A protocol for visual quality control applied to the Reproducible Brain Charts database

Nathalia Bianchini Esper[1], Sydney Covitz[2], Alberto Castro[1], Giovanni Abrahão Salum[1], Hamsanandini Radhakrishnan[2], Jon Clucas[1], Kahini Mehta[2], Matt Cieslak[2], Mauricio Scopel Hoffmann[3], Steve Giavasis[1], Gregory Kiar[1], Michael Milham[1], Theodore D. Satterthwaite[2], Alexandre Rosa Franco[1]

[1]Child Mind Institute, [2]University of Pennsylvania, [3]Universidade Federal de Santa Maria

Motivation

- Data quality is one of the most important brain imaging confounding factors[1].
- Efforts are being made to improve the quality of MRI data by developing imaging sequences that adapt to participant head motion[2] or by using automated methods to perform quality control[3].
- However, manual visual ratings remain the gold standard for data quality control (QC).
- The Reproducible Brain Charts (RBC) initiative aims to harmonize phenotypic and neuroimaging data to build highly reproducible human brain development growth charts. This initiative contains data from five large neuroimaging studies[4-8].
- Here we present a workflow for trained raters to perform visual quality control of the anatomical images from the RBC initiative.

Materials and Methods

Imaging rating protocol

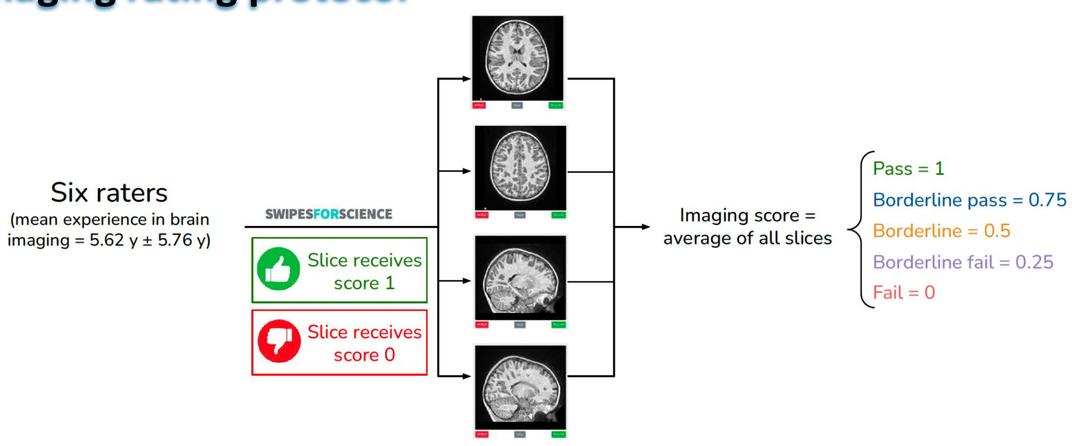

Figure 1: All the T1w volumes were evaluated using Swipes for Science(SfS)[9]. Slices of the volumes were presented to the raters in random order. Each rater reviewed four slices per participant (two axial and two sagittal). A "pass" rating on SfS would be given if an slice was deemed of sufficient quality for skullstriping and segmentation of CSF, WM, and GM. Final participant's score is the average of ratings.

Number of participants and age range per study

BHRC[7]	CCNP[6]	HBN[4]	NKI[5]	PNC[8]	Total of participants
825	190	2,216	2,176	1,153	6,959
5 - 14 yo	6 - 17 yo	5 - 21 yo	6 - 86 yo	8 - 23 yo	5 - 83 yo

Rater training and QC protocol

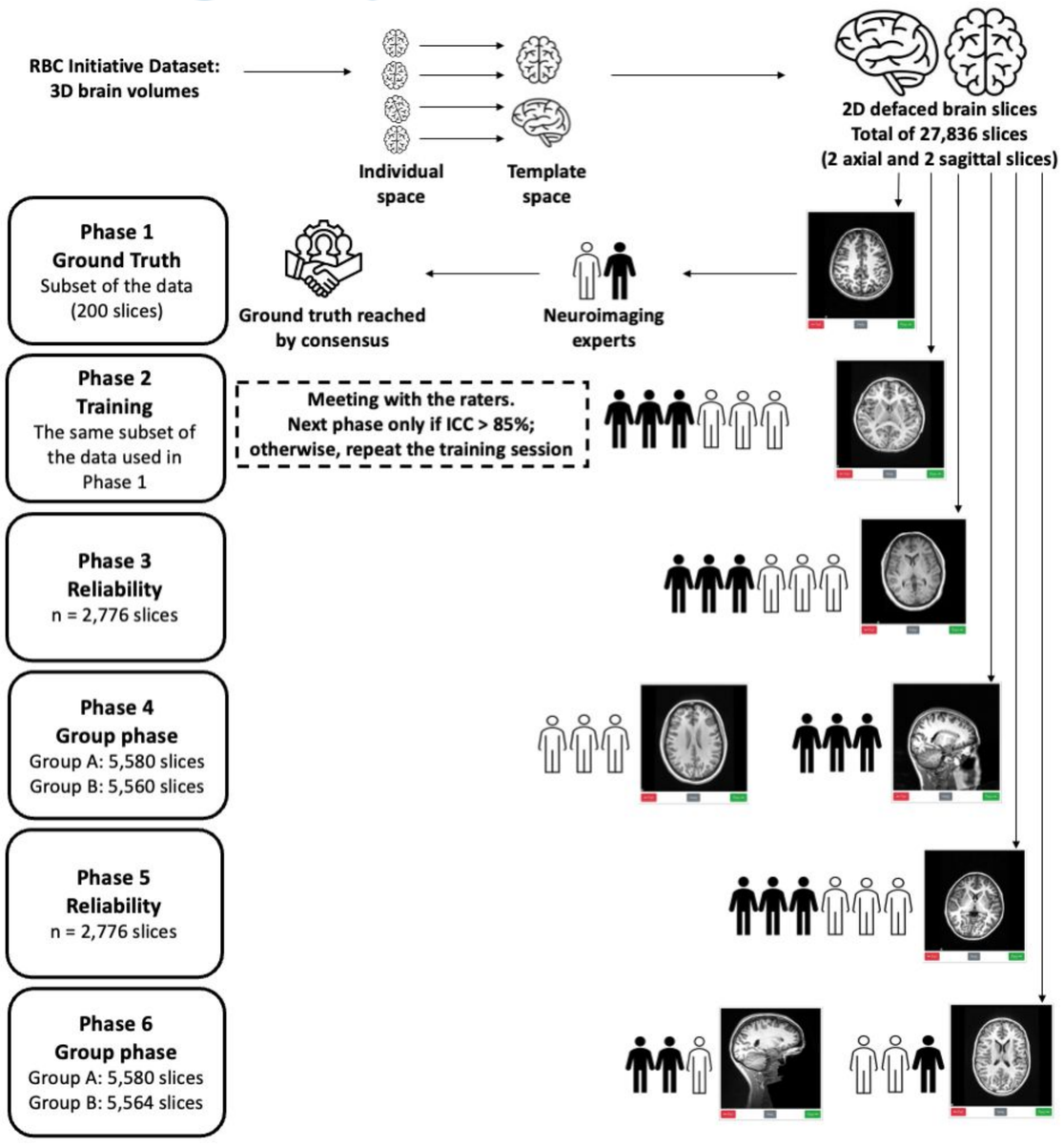

Figure 2: Overview of the visual quality control procedure used for rating the images of the RBC initiative. First, all the data were registered to the MNI152 template using a rigid body transformation. Next, the 3D MRI scans were converted into 2D axial and sagittal slices, which were divided into five groups of images and loaded onto SfS[9], a web application to crowdsource the quality ratings. Two senior raters were responsible for creating the ground-truth on a set of 200 slices (Phase 1), and these slices were used to train all the raters prior to starting QC evaluation (Phase 2). After each of the six raters reached at least 85% of concordance with the ground-truth set, they could start rating the dataset following the phase's order (Phases 3-6). In phase 6, we switched one rater from each group.

Results

Result 1: Reliability across raters

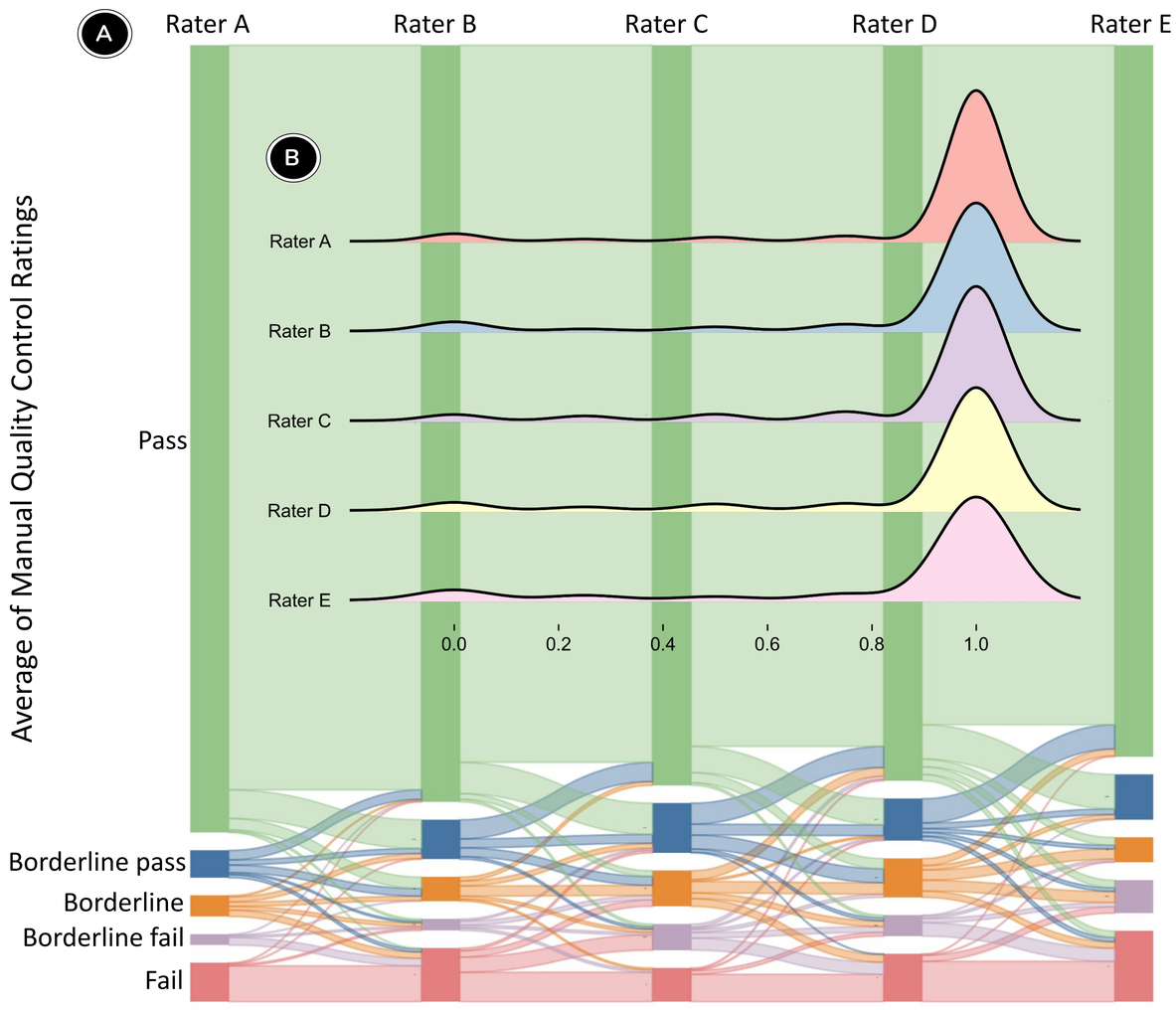

Figure 3: Reliability results across raters. A) Sankey diagram showing the average of manual quality control per subject for each rater. Rater F was excluded from all analyses because their rigorous rating. 1030/1388 participants (74%) had average rating of 1 (pass) and 34/1388 (2.8%) had average rating of 0 (fail). B) Distribution plot of average ratings per rater on the top of Sankey plot. The intraclass correlation coefficient across raters for the reliability analysis was > 0.8 (see GitHub repository for more details; QR code on the bottom left of this poster).

Result 2: Relation between image average rating and study site

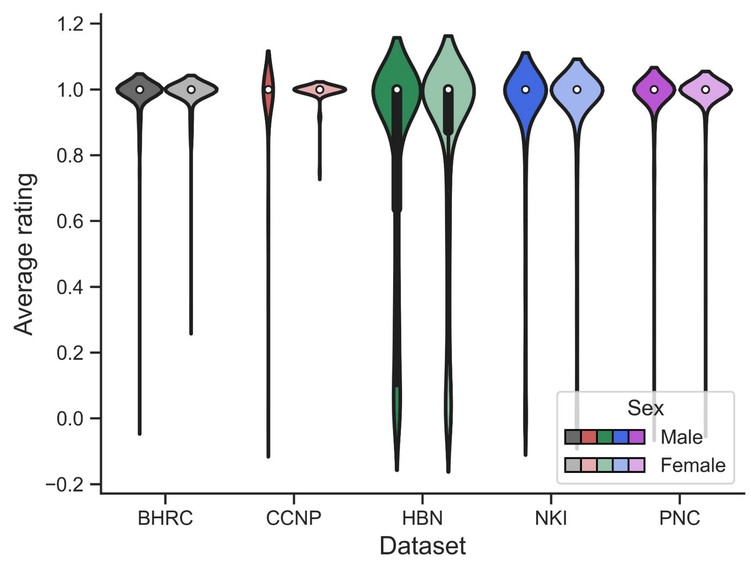

Figure 4: Image average rating distribution per study site. HBN is the site with most exemplars (2,216 participants) and with the widest quality of T1w images.

Result 3: Relation between image average rating and age

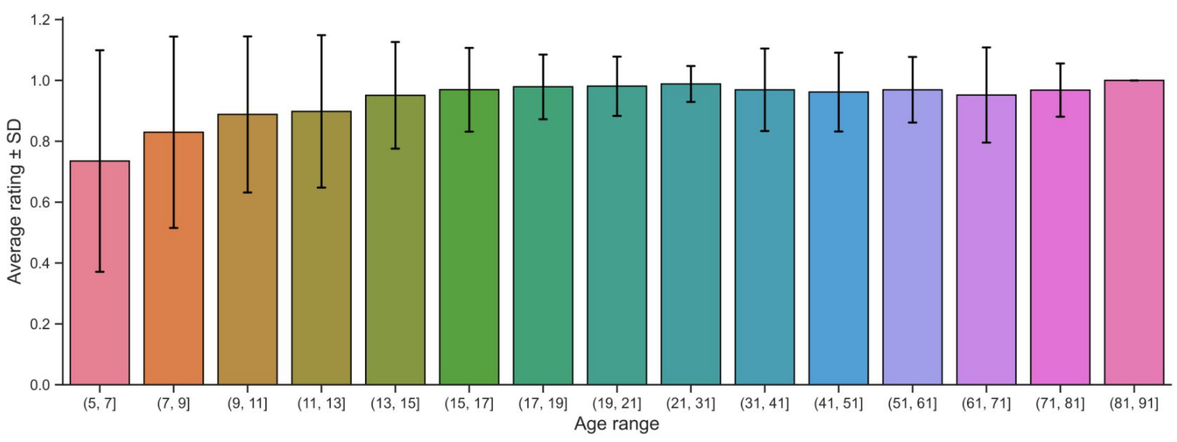

Figure 5: Barplot graph showing the image average rating distribution per age range. Younger participants [5 - 13 yo] tend to have a lowest average rating and a wider standard deviation compared to the older participants.

Take home message

- Raters are reliable.
- Associate manual QC with Euler number to find and acceptable threshold to automatically QCing the data.
- This manually-annotated data from diverse populations, scanners, and sequences is an important data resource for the neuroimaging community and could accelerate the development of automated tools for assessing data quality.
- The RBC initiative dataset is expected to be release before the end of the year.

Contact information:
Nathalia Bianchini Esper - Child Mind Institute (nathalia.esper@childmind.org)
Theodore D. Satterthwaite - University of Pennsylvania (sattertt@pennmedicine.upenn.edu)
Alexandre Rosa Franco - Child Mind Institute (alexandre.franco@childmind.org)

Acknowledgments:
This work was supported by the Child Mind Institute and NIH 5R01MH120482-03.

References:
1. Gilmore et al., 2021
2. Tisdall et al., 2016
3. Rosen et al., 2018
4. Alexander et al, 2017
5. Nooner et al., 2012
6. Liu et al., 2021
7. Salum et al., 2015
8. Satterthwaite et al., 2016
9. Keshavan et al., 2019