PAASP Partnership for Assessment and Accreditation of Scientific Practice

ECNP Amsterdam 2025

Campfire session: Data quality

Arnoud HJ Herremans, Anton Bespalov

Arnoud.herremans@paasp.net anton.bespalov@paasp.net

Introduction

High-quality data and transparent research practices have become central concerns in biomedical, clinical, and translational research. Failures in reproducibility, insufficient documentation, selective reporting, and lack of data sharing have eroded trust in scientific findings and slowed progress. The so-called "replication crisis" has prompted multiple reforms in publishing, funding, and research training (e.g. Resnik et al. on reproducibility and research integrity) David B. Resnik & Adil E. Shamoo (2017) PMC+2NCBI+2. In parallel, reporting guidelines (CONSORT, STROBE, ARRIVE, etc.) and transparency initiatives (e.g. the EQUATOR Network) aim to improve the completeness and clarity of how research is reported — effectively as a mechanism to improve "data quality" as communication and reuse assets.

However, much of the published discourse is top-down: guidelines, meta-research, policy proposals. Less is known about what working researchers perceive as the obstacles and enablers of data quality in their day-to-day practice. The ECNP campfire session on "Data Quality" was convened to surface those perspectives, explore what participants believe good research quality entails, identify threats, and generate collective proposals for improvement.

The goals of this session (and thus of this report) were:

1. To map the themes and concerns that active researchers see as central to data quality, in their own contexts (preclinical, clinical, translational).
2. To relate those themes to the existing literature on reproducibility, transparency, data sharing, and research integrity.
3. To derive actionable lessons, take-home messages, and recommendations for next steps — in terms of practice, training, incentive structures, and policy.

Method

Format & Setting

During the ECNP 2025 meeting in Amsterdam, a "campfire" (small-group, rotating discussion) format was used. Attendees were divided into **four rounds** (i.e. each group rotated through different discussion prompts). At each round, participants discussed a

particular question and noted key points and quotes. After all rounds, groups reconvened to report back to the plenary, summarizing their notes and highlighting quotations that captured their views.

Discussion Prompts

The prompts were:

1. **What is "good research quality" to you?**
2. **What are the threats to good research quality?**
3. **What should happen to improve research quality?**
4. **What motivates you to take action toward better quality?**

These four thematic rounds guided the discussion and the reporting.

Synthesis & Analysis

After the session, the facilitators collated all group notes and quotations. We thematically coded the responses (open coding), grouping similar ideas. Then we cross-referenced those emergent themes with existing literature and meta-research on reproducibility, transparency, and data quality. In the outcome section below, we present the distilled themes and illustrative quotations, and in the discussion interpret them in light of published evidence.

Outcome of the Session

Below is a summary of the main themes identified from the participant discussions, grouped by prompt, augmented with illustrative quotes and cross-theme summaries.

1. "What is good research quality?"

Emergent subthemes:

- **Adequate sample size and replication / reproducibility**

  "Sample size can be a problem both in clinical and preclinical research"
  "It is rare to see replicated results; difficulty with reproducibility of research; no independent replication"

- **Use of guidelines, transparent methodology, documentation & code sharing**

  "Good research quality must involve guidelines (e.g. ARRIVE)"
  "There is a lack of proper documentation of scripts"
  "Code and protocol sharing should be encouraged"
  "Lack of transparent methodology and results sections"

- **Relevance, external validity, meaning to target group**

  "Something meaningful to your target group"
  "Results are externally valid"

- **Active control / intervention rigor**

  "Active intervention control conditions"

- **Data transparency, data sharing, and openness**

  "Data transparency"
  "Data sharing"

- **Training, methodological competence, and institutional support**

  "Lack of people training in methodology"
  "Poor data management"

- **Bias, selection / reporting, authorship issues**

  "Selection bias (e.g. excluding suicidel [sic])"
  "No authorship (individual)"

These responses underscore that participants view research quality not just in statistical or technical terms, but as deeply tied to process, transparency, and societal relevance.

2. "What are the threats to good research quality?"

Main themes:

- **Publication pressure / perverse incentives**

  "Journals not publishing negative results or only exciting results"
  "Pressure to publish … need to advance … having a job … reputation"
  "Speed of publication"

- **Biases and conflicts of interest**

  "Conflict of interest; always because stakes involved/bias"
  "Personal bias influencing the writing"

- **Lack of training or methodological rigor**

  "Lack of training in research methodology"
  "Not admitting that the original finding was likely a false positive"
  "Not sticking to pre-registration analysis"
  "Lack of clear hypothesis"

- **Poor reporting and selective omission**

  "Poor reporting of insignificant negative findings"
  "Leaving out limitations"
  "Reporting bias"

- **Inter-lab communication, fragmentation, over-specialization**

  "Interlab lack of communication"
  "Over-specialization, no one has an overview"

- **Institutional, structural, and cultural constraints**

  "Institutional support"
  "The system"
  "Standardization is a threat; lack of standardization is a threat too"

The recurring emphasis is that systemic incentives, norms, and lack of infrastructural support pose as much risk as purely methodological issues.

3. "What should happen to improve research quality?"

Suggested solutions and reforms:

- **Redesign publishing and peer review models**

  "Get rid of journals"
  "Publish the study protocol and then the results, regardless of findings"
  "Pay or acknowledge the reviewers"

- **Pre-registration, registered reports, mandatory protocol registration**

  "Mandatory pre-registration"
  "Peer to peer mentoring in updating research methodology"

- **Replication and "slow science" approach**

  "Dedicate time and funding to replication of research"
  "Do more slow science, less emphasis on publishing or perishing"
  "Allocate more time to complete a good research project"
  "Give researchers more time"

- **Interdisciplinary coordination and integration**

  "Set up interdisciplinary research with great coordinators/facilitators"

- **Statistical support, methodological consultation**

"Minimal demand of statistical counseling, preferably at the stage of process"
"Adding experimental bias in the statistical model"

- **Reform incentives and metrics**

  "Get rid of impact factor of journals, to minimize focusing on news-worthiness"
  "Article processing charges should be reduced or scraped"

- **Mentorship, culture change, community norms**

  "Peer to peer mentoring in updating research methodology"
  "Change philosophy from money-generating results to empirical driven questions"

These proposals reflect a mix of structural, cultural, and methodological change.

4. "What motivates you to take action?"

Motivational themes:

- **Desire to produce meaningful, trustworthy research**

  "Make my research meaningful"
  "Research should be filled by a desire to make a difference"

- **Frustration with the opacity and unreliability of published studies**

  "Many study results are untrustworthy, or you do not know what to trust"
  "Start in your own circle of influence to aid reproducibility"

- **Career incentives (though ambivalent)**

  "Reach high IF journals"

- **Personal growth, learning, community**

  "Researchers should develop a critical way of analysis"
  "What we just did (having these conversations together)"

- **Tools and practices**

  "Preregister study"
  "Publish in open science papers"
  "Bayes statistics"

- **Social mission, data utility**

  "Combining your skills with meaningful things for society (help people)"
  "Data need to be used"

Thus, motivation is a blend of intrinsic drive (doing good science) and extrinsic pressures (publishing, reputation), but participants express a stronger conviction for intrinsic motivation as more sustainable.

Discussion / Interpretation of the Results

Thematically, what stands out

1. **Transparency, documentation, openness**
   Many responses revolve around the need for full transparency in methods, documentation of code and scripts, data sharing, and clear reporting. This aligns well with the broader literature: transparency is considered one of the "pillars" of trustworthy empirical research, alongside credibility and reproducibility. Maria R. Jones, Kristoffer Bjärkefur, Luíza Cardoso de Andrade, Benjamin Daniels (2021) BioMed Central+3worldbank.github.io+3Wolters Kluwer+3 The Transparency and Openness Promotion (TOP) guidelines represent a codified framework for this (data, methods, materials, preregistration, replication) Partricia K. Baskin, Robert A. Gross. (2019) Wolters Kluwer+1.

   In participant quotes, lack of code/documentation, nontransparent methodology, and poor reporting were flagged repeatedly. That suggests a gap remains between the ideals of open science and everyday practice.

2. **Incentive structures and publication culture as root challenge**
   The dominant threat perceived by participants is the "publish or perish" regime: selecting for positive or "exciting" results, speed over rigor, and marginalizing negative or replication studies. This view echoes the meta-research literature: publication bias, selective reporting, and perverse incentives are widely regarded as central causes of irreproducibility (e.g. in the National Academies report) The national academy of science, engineering, medicine (2019) NCBI+2PMC+2.

   Several participants even advocated radical reforms (e.g. "get rid of journals"). While those may be aspirational, they highlight how deeply frustration with existing publishing norms runs.

3. **Methodological training, competence, and infrastructure gaps**
   Repeated mention of inadequate training in methodology, lack of statistical consultation, and poor data management suggests that many researchers feel underprepared—or unsupported—in implementing good practices. In the literature, inadequate methodological training is often blamed for careless design, analysis, and reporting errors. Patrick D. Nuhoho, Michael A. Offeh (2021) BioMed Central+2NCBI+2

   Without better training and infrastructural support (e.g. data management platforms, code versioning, reproducible pipelines), ideals of transparency may not translate into actual practice.

4. **Replication, slow science, and pre-commitment**
   Participants' calls for more replication, pre-registration, and slow science align well with current proposals in reproducibility reform. The idea of committing to a plan in advance (pre-analysis, registered reports) is thought to reduce bias and p-hacking. Maria R. Jones, Kristoffer Bjärkefur, Luíza Cardoso de Andrade, Benjamin Daniels (2021) worldbank.github.io+2NCBI+2

   Encouragingly, participants recognized the importance of replication—not merely as a formal check but as an integral part of the research ecosystem.

5. **Motivation and culture change**
   The participants' expressed motives emphasize intrinsic values: wanting to make meaningful contributions, be trustworthy, and help society. This matches the view in meta-science that sustainable change must include culture and norms, not just rules. David B. Resnik & Adil E. Shamoo (2017) PMC+2BioMed Central+2

   The fact that the session itself (conversation, peer reflection) was cited as motivating suggests that creating spaces for these dialogues is itself a lever for change.

Relating to the published literature on data quality

From the literature, "data quality" is often discussed in terms of dimensions (accuracy, consistency, completeness, timeliness, believability, etc.) and "fitness for purpose" — i.e. data quality is context-dependent. Data Quality: researsch data management (2025) rdm.mpdl.mpg.de+1 In scientific research, data quality also extends to adequate documentation, metadata, provenance, and reusability. Data Quality: researsch data management (2025) rdm.mpdl.mpg.de+1

What the participants surfaced is that they interpret "data quality" less in narrow technical dimension terms (e.g. missing values) and more in terms of the entire research pipeline: transparency, reproducibility, reporting, sharing, and cultural/institutional enablers. This broader perspective is consistent with modern approaches in open science and reproducibility initiatives, which frame "data quality" as inseparable from the process of producing, documenting, and sharing data.

The literature also emphasizes that data quality cannot be treated in isolation: it depends on design, collection, cleaning, documentation, and downstream use. Maria R. Jones, Kristoffer Bjärkefur, Luíza Cardoso de Andrade, Benjamin Daniels (2021) worldbank.github.io+1 Thus, the participants' pointing to training, protocols, code sharing, and documentation is well aligned with the state-of-the-art view.

One gap that participants less explicitly raised is the role of data quality monitoring tools (automated or semi-automated), validation pipelines, and continuous quality checks (e.g. real-time data quality dashboards). In the informatics / data engineering literature, there is substantial discourse about data profiling, monitoring tools, anomaly detection, pipeline validation, etc. Lisa Ehrlinger, Elisa Rusz, Wolfram Wöss (2019) arXiv Whether and how such tools are applicable in clinical / translational research is a topic to explore further.

Another frontier in the literature is *continuous reproducibility* and integration of software engineering practices (e.g. version control, continuous integration of analysis pipelines) to improve reproducibility and guard against error drift. Venkat S. Malladi, Maria Yazykova, Olesya Melnichenko, Yulia Dubinina (2024) arXiv+1

Thus, your session results resonate strongly with the published discourse, while also highlighting that many barriers remain in daily practice: training, culture, incentives, and infrastructure.

Key Learnings & Take-Home Messages

From this reflection, we identify the following key learnings and take-home messages:

1. **"Data quality" is understood by practitioners broadly**
   Participants see data quality not just as absence of error, but as transparency, documentation, reproducibility, and societal relevance.
2. **Incentives and culture are as critical as technical fixes**
   Many of the threats identified are systemic (publication pressure, bias, institutional norms). Without addressing incentive structures, even the best technical reforms may not be widely adopted.
3. **There is a gap between ideal guidelines and everyday practice**
   Many researchers lack methodological training, statistical support, or infrastructure to implement transparency, code sharing, and reproducible pipelines.
4. **Pre-commitment (pre-registration, registered reports) and replication are strongly endorsed**
   These are viewed not just as optional extras but as central to credible science.
5. **Motivation is often intrinsic — but extrinsic pressures conflict**
   Researchers want to do good science, but career pressures pull them toward safer, publishable results. Aligning extrinsic incentives (funding, publishing norms) with integrity is essential.
6. **Dialogue, community, and peer reflection are themselves drivers of change**
   The session itself, and the mutual sharing of concerns, was cited as motivating — suggesting that creating forums for such reflection may catalyze culture shift.

Recommendations & Next Steps

Based on the outcomes and the relation to the literature, here are proposed recommendations and next steps:

1. **Training, mentoring, and capacity building**
   - Develop modular courses/workshops (in statistical methods, reproducible workflows, data management) tailored for clinical/preclinical researchers.

- Establish **peer mentoring or "methodology clubs"** within institutions (as some participants suggested).
- Encourage cross-disciplinary methodological support (e.g. involving biostatisticians, data scientists) early in project design.

2. **Infrastructural and tooling support**
   - Provide institutional infrastructure: version control systems (e.g. GitHub, GitLab), computing environments that preserve provenance, and data management platforms.
   - Explore adoption of **continuous analysis pipelines** or reproducibility pipelines (automated testing, versioning) as in software engineering paradigms. Venkat S. Malladi, Maria Yazykova, Olesya Melnichenko, Yulia Dubinina (2024) arXiv+1
   - Incorporate **data quality monitoring tools** or validation checks (e.g. data profiling, anomaly detection) to catch issues early. Lisa Ehrlinger, Elisa Rusz, Wolfram Wöss (2019) arXiv
3. **Embed transparency and reproducibility in project design**
   - Make pre-registration or registered reports the default (or strongly encouraged) in funding calls, institutional review boards, or departmental norms.
   - At protocol stage, plan for full documentation, code sharing, metadata, versioning, and archiving.
4. **Incentive and policy reform**
   - Encourage funding agencies and institutions to reward replication, rigorous methodology, transparency, not just high-impact outputs.
   - Promote adoption of metrics like **TOP Factor** (Transparency and Openness Promotion) to assess journal policies favoring reproducibility, rather than purely citation-based metrics. David Mellor (2020) cos.io+1
   - Advocate for journals to adopt transparency policies: e.g. requiring data, materials, code availability statements, verifying data, registered reports.
   - Consider reforming peer review: e.g. open peer review, reviewer acknowledgment or remuneration, decoupling prestige from novelty.
5. **Facilitate replication efforts and "slow science" time**
   - Allocate dedicated funding lines or time for replication studies (which participants strongly requested).
   - Encourage "slow science" — valuing depth over quantity, allowing longer project times to ensure rigor.
6. **Create forums for community reflection and norm building**
   - Host follow-up workshops, journal clubs, or "campfire" sessions regularly to sustain conversation and reinforce norms.
   - Form working groups across ECNP (or similar societies) tasked with drafting guidelines, checklists, or best practice frameworks tailored to neuroscience/psychiatry.
   - Share successful case studies of reproducible research in the field to serve as exemplars.
7. **Evaluate and iterate**
   - Over time, monitor whether interventions (e.g. training, infrastructure) lead to measurable improvements in reporting transparency, code sharing, replicability.

- Use metrics such as the Research Transparency Index or audit adherence to transparency policies. Brian Weimerskirch, Carrie Baker (2025) ScienceDirect
- Conduct periodic surveys or reflections to gauge whether cultural norms are shifting.

References:

David B. Resnik & Adil E. Shamoo (2017) Reproducibility and Research Integrity, Accountability in Research, 24:2, 116-123

equator-network.org (2025) EQUATOR Network | Enhancing the QUAlity and Transparency Of Health Research

Maria R. Jones, Kristoffer Bjärkefur, Luíza Cardoso de Andrade, Benjamin Daniels (2021) The DIME analytics data handbook. Conducting reproducible, transparent, and credible research. Chapter 1.

Partricia K. Baskin, Robert A. Gross. (2019) Transparency in research and reporting. www.wolterskluwer.com/en/expert-insights.

The national academy of science, engineering, medicine (2019) Reproducibility and replicability in science, the National Academies Press.

Patrick D. Nuhoho, Michael A. Offeh (2021) Reproducibility and researxh integrety: the role of scientists and institutions. BMC research notes.

Eszter Hargittai (2018) Potential biases in big data: omitted voices on social media. Social Science Review vol38(1)

Lisa Ehrlinger, Elisa Rusz, Wolfram Wöss (2019) A survey of data quality measurement and monitoring tools. Arxiv.org

Venkat S. Malladi, Maria Yazykova, Olesya Melnichenko, Yulia Dubinina (2024) Continuous analysis: evolution of software engineering and reproducibility for science. Arxiv.org

Data Quality: researsch data management (2025) Max planck institution

David Mellor (2020) New measure rates quality of research journal's policies to promote transparency and reproducibility. Center for open science cos.io

Herman Agnuinis, Zhuyi a. Li, Maw D. Foo (2024) The research transparency index. The leadership quaterly 35.

Six factors affecting reproducibility in life science research and how to handle them. (2025) Nature Portfolio, Springer nature

Katie Harron, Eric Benchimol, Sinead Langan (2018) Using the RECORD guidelines to improve transparent reporting of studies based on routinely collected data. International journal of population data science.

Delwen L Franzen, Maia Salholz-Hillel, Stephanie Müller-Ohlraun, Daniel Strech (2025) Improving research transparency with individualized report cards: A feasibility study in clinical trials at a large university medical center. Medical Research Methodology.

Jack Nash (2025) The importance of reproducibility in open research. Open research blog.

Yemi Macaulay, Ester Bellavia, Rachel Brophy, Angela Coulter, Ben Glampson (2025) Improving transparency in data access processes: developing best practice standards and promoting system-wide change through a competitive funding call. International journal of population data science 10(2).

Brian Weimerskirch, Carrie Baker (2025) Data transparency in healthcare quality improvement. Current problems in pediatric and adolescent health care 55(4)

Make your research more reproducible (2025) F1000 Open research service

Arnim Bleier (2025) What is computational reproducibility? Gesis guides to digital behavioral data.

Meredith Loui, Alfredo Morabia (2025) Ensuring quality in scientific publication through inclusivity and transparency. American journal of public health 115(7).

M. Berger, W. Crown, G. Daniel, H. Eichler, W. Goettsch, J. Graff, J. Guerino, P Jonsson, N.M. Lederer, B. Monz, C.D. Mullins, S. Schneeweiss, D. van Brunt, S.V. Wang, R.j. Willke (2020) Improving Transparency to Build Trust in Real-World Secondary Data Studies for Hypothesis Testing, Why, What, and How Recommendations and a Roadmap from the Real-World Evidence Transparency Initiative. National Pharmaceutical Council.

- **Links to References:**

PMC

Reproducibility and Research Integrity - PMC

by DB Resnik · 2016 · Cited by 150 — Reproducibility—the ability of independent researchers to obtain the same (or similar) results when repeating an experiment or test—is one of the hallmarks ...

-

equator-network.org

EQUATOR Network | Enhancing the QUAlity and ...

Your one-stop-shop for writing and publishing high-impact health research. Find reporting guidelines, improve your writing, join our courses, run your own ...

-

Worldbank.github.io

Chapter 1 Conducting reproducible, transparent, and credible ...

Credibility, transparency, and reproducibility are three pillars of a high-quality empirical research project. The steps and outputs discussed in this ...

-

Wolters Kluwer

Transparency in research and reporting

1 december 2019 — by PK Baskin — The guidelines cover eight standards of transparency in the research process, with three levels of transparency for each standard, so journals can adopt ...

-

NCBI

Summary - Reproducibility and Replicability in Science

One of the pathways by which scientists confirm the validity of a new finding or discovery is by repeating the research that produced it.

BioMed Central

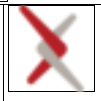
Reproducibility and research integrity: the role of scientists ...

by P Diaba-Nuhoho · 2021 · Cited by 71 — Reproducibility and research integrity are essential tenets of every scientific study and discovery. They serve as proof that an established and documented ...

rdm.mpdl.mpg.de

Data Quality

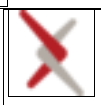
by D Quality — Data quality emphasis on a user perspective has two implications for scientific researchers: reproducibility and in some cases replicability of their own data ...



arXiv

A Survey of Data Quality Measurement and Monitoring Tools

18 juli 2019



arXiv

Continuous Analysis: Evolution of Software Engineering and Reproducibility for Science

4 november 2024

cos.io

New Measure Rates Quality of Research Journals' Policies ...

Feb 10, 2020 — Policies promoting transparency and reproducibility make it easier to evaluate research quality.” TOP Factor assesses the journal's policies, ...

ScienceDirect

The research transparency index

by H Aguinis · 2024 · Cited by 12 — RTI provides feedback on the transparency of manuscripts describing quantitative research across key research stages: theory, design, measurement, data analysis ...

Meer

nature.com

Six factors affecting reproducibility in life science research ...

Scientific advancement depends on a strong foundation of data credibility. However, scientific findings in biomedical research are not always reproducible.

- nih.gov

  Using the RECORD guidelines to improve transparent ...

  by K Harron · 2018 · Cited by 25 — Reporting guidelines such as CONSORT for clinical trials and STROBE for observational studies aim to improve transparency, allowing identification of potential ...

- nih.gov

  Improving research transparency with individualized report ...

  by DL Franzen · 2025 · Cited by 1 — The primary objective of this study was to develop an extensible tool to provide individualized feedback and guidance for improved transparency ...

- wellcomeopenresearch.org

  The importance of reproducibility in open research - Blog

  Jan 15, 2025 — Reproducibility can demonstrate that research results are not due to bias or chance, which is vital for accurate and reliable results.

- ijpds.org

  Improving transparency in data access processes

  by Y Macaulay · 2025 — Transparency in the use of data for research benefits the public and researchers by fostering trust and enabling efficient data sharing. Public ...

sciencedirect.com

Data Transparency in Healthcare Quality Improvement

by B Weimerskirch · 2025 — The principles of data transparency include making the source and validation methods of the data clearly evident as well as ensuring equitable, ...

- f1000.com

  Make your research more reproducible today

  Reproducibility is the degree to which other researchers can achieve the same results using the same dataset and analysis as the original researchers.

- 

  gesis.org

  What is Computational Reproducibility?

  by A BLEIER · 2025 — This guide focuses on computational reproducibility which can be defined as the ability to consistently arrive at numerical comparable findings ...

- 

  wikipedia.org

  Reproducibility

  Reproducibility, closely related to replicability and repeatability, is a major principle underpinning the scientific method.

-

aphapublications.org

Ensuring Quality in Scientific Publication Through Inclusivity ...

by M Loui · 2025 — Transparency can be advanced through collecting stakeholder feedback, analyzing review disparities, publishing data on review timelines and ...

npcnow.org

Improving Transparency to Build Trust in Real-World ...

Sep 16, 2020 — A position paper by the RWE Transparency Initiative describes a plan for improving the transparency of the rese