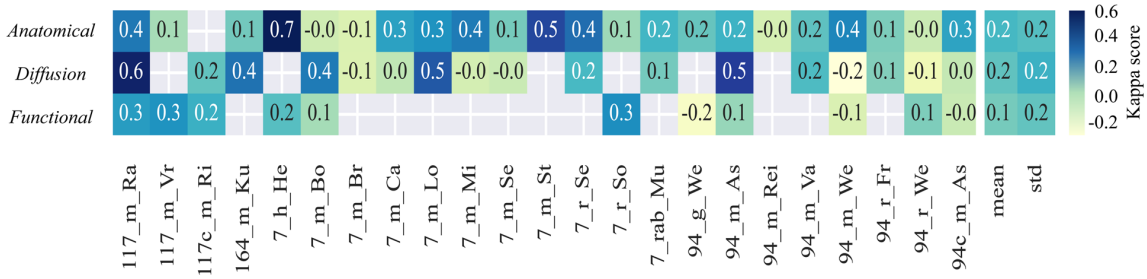



**Fig. 9.** Inter-rater agreement. Fleiss kappa scores were calculated separately for each dataset, serving as an indicator of inter-rater agreement. The last two columns present the mean and standard deviation (std).

of sensitivity, specificity, and accuracy, considering various threshold scenarios between the AIDAqc and human raters (Supplementary Fig. 5). Here, a threshold relates to a minimal number of positive votes. The threshold = 3 for AIDAqc voting refers to the majority vote of 3, that is, at least three algorithms labeled the data to be of bad quality. Similarly, the threshold = 3 for the manual raters indicates that a file was considered of bad quality by at least three of the expert raters. The highest sensitivity ($0.20 \pm 0.28$), that is, out of all the actual "bad-quality" datasets, how many were identified by AIDAqc as "bad-quality" data, was reached with the AIDAqc threshold set to 1 and the manual rater threshold set to 4. For this threshold combination, the overall classification accuracy ($0.79 \pm 0.13$) and specificity ($0.81 \pm 0.12$), that is, identification as "no bad quality," were high. Despite these efforts, the low inter-rater agreement rendered the results less definitive.

### 3.5.2. *Revised post hoc validation*

In response to the low inter-rater agreement observed in the initial validation phase, a second, more focused validation approach was implemented. In this revised strategy, the same experts were tasked with validating only those images flagged as high-threshold outliers (voting thresholds of 4 and 5) by AIDAqc. The precision measurements, which reflect the proportion of images correctly identified as outliers by AIDAqc and confirmed by the experts, showed an average precision of $70.72 \pm 9.9\%$, with an adjusted precision of 72.23% when unsure cases were excluded.

### 3.6. Relevance of outlier removal for data postprocessing

To assess the importance of identifying poor-quality data through AIDAqc in subsequent postprocessing steps, we processed functional and diffusion datasets from two representative control groups, 94c_m_As and 94_m_As, using our in-house software AIDAmri, as established in a prior study (Pallast et al., 2019). These datasets were chosen because they represent control animals, that is, no disease or intervention model, to provide a consistent baseline for assessing functional and structural specificity. In terms of SNR, the functional specificity calculated for the merged dataset (Fig. 10a) illustrates how connectivity metrics can differ significantly even when images appear visually similar at first glance. We assessed functional specificity by computing correlations between specific and nonspecific ROIs. The X-axis of the scatter plot in Figure 10a details the specific ROIs, which include the left and right primary somatosensory barrel fields, expected to show high correlations. The Y-axis represents the nonspecific ROI, formed by combining the anterior cingulate area with the left primary somatosensory barrel field, expected to show lower correlations (Grandjean et al., 2023). High values for the nonspecific ROI might suggest problems either in the unprocessed data or during the processing steps. Adjacent to the scatter plot, the subject with a majority vote of 5 and one other subject not flagged by AIDAqc are shown representatively. In case of a majority vote of 5, a complete loss of signal in the last 25 repetitions was detected, potentially due to a gradient malfunction, leading to incorrect bias field correction and poor atlas registration (Supplementary Material Fig. 7). Among the subjects from the merged datasets, this subject's data confirm that image quality issues predominantly originate from the data itself, not the processing steps, as indicated by AIDAqc's findings. This emphasizes that while high correlations for nonspecific ROIs could stem from various sources, in this instance, the problems are specifically related to data

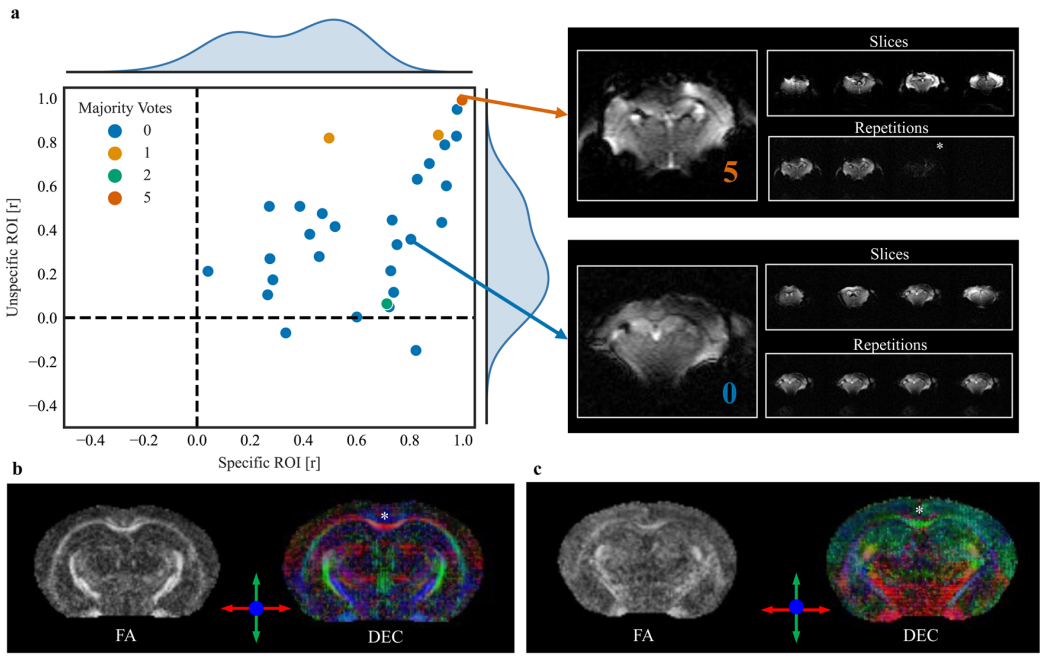



**Fig. 10.** Functional specificity and diffusion tensor imaging analysis for quality assessment. (a) Functional specificity analysis comparing the functional connectivity with the primary somatosensory barrel field relative to the contralateral homotopic area and the anterior cingulate area (ACA). The X-axis represents the correlation to the specific ROI (contralateral primary somatosensory barrel field), and the Y-axis represents the correlation to the unspecific ROI (ACA). Expected results include high correlations for the specific ROI and low correlations for the nonspecific ROI. Data points across the dataset reveal correlation values for each ROI, with outliers, identified by AIDAqc with a majority vote, showing unusually high correlations for the nonspecific ROI. This suggests that these outliers likely represent poor-quality data, which can adversely affect functional connectivity measures. (b, c) Two different examples of diffusion images are classified as good versus bad quality using AIDAqc. Whereas similar anatomical structures can be identified in the FA map, the diffusion tensor (displayed as direction encoded color map) shows very different and anatomically incorrect tensor orientations, for example, in the corpus callosum (*) connecting both hemispheres with horizontal (red) fibers crossing the brain midline.

quality. In diffusion tensor imaging data (Fig. 10b, c), artifacts may not be detectable by examining single slices or directions alone. Instead, diffusion analysis can produce markedly different results even when visual image quality appears only slightly varied. For instance, in the data identified as an outlier (Fig. 10c), marked by a majority vote of 3 out of 5 algorithms, the diffusion tensor distribution showed a more random pattern and less anatomical coherence compared with the good-quality dataset. Notably, key structures like the corpus callosum, a major fiber tract connecting both hemispheres, were detected in the good-quality image but not in the outlier data. These observations are supported by a comprehensive qualitative comparison across all subjects from the two datasets, showing a higher majority vote of 2 and 3 related to suboptimal FA (fractional anisotropy) maps. Extended qualitative comparisons of diffusion-weighted images and corresponding quality metrics are detailed in Supplementary Figure 6.

## 4. DISCUSSION

AIDAqc was developed with the aim of automatically identifying bad-quality data from large and heterogeneous small animal MRI datasets. Although classification was implemented as an automated selection using statistics and machine learning, it should not act as a black box. On the contrary, it was designed to report the classification transparently and flag poor-quality data that need further inspection by the user, for example, it can still be useful for a particular purpose.

To achieve these goals, we collaborated with n = 19 international MRI laboratories, which shared mouse, rat, and gerbil datasets. This collaboration ensured a diverse testing environment, encompassing a wide range of scanner hardware, imaging protocols, and experimental conditions (a detailed summary of the datasets can be found in Supplementary Material Fig. 1 and Table 1). In an iterative process, the AIDAqc workflow was adjusted and optimized to facilitate the processing of all acquired datasets. The findings of this study highlight the potential for improvement in quality control, as well as the benefits of automating the process, thereby avoiding manual interventions and reducing related errors. The AIDAqc pipeline was developed in Python as an open-source project, to allow full flexibility to include other metrics or automated calculations and integrate AIDAqc in other pipelines by the small animal imaging community.

### 4.1. The importance of QC/QA in small animal MRI

Conceptually, QC and QA describe ongoing efforts to ensure the stability of the MRI scanner and the assurance of homogeneous image quality (Sreedher et al., 2021). Small animal MRI has not adopted clinical and industry standards so far, to a certain extent because of the belief that such standards would limit scientific freedom and innovation (Wijnen et al., 2023). In contrast, we argue here in line with the community-driven standardization initiatives as part of the imaging societies ESMI, ESMRMB, and ISMRM, which standards are required to enhance the reliability and reproducibility of all neuroimaging studies and are necessary in the translation of preclinical to clinical protocols.

A central element of clinical QC is checkpoints defined before the project starts, to avoid poorer quality data passing on to the next processing stage (Strother, 2006). Such strict exclusion criteria limit retrospective cherry-picking and avoid bad-quality data to introduce false positive and/or negative results (Niso et al., 2022). However, most animal studies still favor an explorative approach, for which a retrospective selection of good-quality data as implemented in AIDAqc is more meaningful. Similarly, for reuse of data and, more importantly, for reproducibility of existing studies that rely on large databases (e.g., OpenNeuro), retrospective identification of image quality in general and, in particular, signal-to-noise ratio (SNR) and image artifacts is critical. AIDAqc fills an existing gap of efficient software tools and overcomes manual bad-quality selection, which in most cases is bound to an individual decision by the researcher. Importantly, AIDAqc does not contradict previous standardization efforts but complements the development of reliable preclinical MR assays that lead to comparable results across laboratories (Doblas et al., 2015; Waterton et al., 2019).

### 4.2. Outlier detection based on quality metrics

#### *4.2.1. Rationale of selected quality metrics*

The selection of bad-quality data for different MR sequences cannot be accurately done by a single quality measure. Similar to QC tools for human MRI (Esteban et al., 2017; Raamana, 2018; Williams et al., 2023), AIDAqc relies on the combination of multiple automatically derived image quality measures, that is, (t)SNR, motion, and ghosting. The combination ensures high versatility for multiple MR sequences and mitigates known disadvantages for a single quantitative assessment of image quality, as described for SNR, for example (Erdogmus et al., 2004). We have selected basic quality metrics well established in the MRI community. As a novel aspect, we introduce automated calculations, which do not require manually drawn regions of interest. With simulated image artifacts, we have shown that these

calculations provide a good differentiation from image data without artifacts.

It should be noted that this approach could be extended in future versions to provide more information about the type of artifact (e.g., the discrimination of discrete vs. Nyquist ghosts) and include other metrics described in the literature, for example, the contrast-to-noise ratio (CNR), root-mean-squared error (RMSE), coefficient of joint variation of gray matter and white matter, full-width half-maximum estimation of the blurriness of the image, and the overlap of tissue probability maps estimated from the image and a template (Chow & Paramesran, 2016; Esteban et al., 2017). These quality metrics were not included as the information content either overlaps with integrated metrics (e.g., SNR and CNR) or segmentation and template registration, respectively, are required. We decided against a segmentation or template registration process as this would increase processing time and unnecessarily limit the input images to a specific format and orientation. Other restrictions should be considered as well, such as in the case of RMSE where the reliance on a specific reference image or ground-truth data would pose practical challenges due to the diverse nature of MRI datasets and the absence of universally applicable reference images. In the case of CNR, where the signals from two regions of interest are subtracted and divided by noise, automated analysis is theoretically feasible. However, generalizing this approach across different datasets is challenging due to variations in imaging sequences and scan orientations.

#### *4.2.2. Synergistic application of two SNR methods*

In this study, we compared two distinct methods for assessing SNR without the need to manually label regions of interest: the noise estimation method developed by Chang et al. (2005) and the standard SNR method extended with an automatically created center of intensity to measure the signal and cuboids in the corners to measure noise. The results of the quantitative comparison across all datasets revealed the strengths and limitations of each approach, which can be complementary and used as an advantage to enhance the reliability of SNR measurements through a synergistic combination as it is done in AIDAqc. The standard method excels in capturing spatial variations in image quality, particularly in regions of interest where signal integrity is paramount, while Chang's SNR method provides a statistical framework for noise estimation. For datasets with a high probability that the subject is not in the center of the field of view, results showed that the Chang approach is more trustworthy. In the comparison of five randomly selected datasets (7_h_He, 7_m_Se, 7_rab_Mu, 94_m_Va, and 117_m_Ra), the distribution along the standard method is higher than the SNR-Chang method across each dataset. This variation in standard deviation across different datasets suggests that SNR-Chang would better distinguish data from different sources in a scenario combining different studies. Also through the simulation tests of noise and specifically applying the gamma function to change intensity and contrast, mimicking shadows in MRI images, we identified the Chang method to be more responsive to general intensity changes compared with merely adding different types of noise. However, by focusing on one modality or similar images of a dataset, the SNR standard can create a better distinction. Whether this distinction is desired or not depends on the research question at hand. Therefore, it is beneficial to consider the SNR methods first from a synergistic viewpoint. Overall, we found that this synergistic approach enhances the accuracy and robustness of SNR measurements, and subsequently, the machine-learning algorithms can use both features as their input, providing researchers with a more comprehensive understanding of image quality, particularly in neuroimaging studies where image quality directly impacts the validity and interpretability of results. The SNR assessment was also validated qualitatively and quantitatively by simulating noise.

#### *4.2.3. Detection of motion and ghosting using mutual information*

Motion was quantified with a novel application of mutual information (MI). This approach was validated by introducing random motion into the dataset. Incorporating MI for assessing motion severity and ghosting was inspired by the recognized effectiveness of MI as an image registration metric (Maes et al., 2003). The strength lies in the ability of MI to capture nonlinear relationships and tolerate intensity variations. These intensity variations are particularly prevalent in preclinical animal models, where images often feature strokes and lesions. Further investigation into the effectiveness of mutual information in addressing motion severity and ghosting, considering various artifact characteristics and preprocessing techniques, would be valuable for optimizing artifact assessment. In future studies, this method could be compared and complemented by other machine-learning approaches to detect motion (Fantini et al., 2021; Lorch et al., 2017), which have so far only been used in human MRI. The discrimination of the types of ghosts is also planned in future versions of AIDAqc.

### 4.3. Majority voting

As a novel concept, we introduce a statistical- and machine-learning-based outlier detection method in the form of majority voting. In the purely statistical way to

determine outliers based on the interquartile range (Vinutha et al., 2018), a single feature (single variate) of image quality is used. Incorporating a multivariate outlier detection algorithm can lead to more accurate and general results. Using multiple algorithms together and creating a majority vote for an outlier can even stabilize the results further as each algorithm is selective for different types of quality measures and data. With simulated image artifacts, we have shown the robustness and reliability of the majority voting approach in identifying outliers.

The chosen machine-learning algorithms ocSVM, IF, LOF, and EE have been successfully used, for example, in MRI-based tumor classification, and tissue matter segmentation (Mayer et al., 2022; Mohammed et al., 2021; Ong et al., 2022; Zhang, 2011). However, the concept of bringing all of these outlier detectors together and using them as part of a majority vote in QC has, to our knowledge, not been used before. The adaptable and open-source structure of AIDAqc enables the seamless integration or removal of ML algorithms to adapt to specific requirements and new research results. Exploring the integration of additional algorithms such as autoencoders and robust principal component analysis into the majority approach could offer avenues for enhancing its efficacy. These algorithms bring unique capabilities such as nonlinear feature representation and an enhanced resilience to noisy or incomplete data, potentially enriching the diversity of perspectives considered in the decision-making process. However, it is essential to acknowledge that such augmentation might incur heightened computational demands and processing time. Nonetheless, investigating the interplay of diverse machine-learning techniques within the majority voting framework holds promise for advancing its performance.

We noticed that the thresholds to set the minimal number of manual raters and outlier algorithms rating data to be "bad quality," respectively, have a huge impact on the validation of the classification. To understand the role of AIDAqc in outlier detection, it is essential to recognize its capability to identify subjects who deviate from the provided cohort. This does not necessarily imply that flagged subjects are inherently of poor quality or that the specific artifacts cannot be reduced during postprocessing, for example, using motion correction. In a scenario where a cohort consists of subjects selected from two or more different datasets with different acquisition parameters, AIDAqc would not classify an entire group as an outlier, but rather identify individual subjects who have characteristics that differ from their respective group. This targeted outlier detection is a significant advantage of AIDAqc, rooted in its machine-learning algorithms tailored for outlier identification. Traditional statistical approaches often falter in such complex scenarios. Therefore, when designing studies or research questions involving data from multiple sources, leveraging AIDAqc across all datasets *collectively* becomes justified, serving to refine the application of this tool according to the specific research objectives.

In practice, determining the majority vote threshold should involve subsequent checking of the data. As illustrated in the example of a functional scan with a high majority vote indicating a potential issue (Fig. 10a), detailed inspection revealed that the high voting was related to disruptions in the last 25 repetitions due to a potential gradient malfunction. Removing these affected repetitions could rectify the issue, illustrating how the majority voting approach assists in pinpointing specific problems for targeted corrections.

When using AIDAqc to detect outliers, a balance between sensitivity and specificity is critical. Users retain the flexibility to adjust the threshold for the majority voting based on the desired balance between these two metrics. For example, opting for a threshold of a majority vote of 3 or higher may lead to an increase of false positives, thereby reducing specificity, but could potentially decrease false negatives, consequently increasing sensitivity. Conversely, selecting a more stringent threshold, like only those subjects with a majority vote of 5, might further diminish false negatives, increasing the specificity but lowering the sensitivity. After the initial process, researchers can simply focus on the data with the highest majority votes or adjust the threshold based on their preferences.

### 4.4. Validation of outlier detection

#### *4.4.1. Initial validation approach*

Visual criteria of image quality are not standardized and are strictly user dependent. It was, therefore, not surprising to see a significant difference in user ratings for the individual datasets. However, we questioned the rather poor confirmation of AIDAqc outlier detection by the expert raters. Finding a common ground truth proved to be a very challenging task. We noticed that very important noise and image artifact components in the datasets remained hidden for manual raters and can only be identified by automated algorithms scanning the full 2D/3D/4D image volume. Visually detectable artifacts might exist in other slices or in the case of functional and diffusion scans, artifacts might exist at a different time point and diffusion direction, respectively. A tSNR map, which was not part of the manual evaluation, might have been sufficient for manual raters to see the difference between two rs-fMR images (Fig. 11a). Similarly, it is very unlikely to detect motion and ghosting from inspection of single

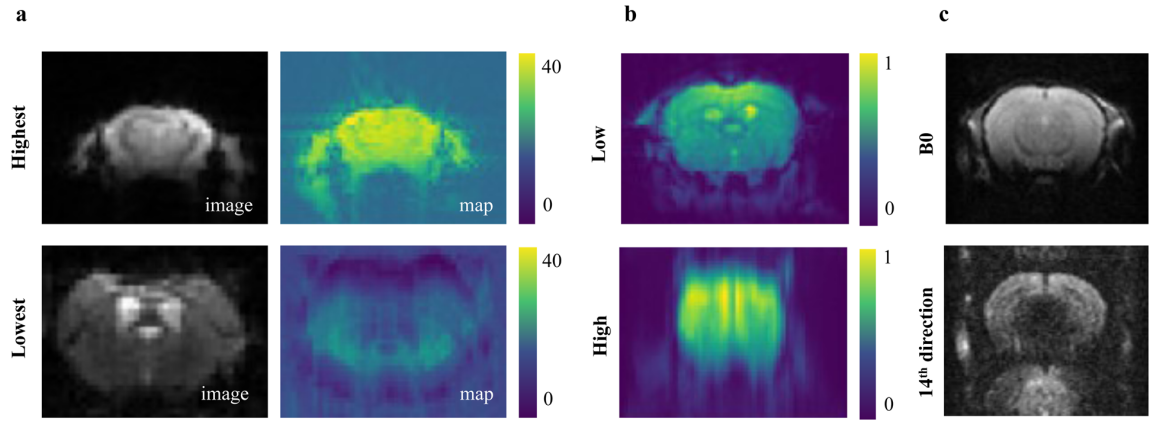



**Fig. 11.** (a) rs-fMR images and their corresponding tSNR maps selected by the highest and lowest values calculated by AIDAqc. (b) Low and high motion examples for rs-fMR image volumes averaged over time to visualize the motion. (c) Ghosting artifacts in the 14th diffusion direction of the 4D diffusion image volume of one subject. Ghosting artifacts can be seen in the lower image, where the upper image corresponding to another the first diffusion direction of the same image volume was rated by the users, and as it does not show any particular problem, it was not rated as a bad-quality image. (a, b) were selected based on the highest (best) and lowest (worst) values between all datasets from Table 1.

images, as the motion between images requires setting a temporary reference image (Fig. 11b), and ghosting might appear only in specific diffusion directions (Fig. 11c). That was also reflected in the sensitivity, which was by trend highest in anatomical compared with diffusion and functional datasets, that is, in the data which can be rated best manually. We noticed that the validation of the automated outlier detection based on the single images, produced as part of the AIDAqc output, remains the most feasible but not optimal scenario, especially for 4D scans. We, therefore, come to the same conclusion as others in human MRI that the manual evaluations are impractical or infeasible (Bedford et al., 2023; Chow & Paramesran, 2016), as it is not possible and not expedient to go through all the individual images for very large datasets.

Therefore, understanding the optimal combinations of thresholds for both AIDAqc and manual raters was crucial. For each sequence type (anatomical, diffusion, and functional scans), we identified the best combinations of thresholds for AIDAqc and manual raters. For the most important measure, sensitivity, that is, the rate of "bad-quality" data identified by AIDAqc compared with the manual raters, the optimal threshold for AIDAqc was 1. This might be interpreted as a logical choice for several reasons: (i) the error rate of AIDAqc is not fluctuating as it is the case for the objective classification by the manual raters, (ii) AIDAqc systematically evaluates every voxel of an image volume, each AIDAqc rater can be considered to have its unique strengths and weaknesses in identifying outliers. In contrast, manual raters, relying on visual inspection, do not have the same systematic approach. Thus it is worth checking any dataset, even if it only got one vote to be an outlier from AIDAqc. In contrast, manual rater thresholds exhibited a favorable trend for accuracy but not for specificity and sensitivity, likely influenced by poor inter-rater agreement.

#### *4.4.2. Revised post hoc validation*

Consequently, we transitioned to a more focused validation method, as elaborated in the second stage of our process, providing a more targeted validation approach. indicating some variability in their assessments, yet demonstrating a generally consistent agreement with AIDAqc. This approach provided a more nuanced insight into AIDAqc's outlier detection capabilities. By focusing on the high-confidence outlier predictions and considering the more decisive assessments of the experts, this method offered a better understanding of the effectiveness in identifying true cases of poor-quality datasets. The observed standard deviation underscores the inherent subjectivity in manual assessments, reinforcing the value of automated tools in systematically evaluating image quality.

### 4.5. Limitations

#### *4.5.1. Datasets*

To validate the reliability of bad-quality detection using AIDAqc, we collected largely heterogeneous data in terms of animal, MRI hardware, and sequence. Unlike previous standardization efforts that focused on homogenization of data collection (Doblas et al., 2015;

Grandjean et al., 2023; Waterton et al., 2019), wide variability was desired for probing the versatility of AIDAqc. Nevertheless, the heterogeneous datasets imposed a challenge for the analysis. This included automated sequence type recognition and the correct sorting into anatomical, diffusion, and functional scans, respectively, which were corrupted by nonstandard abbreviations for MR sequences (see Supplementary Fig. 9a, b, d). As most sequences can be used for multiple purposes, for example, echo planar imaging (EPI) in both diffusion and functional imaging requires more in-depth metadata extraction to identify the actual type of scan. Such metadata extraction is in the current AIDAqc version only possible for BIDS or Bruker data but not for the NIfTI format as it contains only very limited metadata. In future versions, compatibility with DICOM data could be implemented because of its wide distribution in human MRI as well as other vendor-specific raw data next to the Bruker format to increase the versatility. AIDAqc uses an integrated list of strings of the most common nomenclature as keywords, for example, *turbo*, *rare*, *rest*, *diff*, *rs-*, *func*, *anat*, *struc*, *dwi*, and *dti*. Additionally, for DWI and functional images in the raw Bruker format, AIDAqc checks whether a gradient/b-value table is available to make the distinction. We cannot exclude, however, the possibility that automated parsing and sorting may fail for other unconventional sequence names. In the case of BIDS structure, for example, in the *7_rab_Mu* and *94_r_Fr* datasets, the NIfTI format is complemented with metadata in a standardized json file. To facilitate standardization, we highly recommend the BIDS format as it solves the issue of variable sequence names and dataset hierarchy of groups and time points as well as metadata information.

Another challenge was imposed by a corrupted image matrix logic in the x-y-z regime, for example, in the form of 256 x 40 x 256 imposing a x-z-y structure with z being the number of slices and x-y being the image dimension. This occurred in single datasets in both the raw data and by converting the data into the NIfTI format and required manual correction. Manual spatially reorganization is also necessary for noncoronal acquired datasets (see Supplementary Fig. 9c, e, g). This step is critical, as the subsequent processing depends on the correct association of 2D/3D/4D sequences and scans. In future AIDAqc versions, automatic extraction of the actual image orientation and position of the animal would reduce this potential error and manual data preparation for the users. Another potential limitation is the fixed parameter setting in some of the outlier detection algorithms, which should be optimized if AIDAqc is to be used for completely different datasets.

#### *4.5.2. Automated feature calculations*

Unlike most approaches to calculating SNR region based, which requires manual user input to draw the regions of interest, here two fully automated SNR calculations were implemented. This way, there is no need for the user to manually define regions, which makes the calculation less error prone and faster. We noticed, however, that the automated SNR calculation has some limitations. The standard SNR calculation is sensitive to the location of the subject in the field of view: (i) when the subject fills most of the field of view, especially the corners, the cuboids for calculating the standard deviation cover not only noise (Fig. 2), thus decreasing the final SNR, and (ii) when the central sphere used to calculate the mean signal covers regions with artifactually high or low signal, for example, ventricles or blood-rich areas. In such cases, the resulting SNR value might be misleading and we suggest verifying the center of mass in the images to be representative of the signal in the image. We also noticed that the alternative Chang method has limitations when it comes to the ratio of noise-to-tissue in the field of view. In particular, different SNR ranges create different circumstances for the ratio of noise-to-tissue for a correct estimation of the standard deviation of noise in the image (Chang et al., 2005). If the ratio is small, SNR-Chang overestimates the true SNR as the quantitative comparison with the low SNR-Standard in line with the visual comparison showed. It should be noted that SNR-Chang, different to SNR-Standard (which uses the b0 images only), creates an average SNR across all diffusion directions except b0, resulting in a less interpretable value for multiple diffusion weightings (i.e., b-values).

To retain the high versatility of AIDAqc, both SNR algorithms were included, and for most cases, the majority voting approach still resulted in a stable detection of low-quality images.

Similar to the SNR calculation, the motion detection was successfully automated in AIDAqc without the need to manually draw regions or perform separate tissue segmentations. Nevertheless, the user should be aware that the mutual information calculation cannot distinguish between animal and hardware motion if both have similar effects, for example, producing a drift in the image stack. An extreme example dataset is *117_m_Ri* with the highest motion severity of all functional datasets. However, in this case, the animals were paralyzed and artificially ventilated, which practically reduces animal motion to a minimum (Bukhari et al., 2017). The high motion severity can be explained by a hardware limitation of the used gradient system in the 11.7T Bruker scanner, creating thermal artifacts at a high-duty cycle, that is, long acquisitions of thin slices with short repetition time.

#### 4.5.3. *Validation approaches*

In an ideal validation scenario, manual raters would examine every slice, across all time points, repetitions, and diffusion directions, mirroring the extensive analysis performed by AIDAqc. However, such a level of manual scrutiny is impractical and leads to highly variable results as shown in the inter-rater comparison. It underscores the necessity for automated tools and the challenges faced in achieving comprehensive quality control. The full transparency of sharing the code, reporting results, and related data was implemented to ensure future improvements are implemented by the community. In an alternative validation scenario, the tSNR maps or mutual information plots could be provided for the expert raters as an additional level of quality assessment.

Future AIDAqc developments should strike an optimal balance between sensitivity and precision. High sensitivity ensures the tool's ability to detect the majority of outliers or poor-quality data, which is vital for reliable neuroimaging analysis. Simultaneously, high precision is equally important, as it assures users that the data flagged as outliers are indeed of inferior quality. To further increase the precision, alternative approaches could be explored, that is, feeding the images directly without the calculation of quality features into the machine-learning algorithms. However, such an approach would be time consuming and require high computational power that maybe not every user can afford.

## DATA AND CODE AVAILABILITY

Datasets for testing and validating AIDAqc (Table 1) were collected from 19 international MRI laboratories. These datasets (CC BY-NC-SA 4.0 license) organized according to the scheme outlined in Kalantari et al. (2023) can be accessed via: https://doi.gin.g-node.org/10.12751/g-node.q82cjj/. This DOI contains all datasets and results of this publication. Datasets remained unchanged during the revision process, while the *output* and *code* folder contained additional content representing changes during the revision. Access to the live datasets' content can be facilitated by clicking on the "Browse Repository" button on the GIN DOI, where further details are available. The published data listed in Supplementary Table 2 can be obtained through the provided links/references in the table.

## AUTHOR CONTRIBUTIONS

The names of the authors appear as in the list of authors. Aref Kalantari: Lead in manuscript writing, lead software developer, development, testing, and debugging, MRI expert manual rater, data contribution and acquisition; Mehrab Shahbazi, Marc Schneider, Victor Vera Frazão: Software development, testing, and debugging; Adam C. Raikes: MRI Expert manual rater, manuscript writing, reviewing, commenting, data contribution, and acquisition; Avnish Bhattrai, Yujian Diao, Lorenzo Carnevale, Bart A. A. Franx, Francesco Gammaraccio, Lisa-Marie Goncalves, Susan Lee, Esther M. van Leeuwen, Annika Michalek, Alejandro Rivera Olvera, Daniel Padro, Mohamed Kotb Selim, Federico Varriano, H. Elliott Albers, Santiago Canals, Roberta Diaz Brinton, Rick M. Dijkhuizen, Elisenda Eixarch, Gianluigi Forloni, Khan Hekmatyar, Russell E. Jacobs, Giuseppe Lembo, Dario Livio Longo, Naomi S. Sta Maria, Edoardo Micotti, Pedro Ramos-Cabrer, Wilfried Reichardt: Data contribution and acquisition; Susanne Mueller: MRI Expert manual rater; Annette van der Toorn: Data contribution and acquisition, revising, editing; Roël Vrooman, Eike Budinger, Silvia De Santis, Nyoman D. Kurniawan, Guadalupe Soria: Data contribution and acquisition, reviewing, editing; Patricia Wenk, Ileana Jelescu: Data contribution and acquisition, manuscript writing, reviewing, commenting; Philipp Boehm-Sturm: MRI Expert manual rater, data contribution and acquisition; Joanes Grandjean: MRI Expert manual rater, manuscript writing, reviewing, commenting, data contribution and acquisition; Emma Muñoz-Moreno: Data contribution and acquisition, reviewing, commenting; Giovanna D. Ielacqua: Manuscript writing, reviewing, commenting, MRI expert manual rater; Markus Aswendt: Colead in manuscript writing, project conceptualization, MRI expert manual rater, data contribution, and acquisition.

## FUNDING

*This work was funded by the Friebe Foundation (T0498/28960/16)*, COMULISglobe global Showcase Project supported by the Chan-Zuckerberg-Initiative, and the *Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—Project-ID 431549029—SFB1451*. For Mario Negri Institute for Pharmacological Research IRCCS and IRCCS INM Neuromed, Department of Angio-CardioNeurology and Translational Medicine, the project was funded by the Italian Ministry of Health (RRC-2016-2361095, RRC-2017-2364915, RRC-2018-2365796, and RCR-2019-23669119_001, and RCR 2020-23670067) and the Ministry of Economy and Finance (CCR-2017-23669078). The operation of 16.4 T MRI at the Centre for Advanced Imaging at the University of Queensland was supported by the Australian National Imaging Facility (NIF) and the Queensland NMR Network (QNN). For IBB-CNR and the University of Turin, the work has received funding from the European Union's Horizon 2020 research

and innovation program "EOSC-Life—Providing an open collaborative space for digital biology in Europe" under Grant agreement No 824087. The Italian Ministry for Education and Research is gratefully acknowledged for its yearly FOE funding to the Euro-BioImaging Multi-Modal Molecular Imaging Italian Node. The operation of the 7T MRI at the University of Arizona's Translational Bioimaging Resource was supported by NIH S10 OD025016 and projects utilizing the system were supported by the NIH/National Institute on Aging P01AG026572 and R01AG057931 to Dr. Brinton as well as the Center for Innovation in Brain Science to Dr. Brinton. CIC biomaGUNE in San Sebastian received funding from the Spanish Research Agency (Grant PID2020-118546RB-I00). Radboud UMC imaging was supported by the Horizon Europe programs CANDY under grant agreement nos. 847818 and the Dutch Research Council grant OCENW.KLEIN.334 to Dr Grandjean. The Instituto de Neurociencias received funding from the Spanish Research Agency (PID2021-128158NB-C21 [to S.C.], PID2021-128909NA-I00 [to S.D.S.] and CEX2021-001165-S [to S.C. and S.D.S.]), from the Spanish Generalitat Valenciana Government (PROMETEO/2019/015 [to S.C.] and CIDEGENT/2021/015 [to S.D.S.]), "la Caixa" Foundation (ID 100010434), fellowship code LCF/BQ/DI18/11660067 and Marie Skłodowska-Curie- COFUND agreement (Grant No. 713673 [to M.K.S.]). Varian 14T MRI data were supported by the CIBM Center for Biomedical Imaging of the UNIL, CHUV, EPFL, HUG, and UNIGE, and by the SNSF Eccellenza PCEFP2_194260 [to I.J.]. Data collection and rating at Charité-Universitätsmedizin Berlin was supported by the DFG (project BO 4484/2-1 and EXC-2049-390688087 NeuroCure), the German Federal Ministry of Education and Research (BMBF) under the ERA-NET NEURON scheme (01EW1811 and 01EW2305) and by Charité 3R | Replace, Reduce, Refine [to P.B.S. and S.M.]. Magdeburg LIN imaging received funding from the DFG (CRC 1436, project B06, 425899996) and the Leibniz Society Cooperative Excellence (K430/2021 LeaRNetS). The Translational Neuroimaging group at the University Medical Center Utrecht received funding through the PPP Allowance made available by the Top Sector Life Sciences & Health to stimulate public–private partnerships (LSHM17020).

## DECLARATION OF COMPETING INTEREST

The authors declare that they have no conflict of interest.

## SUPPLEMENTARY MATERIALS

Supplementary material for this article is available with the online version here: https://doi.org/10.1162/imag_a_00317

## REFERENCES

Ashrafuzzaman, M., Das, S., Jillepalli, A. A., Chakhchoukh, Y., & Sheldon, F. T. (2020). Elliptic envelope based detection of stealthy false data injection attacks in smart grid control systems. In *2020 IEEE Symposium Series on Computational Intelligence (SSCI)*, Canberra, ACT, Australia, December 1 (pp. 1131–1137). IEEE. https://doi.org/10.1109/ssci47803.2020.9308523

Axel, L., Summers, R. M., Kressel, H. Y., & Charles, C. (1986). Respiratory effects in two-dimensional Fourier transform MR imaging. *Radiology*, *160*(3), 795–801. https://doi.org/10.1148/radiology.160.3.3737920

Backhausen, L. L., Herting, M. M., Buse, J., Roessner, V., Smolka, M. N., & Vetter, N. C. (2016). Quality control of structural MRI images applied using FreeSurfer—A hands-on workflow to rate motion artifacts. *Frontiers in Neuroscience*, *10*, 558. https://doi.org/10.3389/fnins.2016.00558

Bedford, S. A., Ortiz-Rosa, A., Schabdach, J. M., Costantino, M., Tullo, S., Piercy, T., Lai, M.-C., Lombardo, M. V., Di Martino, A., Devenyi, G. A., Chakravarty, M. M., Alexander-Bloch, A. F., Seidlitz, J., Baron-Cohen, S., Bethlehem, R. A. I., & Lifespan Brain Chart Consortium. (2023). The impact of quality control on cortical morphometry comparisons in autism. *Imaging Neuroscience*, *1*, 1–21. https://doi.org/10.1162/imag_a_00022

Bianciardi, M., Fukunaga, M., van Gelderen, P., Horovitz, S. G., de Zwart, J. A., Shmueli, K., & Duyn, J. H. (2009). Sources of functional magnetic resonance imaging signal fluctuations in the human brain at rest: a 7 T study. *Magnetic Resonance Imaging*, *27*(8), 1019–1029. https://doi.org/10.1016/j.mri.2009.02.004

Budiarto, E. H., Erna Permanasari, A., & Fauziati, S. (2019). Unsupervised anomaly detection using K-means, local outlier factor and one class SVM. In *2019 5th International Conference on Science and Technology (ICST)*, Yogyakarta, Indonesia, July (pp. 1–5). IEEE. https://doi.org/10.1109/icst47872.2019.9166366

Budrys, T., Veikutis, V., Lukosevicius, S., Gleizniene, R., Monastyreckiene, E., & Kulakiene, I. (2018). Artifacts in magnetic resonance imaging: How it can really affect diagnostic image quality and confuse clinical diagnosis? *Journal of Vibroengineering*, *20*(2), 1202–1213. https://doi.org/10.21595/jve.2018.19756

Bukhari, Q., Schroeter, A., Cole, D. M., & Rudin, M. (2017). Resting state fMRI in mice reveals anesthesia specific signatures of brain functional networks and their interactions. *Frontiers in Neural Circuits*, *11*, 5. https://doi.org/10.3389/fncir.2017.00005

Buonocore, M. H., & Gao, L. (1997). Ghost artifact reduction for echo planar imaging using image phase correction. *Magnetic Resonance in Medicine*, *38*(1), 89–100. https://doi.org/10.1002/mrm.1910380114

Chang, L.-C., Rohde, G. K., & Pierpaoli, C. (2005). An automatic method for estimating noise-induced signal variance in magnitude-reconstructed magnetic resonance images. *Medical Imaging 2005: Image Processing*, *5747*, 1136–1142. https://doi.org/10.1117/12.596008

Chen, N.-K., & Wyrwicz, A. M. (2004). Removal of EPI Nyquist ghost artifacts with two-dimensional phase correction. *Magnetic Resonance in Medicine*, *51*(6), 1247–1253. https://doi.org/10.1002/mrm.20097

Chow, L. S., & Paramesran, R. (2016). Review of medical image quality assessment. *Biomedical Signal Processing and Control*, *27*, 145–154. https://doi.org/10.1016/j.bspc.2016.02.006

Doblas, S., Almeida, G. S., Blé, F.-X., Garteiser, P., Hoff, B. A., McIntyre, D. J. O., Wachsmuth, L., Chenevert, T. L.,

Faber, C., Griffiths, J. R., Jacobs, A. H., Morris, D. M., O'Connor, J. P. B., Robinson, S. P., Van Beers, B. E., & Waterton, J. C. (2015). Apparent diffusion coefficient is highly reproducible on preclinical imaging systems: Evidence from a seven-center multivendor study. *Journal of Magnetic Resonance Imaging*, *42*(6), 1759–1764. https://doi.org/10.1002/jmri.24955

Erdogmus, D., Larsson, E. G., Yan, R., Principe, J. C., & Fitzsimmons, J. R. (2004). Measuring the signal-to-noise ratio in magnetic resonance imaging: A caveat. *Signal Processing*, *84*(6), 1035–1040. https://doi.org/10.1016/j.sigpro.2004.03.006

Esteban, O., Birman, D., Schaer, M., Koyejo, O. O., Poldrack, R. A., & Gorgolewski, K. J. (2017). MRIQC: Advancing the automatic prediction of image quality in MRI from unseen sites. *PLoS One*, *12*(9), e0184661. https://doi.org/10.1371/journal.pone.0184661

Etman, H. M., Mokhtar, A., Abd-Elhamid, M. I., Ahmed, M. T., & El-Diasty, T. (2017). The effect of quality control on the function of magnetic resonance imaging (MRI), using American College of Radiology (ACR) phantom. *The Egyptian Journal of Radiology and Nuclear Medicine*, *48*(1), 153–160. https://doi.org/10.1016/j.ejrnm.2016.12.003

Fantini, I., Yasuda, C., Bento, M., Rittner, L., Cendes, F., & Lotufo, R. (2021). Automatic MR image quality evaluation using a Deep CNN: A reference-free method to rate motion artifacts in neuroimaging. *Computerized Medical Imaging and Graphics*, *90*, 101897. https://doi.org/10.1016/j.compmedimag.2021.101897

Gilmore, A. D., Buser, N. J., & Hanson, J. L. (2021). Variations in structural MRI quality significantly impact commonly used measures of brain anatomy. *Brain Informatics*, *8*(1), 7. https://doi.org/10.21203/rs.3.rs-103604/v1

Godenschweger, F., Kägebein, U., Stucht, D., Yarach, U., Sciarra, A., Yakupov, R., Lüsebrink, F., Schulze, P., & Speck, O. (2016). Motion correction in MRI of the brain. *Physics in Medicine and Biology*, *61*(5), R32–R56. https://doi.org/10.1088/0031-9155/61/5/r32

Grandjean, J., Desrosiers-Gregoire, G., Anckaerts, C., Angeles-Valdez, D., Ayad, F., Barrière, D. A., Blockx, I., Bortel, A., Broadwater, M., Cardoso, B. M., Célestine, M., Chavez-Negrete, J. E., Choi, S., Christiaen, E., Clavijo, P., Colon-Perez, L., Cramer, S., Daniele, T., Dempsey, E., … Hess, A. (2023). A consensus protocol for functional connectivity analysis in the rat brain. *Nature Neuroscience*, *26*(4), 673–681. https://doi.org/10.1038/s41593-023-01286-8

Grandjean, J., Schroeter, A., Batata, I., & Rudin, M. (2014). Optimization of anesthesia protocol for resting-state fMRI in mice based on differential effects of anesthetics on functional connectivity patterns. *NeuroImage*, *102*, 838–847. https://doi.org/10.1016/j.neuroimage.2014.08.043

Havsteen, I., Ohlhues, A., Madsen, K. H., Nybing, J. D., Christensen, H., & Christensen, A. (2017). Are movement artifacts in magnetic resonance imaging a real problem?—A narrative review. *Frontiers in Neurology*, *8*, 232. https://doi.org/10.3389/fneur.2017.00232

Henkelman, R. M. (1985). Measurement of signal intensities in the presence of noise in MR images. *Medical Physics*, *12*(2), 232–233. https://doi.org/10.1118/1.595711

Jackson, E., Bronskill, M., Drost, D., Och, J., Pooley, R., Sobol, W., & Clark, G. (2010). *Tg 1 —Acceptance Testing and Quality Assurance Procedures for Magnetic Resonance Imaging Facilities*. American Association of Physicists in Medicine (A A P M). https://doi.org/10.37206/101

Kalantari, A., Szczepanik, M., Heunis, S., Mönch, C., Hanke, M., Wachtler, T., & Aswendt, M. (2023). How to establish and maintain a multimodal animal research dataset using DataLad. *Scientific Data*, *10*(1), 357. https://doi.org/10.1038/s41597-023-02242-8

Kalthoff, D., Seehafer, J. U., Po, C., Wiedermann, D., & Hoehn, M. (2011). Functional connectivity in the rat at 11.7T: Impact of physiological noise in resting state fMRI. *NeuroImage*, *54*(4), 2828–2839. https://doi.org/10.1016/j.neuroimage.2010.10.053

Kaufman, L., Kramer, D. M., Crooks, L. E., & Ortendahl, D. A. (1989). Measuring signal-to-noise ratios in MR imaging. *Radiology*, *173*(1), 265–267. https://doi.org/10.1148/radiology.173.1.2781018

Khan, S. U., Ullah, N., Ahmed, I., Ahmad, I., & Mahsud, M. I. (2019). MRI imaging, comparison of MRI with other modalities, noise in MRI images and machine learning techniques for noise removal: A review. *Current Medical Imaging Reviews*, *15*(3), 243–254. https://doi.org/10.2174/1573405614666180726124952

Kilkenny, C., Browne, W., Cuthill, I. C., Emerson, M., & Altman, D. G. (2010). Animal research: reporting in vivo experiments: the ARRIVE guidelines. *British journal of pharmacology*, *160*(7), 1577–1579. https://doi.org/10.1111/j.1476-5381.2010.00872.x

Lorch, B., Vaillant, G., Baumgartner, C., Bai, W., Rueckert, D., & Maier, A. (2017). Automated detection of motion artefacts in MR imaging using decision forests. *Journal of Medical Engineering & Technology*, *2017*, 4501647. https://doi.org/10.1155/2017/4501647

Maes, F., Vandermeulen, D., & Suetens, P. (2003). Medical image registration using mutual information. *Proceedings of the IEEE. Institute of Electrical and Electronics Engineers*, *91*(10), 1699–1722. https://doi.org/10.1109/jproc.2003.817864

Mannheim, J. G., Kara, F., Doorduin, J., Fuchs, K., Reischl, G., Liang, S., Verhoye, M., Gremse, F., Mezzanotte, L., & Huisman, M. C. (2018). Standardization of small animal imaging-current status and future prospects. *Molecular Imaging and Biology*, *20*(5), 716–731. https://doi.org/10.1007/s11307-017-1126-2

Marcus, D. S., Harms, M. P., Snyder, A. Z., Jenkinson, M., Wilson, J. A., Glasser, M. F., Barch, D. M., Archie, K. A., Burgess, G. C., Ramaratnam, M., Hodge, M., Horton, W., Herrick, R., Olsen, T., McKay, M., House, M., Hileman, M., Reid, E., Harwell, J., … WU-Minn HCP Consortium. (2013). Human Connectome Project informatics: Quality control, database services, and data visualization. *NeuroImage*, *80*, 202–219. https://doi.org/10.1016/j.neuroimage.2013.05.077

Mayer, R., Turkbey, B., Choyke, P., & Simone, C. B., 2nd. (2022). Assessing and testing anomaly detection for finding prostate cancer in spatially registered multi-parametric MRI. *Frontiers in Oncology*, *12*, 1033323. https://doi.org/10.3389/fonc.2022.1033323

Mohammed, E. A., Shakeri, E., Shakeri, H. A. Z., Crump, T., & Far, B. (2021). Important features identification for prostate cancer patients stratification using isolation forest and interactive clustering method. In *2021 IEEE 22nd International Conference on Information Reuse and Integration for Data Science (IRI)*, Las Vegas, NV, United States, 10–12 August (pp. 334–341). IEEE. https://doi.org/10.1109/iri51335.2021.00052

Niso, G., Botvinik-Nezer, R., Appelhoff, S., De La Vega, A., Esteban, O., Etzel, J. A., Finc, K., Ganz, M., Gau, R., Halchenko, Y. O., Herholz, P., Karakuzu, A., Keator, D. B., Markiewicz, C. J., Maumet, C., Pernet, C. R., Pestilli, F., Queder, N., Schmitt, T., … Rieger, J. W. (2022). Open and reproducible neuroimaging: From study inception to

publication. *NeuroImage*, *263*, 119623. https://doi.org/10.1016/j.neuroimage.2022.119623

Ong, K., Young, D. M., Sulaiman, S., Shamsuddin, S. M., Mohd Zain, N. R., Hashim, H., Yuen, K., Sanders, S. J., Yu, W., & Hang, S. (2022). Detection of subtle white matter lesions in MRI through texture feature extraction and boundary delineation using an embedded clustering strategy. *Scientific Reports*, *12*(1), 4433. https://doi.org/10.1038/s41598-022-07843-8

Osborne, D. R., Kuntner, C., Berr, S., & Stout, D. (2017). Guidance for efficient small animal imaging quality control. *Molecular Imaging and Biology*, *19*(4), 485–498. https://doi.org/10.1007/s11307-016-1012-3

Pallast, N., Diedenhofen, M., Blaschke, S., Wieters, F., Wiedermann, D., Hoehn, M., Fink, G. R., & Aswendt, M. (2019). Processing pipeline for atlas-based imaging data analysis of structural and functional mouse brain MRI (AIDAmri). *Frontiers in Neuroinformatics*, *13*, 42. https://doi.org/10.3389/fninf.2019.00042

Percie du Sert, N., Hurst, V., Ahluwalia, A., Alam, S., Avey, M. T., Baker, M., ... & Würbel, H. (2020). The ARRIVE guidelines 2.0: Updated guidelines for reporting animal research. *Journal of Cerebral Blood Flow & Metabolism*, *40*(9), 1769–1777. https://doi.org/10.1177/0271678X20943823

Price, R., Allison, J., Clarke, G., Dennis, M., Hendrick, R. E., Keener, C., Masten, J., Nessaiver, M., Och, J., & Reeve, D. (2015). *Magnetic Resonance Imaging (MRI) Quality Control Manual*. American College of Radiology. https://www.acr.org/-/media/ACR/Files/Clinical-Resources/QC-Manuals/MR_QCManual.pdf

Raamana, P. R. (2018). *VisualQC: Assistive tools for easy and rigorous quality control of neuroimaging data*. Zenodo. https://doi.org/10.5281/ZENODO.1211365

Reeder, S. B., Atalar, E., Bolster, B. D., Jr, & McVeigh, E. R. (1997). Quantification and reduction of ghosting artifacts in interleaved echo-planar imaging. *Magnetic Resonance in Medicine*, *38*(3), 429–439. https://doi.org/10.1002/mrm.1910380312

Seabold, S., & Perktold, J. (2010). Statsmodels: Econometric and statistical modeling with python. In *Proceedings of the 9th Python in Science Conference*. Python in Science Conference, Austin, Texas (pp. 92–96). https://doi.org/10.25080/majora-92bf1922-011

Sreedher, G., Ho, M.-L., Smith, M., Udayasankar, U. K., Risacher, S., Rapalino, O., Greer, M.-L. C., Doria, A. S., & Gee, M. S. (2021). Magnetic resonance imaging quality control, quality assurance and quality improvement. *Pediatric Radiology*, *51*(5), 698–708. https://doi.org/10.1007/s00247-021-05043-6

Storey, P., Chen, Q., Li, W., Edelman, R. R., & Prasad, P. V. (2002). Band artifacts due to bulk motion. *Magnetic Resonance in Medicine*, *48*(6), 1028–1036. https://doi.org/10.1002/mrm.10314

Strother, S. C. (2006). Evaluating fMRI preprocessing pipelines. *IEEE Engineering in Medicine and Biology Magazine: The Quarterly Magazine of the Engineering in Medicine & Biology Society*, *25*(2), 27–41. https://doi.org/10.1109/memb.2006.1607667

Tavares, A. A. S., Mezzanotte, L., McDougald, W., Bernsen, M. R., Vanhove, C., Aswendt, M., Ielacqua, G. D., Gremse, F., Moran, C. M., Warnock, G., Kuntner, C., & Huisman, M. C. (2023). Community survey results show that standardisation of preclinical imaging techniques remains a challenge. *Molecular Imaging and Biology*, *25*(3), 560–568. https://doi.org/10.1007/s11307-022-01790-6

Thompson, P. M., Jahanshad, N., Ching, C. R. K., Salminen, L. E., Thomopoulos, S. I., Bright, J., Baune, B. T., Bertolín, S., Bralten, J., Bruin, W. B., Bülow, R., Chen, J., Chye, Y., Dannlowski, U., de Kovel, C. G. F., Donohoe, G., Eyler, L. T., Faraone, S. V., Favre, P., … ENIGMA Consortium. (2020). ENIGMA and global neuroscience: A decade of large-scale studies of the brain in health and disease across more than 40 countries. *Translational Psychiatry*, *10*(1), 100. https://doi.org/10.1038/s41398-020-0705-1

van der Walt, S., Schönberger, J. L., Nunez-Iglesias, J., Boulogne, F., Warner, J. D., Yager, N., Gouillart, E., & Yu, T. (2014). scikit-image: image processing in Python. *PeerJ*, *2*, e453. https://doi.org/10.7717/peerj.453

Vinutha, H. P., Poornima, B., & Sagar, B. M. (2018). Detection of outliers using interquartile range technique from intrusion dataset. In S. Satapathy, J. Tavares, V. Bhateja, & J. Mohanty (Eds.), *Information and Decision Sciences* (pp. 511–518). Springer. https://doi.org/10.1007/978-981-10-7563-6_53

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., … SciPy 1.0 Contributors. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, *17*(3), 261–272. https://doi.org/10.1038/s41592-020-0772-5

Waterton, J. C., Hines, C. D. G., Hockings, P. D., Laitinen, I., Ziemian, S., Campbell, S., Gottschalk, M., Green, C., Haase, M., Hassemer, K., Juretschke, H.-P., Koehler, S., Lloyd, W., Luo, Y., Mahmutovic Persson, I., O'Connor, J. P. B., Olsson, L. E., Pindoria, K., Schneider, J. E., … Schütz, G. (2019). Repeatability and reproducibility of longitudinal relaxation rate in 12 small-animal MRI systems. *Magnetic Resonance Imaging*, *59*, 121–129. https://doi.org/10.1016/j.mri.2019.03.008

Welvaert, M., & Rosseel, Y. (2013). On the definition of signal-to-noise ratio and contrast-to-noise ratio for FMRI data. *PLoS One*, *8*(11), e77089. https://doi.org/10.1371/journal.pone.0077089

Wijnen, J. P., Seiberlich, N., & Golay, X. (2023). Will standardization kill innovation? *MAGMA*, *36*(4), 525–528. https://doi.org/10.1007/s10334-023-01115-w

Williams, B., Hedger, N., McNabb, C. B., Rossetti, G. M. K., & Christakou, A. (2023). Inter-rater reliability of functional MRI data quality control assessments: A standardised protocol and practical guide using pyfMRIqc. *Frontiers in Neuroscience*, *17*, 1070413. https://doi.org/10.3389/fnins.2023.1070413

Wood, M. L., & Henkelman, R. M. (1985). MR image artifacts from periodic motion. *Medical Physics*, *12*(2), 143–151. https://doi.org/10.1118/1.595782

Yang, Q. X., Posse, S., Le Bihan, D., & Smith, M. B. (1996). Double-sampled echo-planar imaging at 3 Tesla. *Journal of Magnetic Resonance. Series B*, *113*(2), 145–150. https://doi.org/10.1006/jmrb.1996.0167

Zaitsev, M., Maclaren, J., & Herbst, M. (2015). Motion artifacts in MRI: A complex problem with many partial solutions. *Journal of Magnetic Resonance Imaging: JMRI*, *42*(4), 887–901. https://doi.org/10.1002/jmri.24850

Zarrar, S., Steven, G., Qingyang, L., Yassine, B., Chaogan, Y., Zhen, Y., Michael, M., Pierre, B., & Cameron, C. (2015). The preprocessed connectomes project quality assessment protocol—A resource for measuring the quality of MRI data. *Frontiers in Neuroscience*, *9*, 47. https://doi.org/10.3389/conf.fnins.2015.91.00047

Zhang, N. (2011). *Feature selection based segmentation of multi-source images: Application to brain tumor segmentation in multi-sequence MRI*. https://theses.hal.science/tel-00701545v1

# Chapter 4

# Proportional recovery in mice with cortical stroke

*Aref Kalantari, Carolin Hambrock, Christian Grefkes, Gereon R. Fink & Markus Aswendt*

RESEARCH PAPER

Experimental Neurology (2025)

The third and final study was based on behavioral and imaging data. It tested the clinical Proportional Recovery Rule (PRR) in 125 C57BL/6J mice with photothrombotic cortical strokes and introduced a novel Mouse Recovery Rule (MRR) for an enhanced recovery phenotype. T2-weighted MRI scans acquired on day 3 were registered to the Allen Mouse Brain Atlas to quantify lesion volume and topology. Deficits (baseline to day 3) and recovery (day 3 to day 28) formed two clusters: approximately 47% of mice followed the clinical PRR, recovering 70% of baseline function, while 53% surpassed this and were classified under the newly proposed MRR with a recovery slope of 0.93. Significant behavioral divergence appeared from day 14 onward ($p < 0.05$). MRI lesion mapping confirmed no significant intergroup differences in lesion size or location. These results show that the clinical PRR applies to a subset of mice but that a distinct, enhanced recovery phenotype captured by the MRR is present, underscoring the value of species-specific recovery models in preclinical stroke research.

Experimental Neurology 386 (2025) 115180

Contents lists available at ScienceDirect

Experimental Neurology

journal homepage: www.elsevier.com/locate/yexnr

Research paper

# Proportional recovery in mice with cortical stroke

Aref Kalantari [a], Carolin Hambrock [a], Christian Grefkes [c], Gereon R. Fink [a,b], Markus Aswendt [a,*]

[a] University of Cologne, Faculty of Medicine and University Hospital Cologne, Department of Neurology, Cologne, Germany
[b] Cognitive Neuroscience, Institute of Neuroscience and Medicine (INM-3), Research Centre Juelich, Germany
[c] Goethe University Frankfurt, University Hospital, Department of Neurology, Frankfurt am Main, Germany

ARTICLE INFO

*Keywords:*
Motor deficit
Mouse model
Ischemic brain lesion
Lesion size
Behavior test
MRI

ABSTRACT

The proportional Recovery Rule (PRR) has been frequently used to predict recovery of lost motor function in acute stroke patients. However, it still needs to be explored whether the same concept applies to preclinical, i.e. animal models of stroke recovery. To address this question, we investigated behavioral data from 125 adult male C57Bl/6 J mice with photothrombotic strokes in the sensorimotor cortex. Lesion size and location were determined in the first week using in vivo T2-weighted MRI. Motor recovery was evaluated repeatedly over four weeks using the cylinder, grid walk, and rotating beam test. Recovery trajectories were analyzed using a newly formulated Mouse Recovery Rule (MRR), comparing it against the traditional PRR. Initial findings indicated variable recovery patterns, which were separated using a stepwise linear regression approach resulting in two clusters: 47 % PRR and 53 % MRR. No significant correlation was found between recovery patterns and lesion size or location, suggesting that other biological factors drive individual differences in recovery. Of note, in the MRR cluster, animals recovered to 90 % of their initial behavioral state within the first four weeks post-stroke, which is higher than the 70 % recovery usually reported in human PRR studies. This study demonstrates the complexity of translating the PRR to stroke recovery models in mice and underscores the need for species-specific recovery models. Our findings have implications for designing and interpreting therapeutic strategies for stroke recovery in preclinical settings, with the potential to improve the predictive accuracy of stroke recovery assessments.

## 1. Introduction

The mechanisms driving recovery from focal brain lesions as in stroke are still incompletely understood despite intensive research over the past decades (Grefkes and Fink, 2020). A major problem is that recovery profiles in preclinical models often differ from those observed in patient populations, thereby hampering interspecies comparisons and translation. An important and widely used concept in human stroke recovery research is the proportional recovery rule (PRR) which posits that the magnitude of recovery from motor impairment is approximately 70 % of the initial impairment (Prabhakaran et al., 2008). However, applying the PRR across studies has yielded inconsistencies, necessitating methodological caution. A common approach in applying the PRR involves fitting a linear regression model that relates initial impairment to change in impairment. Concerns have been raised regarding the removal of 'non-recoverers' or 'non-fitters', regression diagnostics, and potential nonlinear associations between recovery and initial impairment (Kundert et al., 2019; Hope et al., 2019; Bowman et al., 2021; Bonkhoff et al., 2022). There is a mismatch between clinical studies supporting the generalizability of the PRR in larger cohorts (Kundert et al., 2019), and only a handful of preclinical studies in monkeys and rats (Nashed et al., 2024; Jeffers et al., 2018).

Animal models, particularly rat and mouse models, are believed to be strongly limited in the application and interpretation of the PRR (Balkaya and Cho, 2019). To our knowledge, there is only a single study of the PRR in a large cohort of $n = 593$ male rats using a unilateral endothelin-1 lesion model and functional assessment using the Montoya staircase-reaching task (Jeffers et al., 2018). In contrast to human studies, a smaller part of approx. 30 % were identified as fitters, characterized by smaller infarct volumes and initial post-stroke impairments compared to non-fitters. This finding suggested that the existing motor measures, unlike the FM-UE used in humans, might not reflect the actual impairment adequately. The term "recovery" in rodent studies often refers to improved performance on tasks that do not distinguish spontaneous – or treatment-induced recovery from compensation (Corbett et al., 2017). In contrast to the standardized Fugl-Meyer upper limb

* Corresponding author at: University Hospital Cologne, Department of Neurology, Kerpener Strasse 62, 50937 Cologne, Germany.
*E-mail address:* markus.aswendt@uk-koeln.de (M. Aswendt).

https://doi.org/10.1016/j.expneurol.2025.115180
Received 15 October 2024; Received in revised form 6 January 2025; Accepted 2 February 2025
Available online 4 February 2025
0014-4886/© 2025 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

score (FM-UE) in humans, scoring systems in mice, such as the Bederson scale of forelimb flexion (Bederson et al., 1986), are somewhat subjective, less standardized, and not widely used. For example, compensatory responses, e.g., avoiding impaired limb use or using a different movement pattern, are hard to distinguish from "true" recovery (Jones, 2017), especially when with single rodent motor tests focused on a specific movement (e.g. gait). Therefore, applying a battery of motor tests, including spontaneous limb movements (e.g. in the cylinder test), is recommended to detect compensatory mechanisms (Corbett et al., 2017).

To address the gap of PRR validation in a stroke mouse model, we transformed three sensorimotor tests, well established to assess motor recovery after cortical stroke, i.e., the cylinder, grid walk, and rotating beam test, applied over four weeks. We measured the ratio of mice following the PRR and determined a group with accelerated recovery in relation to stroke lesion size and location.

## 2. Methods

### 2.1. Experimental model

The experimental protocol was designed according to the IMPROVE and ARRIVE principles. Adult male C57Bl/6 J mice (9–15 weeks old, The Jackson Laboratory) were randomly assigned to the stroke and sham groups. For this study, several experiments with the same stroke mouse model were pooled based on an initial pre-screening including only mice with longitudinal data (MRI and behavior) from at least three time points. See detailed description in **Supplementary Material**.

### 2.2. Deficit score calculation

To accurately quantify the behavioral state of the mice, a single deficit score was introduced, which combines the results of the rotating beam, cylinder, and grid walk tests. Because these tests depict different types of sensorimotor deficits, which are analyzed differently and therefore have different scales, z-transformation was used to normalize the scores. This way it was possible to average the results of each mouse per test into one composite score, referred to as the deficit score.

### 2.3. Application of the proportional recovery rule

To describe mathematically the proportional recovery rule (PRR) based on the Fugl-Meyer Assessment of the Upper Extremity (FMAUE) score, we used the definition of Kundert et al. (Kundert et al., 2019):

1. **Initial Impairment (FMAUEii)**: In Eq. (1a), FMAUEii represents the initial impairment, which is calculated by subtracting the initial measurement of the FMAUE (FMAUEi) obtained shortly after the stroke from 66 - the maximum possible score.

$$FMAUE_{ii} = 66 - FMAUE_i \tag{1a}$$

2. **Change in FMAUE (ΔFMAUE)**: Eq. (1b) defines ΔFMAUE as the change in the FMA-UE score. It is determined by subtracting the final measurement of the FMA-UE score (FMAUEf) from the initial impairment score (FMAUEii).

$$\Delta FMAUE = FMAUE_f - FMAUE_{ii} \tag{1b}$$

3. **Proportional Recovery Rule (PRR)**: Eq. (1c) illustrates the change in FMA-UE according to the proportional recovery rule. This rule posits that at three months post-stroke, patients should recover approximately 70 % of their maximum potential improvement, with the slope around 0.7.

$$\Delta FMAUE = slope^{*}FMAUE_{ii} + intercept, slope = 0.7 \tag{1c}$$

### 2.4. Introduction of the mouse proportional recovery rule (MRR)

To assess the applicability of the proportional recovery rule (PRR) in a mouse model, we here introduce the mouse recovery rule (MRR). The PRR, commonly applied in human studies, was evaluated to determine its relevance to behavioral data in mice. To this end, a new rule specific to mice was defined, referred to as the mouse recovery rule (MRR). Several aspects needed to be considered in translating the PRR calculation to the mouse model. First, the behavior tests measure a sensorimotor deficit, thus the logic is reversed compared to the FMA-UE score. Another difference, or even an advantage, is that true baseline behavioral data, i.e., before stroke, are available. Accounting for these considerations, the equations were translated as follows:

1. **Initial Impairment (DSii)**: Eq. 1 uses the difference between the healthy state (FMA-UE ~66 or ceiling value) and the FMA-UE score shortly after the stroke. This is translated in Eq. (2a) by replacing the ceiling value of 66 in the FMA-UE score with the baseline value ($DS_{BL}$) of the deficit score, representing the healthy state, and the deficit score on day 3 ($DS_{P3}$) post-stroke, representing the deficit in the (sub-)acute phase.
2. **Change in Deficit Score (ΔDS):** Eq. (2b) defines the deficit score as the difference between the initial impairment ($DS_{ii}$) and the deficit score at 28 days post-stroke ($DSP_{28}$), also referred to as the "Change Observed" throughout the text (Fig. 3).
3. **Proportional Recovery Rule (PRR) for the Deficit Score:** In Eq. (2c), the change in the deficit score follows the same logic as Eq. (1c), with a slope and an intercept. Instead of defining a fixed value of 0.7, this value was regressed based on our data.

$$DS_{ii} = DS_{P3} - DS_{BL} \tag{2a}$$

$$\Delta DS = DS_{ii} - DS_{P28} \tag{2b}$$

$$\Delta DS = slope^{*}DS_{ii} + intercept \tag{2c}$$

### 2.5. Iterative cluster analysis

A two-step approach was developed, which consisted of individual regression analysis and iterative clustering refinement. First, the change observed in motor deficit (Eq. (2b)) was plotted against the initial impairment (Eq. (2a)) for each subject, and a linear fit was applied to the data to determine the slope and intercept of the regression line. This approach allowed for identifying recovery patterns similar to those observed in the human Proportional Recovery Rule (PRR) and Mouse Recovery Rule (MRR). Specifically:

1. **Initial Regression and Clustering:** The MRR was derived with a slope and intercept, while the PRR had a fixed slope of 0.7 with an undefined intercept. The intercept from the MRR fit was used to define the PRR line, and both lines were plotted for comparative analysis. Subjects were then initially clustered based on their adherence to these recovery lines. Clustering was performed by calculating the Euclidean distance between each subject's data point and the two recovery lines (PRR and MRR). Subjects were assigned to the cluster of the line to which they were closer based on this Euclidean distance.
2. **Iterative Refinement:** In the second step, an iterative process was employed to refine the clustering. This process involved recalculating the slope and intercept of the MRR line using only the data points assigned to the MRR cluster. The updated intercept was then applied to redefine the PRR line. The Euclidean distance between each data point and both lines was recalculated, and the clustering was updated iteratively. This refinement process continued until convergence was achieved, meaning clustering assignments stabilized and did not change significantly between iterations.

3. **Outlier Detection and Removal:** Outliers were identified and removed using the interquartile range (IQR) rule to ensure the accuracy of the clustering. This process was applied initially and after each iteration to prevent extreme values from skewing the results.

Utilizing Euclidean distance, this approach ensured that the clustering was based on a comprehensive measure of proximity to the defined recovery lines, reflecting both horizontal and vertical distances. The iterative process allowed for precise refinement of the recovery line parameters and the accurate classification of subjects into the PRR or MRR clusters.

### 2.6. Statistics

Statistical tests for analyzing the dynamic changes over time with consideration of multiple groups were conducted using *Python 3.10.7* and the *statsmodels* and *scipy.stats* libraries. A parametric mixed model analysis was employed to test the null hypothesis (i.e. no difference between groups or time points). Post-hoc tests were conducted for multiple comparisons: Tukey to identify differences between each time point within each group, and a Šidák to assess differences between groups at specific time points.

Effect sizes between PRR and MRR groups were also calculated using Cohen's d at each time point to quantify group differences. Cohen's d measures the magnitude of the effect size, with values of 0.2 considered small, 0.5 medium, and 0.8 or higher large. Additionally, a power analysis was conducted for each time point to determine the adequacy of the cohort size for detecting significant differences. Statistical power reflects the probability of correctly rejecting the null hypothesis (i.e., detecting a true effect) and is generally considered acceptable when it is 0.8 or higher. The power analysis was performed assuming a two-group comparison with the observed sample sizes ($N_{MRR}$ and $N_{PRR}$) at each time point.

## 3. Results

### 3.1. Variability in recovery after stroke: overall patterns

The three behavioral tests, rotating beam, cylinder, and grid walk test, showed spontaneous recovery in stroke mice (Fig. 1) with considerable individual variability (**Supplementary Fig. S2**). The longitudinal data was assessed over time and between groups using a mixed-effects model approach (**Supplementary Table S1**). The tests differed in their sensitivity to detect differences between stroke and sham groups at selected time points. The rotating beam test primarily distinguished stroke from sham mice in the acute phase (day 3 $p < 0.001$, day 7 $p < 0.01$), while the cylinder test, grid walk test, and deficit scores all showed differences between stroke and sham animals up to the chronic phase (days 3, 7, 21, and 28 $p < 0.001$). Regarding time differences, the deficit scores, rotating beam test, cylinder test, and grid walk test revealed significant deficits up to day 28 post-stroke compared to baseline in the stroke group ($p < 0.001$), with the exception of the rotating beam test (only up to day 21 with $p < 0.001$). For the sham group, no significant differences were observed in any of the tests or the deficit scores between any of the time points and baseline.

### 3.2. Iterative cluster analysis

To analyze the PRR, the z-transformed values of the three tests were used as the deficit score to plot the change observed between 28 days post-stroke (P28) and initial impairment (Eq. (2b)) versus the initial impairment itself (Eq. (2a)). The linear fit changed with each iteration until convergence between the 4th and 5th iteration. Further interactions did not change the slope or the intercept for the MRR fit further (Fig. 2). The intercept for both clusters converged at −0.72. The goodness of fit for both MRR and PRR was high (R-squared of 0.93 and 0.80, respectively). To determine if the slopes between the PRR and MRR were significantly different, an F-test was performed. The results showed that the overall slopes were not identical, with an F-value of 4.2 (DFn = 1, DFd = 91) and $p < 0.05$ (**Supplementary Table 3**).

The initial cohort consisted of 125 subjects, however, the final analysis was performed on 95 mice due to the exclusion of subjects where critical time points were missing. Specifically, missing data on time points needed for the calculation of the initial impairment and the observed change led to this reduction in animal numbers. The final clustering resulted in 47.36 % ($N = 45$) mice in the PRR and 52.64 % ($N = 50$) in the MRR cluster. Mice in the PRR cluster exhibited the predefined fixed slope of 0.7, while mice in the MRR cluster displayed a slope of 0.93, indicative of a distinct recovery trajectory.

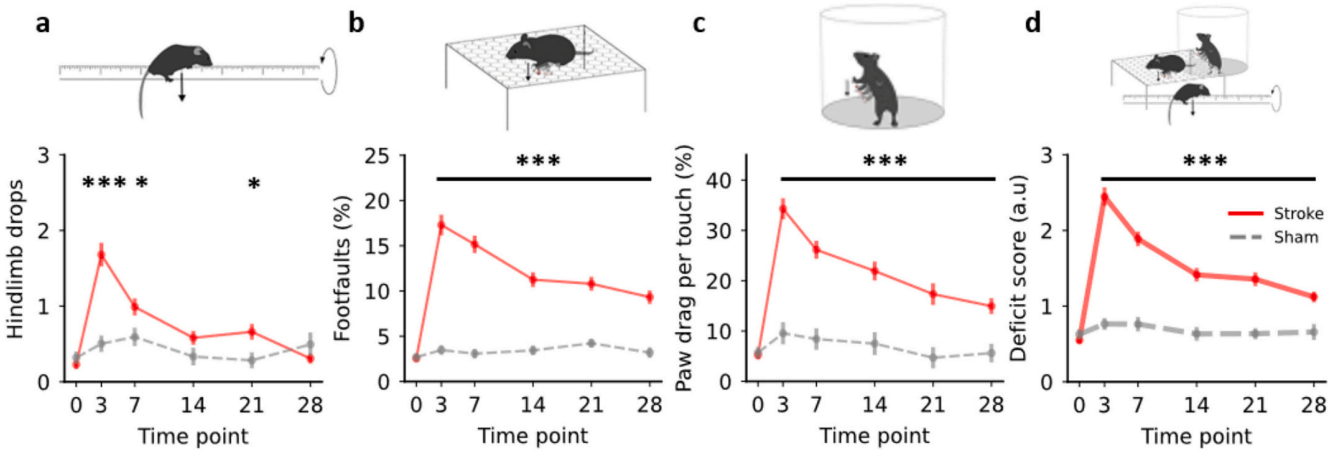



**Fig. 1.** Dynamics of spontaneous amelioration of specific motor deficits four weeks after cortical stroke. (a) Rotating beam tests: number of hindlimb drops recorded when the mouse walks over a rotating beam towards the homecage. (b) Grid walk test: number of footfaults related to the total number of steps during spontaneous walking on an elevated grid. (c) Cylinder test: number of paw drags per touch with the affected paw at the cylinder wall. (d) Deficit score, calculated by averaging the z-transformed values of each test (a-c). Data in graphs are shown as mean ± SEM. Significant differences between stroke and sham at the same time points are shown as $*p \leq 0.05$, $**p \leq 0.01$, and $***p \leq 0.001$, respectively. Mixed model analysis was used to assess significance, followed by post hoc Tukey tests for within-group time point comparisons and Šidák multiple comparison tests for between-group significant time point differences.

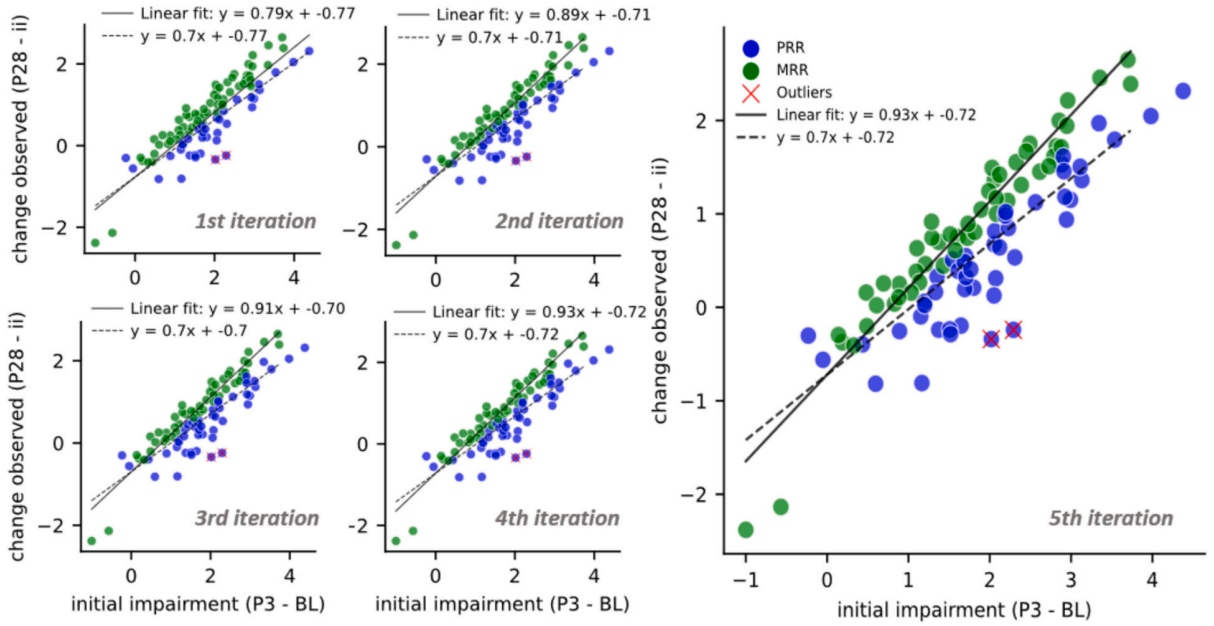



**Fig. 2.** Iterative cluster analysis process for the mouse recovery rule (MRR) and proportional recovery rule (PRR). For the calculated deficit scores, the change observed (difference between 28 days poststroke and initial impairment) for each subject was plotted against their initial impairment (difference between Acute phase or P3 and baseline values), and a linear fit was applied to determine the slope and intercept for the MRR. The PRR was assigned a fixed slope of 0.7, with the intercept derived from the MRR fit. Subjects were initially categorized into two clusters based on their proximity to the PRR or MRR line. An iterative process refined clustering by recalculating linear fits and updating cluster assignments until convergence. The figure illustrates the convergence of this iterative process. By the 5th iteration, the slope and intercept of the linear fit stabilized at 0.93 and − 0.72, respectively. This stability indicates no further changes in these parameters, confirming the consistency of cluster assignments. The goodness of fit for the MRR and PRR was evaluated, resulting in an R-squared of 0.9316 for the MRR fit and 0.8071 for the PRR fit. This analysis was performed on a cohort of 38 subjects, 22 of whom were assigned to the MRR group and 16 to the PRR group.

### 3.3. *Variability in recovery after stroke: cluster-specific trajectories*

Clustered behavior data were statistically compared over time and between groups using a mixed-effects model approach (**Supplementary Table S2**). Both the PRR and MRR clusters showed reduced deficits across all behavioral tests, with significant improvements noted from day 0 to day 3, day 7, day 14, day 21, and day 28 post-stroke across all tests ($p < 0.01$), except for the rotating beam test in the PRR cluster from day 0 to day 28, which showed no significant recovery (Fig. 3). The difference in recovery between day 21 and day 28 was less pronounced, showing a significant change only in the MRR cluster across cylinder tests and deficit scores ($p < 0.05$), suggesting variable recovery trajectories between clusters (**Supplementary Fig. 2**). PRR and MRR started from similar deficit levels (no significant differences at day 3), however, MRR mice recovered faster and more thoroughly, as the comparison with PRR-fitting mice was significant at 14 ($p < 0.05$) and up to 29 days ($p < 0.001$). Except for the grid walk test ($p < 0.001$), the MRR deficit was not significantly different to sham mice at 28 days, In contrast, the PRR group remained with a significantly higher deficit compared to sham mice ($p < 0.001$, except for the rotating beam test with $p > 0.05$).

### 3.4. *Cohort size and group differences*

To ensure the appropriate cohort size and to validate the observed group differences between PRR and MRR, Cohen's d was used to quantify effect sizes and to perform a power analysis at each time point

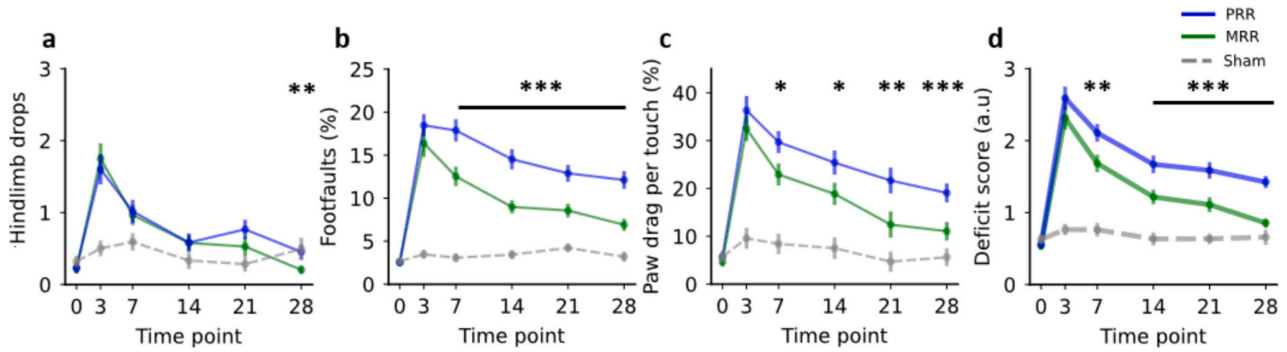



**Fig. 3.** Dynamics of behavioral test performance over time for each group after cluster assignments. (a) Rotating beam test: number of hind limb drops recorded for a walking task on a rotating beam. (b) Grid walk test: number of footfaults calculated as a percentage of total footsteps. (c) Cylinder test: number of paw drags, calculated as a percentage per touch with the affected paw at the cylinder wall. (d) Deficit score, calculated by averaging the z-transformed values of each test. Data are shown as mean ± SEM. Significant differences between the two cluster groups (PRR and MRR) at the same time points are indicated by $^{*}p \leq 0.05$, $^{**}p \leq 0.01$, and $^{***}p \leq 0.001$. Mixed model analysis was used to assess significance, followed by post hoc Tukey tests for within-group time point comparisons and Šidák multiple comparison tests for between-group significant time point differences.

**Table 1**
Effect sizes and statistical power across time points for PRR and MRR groups. Cohen's d and statistical power values are shown for each time point to quantify differences between PRR and MRR groups and evaluate the adequacy of the cohort size.

| Time points (days) | Cohens d | $N_{MRR}$ | $N_{PRR}$ | Power |
|---|---|---|---|---|
| 0 | 0.0755 | 52 | 54 | 0.0668 |
| 3 | 0.2836 | 55 | 50 | 0.3139 |
| 7 | 0.6077 | 55 | 51 | 0.8846 |
| 14 | 0.832 | 47 | 36 | 0.9789 |
| 21 | 0.9867 | 33 | 35 | 0.9765 |
| 28 | 1.6874 | 51 | 45 | 1 |

(Table 1). The results indicate small effect sizes and low power at baseline (P0, Cohen's d = 0.0755, Power = 0.0668) and the acute phase (P3, Cohen's d = 0.2836, Power = 0.3139), consistent with minimal divergence between PRR and MRR groups at these early stages (Fig. 3D). From day 7 after stroke onward, effect sizes increased substantially, with medium to large Cohen's d values (day 7 = 0.6077, day 28 = 1.6874) and high power (≥0.88). These findings confirm that the cohort size was sufficient to detect meaningful differences in recovery trajectories, particularly during the intermediate and chronic phases of recovery. This validation supports the robustness of the clustering approach and the meaningful distinction between PRR and MRR recovery trajectories.

### 3.5. Role of stroke lesion size and location

To verify that the differences in functional recovery were not driven by lesion size or location, a comprehensive qualitative voxel- and quantitative atlas-based analysis of the stroke lesion as determined using T2-weighted MRI was conducted (Fig. 4). The semi-automatically defined lesion masks were registered to the Allen Mouse Brain Atlas (Fig. 4a). The qualitative voxel-wise comparison showed no differences between the MRR and PRR cluster compared to the combined group in terms of lesion pattern and lesion extent (Fig. 4b). This was also reflected in a similar lesion volume for the MRR (15.32 + −10.61) and PRR (18.57 + −10.47) group (Fig. 4c). Further, there was no difference in the lesion location as quantified by the percentage of lesion mask overlap with brain atlas regions (Fig. 4d).

The clustering based on the linear fit approach (Fig. 2d) contained two outliers as detected based on the interquartile range rule. These outliers showed a relatively high initial impairment and a negative change observed, i.e., poor improvement over time. MRI data was available for only one of them, which revealed a stroke volume of approximately 28.36 $mm^3$. This is notably larger than the average stroke volume of 16.31 ± 10.69 $mm^3$ observed across all data points in both

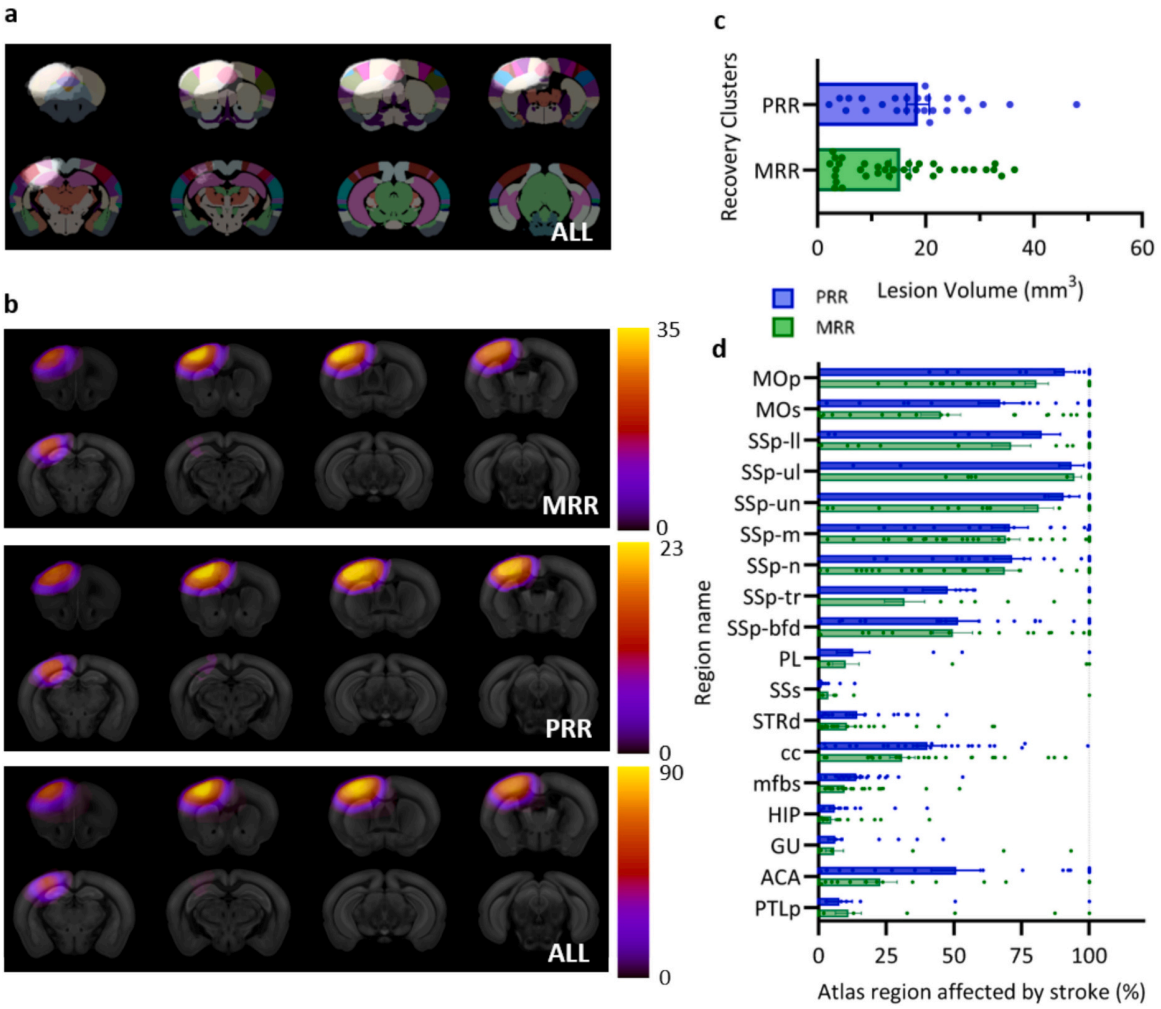



**Fig. 4.** The role of stroke lesion size, and location. (a) Stroke incidence map and modified Allen Brain Atlas (48 regions) overlay. (b) Incidence maps of cortical strokes for MRR, PRR and all mice showing no visual difference. (c) No lesion volume differences between the two recovery clusters MRR and PRR (unpaired *t*-test, $p > 0.05$). (d) There was no difference in the stroke infarct area per atlas region between the two recovery clusters MRR and PRR (mixed model analysis with subsequent Sidak multiple comparisons, $p > 0.05$).

clusters. Additionally, the two lowest points in the MRR cluster, in terms of the *change observed* value, could be considered outliers. However, based on the interquartile range rule, this was not the case. These subjects exhibited lesion sizes of 32.6 $mm^3$ and 36.4 $mm^3$, respectively, again indicating relatively large strokes.

## 4. Discussion

We present a new approach combining three different motor tests to improve the reliability of motor assessment in mice and to compensate for differences in sensitivity over a longer time frame (Balkaya and Cho, 2019). The results revealed that approximately 40 % of the mice with cortical strokes followed the PRR and 60 % followed a new recovery trajectory with a slope of approximately 0.9, which we termed the mouse recovery rule (MRR). MRR-fitting mice recover 90 % of their initial impairment within the first four weeks post-stroke. Notably, the different recovery trajectories were not related to lesion size or location differences. This study provides new insights into the translational applicability of the PRR to mouse studies and emphasizes the individual differences in recovery mechanisms present beyond lesion size and location.

### 4.1. Classification of recovery trajectories

Identifying recovery patterns after stroke is crucial for advancing rehabilitation strategies, with the Proportional Recovery Rule (PRR) representing one established approach (Prabhakaran et al., 2008). Expanding on the classification of recovery trajectories, Van der Vliet et al. developed a longitudinal mixture model using an exponential recovery function (Rick et al., 2020), challenging the PRR's assumptions and addressing criticisms of its predictive oversimplification. Furthermore, Nashet's application of a K-means clustering approach on non-human primate stroke data (Nashed et al., 2024) might provide a deeper separation of the various sources of different recovery patterns. Additionally, Jeffers et al. adapted the PRR to rodent models (Jeffers et al., 2018), categorizing subjects into 'fitters,' who follow the PRR, 'non-fitters,' who do not adhere to the PRR, and 'decliners,' who do not recover at all. This categorization demonstrates the PRR's applicability across species and underscores the variability in recovery trajectories. In our study, we observed similar classifications, with "non-fitters" termed as MRR exhibiting better outcomes than the PRR group, which aligns with the previously defined 'fitters'. In contrast to Jeffers' study where 'non-fitters' experienced poorer recovery, the MRR group showed significantly better recovery outcomes. This observation led to the proposal of the MRR as a novel predictive model tailored to mouse recovery post-stroke.

### 4.2. Behavior tests to assess spontaneous recovery after stroke

In line with previous studies, mice with cortical strokes induced by photothrombosis recovered substantially during the first weeks after stroke (Yu et al., 2015; Clarkson et al., 2013; Götz et al., 2023). As expected, there were considerable differences in the recovery rate and the sensitivity in detecting long-term motor deficits, which can be attributed to the different weighting ranging from general motor coordination (e.g., gait) to detailed motor output (e.g., correct paw placement on the grid). The results revealed that the rotating beam test is highly susceptible to early balance and coordination deficits but less so for long-term ones. In contrast, the cylinder test demonstrated a more gradual recovery, with significant differences between the stroke and sham groups observed at multiple time points, indicating its sensitivity to persistent forelimb use asymmetry. The grid walk test, assessing foot faults, revealed the highest sensitivity to sensorimotor impairments, with significant differences between the stroke and sham groups maintained until 3 weeks post-stroke, highlighting its efficacy in detecting acute and chronic deficits in motor coordination and proprioception.

The combination of these behavioral tests, which all showed a significant impairment at 4 weeks compared to baseline, allowed us to rule out a primary concern that most of the behavioral tests lose sensitivity within weeks so that the PRR cannot hold for animal models of stroke (Balkaya and Cho, 2019). In contrast, our clustering analysis revealed that a substantial number of mice (40 %) adhere to the proportional recovery rule, suggesting that this rule may have relevance beyond clinical settings. In comparison to the previous rat MCAO (middle cerebral artery occlusion) study by (Jeffers et al., 2018), which reported that 30 % of rats conformed to the PRR rule, our findings highlight that a comparable proportion of mice in our study also adhere to the PRR. The identification of a distinct cluster with its own regression pattern highlights the heterogeneity of post-stroke recovery and underscores the importance of individualized treatment approaches.

While our study focuses on the rodent recovery framework, it is important to contextualize the parallels and differences between recovery assessments in humans and rodents. Proportional recovery initially defined for patients with upper limb impairments (assessed using the Fugl-Meyer Assessment of the Upper Extremity, FMAUE), it has been later proven in lower limb impairment (assessed using the FMA lower extremity, FMALE) (Smith et al., 2017). Table 2 provides a conceptual comparison between the FMAUE/FMALE-based PRR used in human studies and the MRR (based on the combination of three behavior tests) developed in this study. This comparison highlights the similarities and differences in motor domains assessed, scoring methodologies, and the availability of baseline data, underscoring the translational challenges when applying human recovery concepts to preclinical models. Although the individual behavior test had potential limitations, the combination of three behavior tests proved to be a valuable complement to the FMA.

### 4.3. Implications of variability in recovery after stroke

The analysis of PRR and MRR clusters revealed no significant differences at the acute phase (Fig. 3, time point P3). However, recovery trajectories and endpoints diverged significantly across subsequent time points. Animals in the MRR cluster demonstrated faster and more comprehensive recovery, implying a higher recovery potential. The implications of these findings are significant for preclinical endpoint selection. If recovery patterns are evaluated based solely on PRR trajectories, the faster and more complete recovery seen in MRR animals could be underestimated, potentially leading to inaccurate assessments of therapeutic efficacy. Recognizing individual variability in recovery trajectories is therefore critical for designing robust preclinical studies and improving the translational relevance of findings.

### 4.4. Role of stroke lesion size and location

Experimental studies of stroke recovery have extensively explored the role of lesion size, location, and neural circuit integrity. Lesion size is a common imaging marker to predict or explain stroke outcomes (Sperber et al., 2023). However, as shown in human and rodent studies, the additional analysis of the lesion topology improves the prognostic value (Scheulin et al., 2021; Knab et al., 2023; Tscherpel et al., 2024). Many rodent models use large infarcts, which may not accurately reflect the typical, smaller strokes experienced by humans. This discrepancy can impact the relevance of findings to human stroke conditions. In large population studies the lesion size ranges from 28 to 80 $cm^3$, which translates to approximately 4.5 % to 14 % of the total volume of the ipsilesional hemisphere (Carmichael, 2005). We used photothrombosis to induce homogeneous and well-defined lesions. Our data show a mean of $7 \pm 2.21$ % for the PRR cluster and $5.6 \pm 2.02$ % for the MRR cluster. This aligns with the retrospective analysis of studies using endothelin-1 to induce forelimb sensorimotor cortex and dorsolateral striatum lesions in rats (Jeffers et al., 2018). Approximately 30 % of the rats were identified as fitters of the PRR, which had smaller lesions (5.3 %) than

**Table 2**
Comparison of recovery assessment frameworks in humans vs. rodents.

| Aspect | Human (PRR) | Rodent (MRR) |
|---|---|---|
| Motor domains | **Gross motor skills** (upper and lower limbs)<br>**Fine motor skills** (only indirectly measured).<br>**Coordination and speed** (precision and time to perform movement)<br>**Balance** (static and dynamic)<br>**Motor planning** (only indirectly measured)<br>**Strength** (via movement performance)<br>**Reflexes activity** (neurological integration) | **Gross Motor Skills:** Assessed across all three tests through general movement and locomotion.<br>**Fine Motor Skills:** Evaluated in the grid walk test via precise paw placement.<br>**Coordination:** Addressed by all three tests, (**but not** analyzed in this study, e.g. based limb kinematics).<br>**Balance:** Primarily assessed by the rotating beam test, to a lesser extent also from cylinder and grid walk tests.<br>**Motor Planning:** Captured in the grid walk and rotating beam tests, reflecting anticipatory and corrective movements (**but not** analyzed in this study).<br>**Strength:** Indirectly assessed in the rotating beam test through the ability to maintain stability.<br>**Reflexes activity:** not directly assessed. |
| Baseline data | Not available; instead assumed maximum score for initial impairment. | Pre-stroke baseline behavioral data available for all mice. |
| Scoring methodology | Composite score derived from multiple subtests according to a standardized (manual) assessment (with potential ceiling effects. | Motor deficits manually analyzed using the video recordings. Deficit score using z-transformed data to standardize test outputs. |
| Recovery rule | Recovery is predicted to be ~70 % of the initial impairment | Recovery trajectories vary, including PRR-like and faster recovery patterns. |

the rats that did not show proportional recovery or deteriorated over time (8.3 %). Given a comparable lesion volume, the ratio of PRR fitters differs between up to 80 % (humans), 30 % (rats), and 47 % (mice). Although the lesion volume overlaps across species, the lesion location differs and thus the existing data cannot be directly compared. Effective recovery modeling must consider specific neuronal circuits affected by the stroke (Carmichael, 2005). Byblow et al. studied the role of the integrity of the corticomotor pathway in proportional recovery. They discovered that patients with a viable ipsilesional corticomotor pathway were more likely to experience proportional recovery, while those with compromised pathways exhibited reduced recovery (Byblow et al., 2015). Such a longitudinal characterization of motor fiber integrity using diffusion MRI in rodents is missing. Future studies may include a comprehensive analysis of lesion topology and the differential effects on functional circuits and structural brain connectivity.

### 4.5. Transformation of behavioral to deficit scores

The process of deriving a deficit score through z-transformation involves converting individual scores into z-scores, and integrating these scores to provide a comprehensive measure of motor recovery. This standardized deficit score was analyzed longitudinally and across experimental groups (sham and stroke) to evaluate the impact of stroke and subsequent recovery. This method facilitates a standardized comparison of motor function across different tests and time points, which is crucial for capturing the behavioral recovery trajectory in stroke-affected mice. Using z-transformation is particularly pertinent because this statistical transformation is typically suitable for data with a normal distribution. However, in this study, a normal distribution across the entire cohort of stroke mice was not assumed. Initially, stroke mice scores were expected to be comparable to sham mice at baseline, but immediately post-stroke, a significant deficit was anticipated, leading to a deviation from normal distribution. While normal distribution within each time point is ideal, factors such as variability in test sensitivity and the subjective nature of assessing the quantified values of each test may challenge this assumption. Despite these complexities, merging the results from three tests into a single deficit score helps mitigate the limitations of individual tests by leveraging their combined strengths (Guilloux et al., 2011). This approach is also supported by the finding that the deficit score revealed more robust significant differences. In light of dynamic changes observed from baseline to the chronic phase, applying the z-transformation across the entire cohort is crucial. This approach allows for meaningful comparisons among subjects with varying scores at different tests and ensures consistency across different time points. Applying z-transformation separately for each time point would obscure these longitudinal dynamics, underscoring the importance of a comprehensive transformation approach. Given the variability in behavioral data and the need to avoid fixed boundaries (Bowman et al., 2021) that might constrain interpretation, the z-score transformation was ultimately chosen for its ability to standardize the data while maintaining flexibility. This decision and its implications are further discussed in the subsequent section.

### 4.6. Statistical considerations

#### 4.6.1. Coupling effects and measurement scales

Mathematical coupling can be a significant issue in PRR-related statistical analyses (Chong et al., 2023; Bowman et al., 2021). Others argue that when properly analyzed, the coupling can be viewed as a notational construct rather than a true confound (Chong et al., 2023). Likewise, our analysis might be subjective to mathematical coupling due to the definitions of the initial impairment (deficit at day 3 - deficit at baseline) and the change in deficit score (deficit at day 28 - initial impairment). Future studies may mitigate the effects of mathematical coupling by using alternative modeling approaches that do not rely on the change variable (Y - X) (Bowman et al., 2021), and fitting models to raw performance scores over time (Rick et al., 2020).

Another confound in PRR-related statistics is the ceiling effect of the maximum FM-UE score of 66, which can compress scores at the upper end and skew the results (Bowman et al., 2021). Discarding or transferring data at the ceiling was suggested to reduce bias. In our case, the deficit score is an average of three behavioral assessments, each of which underwent z-transformation to enable averaging. As the z-transformation creates no ceiling or floor effect, this approach eliminates the problem of scaling thresholds. This perspective is crucial for our study, as it suggests that the theoretical concerns raised by mathematical coupling and ceiling effects do not fundamentally undermine the validity of our observed correlations. By carefully addressing these concerns using z-scores and considering the limitations of scales, we ensure the robustness of our analysis and the validity of applying the proportional recovery rule in our mouse model of stroke recovery.

#### 4.6.2. Availability of baseline data

In human studies, the Fugl-Meyer Assessment of the Upper Extremity (FMA-UE) score reaches a maximum possible score of 66 serving as a reference point for initial impairment calculations. However, in our mouse model, we leverage baseline behavioral data obtained before stroke inducement. This allowed a more precise measurement of the initial impairment (DSii), as the actual pre-stroke performance of each mouse could be used rather than an assumed ceiling value, which can help mitigate the coupling effect as it increases variability in the initial impairment making it dependent on two variables.

### 4.7. Limitations and future directions

The MRR was evaluated in adult male C57BL/6 J mice, which received photothrombotic cortical strokes, where lesion size had no impact on the recovery trajectory. An important consideration is whether the MRR can be directly extended to other rodent models. Previous studies have shown that lesion size differs in rats compared to mice and differs between strains (Fluri et al., 2015) and further that deficit and recovery can be different even when comparing only two rat strains (Kunze et al., 2014). Although lesion size had no influence on the recovery trajectory in our study, in stroke models with larger variability in lesion size and topology, e.g. MCAO, this should be re-considered and validated across species and strains (Jeffers et al., 2018). These animal model-specific differences may influence deficit scores and recovery trajectories, requiring adaptation of MRR to account for model-specific characteristics. Incorporating appropriate behavioral tests tailored to the functional deficits of each model will be critical for extending MRR to diverse preclinical settings. Furthermore, it is essential to acknowledge future studies to add an age and sex-matched cohort of mice (and rats). It has been stressed that the preclinical stroke research should include biological variables and clinically relevant features, such as age, sex, diabetes, obesity, and hypertonia (Wolf and Ergul, 2021; Corbett et al., 2017).

One limitation of our behavior tests is the measure of single partially overlapping deficits, which are more prone to compensation rather than a comprehensive score reflecting true recovery (Prabhakaran et al., 2008). However, all tests reflect spontaneous behavior, which is in general less prone to compensatory effects (Corbett et al., 2017). To our knowledge, no comprehensive scoring system has complemented the FMA-UE scoring system developed for mice. The Bederson scale, helpful for distinguishing stroke lesion size in the acute phase, has not been validated to predict recovery outcomes (Bieber et al., 2019).

Although z-scoring was applied to normalize and combine behavioral data into a single deficit score, it does not address potential biases introduced by manual rater variability. Differences in experimental setup or scoring across labs (e.g., lab A vs. lab B) or even within the same lab when different raters are involved can influence results. Recently, deep behavioral phenotyping with kinematic analysis has shown movement patterns related to compensation, like adapting contralateral limb stability and reducing trunk distance from the target prior to reaching, which are common in human stroke patients (Balbinot et al., 2018; Weber et al., 2022). Future studies would benefit from a more detailed translational kinematic assessment to reduce inter-rater variability and enhance the reproducibility of findings across studies. Also increasing the sample size would allow alternative clustering methods, such as the longitudinal mixture model of FMA-UE recovery (Rick et al., 2020).

## 5. Conclusion

In summary, our study demonstrates the translation of the proportional recovery rule from clinical practice to a mouse stroke model. By employing a clustering approach, we delineate distinct recovery trajectories that are not driven by lesion size or location. These findings suggest that alternative recovery mechanisms drove post-stroke recovery in this homogeneous group of mice. The findings contribute to our understanding of stroke pathophysiology and may inform the development of therapeutic interventions to be translated to stroke patients.

## Ethics statement

All data used in this study were acquired in strict adherence to the ARRIVE guidelines for reporting in vivo animal experiments and the IMPROVE guidelines for stroke animal models, as recommended by (Kilkenny et al., 2012) and (Du Sert et al., 2017). The datasets were processed in accordance with comprehensive ethical protocols approved by the *Landesamt für Natur, Umwelt und Verbraucherschutz* North Rhine-Westphalia, Germany, under animal protocol numbers 84–02.04.2016. A461 and 84–02.04.2014.A305. Additional ethical clearance was obtained from the Gothenburg Ethics Committee, Sweden, under animal permit number 1551/2018.

## Sources of funding

This work was funded by the Friebe Foundation (T0498/28960/16) and the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 431549029 – SFB 1451.

## CRediT authorship contribution statement

**Aref Kalantari:** Writing – review & editing, Writing – original draft, Visualization, Formal analysis, Data curation, Conceptualization. **Carolin Hambrock:** Writing – original draft, Methodology, Formal analysis, Data curation. **Christian Grefkes:** Writing – review & editing, Writing – original draft, Project administration, Funding acquisition. **Gereon R. Fink:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Funding acquisition. **Markus Aswendt:** Writing – review & editing, Writing – original draft, Visualization, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization.

## Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Markus Aswendt reports financial support was provided by the Friebe Foundation. Markus Aswendt, Christian Grefkes and Gereon R. Fink report financial support was provided by the Deutsche Forschungsgemeinschaft. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

We acknowledge Frederique Wieters, Veronika J. Fritz, Olivia Käsgen, Niklas Palast, Marc Schneider, and Victor Vera Frazao for their contributions in acquiring and processing data for this study.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.expneurol.2025.115180.

## Data availability

The project dataset containing all data, code, scripts, and documentation can be accessed online https://gin.g-node.org/Aswendt_Lab/2024_Kalantari_PRR.

## References

Balbinot, Gustavo, Schuch, Clarissa Pedrini, Jeffers, Matthew S., McDonald, Matthew W., Livingston-Thomas, Jessica M., Corbett, Dale, 2018. Post-stroke kinematic analysis in rats reveals similar reaching abnormalities as humans. Sci. Rep. 8 (1), 8738.

Balkaya, Mustafa, Cho, Sunghee, 2019. Optimizing functional outcome endpoints for stroke recovery studies. J. Cereb. Blood Flow Metab. 39 (12), 2323–2342.

Bederson, J.B., Pitts, L.H., Tsuji, M., Nishimura, M.C., Davis, R.L., Bartkowski, H., 1986. Rat middle cerebral artery occlusion: evaluation of the model and development of a neurologic examination. Stroke 17 (3), 472–476.

Bieber, Michael, Gronewold, Janine, Scharf, Anne-Carina, Schuhmann, Michael K., Langhauser, Friederike, Hopp, Sarah, Mencl, Stine, et al., 2019. Validity and reliability of neurological scores in mice exposed to middle cerebral artery occlusion. Stroke 50 (10), 2875–2882.

Bonkhoff, Anna K., Hope, Tom, Bzdok, Danilo, Guggisberg, Adrian G., Hawe, Rachel L., Dukelow, Sean P., Chollet, François, Lin, David J., Grefkes, Christian,

Bowman, Howard, 2022. Recovery after stroke: the severely impaired are a distinct group. J. Neurol. Neurosurg. Psychiatry 93 (4), 369–378.
Bowman, Howard, Bonkhoff, Anna, Hope, Tom, Grefkes, Christian, Price, Cathy, 2021. Inflated estimates of proportional recovery from stroke: the dangers of mathematical coupling and compression to ceiling. Stroke 52 (5), 1915–1920.
Byblow, Winston D., Stinear, Cathy M., Alan Barber, P., Petoe, Matthew A., Ackerley, Suzanne J., 2015. Proportional recovery after stroke depends on Corticomotor integrity. Ann. Neurol. 78 (6), 848–859.
Carmichael, S. Thomas, 2005. Rodent models of focal stroke: size, mechanism, and purpose. NeuroRx 2 (3), 396–409.
Chong, Benjamin, Wang, Alan, Stinear, Cathy M., 2023. Proportional recovery after stroke: addressing concerns regarding mathematical coupling and ceiling effects. Neurorehabil. Neural Repair 37 (7), 488–498.
Clarkson, Andrew N., López-Valdés, Héctor E., Overman, Justine J., Charles, Andrew C., Brennan, K.C., Thomas Carmichael, S., 2013. Multimodal examination of structural and functional remapping in the mouse photothrombotic stroke model. J. Cereb. Blood Flow Metab. 33 (5), 716–723.
Corbett, Dale, Carmichael, S.T., Murphy, Timothy H., Jones, Theresa A., Schwab, Martin E., Jolkkonen, Jukka, Clarkson, Andrew N., et al., 2017. Enhancing the alignment of the preclinical and clinical stroke recovery research pipeline: consensus-based core recommendations from the stroke recovery and rehabilitation roundtable translational working group. Int. J. Stroke 12 (5).
Du, Sert Nathalie Percie, Alfieri, Alessio, Allan, Stuart M., Carswell, Hilary V.O., Deuchar, Graeme A., Farr, Tracy D., Flecknell, Paul, et al., 2017. The IMPROVE guidelines (Ischaemia models: procedural refinements of in vivo experiments). J. Cereb. Blood Flow Metab.
Fluri, Felix, Schuhmann, Michael K., Kleinschnitz, Christoph, 2015. Animal models of ischemic stroke and their application in clinical research. Drug Des. Devel. Ther. 9 (July), 3445–3454.
Götz, Jan, Wieters, Frederique, Fritz, Veronika J., Käsgen, Olivia, Kalantari, Aref, Fink, Gereon R., Aswendt, Markus, 2023. Temporal and spatial gene expression profile of stroke recovery genes in mice. Genes 14 (2).
Grefkes, Christian, Fink, Gereon R., 2020. Recovery from stroke: current concepts and future perspectives. Neurol. Res. Pract. 2 (June), 17.
Guilloux, Jean-Philippe, Seney, Marianne, Edgar, Nicole, Sibille, Etienne, 2011. Integrated behavioral Z-scoring increases the sensitivity and reliability of behavioral phenotyping in mice: relevance to emotionality and sex. J. Neurosci. Methods 197 (1), 21–31.
Hope, Thomas M.H., Friston, Karl, Price, Cathy J., Leff, Alex P., Rotshtein, Pia, Bowman, Howard, 2019. Recovery after stroke: not so proportional after all? Brain J. Neurol. 142 (1), 15–22.
Jeffers, Matthew Strider, Karthikeyan, Sudhir, Corbett, Dale, 2018. Does stroke rehabilitation really matter? Part A: proportional stroke recovery in the rat. Neurorehabil. Neural Repair 32 (1), 3–6.
Jones, Theresa A., 2017. Motor compensation and its effects on neural reorganization after stroke. Nat. Rev. Neurosci. 18 (5), 267–280.
Kilkenny, C., Browne, W.J., Cuthill, I.C., Emerson, M., Altman, D.G., 2012. Improving bioscience research reporting: the ARRIVE guidelines for reporting animal research. Osteoarthr. Cartil./OARS Osteoarthr. Res. Soc. 20 (4), 256–260.
Knab, Felix, Koch, Stefan Paul, Major, Sebastian, Farr, Tracy D., Mueller, Susanne, Euskirchen, Philipp, Eggers, Moritz, et al., 2023. Prediction of stroke outcome in mice based on noninvasive MRI and behavioral testing. Stroke 54 (11), 2895–2905.
Kundert, Robinson, Goldsmith, Jeff, Veerbeek, Janne M., Krakauer, John W., Luft, Andreas R., 2019. What the proportional recovery rule is (and is not): methodological and statistical considerations. Neurorehabil. Neural Repair 33 (11), 876–887.
Kunze, Allison, Zierath, Dannielle, Drogomiretskiy, Olga, Becker, Kyra, 2014. Variation in behavioral deficits and patterns of recovery after stroke among different rat strains. Transl. Stroke Res. 5 (5), 569–576.
Nashed, Joseph Y., Shearer, Kaden T., Wang, Justin Z., Chen, Yining, Cook, Elise E., Champagne, Allen A., Coverdale, Nicole S., et al., 2024. Spontaneous Behavioural recovery following stroke relates to the integrity of parietal and temporal regions. Transl. Stroke Res. 15 (1), 127–139.
Prabhakaran, Shyam, Zarahn, Eric, Riley, Claire, Speizer, Allison, Chong, Ji Y., Lazar, Ronald M., Marshall, Randolph S., Krakauer, John W., 2008. Inter-individual variability in the capacity for motor recovery after ischemic stroke. Neurorehabil. Neural Repair 22 (1), 64–71.
Rick, Van Der Vliet, Selles, Ruud W., Andrinopoulou, Eleni-Rosalina, Nijland, Rinske, Ribbers, Gerard M., Frens, Maarten A., Meskers, Carel, Kwakkel, Gert, 2020. Predicting upper limb motor impairment recovery after stroke: a mixture model. Ann. Neurol. 87 (3), 383–393.
Scheulin, Kelly M., Jurgielewicz, Brian J., Spellicy, Samantha E., Waters, Elizabeth S., Baker, Emily W., Kinder, Holly A., Simchick, Gregory A., et al., 2021. Exploring the predictive value of lesion topology on motor function outcomes in a porcine ischemic stroke model. Sci. Rep. 11 (1), 3814.
Smith, Marie-Claire, Byblow, Winston D., Alan Barber, P., Stinear, Cathy M., 2017. Proportional recovery from lower limb motor impairment after stroke. Stroke 48 (5), 1400–1403.
Sperber, Christoph, Gallucci, Laura, Mirman, Daniel, Arnold, Marcel, Umarova, Roza M., 2023. Stroke lesion size - still a useful biomarker for stroke severity and outcome in times of high-dimensional models. NeuroImage Clin. 40 (September), 103511.
Tscherpel, Caroline, Mustin, Maike, Massimini, Marcello, Paul, Theresa, Ziemann, Ulf, Fink, Gereon R., Grefkes, Christian, 2024. Local neuronal sleep after stroke: the role of cortical Bistability in brain reorganization. Brain Stimul. 17 (4), 836–846.
Weber, Rebecca Z., Mulders, Geertje, Kaiser, Julia, Tackenberg, Christian, Rust, Ruslan, 2022. Deep learning-based behavioral profiling of rodent stroke recovery. BMC Biol. 20 (1), 232.
Wolf, Victoria Lea, Ergul, Adviye, 2021. Progress and challenges in preclinical stroke recovery research. Brain Circul. 7 (4), 230–240.
Yu, Cheng-Long, Zhou, Heng, Chai, An-Ping, Yang, Yue-Xiong, Mao, Rong-Rong, Lin, Xu., 2015. Whole-scale neurobehavioral assessments of Photothrombotic ischemia in freely moving mice. J. Neurosci. Methods 239 (January), 100–107.

# Chapter 5

# Discussion

The workflow was shaped by a focus on practical application and strong community integration. Using DataLad and GIN as foundational components, a robust data-management blueprint was established, ensuring structured project development, clear version control, and direct access to FAIR data principles. The complementary work by Pallast et al. [20] on metadata-focused databases synergized with our approach, which emphasized data integrity and reproducibility. Importantly, our infrastructure, combined with clear Standard Operating Procedures (SOPs), enabled us to carry out and publish our translational case study efficiently. Our implementation demonstrated that even within the constraints of hospital IT systems, the use of DataLad and GIN provides an effective strategy for storing, sharing, and tracking small-animal datasets. A structured, reproducible data management system was achieved that enables immediate project initiation after conceptualization, clear historical tracking, and modular dataset development using the YODA principles. The datasets combined raw MRI, behaviour, and histology data, processed through automated pipelines such as AIDAmri [21] and version-controlled via DataLad. Nesting datasets with submodules provided flexibility while maintaining the integrity of the superdataset.

The development of AIDAqc addressed the gap in small-animal MRI quality control. Traditional manual QC approaches were impractical for large datasets and prone to subjective errors made by manual QC detection, especially for diffusion and functional MRI. AIDAqc automated key metrics such as SNR, tSNR, motion severity, and ghosting, using majority-voting outlier detection across multiple machine-learning algorithms, validated against simulated noise and motion, and also through external datasets from 19 institutes. This approach substantially reduced the user burden of manual QC and

ensured higher reproducibility. AIDAqc's practice focused application showed its utility in rapid identification of low-quality scans before further analysis, directly accelerating the timeline of large multicentre studies and supporting robust downstream data analysis, such as connectivity mapping and tractography. AIDAqc offers several key advantages over existing QC toolboxes such as MRIQC [22] and pyfMRIqc [11]. Although MRIQC version 22.0.0 (May 3, 2022) introduced a rodent pipeline, it remains primarily human-centric, evaluating a broad array of metrics using predefined thresholds. In contrast, AIDAqc supports T2-weighted anatomical scans, diffusion tensor imaging, and resting-state fMRI, computing the core metrics of SNR, tSNR, motion, and ghosting before employing four unsupervised, majority-voting machine-learning detectors to identify scans that diverge from the cohort distribution. This multidimensional outlier-detection strategy enables robust handling of mixed cohorts with variable average image quality. Meanwhile, pyfMRIqc focuses exclusively on human fMRI, generating detailed reports without an integrated outlier-detection mechanism. Finally, AIDAqc's native support for raw Bruker data ensures seamless integration into preclinical workflows.

The proportional recovery study in mice addressed a critical gap in translating human recovery models to preclinical systems. Using robust behavioural tests (cylinder, grid walk, rotating beam) combined into a z-transformed deficit score, two distinct recovery clusters were identified. Approximately half of the mice followed a PRR-like trajectory with ∼70% recovery, while the remainder recovered ∼90% within 2–4 weeks, termed the Mouse Recovery Rule (MRR). Importantly, neither lesion size nor location explained these recovery differences, implying that underlying biological or circuit-level drivers warrant further investigation. This work progressed from an initial idea to a published study in a very short timeframe, enabled by RDM and QC as its foundation. A dedicated repository was created on GIN to track the project's evolution, and development advanced rapidly by leveraging numerous pre-existing scripts from other projects that were already directly linked to the behavioral and imaging data via the YODA structure within our RDM framework on GIN and GitHub. QC of the MRI images was performed with minimal time and effort, filtering out low-quality T2-weighted anatomical scans. In terms of the FAIR principles, this project was feasible only because the data and pre-existing scripts were *findable* and *accessible* through open access on GIN and GitHub, *interoperable* due to the modular code design that required only minimal adjustments to enable the

processing pipeline from raw data to final figures for the PPR project. *Reproducibility* is also ensured; anyone can download the open-access repository with minimal technical skills, track its development from the start, and recreate the publication's figures from the raw data. Additional examples demonstrating RDM and QC applications are presented in the following section.

## Examples of Usage

A central theme across these projects is the integration of imaging information with behavioral deficits and recovery after stroke. This overarching goal has guided multiple studies aiming to bridge neuroimaging markers with functional outcomes in rodent models of stroke. In the following, some projects are briefly discussed as examples and successors of the projects presented in this work. One major project focused on the application of diffusion tract masks to investigate secondary white matter degeneration. Specifically, this work aimed to create a spatio-temporal profile of post-stroke white matter degeneration. The analyses revealed that diffusion changes were often not uniform across entire tracts but rather dependent on the distance from the lesion core. Additionally, associations between early white matter integrity and later behavioral outcomes were investigated to identify potential predictive markers of recovery.

Another example involves the investigation of the physiological and pathophysiological evolution around the glial scar after stroke. This work focuses on understanding peri-lesional tissue changes over time to elucidate the mechanisms underlying stroke progression and recovery potential.

In the application of the proportional recovery rule to mouse stroke data, clinical models of proportional recovery were adapted to preclinical data to enhance translational relevance.

Finally, the integration of functional and structural connectivity analyses represents another example of usage. By bridging these modalities, researchers are exploring structural–functional coupling to understand network-level mechanisms underlying behavioral recovery.

All these projects relied on the clear and structured data management strategies and the implementation of automated pipelines such as AIDAqc, which enabled rigorous

quantification, reproducible analyses, and flexible exploration across these complex datasets.

## Limitations

While DataLad and GIN provide a powerful foundation, their adoption requires a significant time investment and institutional support. Our experience confirms that effective use of DataLad is a discipline in itself, requiring structured training to fully grasp version control workflows, which are not intuitive at first encounter, especially for non-IT researchers. Additionally, there appears to be a broader cultural limitation in academia, where version control is often perceived as a "soft skill" that does not require formal training, even among researchers with programming backgrounds. Another limitation is GIN's handling of subdataset DOIs, which contradicts the modular logic of the YODA design. Furthermore, if the data at hand is sensitive and researchers do not wish to share it on external servers, the proposed solution has limitations. However, implementing a local GIN server could resolve such data privacy concerns.

Regarding AIDAqc, some limitations still exist. Changes in the Bruker ParaVision software complicate automated QC for the raw data due to evolving file structures, which require regular updates in AIDAqc to maintain compatibility with raw Bruker data. Future integration of trained large language models could enhance adaptability to such software updates by interpreting and parsing new file structures dynamically. Another limitation is related to image orientation, which remains a widespread issue in preclinical neuroimaging. Currently, no automatic pipeline, including AIDAqc, reliably detects incorrect orientations based on cohort-wide data comparisons. In practice, orientation errors introduced during acquisition require manual flipping and rotation adjustments to match standard templates, which is tedious, time-consuming, and prone to human error, especially in multidimensional modalities that exceed three dimensions, such as DTI and fMRI. Importantly, incorrect orientation affects all derived QC metrics, including SNR and motion calculations, leading to misleading QC results and potential misclassification during automated assessment. Although flagged data still prompts user inspection, full automation is hindered by the lack of integrated orientation verification and correction. Therefore, implementing automatic detection and correction of faulty orientations is highly relevant for AIDAqc to ensure robust, scalable, and fully automated data quality assessment in small animal MRI studies.

Aligning the timeline of stroke recovery in mice versus humans appears to be poorly researched. There seems to be a nonlinear relationship between them, making it difficult to translate the timespans of the acute, subacute, and chronic phases from mice to humans and vice versa in the context of pathophysiological processes. This is an area that requires further investigation.

## Conclusion & Future Prospects

Together, these studies illustrate a seamless continuum from robust data management to automated quality control and rapid translational application. This modular and open approach has accelerated our publication timelines, strengthened collaborative networks, and laid a foundation for reproducible, high-quality preclinical and translational neuroimaging research.

Building on these interconnected studies, our future work will focus on enhancing adaptability to evolving MRI software through machine-learning-driven quality control, including automated orientation correction, and embedding AIDAqc within scanner workflows to enable real-time feedback and immediate scan assessment. Additionally, it is planned to adapt AIDAqc for application to MRI phantom imaging, broadening its utility to include quality control of phantom data, which is essential for scanner calibration, protocol validation, and long-term stability assessments.

Community documentation will be expanded through interactive guides, tutorials, and SOP libraries to promote broader adoption. However, onboarding remains a challenge for researchers without IT training. Therefore, detailed YouTube tutorials and SOPs will be created to lower barriers and support collaborative work, even in decentralized or remote settings. Continued refinement and community integration of these pipelines will further bridge the gap between basic research and clinical application, ultimately strengthening the translational potential of preclinical neuroimaging and preclinical research in general.

# Bibliography


[1] Neil Vargesson. "Thalidomide-induced teratogenesis: History and mechanisms". In: *Birth Defects Research Part C: Embryo Today: Reviews* 105.2 (2015), pp. 140–156.

[2] William Moy Stratton Russell, Rex Leonard Burch, Charles Westley Hume, et al. *The principles of humane experimental technique*. Vol. 238. Methuen London, 1959.

[3] Mark D Wilkinson et al. "The FAIR Guiding Principles for scientific data management and stewardship". In: *Scientific data* 3.1 (2016), pp. 1–9.

[4] Christopher J Markiewicz et al. "The OpenNeuro resource for sharing of neuroscience data". In: *Elife* 10 (2021), e71774.

[5] Timothy H Vines et al. "The availability of research data declines rapidly with article age". In: *Current biology* 24.1 (2014), pp. 94–97.

[6] Mirko Gabelica, Ružica Bojčić, and Livia Puljak. "Many researchers were not compliant with their published data sharing statement: a mixed-methods study". In: *Journal of Clinical Epidemiology* 150 (2022), pp. 33–41.

[7] Kathrin Giehl et al. "Sharing brain imaging data in the Open Science era: how and why?" In: *The Lancet Digital Health* 6.7 (2024), e526–e535.

[8] Vladimir Petrik et al. "Godfrey Hounsfield and the dawn of computed tomography". In: *Neurosurgery* 58.4 (2006), pp. 780–787.

[9] Julia G Mannheim et al. "Standardization of small animal imaging—current status and future prospects". In: *Molecular imaging and biology* 20 (2018), pp. 716–731.

[10] Dustin R Osborne et al. "Guidance for efficient small animal imaging quality control". In: *Molecular imaging and biology* 19 (2017), pp. 485–498.

[11] Brendan Williams et al. “Inter-rater reliability of functional MRI data quality control assessments: A standardised protocol and practical guide using pyfMRIqc”. In: *Frontiers in Neuroscience* 17 (2023), p. 1070413.

[12] Mira Katan and Andreas Luft. “Global burden of stroke”. In: *Seminars in neurology*. Vol. 38. 02. Thieme Medical Publishers. 2018, pp. 208–211. DOI: 10.1055/s-0038-1649503.

[13] European Stroke Organization. *ESO Action Plan*. https://actionplan.eso-stroke.org/. Accessed: March 18, 2025. 2022.

[14] E Zea Longa et al. “Reversible middle cerebral artery occlusion without craniectomy in rats.” In: *stroke* 20.1 (1989), pp. 84–91.

[15] Vivien Labat-Gest and Simone Tomasi. “Photothrombotic ischemia: a minimally invasive and reproducible photochemical cortical lesion model for mouse stroke studies”. In: *Journal of visualized experiments: JoVE* 76 (2013), p. 50370.

[16] Valentine Bouët et al. “Sensorimotor and cognitive deficits after transient middle cerebral artery occlusion in the mouse”. In: *Experimental neurology* 203.2 (2007), pp. 555–567.

[17] Shyam Prabhakaran et al. “Inter-individual variability in the capacity for motor recovery after ischemic stroke”. In: *Neurorehabilitation and neural repair* 22.1 (2008), pp. 64–71.

[18] OpenAI. *ChatGPT: A large language model*. Version GPT-o4-mini-high. Last accessed 2025-07-14. 2025. URL: https://chat.openai.com/.

[19] Universität zu Köln. *FAQ zur Nutzung von ChatGPT und ähnlichen KI-Tools*. Last accessed 2025-07-14. URL: https://verwaltung.uni-koeln.de/stabsstelle02.1/content/faq/data/chatgpt/index_ger.html?utm_source=chatgpt.com.

[20] Niklas Pallast et al. “Cloud-based relational database for multimodal animal data”. In: *Database* 2018 (2018), bay124.

[21] Niklas Pallast et al. “Processing pipeline for atlas-based imaging data analysis of structural and functional mouse brain MRI (AIDAmri)”. In: *Frontiers in neuroinformatics* 13 (2019), p. 42.

[22] Oscar Esteban et al. “MRIQC: Advancing the automatic prediction of image quality in MRI from unseen sites”. In: *PloS one* 12.9 (2017), e0184661.

# Appendix: Supplements

This appendix compiles the following supplementary materials:

- **Supplementary Material I** – How to establish and maintain a multimodal animal research dataset using DataLad
- **Supplementary Material II** – Automated quality control of small animal MR neuroimaging data
- **Supplementary Material III** – Proportional recovery in mice with cortical stroke

# Supplementary Material

## Video guide

DataLad-Cloning.mp4
DataLad-Creating.mp4

## Software Installation

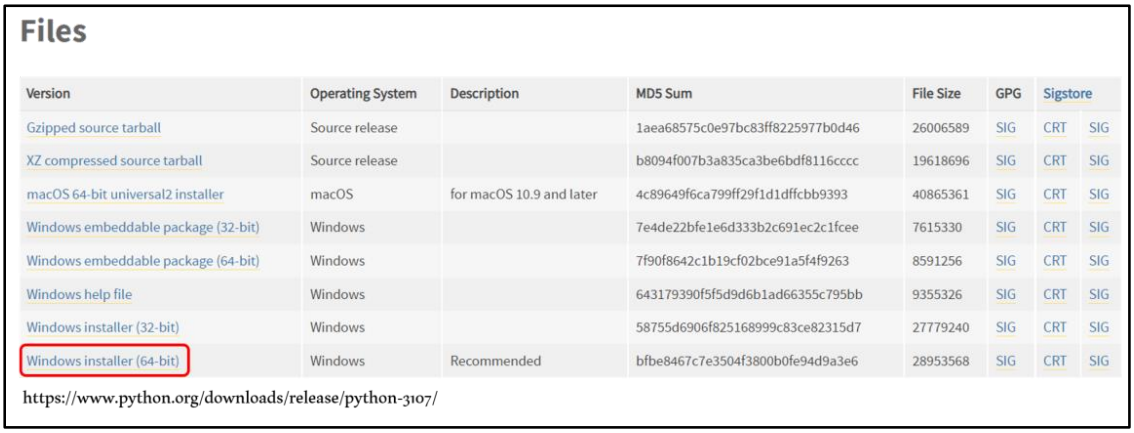
**CRITICAL**

### OS dependencies

All operating systems were built with different purposes and therefore all of them have their special advantages and caveats compared to each other. Here we summarize installation routines for the most popular operating systems, i.e., Windows, Mac, and Linux. A note of caution: pay attention to every detail of the workflow, especially if you want to apply this work to a repository of your own. For the most up-to-date installation instructions, see the related websites. Get familiar with the command line/terminal before starting the installation process (https://ubuntu.com/tutorials/command-line-for-beginners#1-overview. Command line entries are highlighted by `>`.

## WINDOWS

- Python 3.6.x (https://www.python.org/downloads/release/python-3107)



**Files**

| Version | Operating System | Description | MD5 Sum | File Size | GPG | Sigstore | |
|---|---|---|---|---|---|---|---|
| Gzipped source tarball | Source release | | 1aea68575c0e97bc83ff8225977b0d46 | 26006589 | SIG | CRT | SIG |
| XZ compressed source tarball | Source release | | b8094f007b3a835ca3be6bdf8116cccc | 19618696 | SIG | CRT | SIG |
| macOS 64-bit universal2 installer | macOS | for macOS 10.9 and later | 4c89649f6ca799ff29f1d1dffcbb9393 | 40865361 | SIG | CRT | SIG |
| Windows embeddable package (32-bit) | Windows | | 7e4de22bfe1e6d333b2c691ec2c1fcee | 7615330 | SIG | CRT | SIG |
| Windows embeddable package (64-bit) | Windows | | 7f90f8642c1b19cf02bce91a5f4f9263 | 8591256 | SIG | CRT | SIG |
| Windows help file | Windows | | 643179390f5f5d9d6b1ad66355c795bb | 9355326 | SIG | CRT | SIG |
| Windows installer (32-bit) | Windows | | 58755d6906f825168999c83ce82315d7 | 27779240 | SIG | CRT | SIG |
| Windows installer (64-bit) | Windows | Recommended | bfbe8467c7e3504f3800b0fe94d9a3e6 | 28953568 | SIG | CRT | SIG |

https://www.python.org/downloads/release/python-3107/

I. Download the python installer according to figure 2

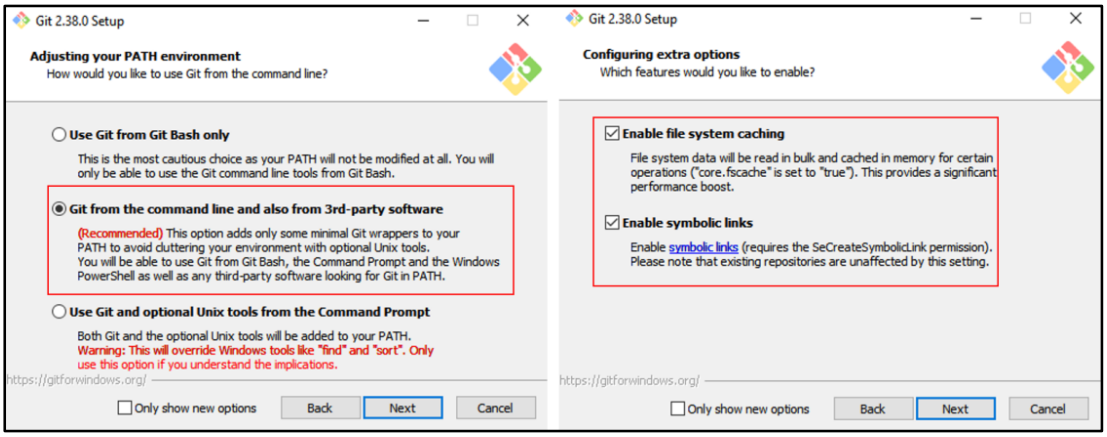
Figure 2: Windows installer (64-bit) or (32-bit)

1

II. When running the installer, make sure to select *Add Python to PATH* option, as this is required for subsequent installation steps and interactive use later on. Other than that, using the default installation settings is just fine.

III. Verify installation by opening the terminal and running the command `python`, the output should look like this: `Python 3.9.7 (default, Sep 16 2021, 16:59:28) [MSC v.1916 64 bit (AMD64)] :: Anaconda, Inc. on win32 Type "help", "copyright", "credits" or "license" for more information.`

- Git (https://git-scm.com/download/win)
  I. Download the Git installer for (64-bit) or (32-bit) windows OS.




II. When installing Git configure the specific settings shown in figure 3

Figure 3: 1) Select Git from the command line and also from 3rd-party software 2) Enable file system caching 3) Enable symbolic links

- Git-annex (Using datalad-installer)
  I. The most convenient way to deploy git-annex is via the DataLad installer. Once Python is available, it can be installed with the Python package manager pip. Open a command prompt and run:
  `> pip install datalad-installer`
  II. Open another command prompt in administrator mode and run:
  `> datalad-installer git-annex -m datalad/packages`
  This will download a recent git-annex, and configure it for your Git installation. The admin command prompt can be closed afterward, all other steps do not need it. For performance improvements, we recommend also setting the following git-annex configuration:
  `> git config --global filter.annex.process "git-annex filter-process"`

- Datalad
  With Python, Git, and git-annex installed, DataLad can be installed, and later also upgraded using pip by running:
  `> pip install datalad`

- 7-Zip (https://7-zip.org/)
  Download it from the 7-zip website (64bit installer when in doubt), and install it into the default target folder.
- nano Text editor (https://files.lhmouse.com/nano-win/)
  I. Download the latest version of the program (.7z file)
  II. Extract the OS-specific version
  III. Copy the `nano.exe` file from `~\pkg_x86_64-w64-mingw32\bin\nano.exe` to `C:\Windows\.`
  IV. Verify installation by running the command `nano` in a new terminal.
  Tip: The caret sign "^" means the control key on the keyboard.

## Mac

On Mac, it is recommended to install a separate version of Python and Datalad in an Anaconda environment. This prevents conflicts with the existing Python version on Mac OS. Make sure you choose the correct installation version, especially if you are using the new Apple M1 chip.

1. Install XCode from the Apple App Store (https://apps.apple.com/us/app/xcode/id497799835)
2. Install the package manager Homebrew (https://brew.sh)
3. Download and install Anaconda (https://docs.anaconda.com/anaconda/install/mac-os/)
   In the terminal: create a new Conda environment for this workflow
   ```
   >conda create -n workflow python=3.10
   ```
   (-n specifies the name of the workflow)
   During the process, Conda will automatically determine the necessary packages to be installed. If asked to proceed, type: `y` (for yes). The installation should finalize with instructions on how to activate/deactivate the environment.
   ```
   >Preparing transaction: done
   >Verifying transaction: done
   >Executing transaction: done
   >#
   ># To activate this environment, use
   >#
   >#     $ conda activate workflow
   >#
   ># To deactivate an active environment, use
   >#
   >#     $ conda deactivate
   ```
   Activate the environment:
   ```
   >conda create -n workflow python=3.10
   ```

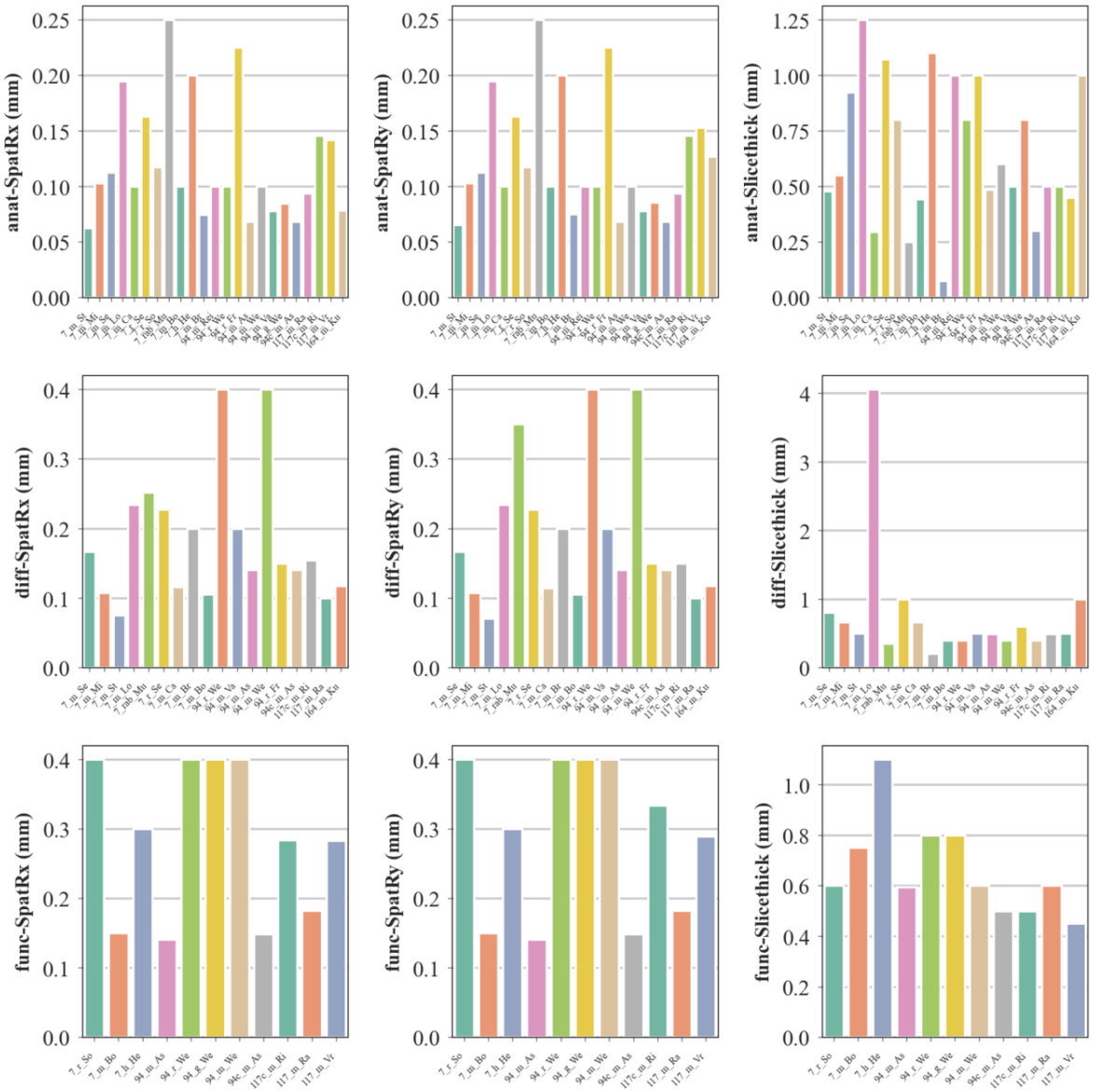
4. Register to GIN and install the command line tool (https://gin.g-node.org/G-Node/Info/wiki/GIN+CLI+Setup#macos)
   ```
   >brew tap g-node/pkg
   >brew install g-node/pkg/gin-cli
   ```

5. Install DataLad
   ```
   >brew install datalad
   ```

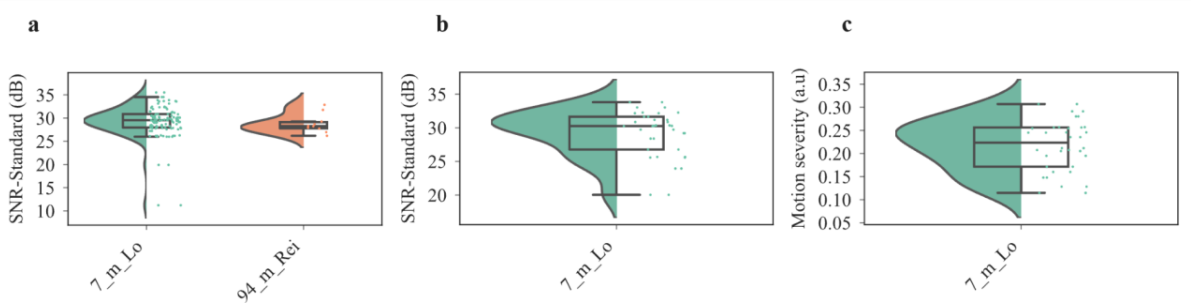
## Linux: (Neuro) Debian, Ubuntu, and similar systems

1. Install DataLad and all dependencies (including git-annex and p7zip)
   ```
   >sudo apt-get install datalad
   ```

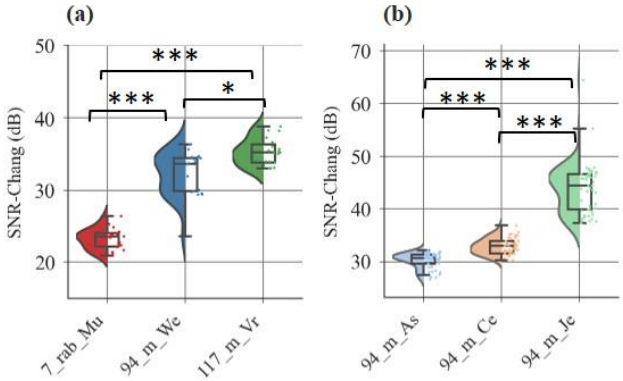
# SUPPLEMENTARY FIGURES AND TABLES



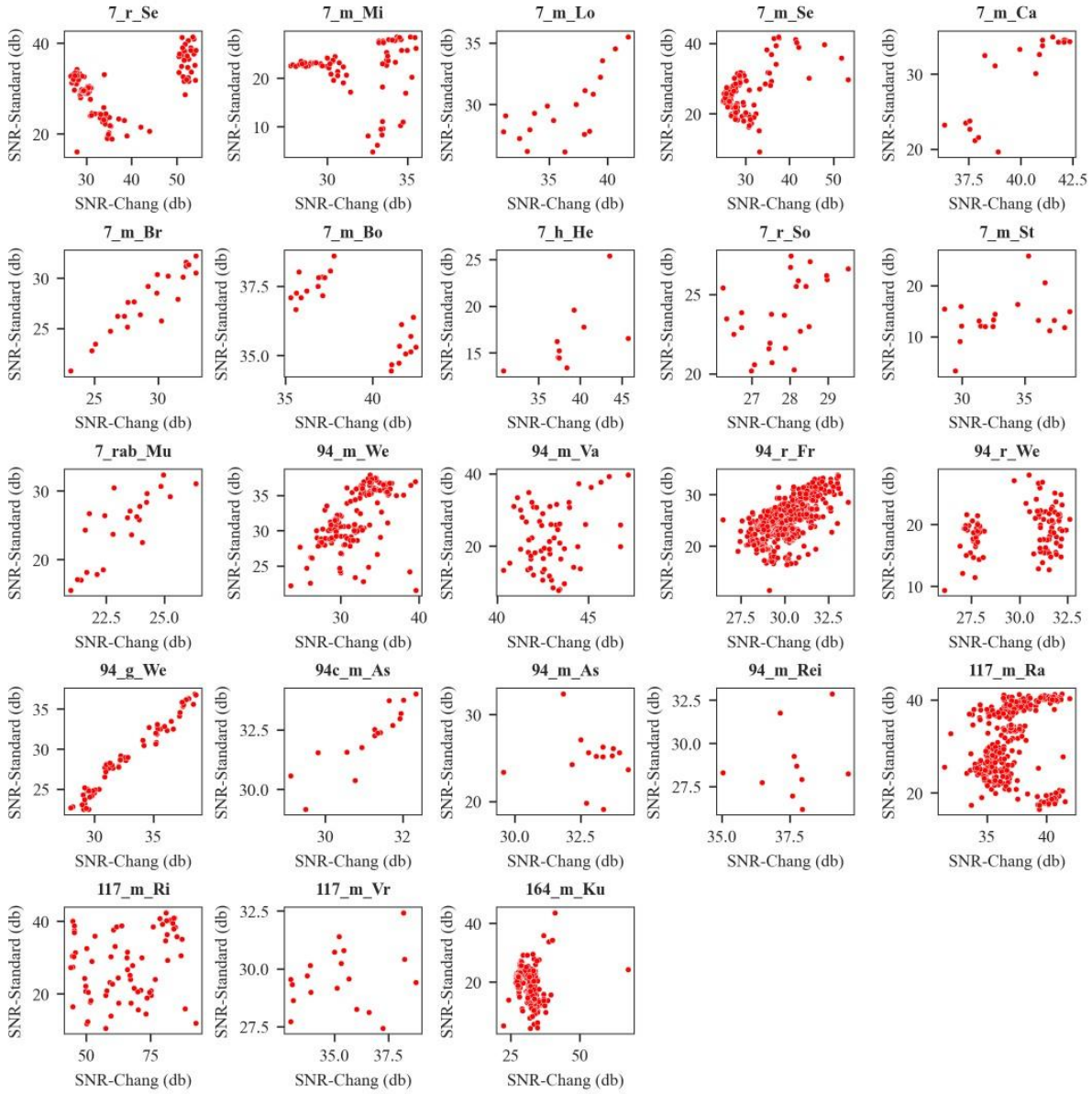

**Supplementary Figure 1: Summary of spatial resolution for all studies.** Spatial resolution in x- and y-direction (SpatRx, SpatRy) and slice thickness for all studies sorted into anatomical (anat), diffusion (diff), and functional (func) scans.




**Supplementary Figure 2: Summary of quality features for anatomical scans.** (a) SNR Standard for anatomical scans. (b) SNR Standard of diffusion scans. (c) Motion severity values for diffusion scans. This plot is the extension of Fig. 5 in the main manuscript for the two abdominal datasets 7_m_Lo and 94_m_Rei.




**Supplementary Figure 3: Comparison of the field strength and voxel size effect on SNR.** Significant differences were observed for SNR values (using the Chang method) between selected datasets with (a) different field strengths and (b) different voxel sizes. Voxel sizes: 7_rab_Mu (0.25x0.25x0.25 mm), 94_m_As (0.068x0.068x0.3 mm), 94_m_Ce (0.078x0.078x0.5), 94_m_Je (0.15x0.15x0.5), 94_m_We (0.1x0.1x0.4 mm), 117_m_Vr (0.142x0.153x0.45 mm). To determine significance, a Kruskal-Wallis test was initially conducted. Subsequently, a post hoc Dunn's multiple comparison test using the Holm–Bonferroni correction for p-value calculation was used. Significant differences between groups are shown as p<0.05 (*), p<0.01 (**), p<0.001 (***).




**Supplementary Figure 4: Correlation of SNR-Standard and SNR-Chang for all anatomical datasets.**

## Role of the rating threshold for the agreement between manual raters and AIDAqc

In the following analysis, the predicted truth by manual raters was compared against the ratings generated by AIDAqc. Datasets with no true positive and/or false negative (TP+FN>0) files were excluded to reduce the bias on the classification statistics, i.e., mitigate the impact of datasets where manual raters did not classify any files as “bad quality” (Table I). Further, different thresholds for the number of manual raters and AIDAqc outlier algorithms were compared.

The following metrics were used for a quantitative comparison:

1. Sensitivity: Out of all the actual “bad quality” datasets, how many were identified by AIDAqc as “bad quality” data?

   $Sensitivity = \frac{TP}{TP+FN}$

2. Specificity: Out of all the “no bad quality” datasets, how many were identified by AIDAqc as such?

   $$Specificity = \frac{TN}{TN + FP}$$

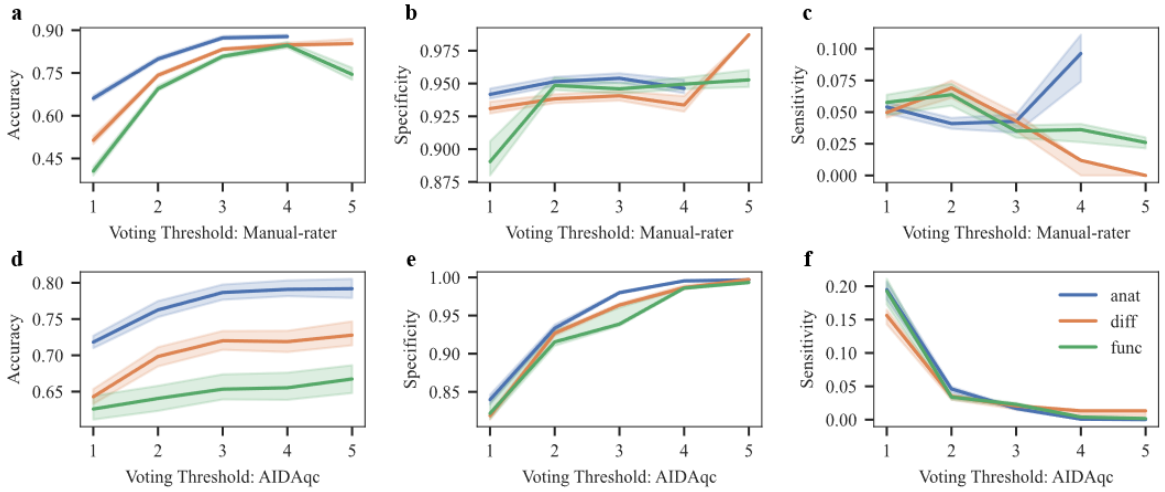
3. Accuracy: Out of all the classifications, what was the rate of correct classifications?

   $Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$

In the first comparison, the threshold for the manual voting was varied and the classification was statistically compared against all combinations of AIDAqc thresholds. The overall accuracy (rate of correct classifications) was relatively high across all manual rater voting thresholds for anatomical (0.77 ± 0.21), diffusion (0.70 ± 0.25), and functional (0.65 ± 0.27) scans (Supplementary Fig. 5a). The accuracy reached a maximum for anatomical and functional scans with the manual rater voting threshold set to four and for diffusion scans with the threshold set to five. For specificity, i.e., “how many true negatives” were identified by AIDAqc, the optimal manual votings threshold was four for anatomical (0.95 ± 0.09) and functional (0.93 ± 0.15) scans, respectively, and five for diffusion scans (0.94 ± 0.08) (Supplementary Fig. 5b). Sensitivity, a measure of how many true positive files were identified by AIDAqc, reached a peak with the manual rating

threshold of two for diffusion (0.05 ± 0.13) and functional (0.05 ± 0.14) scans, respectively, and for anatomical sequences with a threshold set to four (0.77 ± 0.21) (Supplementary Fig. 5c).

In the second comparison, the threshold for the AIDAqc voting was varied and the classification was statistically compared against all combinations of manual thresholds. Here, a consistent trend was observed: accuracy and specificity increased with the number of AIDAqc votes, whereas sensitivity reached the maximum already with the AIDAqc voting threshold set to one (Supplementary Fig. 5d-e). Average accuracy overall values for anatomical, diffusion, and functional scans were (0.77±0.21), (0.70±0.25), and (0.65±0.27) respectively. For specificity, these values were (0.95±0.09), (0.94±0.08) and (0.93±0.15). The highest sensitivity values for AIDAqc were reached with a threshold set to one for all sequence types. The average and std across all values for anatomical, diffusion, and functional were (0.05 ± 0.14), (0.05 ± 0.13), and (0.05 ± 0.14) respectively.




**Supplementary Figure 5: Statistical analysis of the classification.** Panels (a-c) showcase the dynamic interplay among accuracy, specificity, and sensitivity across varying manual rater voting thresholds. Panels (d-f) offer a parallel examination using AIDAqc voting thresholds. Each color represents a distinct sequence (anat - anatomical, diff - diffusion, and func - functional). The 30% confidence interval encapsulates the statistical uncertainty inherent in the plotted metrics. It is noteworthy that the condition (TP+TN)>0 was applied before extracting these values; thus, some datasets ended up with no bad quality data when the manual-rater threshold was set to 5, as evident in the anatomical sequence in (a, c).

## Application of the outlier detection

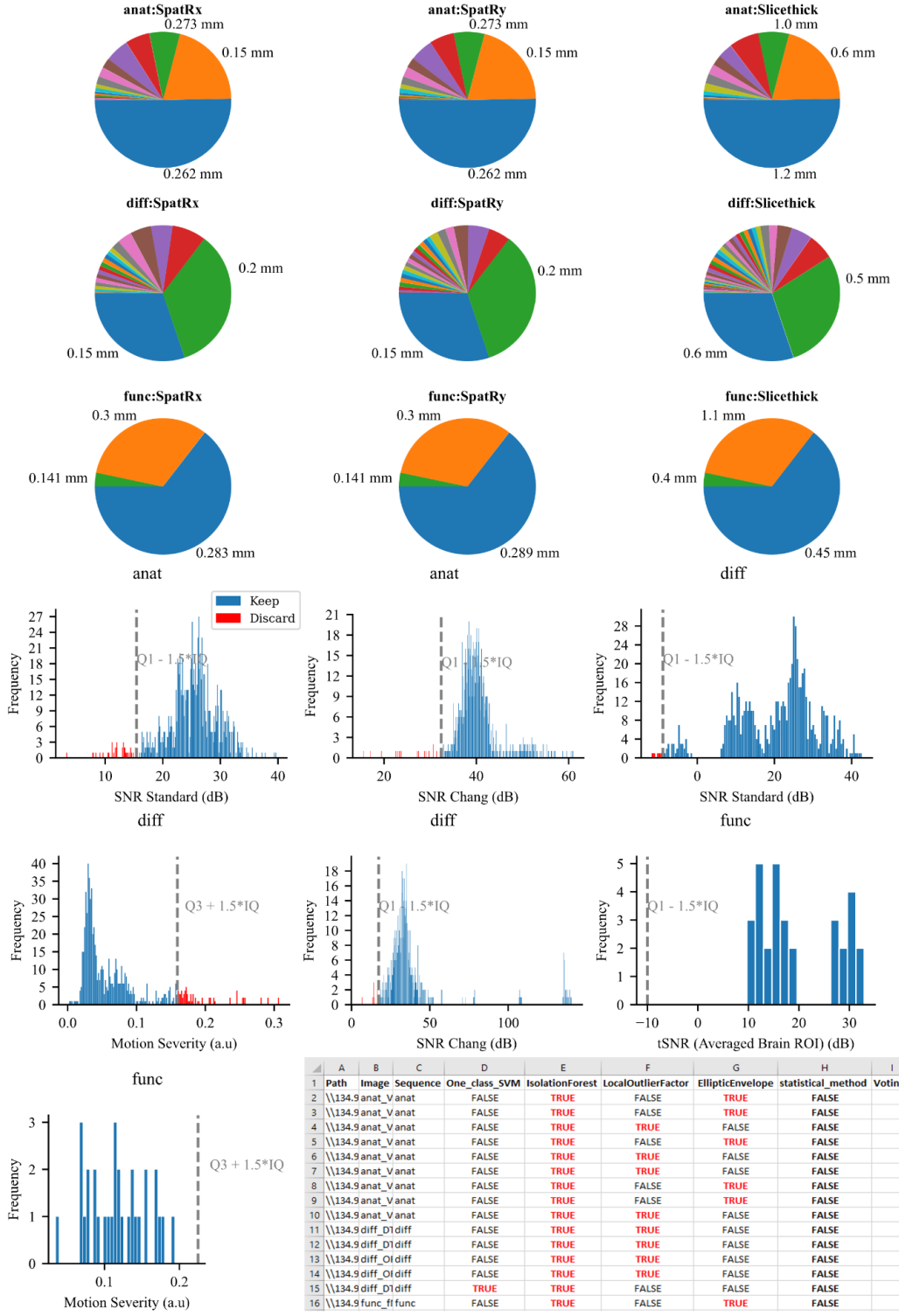
AIDAqc does not directly measure quantitative post-processing values. However, it is also important to note that the quality assessment based on factors like SNR or motion can indirectly affect processes such as brain extraction as shown in Supplementary Fig. 6, for the case with majority vote = 3. Problems in the subsequent processing steps can consequently lead to less precise quantitative values (Fig. 9).

**https://gin.g-node.org/Aswendt_Lab/2023_Kalantari_AIDAqc/src/master/output/SupplementFigures6.pdf**

**Supplementary Figure 6: Original DWI images and corresponding FA map pairs of all subjects from the dataset 94c_m_As and 94_m_As.** The majority vote for the image to be an outlier is written above each DWI image.

**https://gin.g-node.org/Aswendt_Lab/2023_Kalantari_AIDAqc/src/master/output/SupplementFigures7.pdf**

**Supplementary Figure 7: Overview of all subjects and extension of Fig. 10 in the manuscript.** For each subject, 4 different slices, and the registered atlas of those 4 different slices are shown. For the subject with the Majority Vote of 5, it can be seen that the brain extraction has completely failed along with the bias filed correction, leading to a faulty atlas registration.




| | A | B | C | D | E | F | G | H | I |
|---|---|---|---|---|---|---|---|---|---|
| 1 | Path | Image | Sequence | One_class_SVM | IsolationForest | LocalOutlierFactor | EllipticEnvelope | statistical_method | Votings |
| 2 | \\134.9 | anat_V | anat | FALSE | TRUE | FALSE | TRUE | FALSE | 2 |
| 3 | \\134.9 | anat_V | anat | FALSE | TRUE | FALSE | TRUE | FALSE | 2 |
| 4 | \\134.9 | anat_V | anat | FALSE | TRUE | TRUE | FALSE | FALSE | 2 |
| 5 | \\134.9 | anat_V | anat | FALSE | TRUE | FALSE | TRUE | FALSE | 2 |
| 6 | \\134.9 | anat_V | anat | FALSE | TRUE | TRUE | FALSE | FALSE | 2 |
| 7 | \\134.9 | anat_V | anat | FALSE | TRUE | TRUE | FALSE | FALSE | 2 |
| 8 | \\134.9 | anat_V | anat | FALSE | TRUE | FALSE | TRUE | FALSE | 2 |
| 9 | \\134.9 | anat_V | anat | FALSE | TRUE | FALSE | TRUE | FALSE | 2 |
| 10 | \\134.9 | anat_V | anat | FALSE | TRUE | TRUE | FALSE | FALSE | 2 |
| 11 | \\134.9 | diff_D1 | diff | FALSE | TRUE | TRUE | FALSE | FALSE | 2 |
| 12 | \\134.9 | diff_D1 | diff | FALSE | TRUE | TRUE | FALSE | FALSE | 2 |
| 13 | \\134.9 | diff_OI | diff | FALSE | TRUE | TRUE | FALSE | FALSE | 2 |
| 14 | \\134.9 | diff_OI | diff | FALSE | TRUE | TRUE | FALSE | FALSE | 2 |
| 15 | \\134.9 | diff_D1 | diff | TRUE | TRUE | FALSE | FALSE | FALSE | 2 |
| 16 | \\134.9 | func_ff | func | FALSE | TRUE | FALSE | TRUE | FALSE | 2 |

**Supplementary Figure 8: AIDAqc output statistics for all NIfTI datasets** (Table I and Supplementary Table I)**.** Pie diagrams show spatial resolution (in mm) for x, y, z direction. Bar plots show distribution of quality metrics. The table shows outliers not considered a statistical outlier but detected by the other algorithms.


BrkRaw GUI - v0.3.8-rc1
Open File
Open Directory
DataPath
20220726_092255_A260722_A260722_1_1
Refresh
Account nmr
Subject ID A260722
DOB 26 Jul 2022
Type Quadruped
Scan Date 2022-07-26, 09:22:55
Session ID A260722
Sex unknown
Position Supine
Researcher A260722
Study ID 1
Weight 0.001 kg
Entry HeadFirst
Scan ID / Protocol
Slice Axis:: x y z
Selected Scan Info
002::T2_TurboRARE
003::T2_TurboRARE
004::T2_TurboRARE
005::T2_weighted_short_Sth_06
006::MPRAGE_msP
007::LTS_Multidelta2B_Ageing
008::LTS_Multidelta2B_Ageing
009::LTS_Charmed_2B_Ageing
010::B0Map-ADJ_B0MAP
011::1_Localizer
012::LTS_Multidelta2B_Ageing
013::LTS_Multidelta2B_Ageing
014::LTS_Multidelta2B_Ageing
015::LTS_Multidelta2B_Ageing
a
b
c
d
Sequence:
- Bruker:RARE
Protocol:
- T2_TurboRARE
Scan Name:
- T2_TurboRARE (E2)
RepetitionTime:
- 2500 msec
EchoTime:
- 33 msec
FlipAngle:
- 90 degree
PixelBandwidth:
- 140.850 Hz
Dimension:
- 2D
Matrix size:
- 256 x 256 x 23
Number of SlicePacks:
- 1
FOV size:
- 35 x 35 (mm)
Spatial resolution:
- 0.137 x 0.137 x 1.100 (mm)
Scan ID / Protocol
Slice Axis:: x y z
Selected Scan Info
002::2a-RARE_8_coronal
003::2b-RARE_8_transverse
004::2c-RARE_8_sagittal
005::2c-RARE_8_sagittal
006::3-MSMEVTR-15122010
007::4-DWI_2D_15122010-150ms
e
f
Sequence:
- RARE (pvm)
Protocol:
- 2a-RARE_8_coronal
Scan Name:
- 2a-RARE_8_coronal
RepetitionTime:
- 1000 msec
EchoTime:
- 23.334 msec
FlipAngle:
- 180 degree
PixelBandwidth:
- 195.312 Hz
Dimension:
- 2D
Matrix size:
- 256 x 256 x 8
Number of SlicePacks:
- 1
FOV size:
- 20 x 30 (mm)
Spatial resolution:
Scan ID / Protocol
Slice Axis:: x y z
Selected Scan Info
001::01_Localizer
002::02_Localizer_multi_slice
003::03_T2_FLASH_axial
004::04_T2_FLASH_coronal
005::06_fmri_single
006::B0Map-ADJ_B0MAP
007::07_fmri_multi_trigger
008::08_DTI_EPI_36dir
009::09_MRS_water_pfc
010::09_MRS_water_pfc
012::09_MRS_water_pfc
013::09_MRS_water_pfc
014::09_MRS_water_pfc
015::10_MRS_meta_pfc
016::11_MRS_water_thal
017::11_MRS_water_thal
018::20_MRS_meta_thal
019::09_MRS_water_pfc
020::09_MRS_water_pfc
021::20_MRS_meta_thal
g
h
Sequence:
- Bruker:FLASH
Protocol:
- 03_T2_FLASH_axial
Scan Name:
- 03_T2_FLASH_axial (E3)
RepetitionTime:
- 200 msec
EchoTime:
- 3.1 msec
FlipAngle:
- 25 degree
PixelBandwidth:
- 822.368 Hz
Dimension:
- 2D
Matrix size:
- 120 x 100 x 20
Number of SlicePacks:
- 1
FOV size:
- 17 x 15 (mm)
Spatial resolution:
- 0.142 x 0.150 x 0.500 (mm)
Temporal resolution:
- 45600.000 (msec)
Reco ID / DataType
001::MAGNITUDE_IMAGE
Slice
10

**Supplementary Figure 9: Orientatiation in datasets.** Three exemplary subjects from three different datasets from Table I. Multiple TurboRARE sequences with the same sequence, protocol, and scan names (d) but scanned with different orientations (a), (b), wherein the second subject, this was mentioned in the protocol and scan name (f). Subject alignment is supine in (c) whereas in (e) it is set as supine but right to left. In the third subject, both subject alignment and orientation were done in a standard norm (g,h). AIDAqc's current limitation lies in comparing different orientations such as (a), (b), and (c), which may raise questions. It might not lead to outliers if the cohort of subjects has the same sequences, but it still may be questionable and would require further automatic adjustment in terms of how the algorithm extracts the SNR, motion, and ghosting.

## Supplementary Table I: Summary of datasets used for developing AIDAqc

| | Repository | Datasets | Species | Scanner (Vendor, field strength/bore size, electronics, gradient, Tx/Rx coils) | Modalities | Sequences | Data format | Sample size (n) |
|---|---|---|---|---|---|---|---|---|
| 1 | **Aswendt** | **94_m_As** | Mouse brain | Bruker Biospec 94/20 with AVANCE II electronics. Gradient: B-GA12SHP (660 mT/m and slew rates up to 4570 T/m/s). TX/RX: 1H quadrature cryogenic surface coil or 1H mouse brain quadrature surface. | T2w, rs-fMRI, DTI | TurboRARE, GE-EPI, SE-EPI | Bruker raw | 32 |
| 2 | **Aswendt** | **94c_m_As** | Mouse brain | Bruker Biospec 94/20 with AVANCE II electronics. Gradient: B-GA12SHP (660 mT/m and slew rates up to 4570 T/m/s). TX/RX: 1H quadrature cryogenic surface coil or 1H mouse brain quadrature surface. | T2w, rs-fMRI, DTI | TurboRARE, EPI_DTI_BIS,GRE_FMRI | Bruker raw | 32 |
| 3 | **Boehm-Sturm** | **7_m_Bo** | Mouse brain | Bruker Biospec 70/20USR with AVANCE III HD electronics. Gradient: BGA-12S HP (440 mt/m and slew rates up to 3440 T/m/s). TX/RX: 1H quadrature cryogenic surface coil | T2w, DTI | TurboRARE, DTI_EPI_SEG, T2STAR | Bruker raw | 27 |
| 4 | **Carnevale** | **7_m_Ca** | Mouse brain | Bruker Pharmascan 70/16US mm with AVANCE III HD electronics. Gradient: BGA9SHP, Tx: 1H RF circularly polarized 72 mm volume coil. RX: 1H mouse brain quadrature surface. | 3D-T2/DTI | TurboRARE, DTI_EPI | Bruker raw | 20 |
| 5 | **Franx** | **94_r_Fr** | Rat brain | Varian/Agilent 94/210 mm bore. Gradient: 205_120_HD gradient (400 mT/m), Tx: 80 mm Helmholtz coil, Rx:15 mm inductively coupled surface coil receive | T2w, DWI, T1w | multishot EPI spin-echo, 2D gradient echo | Nifti | 722 |
| 6 | **Hekmatyar** | **7_h_He** | Hamster brain | Bruker Biospec 70/20. Gradient: NA. 2x2 1H Rat brain surface array coil 400 mT/mm 12 cm inner diameter) | T2w, rs-fMRI | 2D Single shot spin echo EPI (SE-EPI) | Nifti | 10 |
| 7 | **Kurniawan** | **164_m_Ku** | Mouse brain | Bruker Biospec 164/89 vertical. Gradient: micro 2.5 gradient max 1.5T/m using Great 60 gradient amplifier. Avance II electronics. Tx/Rx: SAW 20 mm mouse head coil. | T2w, DTI | RARE, DWI | Bruker raw | 168 |
| 8 | **Micotti** | **7_m_Mi** | Mouse brain | Bruker Biospec 70/30 with AVANCE III electronics. Gradient: BGA12SL gradient (440 mT/m and slew rates up to 3440 T/m/s). Tx: 1H RF circularly polarized 72 mm volume coil. RX: 1H mouse brain quadrature surface. | T2w, DTI | RARE_POSIZ_AXIAL, DTI_RETE | Bruker raw | 21 |
| 9 | **Muñoz-Moreno** | **7_rab_Mu** | Rabbit brain[1] | Bruker Biospec 70/30 with AVANCE III electronics. Gradient: BGA12S (400 mT/m, 12 cm inner diameter) . Tx: NA. Rx: NA. | T1w, DWI | 7T T1w (MDEFT),DWI | Nifti | 24 |
| 10 | **Brinton** | **7_m_Br** | Mouse brain | Bruker Biospec 70/20 and Bruker 300 1H 2x2 mouse brain surface array or 300 1H 2x2 rat brain surface array | T2w, DWI | TurboRARE, 3D DTI_EPI | Nifti | 20 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **11** | **Ramos-Cabrer** | **117_m_Ra** | Mouse brain | Bruker Biospec USR 117/16 with AVANCe III electronics. Gradient: BGA-9S (750 mT/m strength, and 6600 T/m/s slew rate). Tx: Bruker's 500 1H RF circularly polarized 72 mm volume coil, RX: Bruker's RF 500 1H – Mouse Brain SUC surface coil OR Neoos Biotech's mouse brain surface coil. | T2w, DTI, rs-fMRI | TurboRARE, FLASH, EPI | Bruker raw | 69 |
| **12** | **Reichardt** | **94_m_Rei** | Mouse abdominal | Bruker Biospec 94/20 mm with Avance3 electronics. Gradient: B-GA 12S., 680 mT/m, Tx/RX .1H mouse quadrature volume coil (38mm ID) | T2w | T2_TRARE | Bruker raw | 10 |
| **13** | **Rivera-Olvera** | **117_m_Ri** | Mouse brain | Bruker Biospec USR 117/16, B-GA9S (600 mT/m and slew rates up to 6000 T/m/s) ; TX/RX: 1H quadrature cryogenic surface coil. | T2w, DTI, rs-fMRI | DTIEPI, FLASH, EPI | Bruker raw | 20 |
| **14** | **Sta Maria** | **7_m_St** | Mouse brain | 7T PET-MR system (MR Solutions Ltd, Guildford, UK), bore size: ~ 24 cm, gradient strength: 600 mT/m, coil: 20 mm inner diameter, 18 mm length, quadrature birdcage | T2w, DTI | FSE, FLASH, SE-EPI | Nifti | 17 |
| **15** | **Selim** | **7_m_Se** | Mouse/rat brain | Bruker Biospec 70/30, Bruker 300 1H 2x2 mouse brain surface array or 300 1H 2x2 rat brain surface array | T2w/T1w/DTI | TurboRARE, WEIGHTED_SHORT, SDS | Bruker raw | 20/30 |
| **16** | **Selim** | **7_r_Se** | Mouse/rat brain | Bruker Biospec 70/30, Bruker 300 1H 2x2 mouse brain surface array or 300 1H 2x2 rat brain surface array | T2w/T1w/DTI | TurboRARE, LTS, T2_WEIGHTED_SHORT | Bruker raw | 20/30 |
| **17** | **Soria** | **7_r_So** | Rat brain | 7T Bruker Biospec 70/30 horizontal animal scanner (400 mT/m, 12 cm inner diameter), 1H 2x2 rat brain surface array | T2w, rs-fMRI | 7T T2w (RARE),rs-fMRI | Nifti | 8 (multiple time points) |
| **18** | **Van Leeuwen** | **94_m_Va** | In vivo mouse brain | 9.4 T Varian/Agilent, quadrature birdcage with ID 35 mm, 115_60_HD gradient (1000 mT/m) | T2w, DTI | TurboRARE | Nifti | 131 |
| **19** | **Vrooman** | **117_m_Vr** | In vivo mouse brain | Bruker Biospec USR 117/16, B-GA9S (600 mT/m and slew rates up to 6000 T/m/s) ; Mouse with single loop 10 mm diameter Rx surface coil and 089/072 mm Tx volume resonator. | T2w, GE-EPI | TurboRARE /rs-fMRI | Nifti | 12 (multiple time points) |
| **20** | **Wenk** | **94_m_We** | In vivo mouse/rat/gerbil brain | 9.4 Bruker Biospec USR 94/20, B-GA12SHP (660 mT/m and slew rates up to 4570 T/m/s) ; Mouse with Array coil + 112/086 volume resonator; Gerbil with single-loop coil ID15 + 112/086 volume resonator; Rat with 075/40 volume resonator | T2w, rs-fMRI, fMRI, DTI | T2_ANATOMY, DTI_EPI_ISO, T2STAR, AUDISTIM | Bruker raw | 46/32/25 |
| **21** | **Wenk/Goncalves** | **94_r_We** | In vivo mouse/rat/gerbil brain | 9.4 Bruker Biospec USR 94/20, B-GA12SHP (660 mT/m and slew rates up to 4570 T/m/s) ; Mouse with Array coil + 112/086 volume resonator; Gerbil with single-loop coil ID15 + 112/086 volume resonator; Rat with 075/40 volume resonator | T2w, rs-fMRI, fMRI, DTI | T2_ANATOMY, DTI_EPI_ISO, RSFMRI_EXPERIMENT | Bruker raw | 46/32/25 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **22** | **Wenk/Michalek** | **94_g_We** | In vivo mouse/rat/gerbil brain | 9.4 Bruker Biospec USR 94/20, B-GA12SHP (660 mT/m and slew rates up to 4570 T/m/s) ; Mouse with Array coil + 112/086 volume resonator; Gerbil with single-loop coil ID15 + 112/086 volume resonator; Rat with 075/40 volume resonator | T2w, rs-fMRI, fMRI, DTI | T2_TURBO RARE, NON_ROTATING_EPI | Bruker raw | 46/32/25 |
| **23** | **Longo** | **7_m_Lo** | In vivo mouse abdominal | 7T MicroImaging Bruker Avance III, 1H 30mm TX/RX body volume coil | T2w, DTI | T2 mic_rare, DtiEpi | Bruker raw | 20 |
| [1]ex vivo. Abbreviations: DTI (Diffusion Tensor Imaging), NA (details not available), rs-fMRI (resting-state functional MRI), T2w (T2-weighted MRI), Tx (transmit coil), Rx (Receive coil) | | | | | | | | |

**Supplementary Table II**: Summary of published datasets used for testing AIDAqc

| | Repository | Species /Target | Scanner | Sequences | Data types | Ref. |
|---|---|---|---|---|---|---|
| **1** | **Aswendt** | Mouse | 9.4T Bruker, cryocoil | T2-weighted, rs-fMRI | Bruker raw | [63] https://doi.org/10.17617/3.5p |
| **2** | **Jelescu** | Rat | 14T Varian System | T2-weighted, DTI, rs-fMRI | Nifti | https://openneuro.org/datasets/ds003520/versions/1.0.2 https://openneuro.org/datasets/ds004441/versions/1.0.1 |

**Supplementary Table III**: Confusion matrix visualizing the statistical definition used to classify the data.

| ***Total*** | ***AIDAqc predicted positive (bad quality data)*** | ***AIDAqc predicted positive (no bad quality data)*** |
|---|---|---|
| ***Manual rater assigned positive (bad quality data)*** | **True positive (TP)** | **False negative (FN)** |
| ***Manual rater assigned negative (no bad quality data)*** | **False positive (FP)** | **True negative (TN)** |

# Supplementary Material

All coding, processing, and further analysis started with 125 stroke and 35 sham mice. Among the 125 stroke mice, 67 mice had behavioral data without accompanying MRI, while 58 mice had both behavioral and MRI data. MRI was performed and used to analyze lesion size and location in 90 stroke mice; however, 32 had either incomplete behavioral data or no behavioral data. In total, 58 stroke mice had both behavior and MRI information for comprehensive analysis (**Supplementary Fig. S1**).

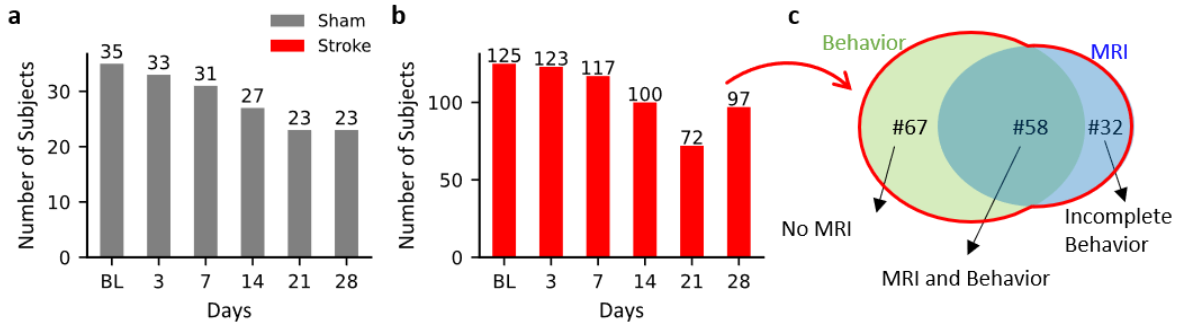



**Figure S1: Overview of data collection, and availability for stroke and sham mice.** Details of data collection for (a) n=35 sham mice and (b) n=125 stroke mice. Behavioral assessments were performed at baseline and post-stroke on days 3, 7, 14, 21, and 28. Lesion size and location were determined using in vivo MRI on days 3 or 7 post-surgery. (c) Distribution of available data for stroke mice of the 125 mice, with 58 having both behavioral and MRI data, and 67 having only behavioral data, taking the initial 125 of the BL as reference. MRI data were obtained for 90 mice, but 32 had incomplete or no behavioral data, as indicated.

Photothrombotic stroke in the sensorimotor cortex was induced in the n=125 mice and sham surgery was applied in the n=35 mice as described before (Götz et al. 2023). Briefly, mice were anesthetized with 4% isoflurane and secured in a stereotactic frame with continuous 2% isoflurane. Temperature was monitored and controlled using a rectal probe and heating pad. After shaving and disinfecting the surgical area, a 1.5 cm incision was made, and the exposed skull was cleaned. A 561 nm laser was aligned to the primary motor cortex (M/L: 2.00 mm and A/P: 0.00 mm), and mice received an intraperitoneal injection of 1000 µg Rose Bengal dye. After 5 minutes, the laser was applied through the skull for 15 minutes at 50 mW. Sham surgery followed the same steps but without laser exposure. The wound was closed post-surgery, and mice recovered in a heated chamber. Analgesia included Tramadol (1 mg/ml) in the drinking water for 3 days before and after surgery, and an intraoperative injection of 4 mg/kg Caprofen. Lesion location and size were determined by an in vivo T2-weighted TurboRare sequence as described before (Aswendt et al. 2021) on day 3 or 7 post-surgery and analyzed using the in-house software AIDAmri (Pallast et al. 2019). Motor recovery was measured at day 0 (before stroke = baseline), and 3, 7, 14, 21, and 28 days post-stroke using the rotating beam, cylinder, and grid walk test (Aswendt et al. 2022). The experimenters were blinded to the animal group (stroke/sham) and the experimenter who performed the stroke surgeries did not perform the MRI and behavior tests.

Unbiased and blinded documentation was performed in an electronic database (Pallast et al. 2018).

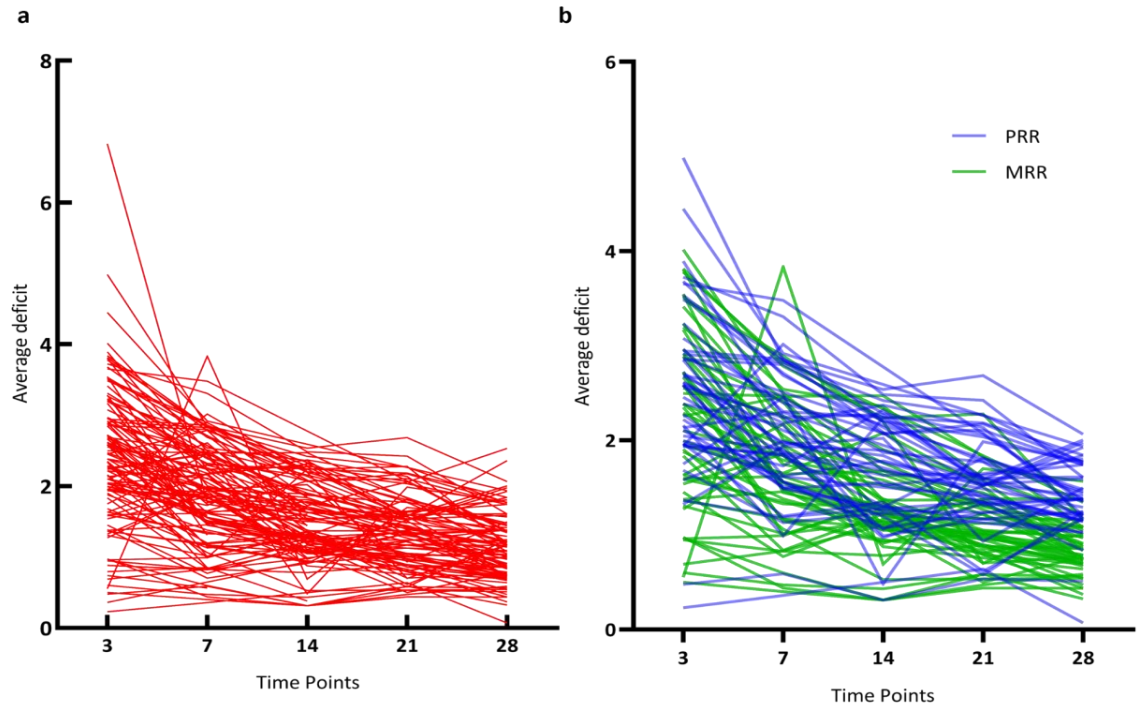



**Figure S2: Individual reduction of motor deficit before and after clustering.** Spaghetti plots for the average deficit of (a) all stroke mice (b) PRR (blue) and MRR (green) mice between 3 and 28 days post-stroke.

**Table S1: Detailed statistics of Figure 2 in the main text**

| Paw drags | | | Hindlimb drop | | | Foot faults | | | Deficit score | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| *Mixed-effect analysis* | | | *Mixed-effect analysis* | | | *Mixed-effect analysis* | | | *Mixed-effect analysis* | | |
| **Comparison** | **Group** | **Sig.** | **Comparison** | **Group** | **Sig.** | **Comparison** | **Group** | **Sig.** | **Comparison** | **Group** | **Sig.** |
| Time | all | *** | Time | all | *** | Time | all | *** | Time | all | *** |
| Group | all | *** | Group | all | *** | Group | all | *** | Group | all | *** |
| Interaction | all | *** | Interaction | all | *** | Interaction | all | *** | Interaction | all | *** |
| *Multiple comparison test: Tukey* | | | *Multiple comparison test: Tukey* | | | *Multiple comparison test: Tukey* | | | *Multiple comparison test: Tukey* | | |
| **Comparison** | **Group** | **Sig.** | **Comparison** | **Group** | **Sig.** | **Comparison** | **Group** | **Sig.** | **Comparison** | **Group** | **Sig.** |
| 0 vs.3 | Stroke | *** | 0 vs.3 | Stroke | *** | 0 vs.3 | Stroke | *** | 0 vs.3 | Stroke | *** |
| 0 vs.7 | Stroke | *** | 0 vs.7 | Stroke | *** | 0 vs.7 | Stroke | *** | 0 vs.7 | Stroke | *** |
| 0 vs.14 | Stroke | *** | 0 vs.14 | Stroke | *** | 0 vs.14 | Stroke | *** | 0 vs.14 | Stroke | *** |
| 0 vs.21 | Stroke | *** | 0 vs.21 | Stroke | *** | 0 vs.21 | Stroke | *** | 0 vs.21 | Stroke | *** |
| 0 vs.28 | Stroke | *** | 0 vs.28 | Stroke | ns | 0 vs.28 | Stroke | *** | 0 vs.28 | Stroke | *** |
| 0 vs.3 | Sham | ns | 0 vs.3 | Sham | ns | 0 vs.3 | Sham | ns | 0 vs.3 | Sham | ns |
| 0 vs.7 | Sham | ns | 0 vs.7 | Sham | ns | 0 vs.7 | Sham | ns | 0 vs.7 | Sham | ns |
| 0 vs.14 | Sham | ns | 0 vs.14 | Sham | ns | 0 vs.14 | Sham | ns | 0 vs.14 | Sham | ns |
| 0 vs.21 | Sham | ns | 0 vs.21 | Sham | ns | 0 vs.21 | Sham | ns | 0 vs.21 | Sham | ns |
| 0 vs.28 | Sham | ns | 0 vs.28 | Sham | ns | 0 vs.28 | Sham | ns | 0 vs.28 | Sham | ns |
| *Multiple comparison test: Šídák* | | | *Multiple comparison test: Šídák* | | | *Multiple comparison test: Šídák* | | | *Multiple comparison test: Šídák* | | |
| **Comparison** | **Group** | **Sig.** | **Comparison** | **Group** | **Sig.** | **Comparison** | **Group** | **Sig.** | **Comparison** | **Group** | **Sig.** |
| Stroke vs.Sham | TP 0 | ns | Stroke vs.Sham | TP 0 | ns | Stroke vs.Sham | TP 0 | ns | Stroke vs.Sham | TP 0 | ns |
| Stroke vs.Sham | TP 3 | *** | Stroke vs.Sham | TP 3 | *** | Stroke vs.Sham | TP 3 | *** | Stroke vs.Sham | TP 3 | *** |
| Stroke vs.Sham | TP 7 | *** | Stroke vs.Sham | TP 7 | ** | Stroke vs.Sham | TP 7 | *** | Stroke vs.Sham | TP 7 | *** |
| Stroke vs.Sham | TP 14 | *** | Stroke vs.Sham | TP 14 | ns | Stroke vs.Sham | TP 14 | *** | Stroke vs.Sham | TP 14 | *** |
| Stroke vs.Sham | TP 21 | *** | Stroke vs.Sham | TP 21 | ** | Stroke vs.Sham | TP 21 | *** | Stroke vs.Sham | TP 21 | *** |
| Stroke vs.Sham | TP 28 | *** | Stroke vs.Sham | TP 28 | ns | Stroke vs.Sham | TP 28 | *** | Stroke vs.Sham | TP 28 | *** |

**Table S2: Detailed statistics of Figure 4 in the main text**

| Paw drags | | | Hindlimb drop | | | Foot faults | | | Deficit score | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| *Mixed-effect analysis* | | | *Mixed-effect analysis* | | | *Mixed-effect analysis* | | | *Mixed-effect analysis* | | |
| **Comparison** | **Group** | **Sig.** | **Comparison** | **Group** | **Sig.** | **Comparison** | **Group** | **Sig.** | **Comparison** | **Group** | **Sig.** |
| Time | all | *** | Time | all | *** | Time | all | *** | Time | all | *** |
| Group | all | *** | Group | all | ** | Group | all | *** | Group | all | *** |
| Interaction | all | *** | Interaction | all | *** | Interaction | all | *** | Interaction | all | *** |
| *Multiple comparison test: Tukey* | | | *Multiple comparison test: Tukey* | | | *Multiple comparison test: Tukey* | | | *Multiple comparison test: Tukey* | | |
| **Comparison** | **Group** | **Sig.** | **Comparison** | **Group** | **Sig.** | **Comparison** | **Group** | **Sig.** | **Comparison** | **Group** | **Sig.** |
| 0 vs.3 | PRR | *** | 0 vs.3 | PRR | *** | 0 vs.3 | PRR | *** | 0 vs.3 | PRR | *** |
| 0 vs.7 | PRR | *** | 0 vs.7 | PRR | *** | 0 vs.7 | PRR | *** | 0 vs.7 | PRR | *** |
| 0 vs.14 | PRR | *** | 0 vs.14 | PRR | ** | 0 vs.14 | PRR | *** | 0 vs.14 | PRR | *** |
| 0 vs.21 | PRR | *** | 0 vs.21 | PRR | ** | 0 vs.21 | PRR | *** | 0 vs.21 | PRR | *** |
| 0 vs.28 | PRR | *** | 0 vs.28 | PRR | ns | 0 vs.28 | PRR | *** | 0 vs.28 | PRR | *** |
| 0 vs.3 | MRR | *** | 0 vs.3 | MRR | *** | 0 vs.3 | MRR | *** | 0 vs.3 | MRR | *** |
| 0 vs.7 | MRR | *** | 0 vs.7 | MRR | *** | 0 vs.7 | MRR | *** | 0 vs.7 | MRR | *** |
| 0 vs.14 | MRR | *** | 0 vs.14 | MRR | ** | 0 vs.14 | MRR | *** | 0 vs.14 | MRR | *** |
| 0 vs.21 | MRR | * | 0 vs.21 | MRR | * | 0 vs.21 | MRR | *** | 0 vs.21 | MRR | *** |
| 0 vs.28 | MRR | ** | 0 vs.28 | MRR | ns | 0 vs.28 | MRR | *** | 0 vs.28 | MRR | *** |
| *Multiple comparison test: Šídák* | | | *Multiple comparison test: Šídák* | | | *Multiple comparison test: Šídák* | | | *Multiple comparison test: Šídák* | | |
| **Comparison** | **Group** | **Sig.** | **Comparison** | **Group** | **Sig.** | **Comparison** | **Group** | **Sig.** | **Comparison** | **Group** | **Sig.** |
| PRR vs. MRR | TP 0 | ns | PRR vs. MRR | TP 0 | ns | PRR vs. MRR | TP 0 | ns | PRR vs. MRR | TP 0 | ns |
| PRR vs. Sham | TP 0 | ns | PRR vs. Sham | TP 0 | ns | PRR vs. Sham | TP 0 | ns | PRR vs. Sham | TP 0 | ns |
| MRR vs. Sham | TP 0 | ns | MRR vs. Sham | TP 0 | ns | MRR vs. Sham | TP 0 | ns | MRR vs. Sham | TP 0 | ns |
| PRR vs. MRR | TP 3 | ns | PRR vs. MRR | TP 3 | ns | PRR vs. MRR | TP 3 | ns | PRR vs. MRR | TP 3 | ns |
| PRR vs. Sham | TP 3 | *** | PRR vs. Sham | TP 3 | *** | PRR vs. Sham | TP 3 | *** | PRR vs. Sham | TP 3 | *** |
| MRR vs. Sham | TP 3 | *** | MRR vs. Sham | TP 3 | *** | MRR vs. Sham | TP 3 | *** | MRR vs. Sham | TP 3 | *** |
| PRR vs. MRR | TP 7 | ns | PRR vs. MRR | TP 7 | ns | PRR vs. MRR | TP 7 | ** | PRR vs. MRR | TP 7 | ns |
| PRR vs. Sham | TP 7 | *** | PRR vs. Sham | TP 7 | ns | PRR vs. Sham | TP 7 | *** | PRR vs. Sham | TP 7 | *** |
| MRR vs. Sham | TP 7 | *** | MRR vs. Sham | TP 7 | ns | MRR vs. Sham | TP 7 | *** | MRR vs. Sham | TP 7 | *** |
| PRR vs. MRR | TP 14 | ns | PRR vs. MRR | TP 14 | ns | PRR vs. MRR | TP 14 | *** | PRR vs. MRR | TP 14 | * |
| PRR vs. Sham | TP 14 | *** | PRR vs. Sham | TP 14 | ns | PRR vs. Sham | TP 14 | *** | PRR vs. Sham | TP 14 | *** |
| MRR vs. Sham | TP 14 | *** | MRR vs. Sham | TP 14 | ns | MRR vs. Sham | TP 14 | *** | MRR vs. Sham | TP 14 | *** |
| PRR vs. MRR | TP 21 | ns | PRR vs. MRR | TP 21 | ns | PRR vs. MRR | TP 21 | *** | PRR vs. MRR | TP 21 | ** |
| PRR vs. Sham | TP 21 | *** | PRR vs. Sham | TP 21 | * | PRR vs. Sham | TP 21 | *** | PRR vs. Sham | TP 21 | *** |
| MRR vs. Sham | TP 21 | ns | MRR vs. Sham | TP 21 | ns | MRR vs. Sham | TP 21 | *** | MRR vs. Sham | TP 21 | *** |
| PRR vs. MRR | TP 28 | ** | PRR vs. MRR | TP 28 | ns | PRR vs. MRR | TP 28 | *** | PRR vs. MRR | TP 28 | *** |
| PRR vs. Sham | TP 28 | *** | PRR vs. Sham | TP 28 | ns | PRR vs. Sham | TP 28 | *** | PRR vs. Sham | TP 28 | *** |
| MRR vs. Sham | TP 28 | ns | MRR vs. Sham | TP 28 | ns | MRR vs. Sham | TP 28 | *** | MRR vs. Sham | TP 28 | ns |

**Table S3: Detailed statistics of Figure 2 in the main text**

| | PRR | MRR |
|---|---|---|
| **Best-fit values** | | |
| **Slope** | 0.7864 | 0.9059 |
| **Y-intercept** | -0.9918 | -0.6534 |
| **X-intercept** | 1.261 | 0.7213 |
| **1/slope** | 1.272 | 1.104 |
| | | |
| **Std. Error** | | |
| **Slope** | 0.04538 | 0.03658 |
| **Y-intercept** | 0.1004 | 0.07162 |
| | | |
| **95% Confidence Intervals** | | |
| **Slope** | 0.6951 to 0.8778 | 0.8323 to 0.9796 |
| **Y-intercept** | -1.194 to -0.7897 | -0.7977 to -0.5092 |
| **X-intercept** | 1.113 to 1.388 | 0.5995 to 0.8310 |
| | | |
| **Goodness of Fit** | | |
| **R squared** | 0.8672 | 0.9317 |
| **Sy.x** | 0.2781 | 0.2731 |
| | | |
| **Is slope significantly non-zero?** | | |
| **F** | 300.4 | 613.4 |
| **DFn, DFd** | 1, 46 | 1, 45 |
| **P value** | <.001 | <.001 |
| **Deviation from zero?** | Significant | Significant |
| | | |
| **Equation** | Y = 0.7864*X - 0.9918 | Y = 0.9059*X - 0.6534 |
| | | |
| **Data** | | |
| **Number of X values** | 95 | 47 |
| **Maximum number of Y replicates** | 1 | 1 |
| **Total number of values** | 48 | 47 |
| **Number of missing values** | 47 | 0 |

Are the slopes equal?
F = 4.216. DFn = 1, DFd = 91
P = .043

If the overall slopes were identical, there was a 4.29% chance of randomly choosing data points with slopes this different. You can conclude that the differences between the slopes were significant.

Because the slopes differ so much, it is impossible to test whether the intercepts differ significantly.

# Acknowledgments

This work is the result of four years of research at the University Hospital Cologne, in the Department of Neurology, within the laboratory of **Dr. Markus Aswendt**. The project was funded by the German Research Foundation (DFG) within the framework of the Collaborative Research Center 1451. With an engineering background, I entered this project with the aim of developing my programming skills within the context of stroke research. Simply put, the project focuses on linking imaging data to stroke recovery — an area where technical and medical disciplines intersect in complex ways.

As the work progressed, I quickly realized that, beyond technical programming tasks, one topic became increasingly central: data management. The vast amount of imaging data, combined with behavioral assessments and biological information, made it clear that good data management is essential to ensure that research remains transparent, reproducible, and efficient. Particularly in the environment of translational research, where collaboration between clinicians, biologists, engineers, and software developers is routine, I observed how different perspectives on data organization often lead to inconsistencies or inefficiencies. Over time, I developed a strong affinity for the different approaches and capabilities that exist in data management. Through this, I also came to recognize that data management is no longer just a supporting aspect — it is becoming an independent research topic in its own right. This is a field I would like to contribute to in the future.

This brings me to expressing my gratitude to my supervisor, **Dr. Markus Aswendt**, for providing the freedom and support to engage with these topics and to develop within the framework of this project. His ability to recognize and encourage the interests of his students, while fostering a constructive and open environment where mistakes are seen as part of the learning process, has been invaluable. I would also like to thank my other PIs, **Prof. Adele Rüger** and **Prof. Michael Schröter**, who were always supportive and helpful in any situation. My gratitude extends to my tutors, **Prof. Joseph Kambeiz**

and **Prof. Thilo van Eimeren**, for their careful guidance and constructive feedback throughout my thesis. Also, special thanks to **Prof. Michael Hanke** and **Dr. Michal Szczepanik**, who taught me most of what I know about data management.

My thanks also go to **Prisca** and **Jan**, whose support made the start of my PhD journey much easier and more enjoyable. I am also grateful to **Viktor**, **Sebastian**, **Luca**, **Carolin**, **Mehrab**, **Franzi**, **Jule**, **Anika**, **Leon**, **Marko**, **Mark**, **Daniel**, **Irem**, **Kimia**, **Vincent**, and **Vera** for the countless scientific and personal conversations that enriched daily life in and outside the office. A special thank you goes to **Motahare** and **Fatemeh**, whose presence and conversations in Farsi brought a comforting sense of home to the workplace.

Next, I want to thank my best friends **Mehdi** and **Farhad**, who were always there for me when I needed them, even across long distances. I am also grateful to my friend **Wesley**, who listened to my research stories with the same enthusiasm that I have for research. Great thanks also to my friends **Yousef**, **Vali**, **Kian**, **Dalir**, **Sahand** and **Behnam** for their support even if they were far away. Also, thanks to **Alireza**, **Siya**, and **Saba**, with whom I created countless memories during these four years in Cologne. A special thanks to **Alireza** for the many hours we spent bouldering together.

My heartfelt thanks go to my family — **Gertrud**, **Hossein**, **Daya**, and **Arezoo** — for their constant encouragement and support. And of course, to little **Aaron**, my nephew — I didn't forget you. I would also like to express my gratitude to my parents-in-law, **Hamid** and **Fatemeh**, and to my brother-in-law, **Mohammad**, for their kindness throughout these years.

Finally, and most importantly, I would like to thank my wife, **Mahdis**, who kept me grounded throughout these years. In her special way, she has always helped me find the right balance, reminding me not to take my work too seriously, but never to neglect it either. For that, I am deeply grateful.

And also :)



# Übersicht der Publikationen

- **How to establish and maintain a multimodal animal research dataset using DataLad**

  **Original Article** – 10.1038/s41597-023-02242-8

  *Scientific Data*, Publiziert in 2023

  **Arbeitsbeteiligung:**

| **Autor** | **Beitrag** |
|---|---|
| Kalantari, A. | Konzeptualisierung, Evaluation, Implementierung, Datenerhebung, Schreiben und Überarbeitung des Manuskripts |
| Szczepanik, M. | Supervision, Überarbeitung des Manuskripts |
| Heunis, S. | Überarbeitung des Manuskripts |
| Mönch, C. | Überarbeitung des Manuskripts |
| Hanke, M. | Supervision, Überarbeitung des Manuskripts |
| Wachtler, T. | Supervision, Überarbeitung des Manuskripts |
| Aswendt, M. | Supervision, Konzeptualisierung, Zusammenarbeit, Überarbeitung des Manuskripts |

- **Automated quality control of small animal MR neuroimaging data**

  **Original Article** – 10.1162/imag_a_00317

  *Imaging Neuroscience*, Publiziert in 2024

  **Arbeitsbeteiligung:**

| **Autor** | **Beitrag** |
|---|---|
| Kalantari, A. | Konzeptualisierung, Datenerhebung, Implementierung, Schreiben und Überarbeitung des Manuskripts |
| Shahbazi, M. | Datenerhebung, Überarbeitung des Manuskripts |
| Schneider, M. | Datenerhebung, Überarbeitung des Manuskripts |
| Lelacqua, G. | Überarbeitung des Manuskripts |
| Raikes, A. C. | Überarbeitung des Manuskripts |
| Vera Frazão, V. | Data curation |
| Aswendt, M. | Supervision, Konzeptualisierung, Zusammenarbeit, Überarbeitung des Manuskripts |

  **Zusätzliche Informationen zu Publikation:**

  Die Publikation *Automated quality control of small animal MR neuroimaging data* ist das Ergebnis einer umfangreichen Zusammenarbeit für unsere AIDAqc-Pipeline und stellt meine bisher umfassendste Arbeit dar. Das Projekt wurde in Kooperation mit 19 internationalen Institutionen durchgeführt und basierte auf 23 verschiedenen Datensätzen, anhand derer unsere Pipeline unter vielfältigen Bedingungen gründlich getestet wurde. Diese Arbeit, veröffentlicht in *Imaging Neuroscience*, gehört zu den umfassendsten Studien zu diesem Thema. Die Auflistung der in der obigen Tabelle genannten Personen umfasst nur diejenigen, die über die Datenerhebung hinaus an der Konzeptualisierung oder anderen Aspekten des Projekts mitgearbeitet haben. Die restlichen Mitwirkenden, die nicht aufgeführt sind - weil dies die Liste sprengen würde – waren primär für die Datenerhebung und die Bereitstellung der Daten verantwortlich. Für einen genaueren Überblick bitte unter "Author contributions" in den Publikationen nachlesen.

- **Proportional recovery in mice with cortical stroke**

  **Research Paper** – 10.1016/j.expneurol.2025.115180

  *Experimental Neurology*, Publiziert in 2025

  **Arbeitsbeteiligung:**

| **Autor** | **Beitrag** |
|---|---|
| Kalantari, A. | Konzeptualisierung, Datenerhebung, Implementierung, Analyse, Schreiben und Überarbeitung des Manuskripts |
| Hambrock, C. | Datenerhebung, Überarbeitung des Manuskripts |
| Grefkes, C. | Überarbeitung des Manuskripts |
| Fink, G. R. | Überarbeitung des Manuskripts |
| Aswendt, M. | Supervision, Konzeptualisierung, Zusammenarbeit, Überarbeitung des Manuskripts |

# Erklärung

Hiermit versichere ich an Eides statt, dass ich die vorliegende Dissertationsschrift selbstständig und ohne die Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Zur sprachlichen Überarbeitung einzelner Passagen dieser Arbeit habe ich KI-basierte Tools wie OpenAI's ChatGPT (Version GPT-o4-mini-high) unterstützend eingesetzt. Alle Stellen – einschließlich Tabellen, Karten und Abbildungen –, die wörtlich oder sinngemäß aus veröffentlichten und nicht veröffentlichten anderen Werken entnommen sind, habe ich als solche kenntlich gemacht. Ich versichere ferner, dass diese Dissertationsschrift noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat, dass sie – abgesehen von unten angegebenen Teilpublikationen – noch nicht veröffentlicht worden ist und dass ich eine solche Veröffentlichung vor Abschluss der Promotion nicht ohne Genehmigung des/der Vorsitzenden des IPHS-Promotionsausschusses vornehmen werde. Die Bestimmungen dieser Ordnung sind mir bekannt. Die von mir vorgelegte Dissertation wurde von PD Dr. Markus Aswendt betreut. Darüber hinaus erkläre ich hiermit, dass ich die „Ordnung zur Sicherung guter wissenschaftlicher Praxis und zum Umgang mit wissenschaftlichem Fehlverhalten" der Universität zu Köln gelesen und bei der Durchführung der Dissertation beachtet habe, und verpflichte mich, die dort genannten Vorgaben bei allen wissenschaftlichen Tätigkeiten einzuhalten und umzusetzen.

Ich versichere, dass ich alle Angaben wahrheitsgemäß nach bestem Wissen und Gewissen gemacht habe und verpflichte mich, jedmögliche, die obigen Angaben betreffenden Veränderungen, dem IPHS-Promotionsausschuss unverzüglich mitzuteilen.

30.10.2026 Aref Kalantari Sarcheshmeh