1

Let's take a break:

Head motion during fMRI tasks is reduced in children and adults

if data acquisition is distributed across sessions or days

5

6

7   Tobias W. Meissner[a,b], Jon Walbrin[c,d], Marisa Nordt[b,e], Kami Koldewyn[c], Sarah Weigelt[a]

8

9

10   [a] TU Dortmund University, Faculty of Rehabilitation Sciences, Department of Vision, Visual
11   Impairments & Blindness, Emil-Figge-Str. 50, 44227 Dortmund, Germany
12

13   [b] Ruhr University Bochum, Faculty of Psychology, Universitätsstr. 150, 44801 Bochum, Germany
14

15   [c] Bangor University, School of Psychology, Developmental Social Vision Lab, Penrallt Road, Bangor,
16   LL57 2AS, Wales, United Kingdom
17

18   [d] University of Coimbra, Faculty of Psychology and Education Sciences, Proaction Lab, Rua Colégio
19   Novo, 3000-115, Coimbra, Portugal
20

21   [e] Stanford University, Psychology Department, 450 Serra Mall, Stanford, CA 94305, USA
22

23

24   **Corresponding author:** Tobias W. Meissner, tobias.meissner@rub.de

25   **Email of co-authors:** JW: jon.walbrin@uc.pt, MN: marisa.nordt@rub.de, KK:
26   k.koldewyn@bangor.ac.uk, SW: sarah.weigelt@tu-dortmund.de
27

Abstract

Head motion remains a challenging confound in functional magnetic resonance imaging (fMRI) studies of both children and adults. Most pediatric neuroimaging labs have developed experience-based, child-friendly standards concerning e.g. the maximum length of a session or the date of mock scanner training. However, it is unclear which factors of child-friendly neuroimaging approaches are effective in reducing head motion. Here, we investigate three main factors including (i) time lag of mock scanner training to the actual scan, (ii) prior scan time, and (iii) task engagement in a dataset of 77 children and 64 adults using a multilevel modeling approach. In children, distributing fMRI data acquisition across multiple same-day sessions reduces head motion. Nevertheless, motion increases over the course of a study, especially in older children. In adults, splitting data acquisition into multiple days, but not same-day sessions, reduces head motion. Moreover, motion is reduced after inside-scanner breaks. In both children and adults, motion increases with run length. Our results suggest that splitting up fMRI data acquisition is an effective tool to reduce head motion in general. At the same time, different ways of splitting up data acquisition benefit children and adults.

Keywords

head motion, fMRI, child, movement, data quality, quality control, MRI acquisition

Highlights

- In children, fMRI data acquisition split into multiple sessions reduces head motion
- However, children's head motion increases over the duration of the study
- In adults, fMRI data acquisition split into several days reduces head motion
- Moreover, adults' motion decreases after inside-scanner breaks
- In both children and adults, motion increases with run length

1 Introduction

"Please remember: Relax and try not to move. Here we go." In functional magnetic resonance imaging (fMRI) experiments, this is often the last thing the researcher says before starting the scanner and the experiment. What is left to do is hoping for good quality data. As MRI is very susceptible to head motion during the scan (Friston et al., 1996), hoping for good quality data is often equivalent to hoping for data with low head motion. The importance of reducing head motion during data acquisition has been impressively documented in several studies showing that head motion can lead to misleading results (in some cases, even after retrospective motion correction; Power et al., 2012; Satterthwaite et al., 2012; van Dijk et al., 2012).

In pediatric neuroimaging studies, the general problem of head motion is especially challenging. Despite children's high motivation to lie still, in most studies children still move more than adults. This poses a problem for data quality in pediatric neuroimaging especially for group comparisons between children and adults. Even with retrospective head motion correction (which is limited in its ability to correct for motion Field et al., 2000; Freire and Mangin, 2001; Friston et al., 1996), group differences in head motion often persist. This usually leads to the exclusion of motion-affected runs or participants from data analysis (e.g. Meissner et al., 2019; Nordt et al., 2018; Walbrin et al., 2019), hence requiring additional time and research funds to achieve an adequately powered study design.

To minimize motion during data acquisition, various solutions have been developed. For structural MRI scans in clinical settings, sedation is often used. However, this is not an option for fMRI studies in research settings due to the need of attentive participants as well as due to the risk of inacceptable health-related side effects (Bie et al., 2010). One set of solutions for fMRI studies aims at restraining head motion physically. Restrictive versions including bite bars (Menon et al., 1997) or thermoplastic face masks (Green et al., 1994) are considered efficient for short scan durations. However, they reduce comfort and are therefore seldomly accepted in pediatric neuroimaging. In contrast to the desired outcome, the discomfort of these methods can also lead to more fidgeting and wiggling in search of more comfort (Zeffiro, 1996). Approaches that use individually 3D-printed

3

79   styrofoam head molds are effective and said to be comfortable (Power et al., 2019), but require

80   additional time and research funds for each participant. A less resource-intensive, yet effective and

81   accepted solution, is to provide tactile feedback about participant's head motion by applying a tape or

82   ribbon across the head coil that touches the participant's forehead (Krause et al., 2019).

83   Aside from physical constraints, pediatric neuroimaging groups often adapt the study

84   procedure to reduce children's motion during data acquisition. The most prominent tool is to precede

85   the actual MRI session(s) with a scanner training session (Raschle et al., 2009; Slifer et al., 1993).

86   Scanner training sessions are usually performed in a mock scanner, a custom-built or purchased

87   facsimile of a real MRI, which lacks the technical capability to acquire actual data, but can play

88   scanner sounds and has a similar setup to the real scanner (e.g. head coil-mounted mirror, response

89   buttons, etc.). During these sessions, behavioral training is used to teach lying still in the scanner bore

90   and responding to the task. There is a general consensus that a scanner training session is beneficial for

91   fMRI data acquisition as it reduces children's anxiety (Durston et al., 2009; Raschle et al., 2009;

92   Rosenberg et al., 1997). While a reducing effect on head motion is also assumed, the current literature

93   has not been able to show this (due to a lack of control groups that did not receive scanner training; but

94   see Barnea-Goraly et al., 2014; Bie et al., 2010; Epstein et al., 2007 for success rates and reduced

95   motion after scanner training).

96   Similarly, many pediatric neuroimaging groups have established experience-based guidelines

97   for child-friendly study designs in terms of the scanning procedure. These usually include keeping

98   runs short, interspersing anatomical scans as task-free inside-scanner breaks while presenting

99   entertaining video clips, and limiting the total scanning session time. Moreover, based on findings in

100   adults (Huijbers et al., 2017), tasks are designed in an engaging and interactive way rather than just

101   requiring passive perception of stimuli. In consequence, a small body of literature has developed that

102   recommends scanning procedures for pediatric neuroimaging (e.g. Greene et al., 2016; Habibi et al.,

103   2015; Raschle et al., 2009). However, as for scanner training, the contribution of presumably child-

104   friendly adaptations in scanning procedures on minimizing head motion has not been investigated so

105   far.

4

106   The present study set out to identify which fMRI study procedures contribute to reduced head

107   motion in fMRI studies. Our aim was to generate data-driven suggestions to optimize study procedures

108   for children and adults separately. To this end we utilized head motion estimates derived from a

109   standard motion correction pipeline applied to 77 children and 64 adults from three fMRI studies at

110   two sites and notes that yielded demographic data, the date of scanner training, and the sequence of

111   data acquisition. Using separate multilevel linear models for children and adults, we investigated the

112   effect of splitting up data acquisition into several days or sessions, the effect of interspersing

113   functional data acquisition with structural runs and video clip breaks, the effect of time between

114   scanner training and the actual scan, and the effect of task engagement.

115

2   Methods

2.1   Definition of session, run, and functional segment

In this article, we make important distinctions between "session", "run", and "functional segment". Session corresponds to a continuous period of time spent inside the scanner. A session begins upon entering the scanner and ends as a participant leaves the scanner. For example, if a participant enters the scanner, takes a break outside the scanner and re-enters the scanner, this would constitute two sessions. Run corresponds to a continuous image acquisition sequence. For example, a participant could complete an experiment with four fMRI runs within a single session. Functional segment corresponds to a period of perpetually consecutive acquisition of functional runs inside the scanner. For example, if a scan procedure involves an anatomical T1-scan, three fMRI runs, two diffusion-weighed imaging (DWI) sequences, and finally four more fMRI runs, this participant has completed two functional segments.

2.2   Participants

Our study included data from two neuroimaging centers (see 2.3, Neuroimaging) and three developmental cognitive neuroimaging studies, a total 680 runs from 77 children and 623 runs from 64 adults. Some of the data has been used to answer questions concerning the neurocognitive visual and social development in children and adults previously (Meissner et al., 2019; Nordt et al., 2018; Walbrin et al., 2019). The final analyzed data set was reduced to 626 runs from 77 children and 469 runs from 54 adults due to several exclusion criteria (see 2.5, Data exclusion). Children's age ranged from $6.78 - 13.01$ years ($M = 9.52$, $SD = 1.68$), adult's age ranged from $18.41 - 35.02$ years ($M = 21.76$, $SD = 2.67$). All participants were healthy, had normal or corrected-to-normal vision, and had been born at term. No participant reported past or current neurological or psychiatric conditions, or had structural brain abnormalities.

2.3 Neuroimaging

139

140   fMRI took place at the Neuroimaging Centre of the Research Department of Neuroscience at

141   Ruhr University Bochum and the Bangor Imaging Centre at Bangor University. At both sites, 3.0T

142   Achieva scanners (Philips, Amsterdam, The Netherlands) and 32-channel head coils were used

143   (Supplementary Figure S1). For all fMRI, we used blood oxygen level dependent (BOLD) sensitive

144   T2*-weighted sequences. Across sites and experiments, the following fMRI acquisition parameters

145   were constant: FOV = 240 mm × 240 mm, matrix size = 80 × 80, voxel size = 3 mm × 3 mm × 3 mm,

146   TR = 2000 ms, TE = 30 ms. However, slice gap (Bangor: 0 mm, Bochum: 0.4 mm), number of slices

147   (Bangor: 32, Bochum: 33), slice scan order (Bangor: ascending, Bochum: ascending interleaved), and

148   flip angle (Bangor: 83°, Bochum: 90°) differed between sites. Visual stimuli were presented via a

149   VisuaStim Digital goggle system (FOV: 30° × 24°, 800 × 600 pixel, Resonance Technology Inc., CA,

150   USA) in Bochum and via a MR-safe monitor and mirror system (FOV: 25.92° × 16.90°, 1920 × 1200

151   pixel; Cambridge Research Systems, UK) in Bangor. Due to the visual isolation inherent with the

152   goggle system, a researcher was present next to the scanner bore opening at all times for immediate

153   verbal contact and motion feedback in Bochum, but not in Bangor.

154   For each study, a certain number of task runs and structural scans were planned. In adults,

155   experimenters followed this protocol, checking in with participants after the completion of each

156   experiment (i.e. after multiple runs) to explain what would happen next and only stopped or diverted

157   from protocol if participants actively reported feeling unwell. In contrast, in children, we checked in

158   with participants after each run to actively inquire about their well-being. Moreover, nearing the end of

159   a session, we also actively asked if they still felt good and ready to do another run. This was done in

160   order to give children the opportunity to express any signs of discomfort, which to our experience

161   children do not necessarily utter spontaneously but sometimes only after encouraging them to be open

162   about it or giving them an explicit opportunity. Thus, we dynamically decided when to break up

163   acquisition into sessions or days, when to intersperse tasks with a structural scan accompanied by an

164   entertaining video, or when to end the study.

7

2.4   Scanner training

All children underwent a scanner training in order to familiarize them with the scanner environment and achieve high-quality scans with as little motion as possible (Supplementary Figure S1). At both scanner sites, pre-recorded audio of MRI acquisition sequences was played during scanner training to simulate the real scanner environment and visual stimulation was achieved through a mirror system targeted at a monitor outside of the mock scanner bore.

In other aspects, training sessions differed between sites. In Bangor, scanner training was conducted right before scanning and entailed lying inside a mock scanner with a motion sensitive electrode placed on the forehead to measure movement across three translation and three rotation axes (MoTrak Head Motion Tracking System; Psychology Software Tools, 2017). Children received visual motion feedback via an on-screen cursor that was controlled by children's head movements. Children were instructed to lie still and keep the cursor in the middle of a target circle, whose diameter allowed for 3 mm head translations in any direction. Once children were able to keep the cursor within this target region for a timed period of 30 seconds, children watched a short animated video. When movement exceeded 3 mm translation, video playback was paused, providing immediate feedback that they had moved too much. Once children were able to watch the video for a period of 2 minutes without a video pause, scanner training was completed.

In Bochum, scanner training was conducted between one to ten days before the MRI study. Training began with sitting on the extended mock scanner bed, watching a short animated video that introduced a cover-story explaining the tasks to be performed inside the scanner, and practicing these tasks (button-press). This was followed by explaining and demonstrating the procedure of entering an MRI scanner with a large puppet. Next, children entered the mock scanner bore, watched a short animated video and performed several practice trials of the tasks presented on a screen that was visible via a mirror system. The researchers gave feedback with respect to the level of movement and task performance throughout the scanner training session. It was established that a researcher would gently touch the children's shin in case of excessive motion as a means of motion feedback. Verbal feedback

8

191   was gradually decreased throughout the session. Once the researcher decided that motion levels were

192   acceptable and the tasks were understood, scanner training was completed.

2.5 Data exclusion

194   Our initial dataset encompassed all completed fMRI runs of two developmental cognitive

195   neuroscience studies conducted in Bochum and one in Bangor. Aborted runs (e.g. due to technical

196   errors) were not included, as they were not retrieved from the MRI-controlling computer and not

197   reliably recorded in handwritten notes. To ensure that our results are valid and interpretable, we

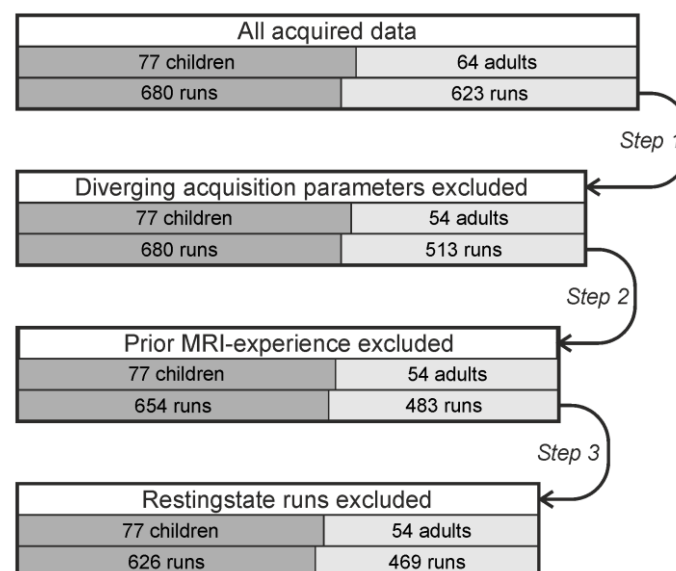198   excluded data that would have biased our analyses (Figure 1).

Figure 1: Data exclusion steps and resulting sample sizes for children (left, dark gray) and adults (right, light gray). Main box widths are in relation to the original total number of runs. Box segmentation mirrors the relative distribution of adult vs. child participants or runs, respectively.

199   In a first step, we excluded ten participants (110 runs) from Bangor that were recorded with

200   different acquisition parameters, i.e. with a slice thickness of 3.5 mm instead of 3.0 mm. This

201   difference in voxel size along the z-axis would affect three out of the six motion estimation parameters

202   (rotation around the y-axis, rotation around the x-axis, and translation along the z-axis) during motion

203   correction, leading to systematically larger motion estimates. In a second step, we excluded 56 runs

204   from four participants that were known to have participated in an earlier MRI study in Bochum. This

205   was to control for possible confounding effects of prior scanning experience on motion. Implementing

206   this control step ensured that we report on the rather typical case—at least for children—of a first-time

207   MRI study participant. Note that we could not rule out the possibility that participants took part in an

208   MRI study of a different lab before, although this is highly unlikely for children and most adults at our

209   neuroimaging centers. In a last step, we excluded 42 resting-state runs. Resting-state runs were only

210   acquired in one of the three studies, only once per participant, and in the majority of cases the run was

211   acquired at the end of a session. Moreover, resting-state runs were the only runs in which participants

212   had their eyes closed during acquisition. This unique set of circumstances makes resting-state runs

213   very likely to bias our results, as there is not enough data to efficiently model this set of circumstances,

214   e.g. as an independent variable.

2.6   Head motion estimation

216   To estimate head motion during fMRI scans, we used the neuroimaging software package

217   BrainVoyager (Version 20.2 for 64bit Windows, RRID: SCR_013057). First, fMRI run series in

218   DICOM format were converted to the proprietary STC and FMR format using BrainVoyager scripts.

219   Then, we applied BrainVoyager fMRI data preprocessing tools in their default settings. That is, for

220   slice scan time correction, we used cubic spline interpolation. For 3D motion correction, we used a

221   trilinear detection and sinc interpolation approach with a maximum of 100 iterations to fit a volume to

222   the reference volume. Resulting motion log files contained six timeseries representing the estimated

223   volume-wise instantaneous translation and rotation for axes x, y, and z in reference to the first volume

224   of the respective experiment.

225   For each of the six motion parameters, we calculated a timeseries of volume-to-volume

226   motion, i.e. the difference between each volume's motion parameter and the previous volume's motion

227   parameter. Rotational motion was converted from degree to millimeter by calculating displacement on

228   the surface of a sphere of 50 mm radius (approximating the distance from the center of the head to the

229   cortex; Power et al., 2012). The framewise displacement of the head position (FD) was calculated as

230   the sum of the absolute volume-to-volume motion values (Power et al., 2012). FD was shown to have

231   a strong association with motion-induced artifacts (Ciric et al., 2017). For each run, we calculated the

232   mean FD.

233   Moreover, for discrete one-minute-intervals (i.e. 30 volumes) within runs, we calculated the

234   percentage of high-motion volumes with FD above the threshold of 0.3 mm (volume-to-volume). This

235   threshold for high-motion volumes was determined by extensive exploration of our data before any

236   statistical analysis, as done previously (Power et al., 2019; Power et al., 2012) and resulted in plausible

237   ratios of high-motion volumes for children (M = 12.80 %, SD = 20.89 %) and adults (M = 4.48 %, SD

238   = 12.10 %). The two head motion measures—mean FD per run and frequency of high-motion volumes

239   per minute of a run—were investigated in separate analyses.

2.7   Predictor variables

241   Our study did not aim to examine age group differences between children and adults, but to

242   provide guidelines for practitioners and researchers who conduct pediatric neuroimaging examinations

243   and/or experiments that might also include adult control groups. Thus, we report separate analyses for

244   children and adults. This approach enabled us to include two more predictor variables (PVs) for

245   children, as adults did not perform scanner training and we did not expect age effects in our range of

246   18-35-year-old adults. For each age group, we assessed two head motion measures with up to eight

247   possible predictors.

2.7.1   Mean FD

249   In our first analysis, we asked which factors would influence the mean motion during a run.

250   Seven PVs were evaluated:

251   - PVs 1-4) *Prior scan time* encoded the time in minutes that a participant had been in

252   the scanner already. That is, the summed scan time of all functional as well as

253   structural runs that had been administered 1) since the beginning of the *functional*

254   *segment*, 2) since the beginning of the *session*, 3) since the beginning of the *day*, or 4)

255   since the beginning of the *study*. (For definitions of *session* and *functional segment*,

256   please refer to section 2.1, Definition of session, run, and functional segment.)

257   For example, imagine the complex case of a participant that comes in for a second day

258   of fMRI scanning (after 60 min. of scanning on the first day), takes a bathroom break

11

259   after the completion of the day's first fMRI experiment (after 15 min.), re-enters the

260   scanner, completes the second fMRI experiment (15 min.), and is now scheduled for

261   an inside-scanner break, during which a diffusion-weighed imaging sequence is being

262   acquired while a video clip is presented (5 min.). The next fMRI run that is acquired

263   would have different values for each of the *prior scan time variables*, e.g. a) *prior*

264   *functional segment scan time* = 0 min., b) *prior session scan time* = 20 min., c) *prior*

265   *day scan time* = 35 min., and d) *prior study scan time* = 95 min.

266   In general, we hypothesized that participants' motion would increase with time spent

267   in the scanner, but could be "reset" to lower motion by breaks. Thus, we tested the

268   effect of four different *prior scan time* variables to determine which—if any—way of

269   breaking up data acquisition reduces subsequent motion. Figure 2 visualizes the

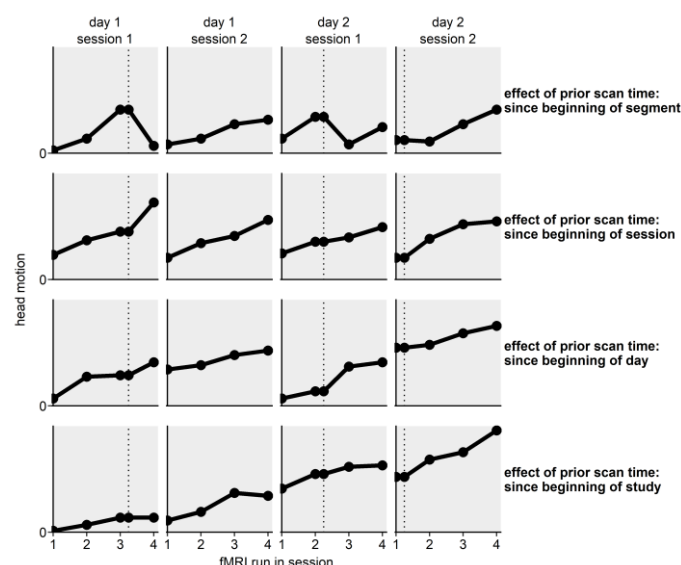270   possible main effects of *prior scan time* variables.

Figure 2: Visualization of the effect of prior scan time with hypothetical data. Each row shows the effect of a different kind of prior scan time. Each column shows a single session, consisting of four fMRI runs. The dotted vertical line and points on the vertical line denote inside scanner breaks during which structural scans were acquired and video clips were presented.

271   To determine the *prior scan time* variables values, we evaluated structured

272   handwritten MRI session notes that stated the sequence of acquired functional and

273   structural runs and breaks between sessions and days for each participant. Then, the

274   duration of each run was retrieved from its DICOM file header. This DICOM header

275   duration always exceeded the product of repetition time and number of volumes, i.e. it

12

276   included scanner- and sequence-specific preparation time. We did not include

277   potential breaks between runs that occur due to normal operation time that is required

278   to start a new paradigm on the stimulation computer or to start the scanner, how-do-

279   you-feel-inquiries, or minor technical issues, because we did not log these events.

280   Further, we did not include aborted runs, because we did not save this data nor took

281   reliable notes, rendering a determination of the duration of these scans impossible.

282   • PV 5) *Age* was defined in years at the first day of scanning and was only investigated

283   for children. This variable was included to explore possible interactions with the other

284   PVs, not for main effects of age, as main effects of age are well established in the field

285   and would only lead to an unnecessarily complex model with less power to detect

286   effects of interest.

287   • PV 6) *Scanner training date* was recorded as the number of days that passed since the

288   scanner training for children only. We hypothesized that a greater time interval

289   between scanner training and actual MRI scan would result in higher motion.

290   • PV 7) *Task engagement* encoded how much active engagement a given fMRI run

291   required. We distinguished if participants just had to passively watch the display, or if

292   participants had to perform a task that included button-pressing. We expected that an

293   active task engagement would result in reduced motion due to enhanced attention and

294   less awareness of a possibly uncomfortable situation, itching, or other distractions.

295   In addition to these main effects, we investigated the possible interactions between the *prior*

296   *scan time* variables and the other variables. Other two-way or three-way interactions were not

297   investigated to avoid overly complex and highly computation-intensive models and as the detection of

298   high-order effects would require a more powerful study design (Simonsohn, 2014).

299   *2.7.2   High-motion volumes*

300   In our second analysis we asked which variables would influence head motion over the course

301   of a run. As decisions on retaining or discarding runs for further analysis are often informed by high-

302   motion volumes within a run rather than a run's mean motion, instead of mean FD, we investigated the

13

303   percentage of high-motion volumes for each minute of a run (i.e. $FD_{volume} > 0.3$ mm; see 2.6., Head

304   motion estimation). Consequently, eight PVs were evaluated:

305   • PVs 1-7) All previous PVs, i.e. *prior functional segment scan time, prior session scan*

306   *time, prior day scan time, prior study scan time, age, scanner training date, task*

307   *engagement*

308   • PV 8) *Minute of run* coded discrete one-minute-intervals within runs, allowing us to

309   investigate the course of high-volume motion occurrence across the course of a run.

310   In addition to main effects, we investigated the possible interactions between *minute of run*

311   and the other variables. Again, we did not investigate other two- or three-way interactions (see 2.7.1,

312   Mean FD)

313   ## 2.8   Statistical model

314   *2.8.1   Hierarchical data structure*

315   We analyzed our data using multilevel linear models (MLMs) due to the hierarchical four- or

316   five-level grouping structure inherent in our data (cf. Engelhardt et al., 2017; Figure 3): Motion

317   estimations for mean FD per run are nested within sessions (level 4), sessions are nested within days

318   (level 3), days are nested within participants (level 2), and participants are nested within studies (level

319   1). Motion estimations for percentage of high-motion volumes per minute of a run are nested within

320   runs, establishing a fifth level. This hierarchical data structure is likely to introduce dependency of
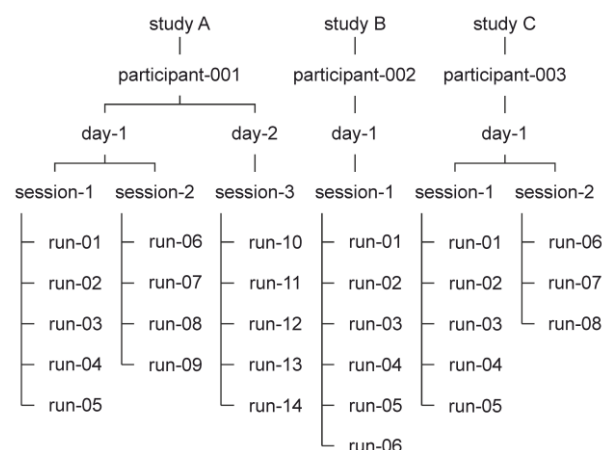

Figure 3: Hierarchical data structure with four levels. For visualization purposes, only one fictional subject is displayed for each study.

321 observations within a grouping variable. For example, motion from two randomly selected runs from

322 the same participant is likely to be much more similar than motion from different participants. In

323 technical terms, the residuals within a grouping variable are likely to be correlated. Thus, the

324 assumption of independent residuals, crucial for parametric tests, is violated. While dependent

325 observations for one level, e.g. within participants, could be handled easily by a conventional

326 repeated-measures ANOVA, this case is a complex nested multilevel dependency structure that cannot

327 be adequately and validly handled by a repeated-measures ANOVA—but can be by an MLM.

328 *2.8.2   MLM creation step 1: Identifying relevant grouping structures*

329   MLMs were created in a data-driven process. First, we assessed the possibility that the

330 grouping variables would introduce dependencies in the data—and thus confirm the need for an MLM.

331 To this end, we calculated the intraclass correlation (ICC, i.e. the proportion of the total variance that

332 is explained by the respective grouping factor) for each grouping level of the model and liberally

333 incorporated all grouping levels into our model that would explain at least 1% of the total variance

334 (i.e. ICC ≥ 0.01, Supplementary Text S1). Next, we used the chi-square likelihood ratio test and

335 Akaike's information criterion (AIC) to test if the model that incorporated the grouping structure

336 actually had a better fit to our data than the model without the grouping structure and included or

337 ignored the grouping structure in all subsequent models accordingly (Supplementary Text S2).

338   ICC analysis for mean FD over the course of a study indicated that in both children and adults,

339 the grouping factors *participant* (children: 0.487, adults: 0.376) and *session* (children: 0.229, adults:

340 0.021) explained a substantial amount of variance in the data, while the grouping factors *study* and *day*

341 did not (both age groups < 0.001). In both children and adults, models that allowed random intercepts

342 for *participant* and *session* fit the data significantly better than a model with fixed intercepts (children:

343 Supplementary Table S3, adults: Supplementary Table S4). Thus, the MLMs were built with random

344 intercepts for *participant* and *session*.

345   ICCs for the percentage of high-motion volumes across the course of a run in both children

346 and adults revealed substantial dependency within the grouping levels *participant* (children: 0.136,

347 adults: 0.292), *session* (children: 0.275, adults: 0.037), and *run* (children: 0.143, adults: 0.309), but not

15

348   within the grouping levels *study* (both age groups: < 0.001) and *day* (children: < 0.001, adults: 0.004).

349   A random intercept model for *participant*, *session*, and *run* fit our data better that a fixed intercept

350   model (children: Supplementary Table S5, adults: Supplementary Table S6). Consequentially, the

351   MLMs were built with random intercepts for *participant*, *session* and *run*.

352   *2.8.3   MLM creation step 2: Stepwise inclusion of relevant predictors*

353   Second, we introduced the PVs, i.e. main effects and selected interactions (fixed effects), into

354   the model stepwise, to test if they improved the model fit significantly. The order of introduction for

355   mean FD and high-motion volumes MLMs is listed in Supplementary Table S1 and Supplementary

356   Table S2, respectively. If a model including a fixed predictor led to a better model fit than the previous

357   model without it, the predictor was included in all subsequent models, otherwise it was left out of

358   subsequent models.

359   For children's mean FD MLM main effects of *prior functional segment time*, *prior session*

360   *time*, *prior day time,* and *scanner training date*, as well as the interaction effects of *age × prior session*

361   *time* and *age × prior study time* were included in the model (Supplementary Table S3). For children's

362   high-motion volumes MLM, main effects of *minute of run*, *prior segment time*, *prior session time*,

363   *prior day time*, and *scanner training date* were added to the model (Supplementary Table S5). For

364   adult's mean FD MLM, main effects of *prior functional segment time*, *prior day time*, and a *prior*

365   *functional segment time × task engagement* interaction were added to the model (Supplementary Table

366   S4). For adult's high-motion volumes MLM main effects of *minute of run*, *prior segment time*, and

367   *prior session time*, were added to the model (Supplementary Table S6).

368   *2.8.4   MLM creation step 3: Assessing the need for participant-specific predictor effects*

369   Third, we tested if the model fit improved further if we let the effect of each included main

370   effect predictor vary over participants (random slopes) and added significant random effects to the

371   model accordingly. Predictors *scanner training date* and *age* were not considered as they are constant

372   for each participant's run. Further, we did not fit random slopes for interaction terms as this prevented

373   the MLMs to converge—possibly because our data did not have enough power to fit a model of such

374   high level of complexity.

16

375   Adding random slopes across participants for all main effects further improved model fit in all

376   MLMs and thus was incorporated in the final models (Supplementary Table S3, S4, S5, S6).

377   **2.9   Software**

378   Data handling and statistical data analysis was performed using R (version 3.6.0, RRID:

379   SCR_001905, R Core Team, 2015) in RStudio (version 1.2.1335; RRID: SCR_000432). For MLMs,

380   we used the nlme package (version 3.1-140, RRID:SCR_015655, Pinheiro et al., 2019). Figures were

381   created using the ggplot2 package (version 3.2.0, RRID:SCR_014601, Wickham, 2016). Regression

382   lines and confidence interval bands for fixed effects visualizations, i.e. predictor effect plots, were

383   created using the effects package (version 4.1-1, Fox and Weisberg, 2018). Marginal R^2 (for fixed

384   effects) and conditional R^2 (for fixed effects and random effects combined) goodness-of-fit was

385   estimated using the r.squaredGLMM function of the MuMIn package (version 1.43.6, Barton, 2019),

386   which is based on methods by Nakagawa and Schielzeth, 2013, Johnson, 2014, and Nakagawa et al.,

387   2017.

388

17

3 Results

389

390   We evaluated which variables would predict head motion during fMRI in children and adults.

391   We assessed two measures of head motion in separate analyses: First, we investigated the course of

392   head motion across an fMRI study in terms of mean framewise displacement per run. Then, we

393   investigated the course of head motion over the course of a run in terms of the frequency of high-

394   motion volumes per minute. Each analysis was performed separately for children and adults.

395   Regression coefficients for all main and interaction effects included in the MLMs were tested against

396   zero using one-sample t-tests with a significance threshold of $\alpha = 0.05$.

3.1 Splitting up data acquisition reduces motion
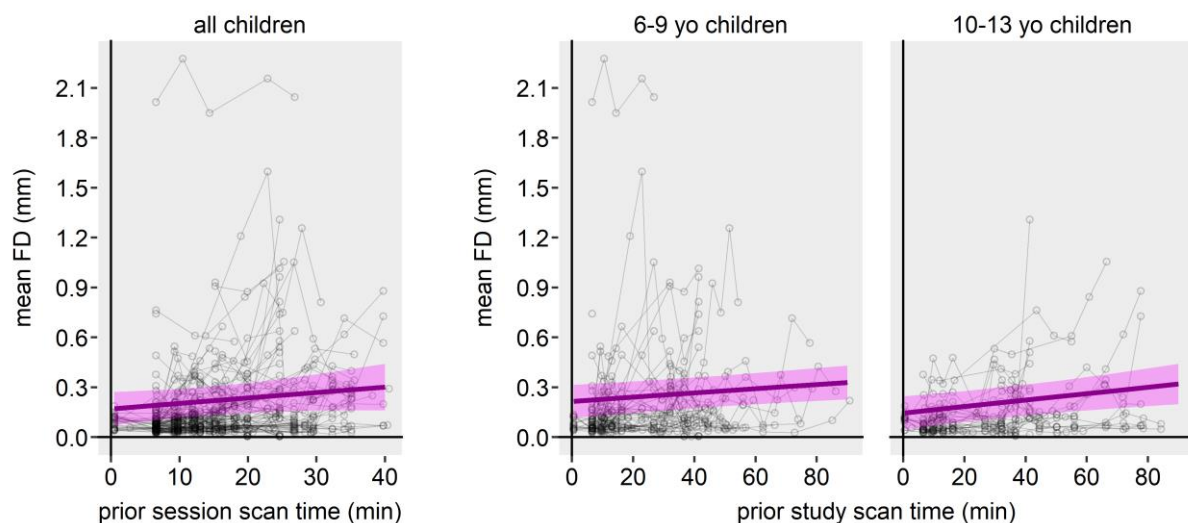
397

3.1.1 Children

398

Figure 4: Predictor effect plots for children's head motion across the course of a study.Thick magenta lines are fitted linear regression lines for the fixed effects predictors denoted on the x-axes. Shaded magenta areas show the 95% confidence band for the regression lines. Each gray circle represents one run. Left: Head motion across the course of a session. Each circle-connecting line represents a participant's session. Participants that were scanned in multiple sessions are represented by multiple lines. Right: Head motion across the course of a study for 6-9-year-old and 10-13-year old children. Age bins were chosen arbitrarily to visualize the age × prior study scan time interaction. Each circle-connecting line represents one participant.

399   Evaluation of fixed effects predictors in the final mean FD MLM revealed that *prior session*

400   *scan time* and a *prior study scan time* × *age* interaction significantly predicted FD in children (Figure

401   4, Table 1). Children showed greater motion with ongoing session length (Figure 4, left). In addition,

18

402   motion increased with study length in older children, but not as much in younger children (Figure 4,

403   right).

Table 1. Fixed effects parameter estimates of final model for children's mean motion across the course of a study.

parameter	β	Lower 95% CI	Upper 95% CI	SE	df	t	p
Intercept	0.05193	0.01980	0.08406	0.01645	509	3.158	.002
Prior functional segment scan time	0.00139	-0.00200	0.00478	0.00173	509	0.801	.424
Prior session scan time	0.01412	0.00220	0.02603	0.00610	509	2.314	.021
Prior day scan time	0.00240	-0.00026	0.00507	0.00136	509	1.761	.079
Scanner training date	0.00532	-0.01831	0.02896	0.01210	509	0.440	.660
Age × prior session scan time	-0.00111	-0.00228	0.00005	0.00060	509	-1.864	.063
Age × prior study scan time	0.00017	0.00005	0.00029	0.00006	509	2.785	.006

Note: CI = confidence interval, SE = standard error, df = degrees of freedom
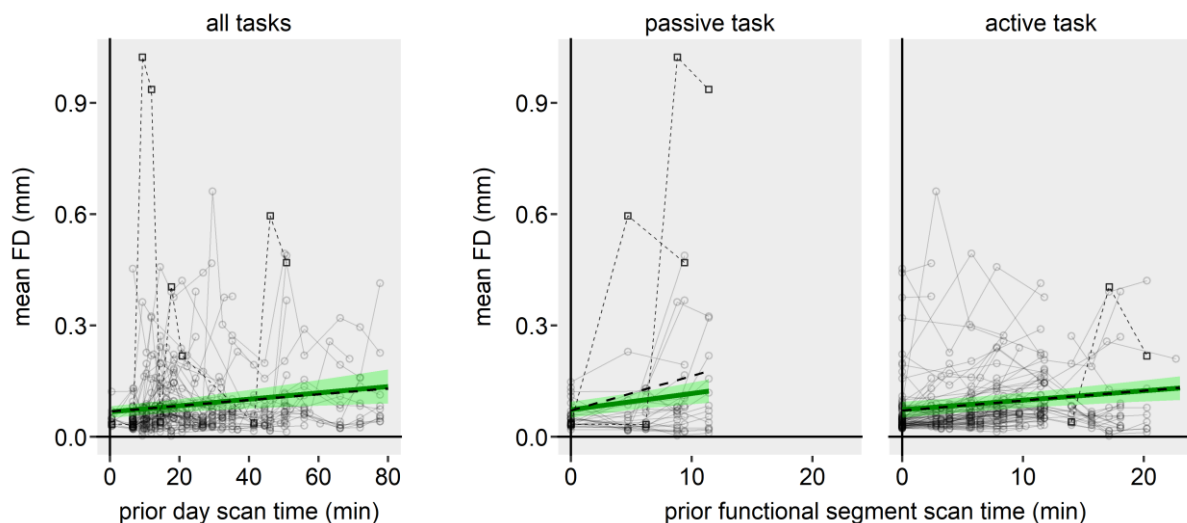
404   *3.1.2   Adults*

Figure 5: Predictor effect plots for adult's head motion across the course of a study. Thick green lines are fitted linear regression lines for the fixed effects predictors denoted on the x-axes. Shaded green areas show the 95% confidence band for the regression lines. Each gray circle represents one run. Left: Head motion across the course of a day. Each circle-connecting line represents a participant's day. Participants that were scanned on multiple days are represented by multiple lines. Right: Head motion across the course of a functional segment for passive and active fMRI tasks. Each circle-connecting line represents one participant's functional segment data acquisition course. Participants that were scanned on multiple days, on multiple sessions, or with multiple functional segments are represented by multiple lines. Black dashed lines and squares show the participant that drove the task engagement x prior functional segment time interaction.

405   For adults, evaluation of fixed effects predictors in the final model revealed that a *prior*

406   *functional segment scan time* and a *prior day scan time* effect significantly predicted motion (Figure 5,

407   Table 2). Adults' motion increased with ongoing *day scan time*. Moreover, the longer functional scans

408   are acquired without an inside-scanner break (e.g. through a structural scan with relaxing video), the

409   more adults moved. Originally, we also found that a *task engagement × prior functional segment scan*

410   *time* interaction significantly predicted motion, i.e. for passive tasks, adults showed a steeper increase

411   in motion than for active tasks (Figure 5, middle, thick black dashed line). However, as it seemed

19

412   possible that this *task engagement × prior functional segment scan time* interaction was driven by an

413   extreme increase in motion at the end of a functional segment in passive tasks in a single participant

414   only (Figure 5, thin black dashed lines and squares), we reanalyzed the MLM without data from this

415   participant. While all other effects remained stable (Figure 5, see overlapping green solid and black

416   dashed lines), the *task engagement × prior functional segment scan time* interaction was rendered

417   insignificant by the exclusion of the participant (Figure 5, see black dashed line outside of green

418   confidence band, Supplementary Table S7).

Table 2. Fixed effects parameter estimates of final model for adult's mean motion across the course of a study.

parameter	β	Lower 95% CI	Upper 95% CI	SE	df	t	p
Intercept	0.05055	0.03355	0.06755	0.00868	376	5.822	< .001
Prior functional segment scan time	0.00253	0.00101	0.00405	0.00078	376	3.258	.001
Prior day scan time	0.00086	0.00028	0.00144	0.0003	376	2.896	.004
Task engagement × Prior functional segment scan time	0.00183	-0.0006	0.00426	0.00124	376	1.476	.141

Note: CI = confidence interval, SE = standard error, df = degrees of freedom

419   *3.1.3   Summary*

420   Our results demonstrate that head motion in children and adults can be reduced by splitting up

421   data acquisition. However, depending on the age group, different strategies seem to be effective. For

422   children, our data suggests a benefit of splitting up data acquisition into multiple, short sessions on the

423   same day and keeping the overall study length as short as possible, especially for older children. In

424   contrast, adults benefited from splitting up data acquisition into multiple days and interspersing

425   experimental runs with inside-scanner breaks, especially for passive tasks.

3.2 High-motion event occurrence increases with run length
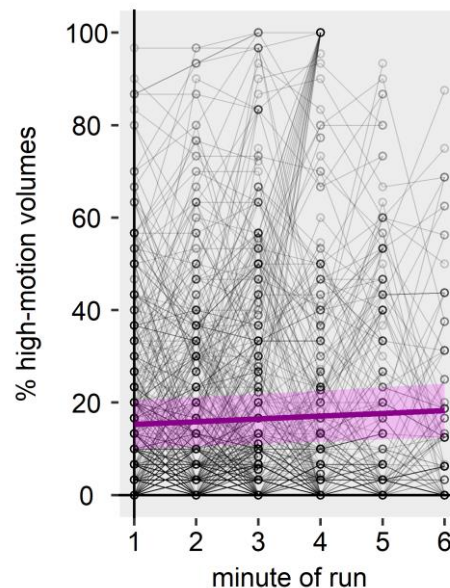
3.2.1 Children

Figure 6: Children's head motion across the course of a run. Each gray circle represents one minute bin. Each circle-connecting line represents a participant's run. Participants that were scanned for more than one functional run are represented by multiple lines. Thick magenta line shows the fitted linear regression line for the minute of run effect. Shaded magenta area show the 95% confidence band for the regression line.

Minute of run significantly predicted motion in children, i.e. children's motion increased with increasing run length (Figure 6, Table 3). For example, in the third minute of a run, the average risk of a high-motion volume was at 16.5 %, while it increased to 18.3 % in the sixth minute of a run. *Prior session scan time* was also a significant predictor of motion, but its effect is not readily interpretable as for every value of *prior session scan time*, several values of *minute of run* existed. Thus, we were only interested in the main effect of *minute of run* and its possible interaction effects.

Table 3. Fixed effects parameter estimates of final model for children's frequency of motion peaks across the course of a run

parameter	β	Lower 95% CI	Upper 95% CI	SE	df	t	p
Intercept	1.13974	-2.87403	5.1535	2.04910	2290	0.556	.578
Minute of run	0.60064	0.15445	1.04683	0.22779	2290	2.637	.008
Prior functional segment scan time	0.22457	-0.01326	0.4624	0.12142	2290	1.85	.064
Prior session scan time	0.09011	-0.17798	0.3582	0.13686	2290	0.658	.510
Prior day scan time	0.40927	0.21623	0.60231	0.09855	2290	4.153	< .001
Scanner training date	0.79343	-0.74224	2.32909	0.78398	2290	1.012	.312

Note: CI = confidence interval, SE = standard error, df = degrees of freedom

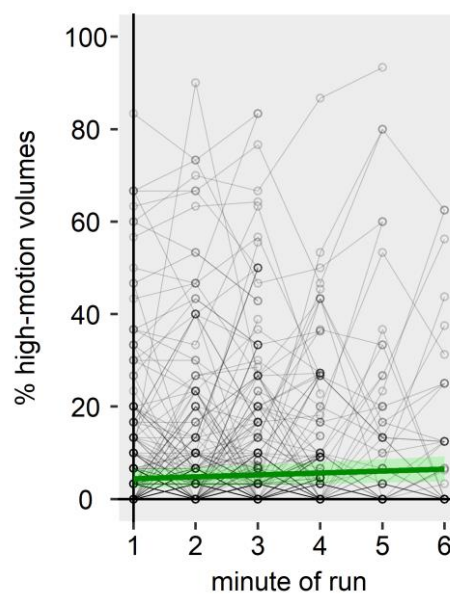21

434   *3.2.2   Adults*

Figure 7: Adult's head motion across the course of a run. Each gray circle represents a one minute bin. Each circle-connecting line represents a participant's run. Participants that were scanned for more than one functional run are represented by multiple lines. Thick green line shows the fitted linear regression line for the minute of run effect. Shaded green area show the 95% confidence band for the regression line.

435   *Minute of run* significantly predicted motion in adults, i.e. adult's motion increased with

436   increasing run length (Figure 7, Table 4). For example, in the third minute of a run, the average risk of

437   a high-motion volume was at 5.3 %, while it increased to 6.5 % in the sixth minute of a run. *Prior*

438   *functional segment scan time* and *prior session* scan *time* were also significant predictors of motion,

439   but their effects are not interpretable as for each of their values, several values of *minute of run*

440   existed. Thus, we were only interested in the main effect of minute of run and its possible interaction

441   effects.

Table 4. Fixed effects parameter estimates of final model for adult's frequency of motion peaks across the course of a run

parameter	β	Lower 95% CI	Upper 95% CI	SE	df	t	*p*
Intercept	0.76618	-1.51000	3.04235	1.16168	1638	0.660	0.510
Minute of run	0.42857	0.10352	0.75362	0.16589	1638	2.583	0.010
Prior functional segment scan time	0.18683	0.00786	0.36579	0.09134	1638	2.045	0.041
Prior session scan time	0.11664	0.04528	0.18800	0.03642	1638	3.203	0.001

Note: CI = confidence interval, SE = standard error, df = degrees of freedom

442   *3.2.3   Summary*

443   Our results indicate that long runs have a negative impact on data quality: The frequency of

444   high-motion events increases with ongoing run length for both children and adults.

445

4   Discussion

446

447   We identified study procedure variables, that predict participant's head motion in three

448   neurodevelopmental fMRI studies including data of 77 children and 64 adults. Using MLMs, we

449   investigated the effect of scanner training date, task engagement, as well as *prior scan time* since the

450   beginning of the *study*, *day*, *session*, *or functional segment*. In children, splitting fMRI data acquisition

451   into multiple sessions reduced motion. However, motion still increased across the course of a study,

452   especially in older children. In adults, splitting fMRI data acquisition into multiple days reduced

453   motion. Moreover, motion was reduced after task-free inside-scanner breaks. In both children and

454   adults, motion increased with run length.

4.1   Splitting up data acquisition reduces motion

455

4.1.1   Children

456

457   Children do not seem to benefit from splitting up studies into several days, as *prior study scan*

458   *time* predicted head motion, especially for older children, but *prior day scan time* did not. To speculate

459   on possible explanations, for children, the initial excitement of participating in an fMRI study and the

460   commitment to do everything right might be less pronounced after the first day. On the second day,

461   they might be less motivated to lie still, e.g. due to the lacking novelty of the situation. Possibly, the

462   similarity of visual appearance or demands of the tasks lead to boredom and more movement on the

463   second day. In this regard, it might be interesting to investigate if a longer period between scanning

464   days, such as four weeks instead of the usual few days, would lead to less motion on day two.

465   Older children showed a larger increase in motion with increasing study length. Initially, this

466   seems unexpected. Intuition would suggest that older children are better at lying still for extended

467   periods of time. However, it seems that this *age × prior study scan time* interaction could be driven by

468   a higher baseline motion of young children at the beginning of the study, which then does not change

469   as much over time. In contrast, older children seem to be able to suppress the urge to move at the

470   beginning of a study quite well, and then gradually relax into motion levels similar to that of younger

471   children.

472   Our results indicate that pediatric neuroimaging studies may benefit from breaking up data

473   acquisition into several sessions on the same day, as children's head motion increases with ongoing

474   session length. So far, no other studies have investigated motion across multiple sessions on the same

475   day. Nevertheless, single-session studies in children have found that motion increases with increasing

476   session length in children (Achterberg and van der Meulen, 2019; Engelhardt et al., 2017). Our study

477   points to an important distinction regarding the kind of breaks that researchers can schedule in

478   children's fMRI scanning protocols. While motion can be effectively reduced by allowing children to

479   exit the scanner, inside-scanner breaks in fMRI data acquisition during which children cannot exit the

480   scanner, but instead watch a video while a structural scan is acquired, do not seem to reduce motion

481   after the break. Possibly, it is important for children to be able to move freely, get face-to-face social

482   interactions, or just relieve the desire to void their bladder, none of which are possible without exiting

483   the scanner. Alternatively, showing a video might have effects opposite to those desired: Because the

484   video is highly engaging, it leads to reduced motion during the break from tasks (Cantlon and Li,

485   2013; Vanderwal et al., 2015). However, due to the high level of engagement in the video, any

486   subsequent normal fMRI task might be perceived as boring due to a contrast effect, leading to higher

487   motion in turn.

488   In our study, we did not find an effect of task engagement on head motion in children. This

489   contradicts previous findings, which showed that children show less motion in engaging tasks in

490   contrast to less engaging tasks or resting state (Engelhardt et al., 2017). This disparity might be due to

491   different definitions and aims of the studies. In our analysis, we defined tasks as engaging if

492   participants had to perform a task that included button-pressing, and we defined tasks as non-engaging

493   if participants just had to passively watch the display. We chose this definition as our aim was to

494   identify aspects in fMRI study designs that can be actively manipulated by the researcher. Engelhardt

495   et al. (2017) defined engaging tasks as fast-paced, cognitively demanding, or socially engaging. These

496   aspects are usually inherent in a task and cannot be changed without changing the research question.

497   For example, experiments on social development in children will need to have a socially engaging

498   component and experiments on the development of working memory or response inhibition will be

499   cognitively demanding. However, while researchers can decide if a given task will require a manual

500   response or not, other than providing a way to track participant's behavior in the scanner this choice

501   alone does not seem to make a difference in motion one might have hoped for.

502   *4.1.2   Adults*

503   For adults, it seems that breaking up data acquisition into several days is especially useful in

504   reducing motion as we found a significant increase in motion with increasing *prior day scan time*, but

505   not with increasing *prior study scan time*. Moreover, our results suggest that adults' head motion

506   remains constant across the course of a session. Thus, the observed increasing motion across the

507   course of a day is presumably due to a higher mean motion in the second session of the day. Thus,

508   adults seem to be able to restrain their head motion across relatively long sessions, but are more likely

509   to increase head motion in a second session on the same day. Therefore, if lengthy studies require a

510   break, it might be advisable to continue scanning on a different day. We are not aware of studies that

511   investigated motion in adults across multiple sessions within the same day. However, our findings

512   connect with the existing literature in so far as the number of completed runs in a single-session study

513   with five runs did not have any effect on motion in a previous study (Huijbers et al., 2017).

514   One possible reason for adult's ability to maintain motion on a constant level within session

515   might be the introduction of breaks from task-based fMRI with structural scans during which

516   participants can watch a video. Although these breaks do not reduce head motion in children, these

517   inside-scanner breaks seem to reduce motion after the break for adults, an effect that is particularly

518   prominent for passive tasks. Thus, passive tasks in adults should be planned at the beginning of a

519   session, or after an inside-scanner break. These findings complement previous findings that showed

520   higher motion for passive tasks or resting state and lower motion for active tasks (Cantlon and Li,

521   2013; Huijbers et al., 2017; Vanderwal et al., 2015). Here, our findings show that motion during

522   passive tasks is not only higher but also increases at a steeper rate with ongoing time after any break.

4.2   High-motion event occurrence increases with run length

523

524   We found that the frequency of motion peaks within a run increases with run length in both

525   children and adults. This suggests that it is preferable to plan rather short runs instead of longer ones.

526   Interestingly, we did not find interactions of *run minute* with any of the *prior scan time* variables. That

527   is, a common expectation—that motion peaks are especially evident at the beginning of the first run of

528   a study or day—does not hold up. In a previous study that investigated the association between total

529   run length and mean run motion in up to 6-7 minute long runs, results were mixed, i.e. a positive

530   relationship in one sample and a negative relationship in the other sample (Engelhardt et al., 2017).

531   ## 4.3   Limitations

532   While our study marks an important contribution towards understanding which factors are

533   effective in obtaining good quality data from fMRI experiments, some limitations of our methodology

534   should be considered. Data acquisition in terms of which task was acquired when, the positions of

535   structural scans, i.e. inside-scanner breaks, and the number of total runs was largely predetermined for

536   each study and participant. However, of course we adapted our planning to the requirements of the

537   given participant, for the greatest part for children only. Therefore, children were able to influence the

538   *prior study time* factors and possibly their head motion.

539   ## 4.4   Future directions

540   It remains unclear which contribution scanner training has on children's head motion.

541   Considering that a simple toy tunnel and a commercial mock scanner resulted in comparable success

542   rates for high-quality structural images (Barnea-Goraly et al., 2014), it might be interesting to see if

543   these results generalize onto strict motion thresholds during longer fMRI studies. Here, future studies

544   should experimentally manipulate the kind of training for children, e.g. mock scanner training vs. toy

545   tunnel training vs. no training. Also, these studies could investigate if short version trainings have

546   comparable effects to extensive trainings sessions and at which age children cease to benefit from

547   scanner training.

548   Aside from the investigated and discussed procedures, technical solutions might help to reduce

549   motion. Recent developments have made real-time motion detection available that can be used to

550   provide immediate feedback to participants and researchers (Dosenbach et al., 2017). In children

551   younger than 10 years, providing immediate feedback has been successful in reducing motion using

27

552   this technology (Greene et al., 2018). Thus, age effects in motion could occur not due to an inability to

553   lie still in younger children, but possibly due to a lower awareness of their own movements. Even if

554   motion is not reduced, this technology provides the possibility to adapt a common fMRI study

555   protocol—based on average-score recommendations like those in this study—to the individual

556   participant's behavior.

557   This possibility might lead to a considerable reduction in cost and improvement of data

558   quality, as our ICC analysis showed that motion differs substantially between *participants* and is

559   relatively similar within participants (despite weak groupings of *sessions* and *runs*). This finding is in

560   line with a small body of studies (including twin-studies) finding that motion during fMRI tasks is a

561   very stable neurobiological trait in children and adults that seems to be heritable, i.e. under strong

562   genetic control (Achterberg and van der Meulen, 2019; Couvy-Duchesne et al., 2014; Engelhardt et

563   al., 2017; van Dijk et al., 2012; Zeng et al., 2014). More generally, fidgeting in children, measured

564   using a psychometric scale, seems to be under strong genetic control, while objectively-measured

565   physical activity seems to be under strong environmental control (Fisher et al., 2010). However, as the

566   environment of an fMRI study is highly stable, it is not surprising that genetically controlled

567   physiological or cognitive traits (e.g. strong swallowing, restlessness) leads to stable motion within

568   participants.

569   ## 4.5   Conclusion

570   The best way of dealing with head motion in fMRI is to prevent it in the first place. Our study

571   shows that motion can be reduced by careful planning of the data acquisition procedure. Breaking up

572   data acquisition into several sessions with the opportunity to leave the scanner is effective in reducing

573   motion in children, while introducing inside-scanner breaks with continued acquisition of structural

574   data is not effective. Thus, for children, it appears feasible to acquire all fMRI data of a study on one

575   day, with no necessity of scanning on separate days. For adults, it is best to acquire data in long

576   sessions that might be interspersed with inside-scanner breaks during which further structural data can

577   be acquired. Moreover, lowest motion in adults will be achieved if only one session per day is

578   planned; i.e. data for extensive studies should be acquired on separate days. For both children and

579   adults, we advise to avoid long runs, as the probability for high-motion peaks increases with ongoing

580   run length.

581

582 # Data and code availability statement

583   Data analysis code as well anonymized motion, demographic, and study procedure data is

584 publicly available at the Open Science Framework (https://doi.org/10.17605/osf.io/jhm5t).

585

586 # Ethics statement

587   The of Ruhr University Bochum Faculty of Psychology and the University of Bangor School

588 of Psychology ethics boards approved the studies, in which the data was acquired (Bochum: proposal

589 no. 167 and 280, Bangor: proposal no. 2016-15743). All participants as well as children's parents gave

590 informed and written consent to participate voluntarily.

591

592 # Conflict of interest statement

593   The authors report no conflict of interest.

594

595 # Acknowledgements

609

Author contributions

611   Conceptualization: TWM; Methodology: TWM; Formal Analysis: TWM; Investigation:

612   TWM, JW, MN; Resources: SW, KK; Data Curation: TWM, JW, MN; Writing—Original Draft:

613   TWM; Writing—Review and Editing: TWM, MN, SW, JW, KK; Visualization: TWM; Supervision:

614   SW, KK; Project Administration: TWM; Funding Acquisition: SW, TWM, JW, MN, KK

615

References

616

617 Achterberg, M., van der Meulen, M., 2019. Genetic and environmental influences on MRI scan

618   quantity and quality. Dev Cogn Neurosci 38, 100667. 10.1016/j.dcn.2019.100667.

619 Barnea-Goraly, N., Weinzimer, S.A., Ruedy, K.J., Mauras, N., Beck, R.W., Marzelli, M.J., Mazaika,

620   P.K., Aye, T., White, N.H., Tsalikian, E., Fox, L., Kollman, C., Cheng, P., Reiss, A.L., 2014. High

621   success rates of sedation-free brain MRI scanning in young children using simple subject

622   preparation protocols with and without a commercial mock scanner--the Diabetes Research in

623   Children Network (DirecNet) experience. Pediatr Radiol 44 (2), 181–186. 10.1007/s00247-013-

624   2798-7.

625 Barton, K., 2019. MuMIn: Multi-Model Inference.: R package version 1.43.6. https://CRAN.R-

626   project.org/package=MuMIn.

627 Bie, H.M.A. de, Boersma, M., Wattjes, M.P., Adriaanse, S., Vermeulen, R.J., Oostrom, K.J., Huisman,

628   J., Veltman, D.J., Delemarre-van Waal, H.A. de, 2010. Preparing children with a mock scanner

629   training protocol results in high quality structural and functional MRI scans. Eur J Pediatr 169 (9),

630   1079–1085. 10.1007/s00431-010-1181-z.

631 Cantlon, J.F., Li, R., 2013. Neural activity during natural viewing of Sesame Street statistically

632   predicts test scores in early childhood. PLoS Biol 11 (1), e1001462.

633   10.1371/journal.pbio.1001462.

634 Ciric, R., Wolf, D.H., Power, J.D., Roalf, D.R., Baum, G.L., Ruparel, K., Shinohara, R.T., Elliott,

635   M.A., Eickhoff, S.B., Davatzikos, C., Gur, R.C., Gur, R.E., Bassett, D.S., Satterthwaite, T.D.,

636   2017. Benchmarking of participant-level confound regression strategies for the control of motion

637   artifact in studies of functional connectivity. Neuroimage 154, 174–187.

638   10.1016/j.neuroimage.2017.03.020.

639 Couvy-Duchesne, B., Blokland, G.A.M., Hickie, I.B., Thompson, P.M., Martin, N.G., Zubicaray, G.I.

640   de, McMahon, K.L., Wright, M.J., 2014. Heritability of head motion during resting state functional

641   MRI in 462 healthy twins. Neuroimage 102 Pt 2, 424–434. 10.1016/j.neuroimage.2014.08.010.

642 Dosenbach, N.U.F., Koller, J.M., Earl, E.A., Miranda-Dominguez, O., Klein, R.L., Van, A.N., Snyder,

643   A.Z., Nagel, B.J., Nigg, J.T., Nguyen, A.L., Wesevich, V., Greene, D.J., Fair, D.A., 2017. Real-

644   time motion analytics during brain MRI improve data quality and reduce costs. Neuroimage 161,

645   80–93. 10.1016/j.neuroimage.2017.08.025.

646 Durston, S., Nederveen, H., van Dijk, S., van Belle, J., Zeeuw, P. de, Langen, M., van Dijk, A., 2009.

647   Magnetic resonance simulation is effective in reducing anxiety related to magnetic resonance

648   scanning in children. J Am Acad Child Adolesc Psychiatry 48 (2), 206–207.

649   10.1097/CHI.0b013e3181930673.

650 Engelhardt, L.E., Roe, M.A., Juranek, J., DeMaster, D., Harden, K.P., Tucker-Drob, E.M., Church,

651   J.A., 2017. Children's head motion during fMRI tasks is heritable and stable over time. Dev Cogn

652   Neurosci 25, 58–68. 10.1016/j.dcn.2017.01.011.

653 Epstein, J.N., Casey, B.J., Tonev, S.T., Davidson, M., Reiss, A.L., Garrett, A., Hinshaw, S.P.,

654   Greenhill, L.L., Vitolo, A., Kotler, L.A., Jarrett, M.A., Spicer, J., 2007. Assessment and prevention

655   of head motion during imaging of patients with attention deficit hyperactivity disorder. Psychiatry

656   Res 155 (1), 75–82. 10.1016/j.pscychresns.2006.12.009.

657 Field, A.S., Yen, Y.F., Burdette, J.H., Elster, A.D., 2000. False cerebral activation on BOLD

658   functional MR images: study of low-amplitude motion weakly correlated to stimulus. Am J

659   Neuroradiol 21 (8), 1388–1396.

660 Fisher, A., van Jaarsveld, C.H.M., Llewellyn, C.H., Wardle, J., 2010. Environmental influences on

661   children's physical activity: quantitative estimates using a twin design. PloS one 5 (4), e10110.

662   10.1371/journal.pone.0010110.

663 Fox, J., Weisberg, S., 2018. Visualizing Fit and Lack of Fit in Complex Regression Models with

664   Predictor Effect Plots and Partial Residuals. J Stat Soft 87 (9). 10.18637/jss.v087.i09.

665 Freire, L., Mangin, J.F., 2001. Motion correction algorithms may create spurious brain activations in

666   the absence of subject motion. Neuroimage 14 (3), 709–722. 10.1006/nimg.2001.0869.

667 Friston, K.J., Williams, S., Howard, R., Frackowiak, R.S.J., Turner, R., 1996. Movement-Related

668   effects in fMRI time-series. Magn Reson Med 35 (3), 346–355. 10.1002/mrm.1910350312.

669 Green, M.V., Seidel, J., Stein, S.D., Tedder, T.E., Kempner, K.M., Kertzman, C., Zeffiro, T.A., 1994.

670   Head movement in normal subjects during simulated PET brain imaging with and without head

671   restraint. J Nucl Med 35 (9), 1538–1546.

672 Greene, D.J., Black, K.J., Schlaggar, B.L., 2016. Considerations for MRI study design and

673 implementation in pediatric and clinical populations. Dev Cogn Neurosci 18, 101–112.

674 10.1016/j.dcn.2015.12.005.

675 Greene, D.J., Koller, J.M., Hampton, J.M., Wesevich, V., Van, A.N., Nguyen, A.L., Hoyt, C.R.,

676 McIntyre, L., Earl, E.A., Klein, R.L., Shimony, J.S., Petersen, S.E., Schlaggar, B.L., Fair, D.A.,

677 Dosenbach, N.U.F., 2018. Behavioral interventions for reducing head motion during MRI scans in

678 children. Neuroimage 171, 234–245. 10.1016/j.neuroimage.2018.01.023.

679 Habibi, A., Sarkissian, A.D., Gomez, M., Ilari, B., 2015. Developmental Brain Research With

680 Participants From Underprivileged Communities: Strategies for Recruitment, Participation, and

681 Retention. Mind Brain Educ 9 (3), 179–186. 10.1111/mbe.12087.

682 Huijbers, W., van Dijk, K.R.A., Boenniger, M.M., Stirnberg, R., Breteler, M.M.B., 2017. Less head

683 motion during MRI under task than resting-state conditions. Neuroimage 147, 111–120.

684 10.1016/j.neuroimage.2016.12.002.

685 Johnson, P.C.D., 2014. Extension of Nakagawa & Schielzeth's R2GLMM to random slopes models.

686 Methods Ecol Evol 5 (9), 944–946. 10.1111/2041-210X.12225.

687 Krause, F., Benjamins, C., Eck, J., Lührs, M., van Hoof, R., Goebel, R., 2019. Active head motion

688 reduction in magnetic resonance imaging using tactile feedback. Hum Brain Mapp 40 (14), 4026–

689 4037. 10.1002/hbm.24683.

690 Meissner, T.W., Nordt, M., Weigelt, S., 2019. Prolonged functional development of the

691 parahippocampal place area and occipital place area. Neuroimage 191, 104–115.

692 10.1016/j.neuroimage.2019.02.025.

693 Menon, V., Lim, K.O., Anderson, J.H., Johnson, J., Pfefferbaum, A., 1997. Design and efficacy of a

694 head-coil bite bar for reducing movement-related artifacts during functional MRI scanning. Behav

695 Res Meth Instrum Comput 29 (4), 589–594. 10.3758/BF03210613.

696 Nakagawa, S., Johnson, P.C.D., Schielzeth, H., 2017. The coefficient of determination R2 and intra-

697 class correlation coefficient from generalized linear mixed-effects models revisited and expanded.

698 J R Soc Interface 14 (134). 10.1098/rsif.2017.0213.

699   Nakagawa, S., Schielzeth, H., 2013. A general and simple method for obtaining R2 from generalized

700   linear mixed-effects models. Methods Ecol Evol 4 (2), 133–142. 10.1111/j.2041-

701   210x.2012.00261.x.

702   Nordt, M., Semmelmann, K., Genç, E., Weigelt, S., 2018. Age-related increase of image-invariance in

703   the fusiform face area. Dev Cogn Neurosci 31, 46–57. 10.1016/j.dcn.2018.04.005.

704   Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., R Core Team, 2019. nlme: Linear and Nonlinear

705   Mixed Effects Models: R package version 3.1-140. http://CRAN.R-project.org/package=nlme.

706   Power, J.D., Barnes, K.A., Snyder, A.Z., Schlaggar, B.L., Petersen, S.E., 2012. Spurious but

707   systematic correlations in functional connectivity MRI networks arise from subject motion.

708   Neuroimage 59 (3), 2142–2154. 10.1016/j.neuroimage.2011.10.018.

709   Power, J.D., Silver, B.M., Silverman, M.R., Ajodan, E.L., Bos, D.J., Jones, R.M., 2019. Customized

710   head molds reduce motion during resting state fMRI scans. Neuroimage 189, 141–149.

711   10.1016/j.neuroimage.2019.01.016.

712   R Core Team, 2015. R: A Language and Environment for Statistical Computing. Vienna, Austria.

713   http://www.R-project.org/.

714   Raschle, N.M., Lee, M., Buechler, R., Christodoulou, J.A., Chang, M., Vakil, M., Stering, P.L., Gaab,

715   N., 2009. Making MR imaging child's play - pediatric neuroimaging protocol, guidelines and

716   procedure. J Vis Exp (29). 10.3791/1309.

717   Rosenberg, D.R., Sweeney, J.A., Gillen, J.S., Kim, J., Varanelli, M.J., O'Hearn, K.M., Erb, P.A.,

718   Davis, D., Thulborn, K.R., 1997. Magnetic resonance imaging of children without sedation:

719   preparation with simulation. J Am Acad Child Adolesc Psychiatry 36 (6), 853–859.

720   10.1097/00004583-199706000-00024.

721   Satterthwaite, T.D., Wolf, D.H., Loughead, J., Ruparel, K., Elliott, M.A., Hakonarson, H., Gur, R.C.,

722   Gur, R.E., 2012. Impact of in-scanner head motion on multiple measures of functional

723   connectivity: Relevance for studies of neurodevelopment in youth. Neuroimage 60 (1), 623–632.

724   10.1016/j.neuroimage.2011.12.063.

725   Simonsohn, U., 2014. [17] No-way Interactions. The Winnower. 10.15200/winn.142559.90552.

726   Slifer, K.J., Cataldo, M.F., Cataldo, M.D., Llorente, A.M., Gerson, A.C., 1993. Behavior analysis of

727   motion control for pediatric neuroimaging. J Appl Behav Anal 26 (4), 469–470.

728   10.1901/jaba.1993.26-469.

729   Slifer, K.J., Koontz, K.L., Cataldo, M.F., 2002. Operant-contingency-based preparation of children for

730   functional magnetic resonance imaging. J Appl Behav Anal 35 (2), 191–194.

731   10.1901/jaba.2002.35-191.

732   van Dijk, K.R.A., Sabuncu, M.R., Buckner, R.L., 2012. The influence of head motion on intrinsic

733   functional connectivity MRI. Neuroimage 59 (1), 431–438. 10.1016/j.neuroimage.2011.07.044.

734   Vanderwal, T., Kelly, C., Eilbott, J., Mayes, L.C., Castellanos, F.X., 2015. Inscapes: A movie

735   paradigm to improve compliance in functional magnetic resonance imaging. Neuroimage 122,

736   222–232. 10.1016/j.neuroimage.2015.07.069.

737   Walbrin, J., Mihai, I., Landsiedel, J., Koldewyn, K., 2019. Developmental Changes in Visual

738   Responses to Social Interactions. bioRxiv. 10.1101/800532.

739   Wickham, H., 2016. ggplot2: Elegrant Graphics for Data Analysis. Springer, Switzerland.

740   Zeffiro, T., 1996. Clinical functional image analysis: artifact detection and reduction. Neuroimage 4 (3

741   Pt 3), S95-100. 10.1006/nimg.1996.0059.

742   Zeng, L.-L., Wang, D., Fox, M.D., Sabuncu, M., Hu, D., Ge, M., Buckner, R.L., Liu, H., 2014.

743   Neurobiological basis of head motion in brain imaging. Proc Natl Acad Sci U S A 111 (16), 6058–

744   6062. 10.1073/pnas.1317424111.

745