Robustness and reliability of cortical surface reconstruction in CIVET and FreeSurfer

Lindsay B. Lewis[1], Claude Y. Lepage[1], Najmeh Khalili-Mahani[1], Mona Omidyeganeh[1], Seun Jeon[1], Patrick Bermudez[1], Alex Zijdenbos[2], Robert Vincent[1], Reza Adalat[1], and Alan C. Evans[1]

[1]McGill Centre for Integrative Neuroscience (http://mcin.ca), Montreal Neurological Institute, McGill University, Montreal, Quebec, Canada
[2]Biospective, Inc., Montreal, Quebec, Canada

INTRODUCTION

In human brain MRI, measures of cortical surface morphometry such as cortical thickness are often utilized to characterize normal as well as abnormal populations. Two well-known pipelines that provide automated structural analysis, CIVET (Kim 2005; Lepage 2017 / Poster 4166) and FreeSurfer (FS, for a review see Fischl 2012), have existed for nearly two decades and remain actively maintained. Validation studies are typically limited in that they rely upon manual segmentations as a gold standard or examine only CIVET (Kabani 2001) or only FS (Rosas 2002; Kuperberg 2003; Cardinale 2014). To our knowledge, two studies (Lee 2006, Redolfi 2015) have directly compared CIVET and FS; however they employed older versions of both pipelines (Lee 2006) or of just CIVET (Redolfi 2015). Furthermore, one should be cautious in using correlations with behavioral measures as a sole benchmark of performance (Redolfi 2015). No recent studies have directly compared reproducibility of the two pipelines.

A reproducibility comparison is of particular timeliness given that recent updates to both CIVET (v2.1) and FS (v6.0) include new high-resolution "hires" options, for which each pipeline takes a differing approach:
- CIVET: cortical surface extraction is performed in stereotaxic ("stx") space. Regardless of input resolution, the user may choose the stx template to be (default) 1mm, or (hires) 0.5mm or 0.75mm.
- FS: cortical surface extraction is performed in native space, at (default) 1mm, or for the *-hires* flag, conformed to the input resolution (<1mm).

We utilize a series of repeated acquisitions on the same subject to investigate reproducibility across comparable processing options for CIVET v2.1 and FS v6.0. Specifically, we assess robustness of the default and hires options for each pipeline (1) against a minor perturbation and (2) against noise.

METHODS

We processed n=91 3T 1mm³ T1w images of a single male subject from the multisite Infant Brain Imaging Study ("IBIS Phantom"), collected while he was 36-42 yrs, with both CIVET v2.1 (2016) and FreeSurfer (FS) v6.0 (2017). Following QC, we retained a total of n=88. Fig 1 contains a flowchart of the processing steps.

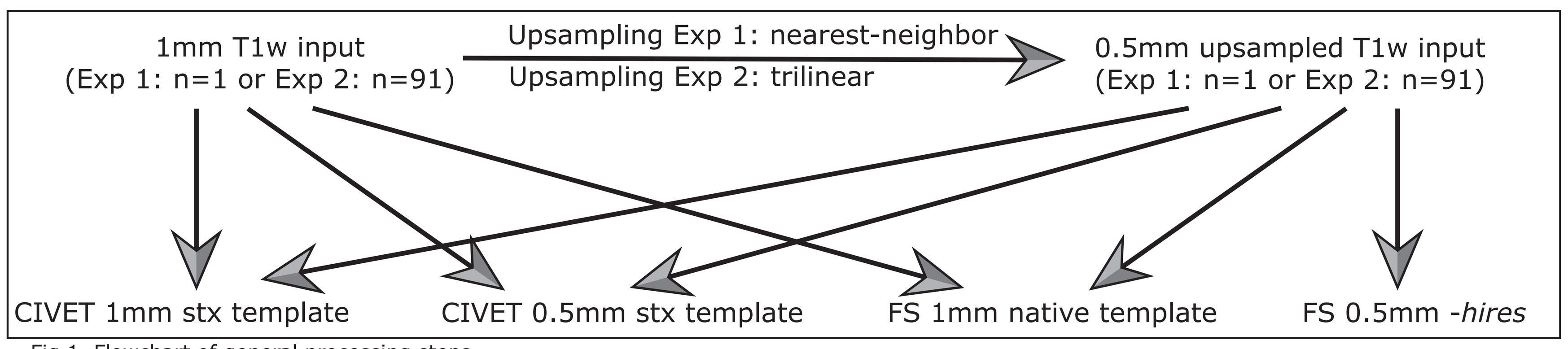

1mm T1w input
(Exp 1: n=1 or Exp 2: n=91)
Upsampling Exp 1: nearest-neighbor
Upsampling Exp 2: trilinear
0.5mm upsampled T1w input
(Exp 1: n=1 or Exp 2: n=91)
CIVET 1mm stx template
CIVET 0.5mm stx template
FS 1mm native template
FS 0.5mm -hires

Fig 1. Flowchart of general processing steps.

Each pipeline was invoked with its optimal configuration (re: CIVET, see Lepage 2017 / Poster 4166):
- CIVET: *hires* surfaces (163842 vertices / hemi), *tfs* (like FS) thickness metric, fwhm=20mm smoothing.
- FS *recon-all* (default 1mm): *-3T* flag.
- FS *recon-all -hires*: *-3T* flag, *mris_inflate -n 55*. Exp 2 only: *-parallel* and *-openmp 24* flags.

In CIVET, we generated a custom ("Phantom") average surface. To facilitate vertex-wise analyses between CIVET and FS, we imported our CIVET Phantom average surface into FS, using it as an alternative to *fsaverage* for registration of multiple scans. As a sanity check, we also show the same FS data resampled to default *fsaverage*.

Exp 1. Robustness against a minor perturbation:
- A minor perturbation in a non-cortical region should have no impact on cortical surfaces.
- In a single T1 (n=1), increased intensity by 1% of a 1mm³ voxel in center of corpus callosum.
- Perturbed the same 1mm³ cube (2^3 0.5mm voxels) for upsampled native T1 at 0.5mm.
- Expectation: zero (or minimal) differences in cortical surface placement of perturbed vs. original.

Exp 2. Reproducibility / robustness against noise:
- Any variability in cortical thickness measurements of n=88 T1s of same subject does not reflect real biological variation, but instead must be algorithmic in nature or due to acquisition noise.
- Individual thickness maps resampled to Phantom average surface (CIVET, FS) or to *fsaverage* (FS only).
- Expectation: minimal vertexwise coefficient of variation (COV).

Vertexwise analyses are performed and visualized via the SurfStat toolbox for MATLAB (Worsley 2009).

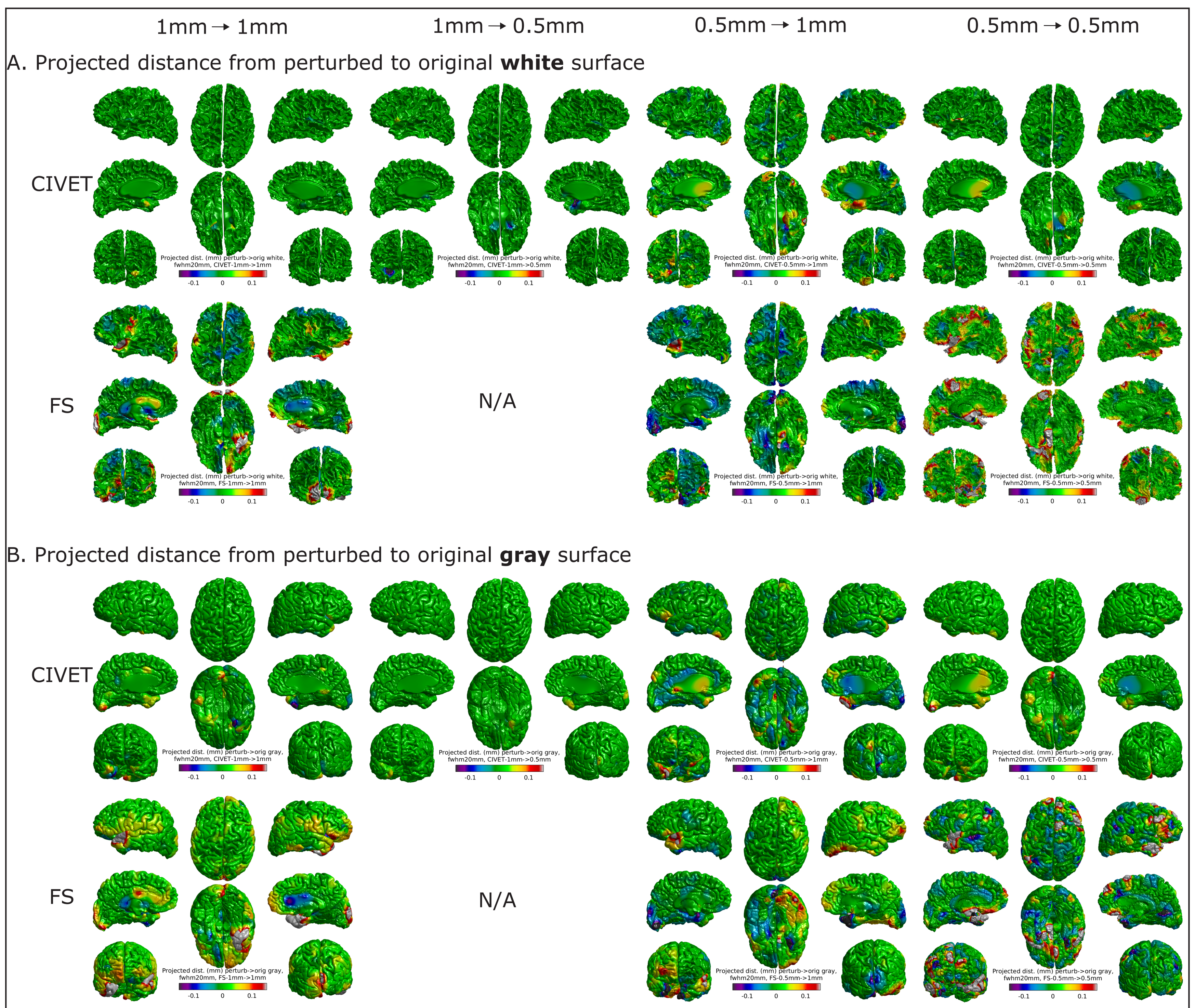

1mm → 1mm
1mm → 0.5mm
0.5mm → 1mm
0.5mm → 0.5mm
A. Projected distance from perturbed to original **white** surface
CIVET
FS
N/A
B. Projected distance from perturbed to original **gray** surface
CIVET
FS
N/A

Fig 2. Projected distance maps (mm) for CIVET v2.1 and FS v6.0 between surfaces derived from an identical input T1 image (n=1), with the exception of a 1% intensity increase perturbation in a single 1mm³ cube in the corpus callosum. Column 1: native input at 1mm submittted to default 1mm pipelines. Column 2: native input at 1mm submittted to CIVET's 0.5mm stereotaxic processing template (there is no analogous condition for FS). Column 3: native input was first upsampled to 0.5mm and then was submitted to default 1mm pipelines. Column 4: native input was first upsampled to 0.5mm and then was submitted to CIVET's 0.5mm stereotaxic processing template or FS' *-hires* pipeline (conforms to input resolution). A. Projected distance from perturbed to original **white** surface. B. Projected distance from perturbed to original **gray** surface. Warm colors indicate that the original lies above the perturbed surface, while cool colors indicate that the perturbed lies above the original surface. Green values near 0 are the ideal.

RESULTS

Robustness against a minor perturbation: Exp 1 results are depicted in Fig 2. Projected distances between the modified and original surfaces are systematically smaller (better) for CIVET v2.1 than FS v6.0.

It should be noted that the FS *-hires* pipeline (rightmost column in Fig 2) demonstrated difficulties in processing our nearest-neighbor interpolated 0.5mm input data. This resulted in longer than usual processing times, and visibly poor quality surfaces. Whatever the quality of surfaces, the reliability results above (simple perturbation of a 1mm³ cube vs. the original) remain of interest.

Reproducibility / robustness against noise: Exp 2 results are depicted in Fig 3. Across most of the cortex, CIVET demonstrates significantly lower variance / higher reliability of cortical thickness than FS, in particular at 0.5mm processing resolution.

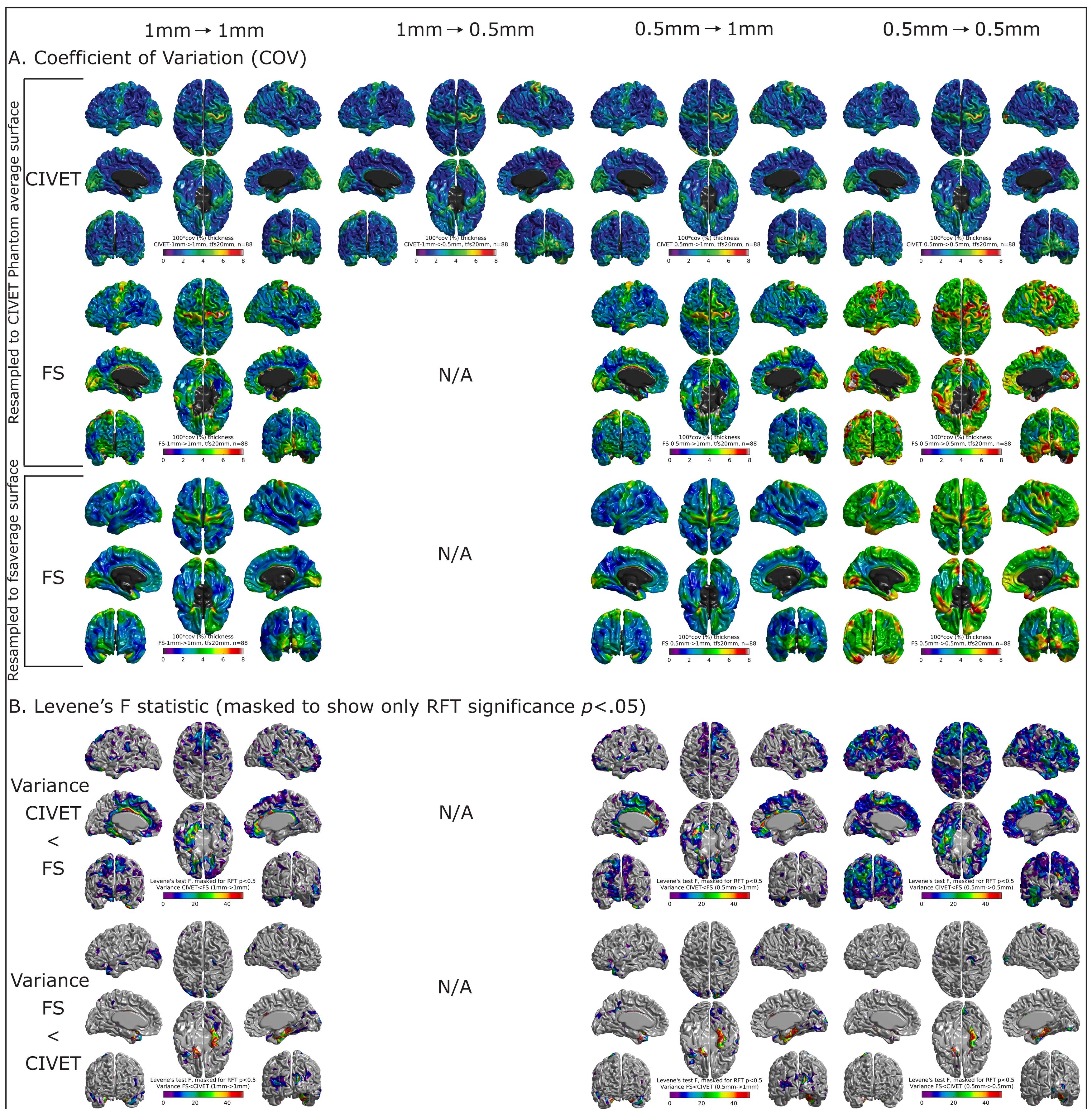

1mm → 1mm
1mm → 0.5mm
0.5mm → 1mm
0.5mm → 0.5mm
A. Coefficient of Variation (COV)
Resampled to CIVET Phantom average surface
CIVET
FS
N/A
Resampled to fsaverage surface
FS
N/A
B. Levene's F statistic (masked to show only RFT significance p<.05)
Variance CIVET < FS
N/A
Variance FS < CIVET
N/A

Fig 3. A. Coefficient of variation (COV) of cortical thickness for n=88 repeated scans on the same subject for CIVET v2.1 and FS v6.0. Columns 1-4 are identical to and are already described in Fig 2 caption. Rows 1 and 2: To enable vertexwise comparisons, both CIVET and FS thickness values were resampled to CIVET's Phantom average surface. Row 3: The same FS thickness values are also shown resampled to default *fsaverage*. B. For the same columns as described above, Levene's F statistic for differences in variance between CIVET and FS, masked to show only vertices where RFT (peak) $p<.05$. Top row: Regions where variance of CIVET is less than that of FS. Bottom row: Regions where variance of FS is less than that of CIVET.

CONCLUSIONS

Widespread cortical changes due to minor non-cortical perturbation: In Exp 1, a 1% intensity modification of a non-cortical 1mm³ cube yielded widespread differences in surface placement across the cortex. The extent of these differences was less in CIVET v2.1 compared to FS v6.0. We chose to demonstrate a simple example prior to effects of resampling with n=1; however, a group analysis is warranted. It also remains possible that CIVET's resampling of the native image to stereotaxic space via trilinear interpolation could have had an attenuating effect on the perturbation.

Regional heterogeneity in cortical thickness reproducibility: In Exp 2, variability of repeated scans on the same subject - which does not reflect real biological variation, but rather noise as a consequence of scanner or software - was reduced / better in CIVET v2.1 compared to FS v6.0 for many regions.

Impact of processing resolution: Given the increasing availability of high-resolution data and processing options, one might expect to obtain better results at 0.5mm than 1mm resolution. Counterintuitively, in FS v6.0, when native input is 0.5mm, variability is higher / worse for the *-hires* pipeline than for the default 1mm pipeline. However, one limitation of the current study is the upsampling of input. We are currently conducting similar analyses with original native inputs of 0.5mm.

Processing time: It is worth noting that the CIVET 0.5mm pipeline takes <24 hrs to complete (on a single core), while FS 0.5mm *-hires* pipeline takes anywhere from 65-120+ hrs (depending on quality of input and whether *-openmp* was used).

Future work: It is critical that these reproducibility analyses be utilized during the ongoing development of structural analysis pipelines in order to minimize susceptibility to noise. Nonetheless, it must be emphasized that high reproducibility does not necessarily imply high validity. To that end, we are currently supplementing our reproducibility analyses with a validation approach derived from the high-resolution BigBrain (Amunts 2013).

ACKNOWLEDGEMENTS

The authors thank Compute Canada (www.computecanada.ca) for usage of the computing facilities.

REFERENCES

Amunts, K. *et al*. (2013), 'BigBrain: An Ultrahigh-Resolution 3D Human Brain Model', Science 340(6139): 1472-1475.
Cardinale, F. *et al*. (2014), 'Validation of FreeSurfer-estimated brain cortical thickness: comparison with histologic measurements', Neuroinformatics 12(4): 535-542.
Fischl, B. (2012), 'FreeSurfer', Neuroimage 62(2): 774-781.
Kabani, N. *et al*. (2001), 'Measurement of cortical thickness using an automated 3-D algorithm: a validation study', Neuroimage 13(2): 375-380.
Kim, J.S. *et al*. (2005), 'Automated 3-D extraction and evaluation of the inner and outer cortical surfaces using a Laplacian map and partial volume effect classification', NeuroImage 27(1): 210-221.
Kuperberg, G.R. *et al*. (2003), 'Regionally localized thinning of the cerebral cortex in schizophrenia', Arch Gen Psychiatry 60(9): 878-888.
Lepage C.Y. *et al*. (2017), 'Human MR Evaluation of Cortical Thickness Using CIVET v2.1' (Poster 4166), OHBM, Vancouver, BC, Canada.
Redolfi, A. *et al*. (2015), 'Head-to-head comparison of two popular cortical thickness extraction algorithms: a cross-sectional and longitudinal study', PLoS One 10(3): e0117692.
Rosas, H.D. *et al*. (2002), 'Regional and progressive thinning of the cortical ribbon in Huntington's disease', Neurology 58(5): 695-701.
Worsley, K.J., *et al*. (2009), 'SurfStat: A Matlab toolbox for the statistical analysis of univariate and multivariate surface and volumetric data using linear mixed effects models and random field theory', OHBM, San Francisco, USA.