Published in final edited form as:

J Child Psychol Psychiatry. 2014 June ; 55(6): 681–684. doi:10.1111/jcpp.12263.

The best and worst of times: Commentary on “Current limitations and future directions in MRI studies of child- and adult-onset developmental psychopathologies”

F. Xavier Castellanos, MD^{1,2} and Yuliya Yoncheva, PhD¹

¹Center for Neurodevelopmental Disorders, NYU Langone Medical Center Department of Child and Adolescent Psychiatry, New York, NY, USA

²Nathan Kline Institute for Psychiatric Research, Orangeburg, NY, USA

In the current issue, Horga and colleagues (Horga, Kaur, & Peterson, 2013) provide a comprehensive overview of the current limitations of magnetic resonance imaging (MRI) of developmental psychopathologies focusing particularly on experimental design. Horga et al. are unsparing in their assessment of the problems that plague current clinical neuroimaging studies. We will not reiterate the long list of deficiencies in the imaging literature, which persist despite its impressive volume (PubMed lists more than 135,000 papers with the terms “magnetic resonance imaging” and “brain”). Human MRI can be dated from 1981 when the first results were reported from a custom-built magnet with a field strength of 0.15 Tesla (T) (Doyle et al., 1981). In the three decades since, hardware and software have improved remarkably, generally by factors of 2-3 fold. For example, subsequent generations of scanners featured increases in field strength to 0.45 T and then to 1.5 T. The 1.5 T scanners were the first high-resolution machines, and they remain the most frequently used clinical instruments worldwide. Higher-field magnets yield stronger signals, and so are preferred for research studies. Current high-field scanners typically operate at 3 T or 4 T, although, leading edge work is increasingly being performed at 7 T. Larger, more powerful machines have been built at 9.4 T, 10.5 T, 11.7 and even 14 T. But stronger signals are also accompanied by stronger image artifacts, and each successive enhancement has required teams of investigators, including MRI physicists, to refine scan sequences to optimize image quality for each technological plateau. This is a non-trivial, non-linear problem, grounded in the intricacies of quantum physics. Thus, the task of extrapolating parameters when the field strength is doubled, such as from 1.5 T to 3 T, cannot be handled by simply halving or doubling specific parameters. Nevertheless, the imaging community has successfully confronted this challenge at each stage, bolstered in no small part by the phenomenon known as Moore's law. The corollary of this observation, that computing power doubles every two years, has held since 1965, and has accounted for the astonishing increases from which data-intensive fields, such as neuroimaging, have benefitted most strikingly.

Those are the good news. Through the “magic” of quantum physics alloyed with accelerating computational resources, the imaging community has been able to obtain

previously unimaginable images of brain structure and function. But as Horga et al. remind us, much room for improvement remains. In their annual research review (Horga et al., 2013), they appropriately focus on the tried-and-true aspects of experimental design that have been relatively neglected in neuroimaging, partly due to the awe provoked by the astonishing technological advances we are witnessing. They point out the dangers of contrasting patient samples to differentially screened healthy controls; the inherent inability of cross-sectional designs to identify true developmental trajectories; the impossibility of differentiating causes from consequences when studying already affected individuals; the inaccessibility of causal mechanisms from correlational studies; and the loss of opportunity inherent in the failure to integrate brain measures of structure, function, connectivity, and metabolism. Cumulatively, these contribute to the so-far disappointing failure of imaging approaches to contribute meaningfully to the process of diagnostic classification or prediction of prognosis in psychiatric disorders.

So where does this leave us? Are our fancy new tools just a more expensive version of the “emperor's new clothes,” too rudimentary for the complexity of the brain? Let us hope not. Instead, while we agree with Horga et al. that neuroimaging approaches merely represent one more type of tool, we note that a long apprenticeship process is to be expected whenever new tools are introduced. Twenty-fold increases in magnet strength are easy to appreciate. The ever-increasing sophistication of openly available software packages for analyzing data (SPM, FSL, AFNI, etc.) should also be celebrated. But those just set the stage for where we find ourselves now: arguably the most difficult phase of the learning process. Despite their seductive appeal, images of the brain are not really images of the brain. They are computer-generated constructions, which are completely dependent for their approximate veracity on the scientific rigor of the experimental design from which they emerge.

This point is made most tellingly with regard to the importance of selecting appropriate comparison groups. Specifically, contrasting patients, who are typically affected by multiple comorbid conditions, with “well” controls screened to exclude any occurrence of current, and often past, disorders (i.e., super-healthy controls), yields results that might largely be ascribable to the multiple differences among the groups, rather than to the diagnostic entity which is the ostensible focus of study. Such misguided attempts at diagnostic purity remain the rule, as they were formerly demanded by peer review committees, and this practice persists despite being a long recognized problem (Schwartz & Link, 1989). Along with this needed reminder by Horga et al., the shift to Research Domains Criteria dimensional approaches <http://www.nimh.nih.gov/research-priorities/rdoc/nimh-research-domain-criteria-rdoc.shtml> from the DSM-centric framework, recently advocated by NIMH, should assist investigators in resisting expectations to contrast patients to healthy controls in primary analyses.

Despite the gratifying benefits of exponentially cheaper computing power, the costs of MRI have remained substantial. Scanning is comparably expensive throughout the world since the major costs are for hardware and provision of an increasingly precious resource, liquid helium. The persistently high cost of high-resolution brain imaging studies (MRI, positron-emission tomography, and magnetoencephalography) contributes to its endemic “power failure” as in neuroscience more broadly (Button et al., 2013). Low statistical power,

especially when studies are expensive, invites cluttering the literature with false positive results. These are extravagantly more likely when analyses are not appropriately corrected for the many thousands of statistical tests that are typically performed. This lesson has been thoroughly embraced by the molecular genetics field, after the resounding failure of candidate gene studies to generate replicable findings. Imaging journals are now increasingly insisting that analyses should be corrected for multiple comparisons; nonetheless tenacious investigators can still find a home for uncorrected results, often by proffering these findings as based on pre-defined regions-of-interest – the imaging version of candidate genes. This is enabled by the proliferation of on-line journals with questionable dedication to peer-review (Bohannon, 2013). One solution to differentiate true *a priori* hypothesis-driven analyses of imaging datasets from unbridled “fishing expeditions” may be for investigators to register their study and corresponding analytical plan as observational studies on ClinicalTrials.gov <http://clinicaltrials.gov/>. As of December 2013, more than 29,000 observational studies of multiple types have been registered there. We hold that such prior registration would enable readers to distinguish tests of well-formulated prior hypotheses from exploratory post-hoc results. At the same time, we acknowledge and defend the utility of full exploration, as long as everyone is aware that any results emanating from such efforts must be treated with extreme caution until fully and independently replicated.

The problem of low statistical power is particularly germane for investigators conducting pediatric psychopathology imaging, as it is aggravated by the challenges of collecting data from young children, especially those with psychiatric disorders, who usually have greater difficulty keeping still in a scanner. Even minor degrees of head motion are turning out to be pernicious, in that motion produces non-systematic errors in the data, which cannot be fully corrected by applying statistical methods or even by recruiting larger samples. One labor-intensive approach that is being more widely adopted for scanning infants and young children is to perform scans during natural sleep. Movement can also occur during sleep, but in general, the yield of such sleep studies is much greater than attempts to image very young children while they are awake. Finally, pediatric data collection is even more heroic when multiple modalities are probed simultaneously. The ability to relate data from complementary techniques, such as EEG and fMRI, is invaluable, especially in sleep and resting state investigations. Recent strides in this direction have begun to map out the development of intrinsic brain circuits during “resting state” imaging (Lüchinger, Michels, Martin, & Brandeis, 2012).

Another response to the problems of low statistical power and the high cost of brain imaging is the increasing support by investigators and funding agencies for open science approaches, including data sharing. Imaging datasets are profoundly multidimensional, and any single imaging dataset can serve as the basis for myriad analyses and publications. Yet, investigators rarely publish more than a few papers from any single dataset, leaving an untold number of potential results latent. Against this background, aggregating results or samples across multiple research centers has recently become a practical, albeit imperfect reality. Two kinds of imaging data are in the lead in terms of large-scale aggregation: structural imaging and “resting state” functional imaging.

With regard to anatomic structure, the continuing refinement and universal accessibility of openly available software such as FreeSurfer has enabled the development of an approach termed Enhancing NeuroImaging Genetics through Meta-Analysis (ENIGMA) <http://enigma.ini.usc.edu/>. By allowing investigators to upload the statistical results of decentralized imaging and genetic analyses, samples exceeding 20,000 datasets have been achieved and have revealed stringently corrected novel genetic variants associated with hippocampal and intracranial volumes (Stein et al., 2012). The same approach is currently being used in an ongoing effort to examine subcortical volumes in Attention-Deficit/Hyperactivity Disorder (ADHD), while attempting to account for demographic and clinical heterogeneity.

The relevance of “resting state” functional imaging was buttressed by demonstrations of the feasibility of aggregating resting state scans across multiple imaging centers despite the lack of *a priori* coordination (Mennes, Biswal, Castellanos, & Milham, 2013). The full release of the aggregated dataset (known as the 1000 Functional Connectomes Project, http://fcon_1000.projects.nitrc.org/) in December 2009 has led to more than 100 peer-reviewed papers published by investigators throughout the world. In the field of ADHD imaging, the ADHD-200 Project highlighted how far we have to go before we can depend on imaging to assist in refining diagnostic decisions, as noted by Horga et al. It has also provided an example of how large-scale aggregated datasets can be used to hone experimental designs. In the paper authored by the members of the ADHD-200 Consortium, led by Damien Fair, 10 different motion correction algorithmic approaches were contrasted (Fair et al., 2012). Three of these were selected for additional analyses, applied to subsets of children with Combined Type ADHD, Predominantly Inattentive Type ADHD, and unaffected, comparison children that were selected to be comparable in age, intelligence, and particularly low levels of head motion. These definitive analyses included 52 children per subgroup, which represented about 20% of the entire aggregate dataset. Despite the ostensible loss of sample size, analyses revealed the first statistically robust distinctions among the DSM-IV ADHD subtypes, as well as significant differences between each of the subtypes and controls (Fair et al., 2012).

Pediatric efforts such as the ADHD-200 Project and the Autism Brain Imaging Data Exchange (ABIDE) [http://fcon_1000.projects.nitrc.org/indi/abide/] have been conducted as community efforts without dedicated external funding. Impelled by the aging of the population, innovative prospective strategies have been implemented to establish open data repositories focusing on the aging brain with a particular focus on dementia. The Alzheimer's Disease Neuroimaging Initiative (ADNI) is listed on more than 379 PubMed entries and has invigorated dementia research to a remarkable degree. ADNI represents a well-funded public-private partnership across multiple research centers.

Another effort to consider closely for lessons that may be applicable to the first three decades of life is the UK Biobank [<http://www.ukbiobank.ac.uk/>]. The UK Medical Research Council has recently provided 9.6 million pounds for a large-scale feasibility study that is intended to lead to studying 100,000 individuals in state-of-the-art imaging of the brain, heart, abdomen, cervical arteries, and low-power X-ray imaging of bones and joints. If the initial pilot phase of several thousand participants is successfully conducted, a total of

three matched dedicated scanners will be purchased and operated for 14 hours per day, 7 days per week for six years to accumulate 100,000 datasets linked to genetics, environmental and health record data. Notably, the funding only provides for data collection. It is expected that analyses will be conducted by investigators accessing data on the “cloud,” with grants supporting the net cost of computing time. This unprecedented effort is powered to reveal gene by environment interactions and correlations and provides a sobering lower estimate of the scale that will likely be needed to effectively examine such issues during childhood, adolescence and early adulthood. It also raises complex ethical challenges that must be confronted especially with regard to mental-health questions, and on how to ensure safeguards to protect vulnerable populations (Amarasinghe et al., 2013).

In light of our still too-rudimentary understanding of the pathophysiology of child- and adult-onset developmental psychiatric conditions, we assert that the greatest ethical imperative for clinical investigators is to accelerate and enhance research efforts across the board. Attending to the lessons thoughtfully laid out by Horga and colleagues (Horga et al., 2013) on how to place research design at the forefront in clinical neuroimaging represents an important step forward. We can anticipate increasingly doing so as a truly global community of scientists, clinicians, patients, family members and advocates. It is always the best of times and the worst of times. Let the best of times prevail.

References

- Amarasinghe M, Tan H, Larkin S, Ruggeri B, Lobo S, Brittain P, Broadbent M, et al. Banking the brain. Addressing the ethical challenges of a mental-health biobank. *EMBO reports*. 2013; 14(5): 400–4. [PubMed: 23598516]
- Bohannon J. Who's afraid of peer review? *Science*. 2013; 342(6154):60–5. [PubMed: 24092725]
- Button KS, Ioannidis JPA, Mokrysz C, Nosek BA, Flint J, Robinson ESJ, Munafò MR. Power failure: why small sample size undermines the reliability of neuroscience. *Nature reviews Neuroscience*. 2013; 14(5):365–76.
- Doyle FH, Pennock JM, Orr JS, Gore JC, Bydder GM, Steiner RE, Young IR, et al. Imaging of the brain by nuclear magnetic resonance. *The Lancet*. 1981; 318(8237):53–57.
- Fair DA, Nigg JT, Iyer S, Bathula D, Mills KL, Dosenbach NUF, Schlaggar BL, et al. Distinct neural signatures detected for ADHD subtypes after controlling for micro-movements in resting state functional connectivity MRI data. *Frontiers in systems neuroscience*. 2012; 6:80. [PubMed: 23382713]
- Horga G, Kaur T, Peterson BS. Current limitations and future directions in MRI studies of child- and adult-onset developmental psychopathologies. *Journal of child psychology and psychiatry*. 2013
- Lüchinger R, Michels L, Martin E, Brandeis D. Brain state regulation during normal development: Intrinsic activity fluctuations in simultaneous EEG-fMRI. *NeuroImage*. 2012; 60(2):1426–39. [PubMed: 22245357]
- Mennes M, Biswal BB, Castellanos FX, Milham MP. Making data sharing work: the FCP/INDI experience. *NeuroImage*. 2013; 82:683–91. [PubMed: 23123682]
- Schwartz S, Link BG. The “well control” artefact in case/control studies of specific psychiatric disorders. *Psychological medicine*. 1989; 19(3):737–42. [PubMed: 2798641]
- Stein JL, Medland SE, Vasquez AA, Hibar DP, Senstad RE, Winkler AM, Toro R, et al. Identification of common variants associated with human hippocampal and intracranial volumes. *Nature genetics*. 2012; 44(5):552–61. [PubMed: 22504417]