ORIGINAL ARTICLE

Cognitive Reserve and Brain Maintenance: Orthogonal Concepts in Theory and Practice

C. Habeck[1], Q. Razlighi[1], Y. Gazes[1], D. Barulli[1], J. Steffener[2], and Y. Stern[1]

[1]Cognitive Neuroscience Division, Department of Neurology, Columbia University, New York, NY 10032, USA and [2]PERFORM Center and Department of Psychology, Concordia University, Montréal, Québec, Canada H4B 1R6

Address correspondence to Christian Habeck. Email: ch629@cumc.columbia.edu

Abstract

Cognitive Reserve and Brain Maintenance have traditionally been understood as complementary concepts: Brain Maintenance captures the processes underlying the structural preservation of the brain with age, and might be assessed relative to age-matched peers. Cognitive Reserve, on the other hand, refers to how cognitive processing can be performed regardless of how well brain structure has been maintained. Thus, Brain Maintenance concerns the "hardware," whereas Cognitive Reserve concerns "software," that is, brain functioning explained by factors beyond mere brain structure. We used structural brain data from 368 community-dwelling adults, age 20–80, to derive measures of Brain Maintenance and Cognitive Reserve. We found that Brain Maintenance and Cognitive were uncorrelated such that values on one measure did not imply anything about the other measure. Further, both measures were positively correlated with verbal intelligence and education, hinting at formative influences of the latter to both measures. We performed extensive split-half simulations to check our derived measures' statistical robustness. Our approach enables the out-of-sample quantification of Brain Maintenance and Cognitive Reserve for single subjects on the basis of chronological age, neuropsychological performance and structural brain measures. Future work will investigate the prognostic power of these measures with regard to future cognitive status.

Key words: aging, brain maintenance, cognitive reserve, fMRI

Introduction

Cognitive Reserve (CR) is a long-standing concept that describes the influence of factors such as life exposures on cognitive performance when considering the relationship between brain structure and function (Stern 2002; Barulli and Stern 2013). That is, CR cannot be captured with structural brain measures and might manifest itself as a residual to, or moderator of, the structure–function relationship (Stern 2002, 2007; Jones et al. 2011).

Proxies of CR have been described in numerous forms, each with independent contributions to a delaying effect on the clinical impairment associated with neurodegenerative disease: education (Stern et al. 1999; Garibotto et al. 2008, 2012), leisure activities (Scarmeas et al. 2001; Crowe et al. 2003; Akbaraly et al. 2009) and occupational attainment (Stern et al. 1995, 1995; Garibotto et al. 2008).

The contribution of these CR-proxies was found to offset the effect of disease-related pathology on cognitive status. The concept of CR does not explicitly consider the possible effect of these CR-proxies on structural brain indicators, although such a relationship is not explicitly ruled out. Recently under the concept of Brain Maintenance (BM), influences of traditional CR-proxies on structural measures of brain health have indeed been described by us and others (Nyberg et al. 2012; Steffener et al. 2016). BM in this conception is essentially a slowing of age-related brain changes, mainly influenced by life style, CR-proxies and possibly genetics.

Both CR and BM are thus relational assessments that cannot be rendered on the basis of the current state of someone's structural brain health, chronological age or neuropsychological functioning alone, but only from the relationships

between these variables. Brain Maintenance assesses to what extent the state of brain health is better or worse than what can be expected given a participant's chronological age. In a similar vein, CR assesses to what extent cognitive function is better or worse than what could be expected given someone's state of brain health.

One goal of this paper is to operationalize simple direct measures for CR and BM, based on linear regressions or nearest-neighbor techniques. This is done in a manner that is independent of verbal intelligence and education, which are usually associated with these concepts. Not employing verbal intelligence and education in the construction of BM and CR leaves them as means for validating the derived measures.

We operationalize BM as brain structural health relative to age-matched peers, comprising gray matter volume, thickness and whiter-matter tract integrity. CR is operationalized as the difference between actual neuropsychological performance and predicted performance on the basis of brain structure. Both measures were psychometrically validated in extensive split-half robustness simulations.

Methods

Participants

Three hundred and sixty eight community-dwelling adults participated in the current study (161 male, 207 female; mean age: 48.9 years; STD (age) = 18.0 years; minimum age = 20; maximum age = 80). Table 1 shows the main demographic characteristics broken down by decade. In our sample, there was a recruitment bias in that older participants had higher verbal intelligence ($R = 0.28$, $p < 0.0001$) and more years of education ($R = 0.14$, $p = 0.0059$).

Structural Neuroimaging

We used 3 structural indices in our calculations: gray matter volume and thickness by regions of interests (ROIs), and fraction anisotropy (FA) by specific white matter tracts.

Participants underwent a T1-weighted MPRAGE scan, acquired on a 3.0 T Philips Achieva MRI scanner. These scans were acquired with TE/TR of 3/6.5 ms and Flip Angle of 8°, in-plane resolution of 256 × 256, field of view of 25.4 × 25.4 cm, and 165–180 slices in axial direction with slice-thickness/gap of 1/0 mm. Each subject's structural T1 scans were reconstructed using FreeSurfer v5.1 (http://surfer.nmr.mgh.harvard.edu/). The accuracy of FreeSurfer's subcortical segmentation and cortical parcellation (Fischl et al. 2002, 2004) has been reported to be comparable to manual labeling. Each subject's white and gray matter boundaries as well as gray matter and cerebral spinal fluid boundaries were visually inspected slice by slice, manual control points were added in the case of any visible discrepancy, and reconstruction was repeated until we reached

satisfactory results within every subject. The subcortical structure borders were plotted by *freeview* visualization tools and compared against the actual brain regions. In case of discrepancy, they were corrected manually. Finally, we computed mean values for 68 cortical ROIs for cortical thickness and cortical volume for each participant to be used in group-level analyses.

Tracts Constrained by Underlying Anatomy (TRACULA), distributed as part of the FreeSurfer v. 5.3 library (Yendiki et al. 2011) uses probabilistic tractography to extract 18 major white matter tracts. The software performs informed automatic tractography by incorporating anatomical information from a training data set, provided by the software, with the FreeSurfer anatomical parcellation of the T1 image of the current data set, thus increasing the accuracy of the white matter tract placement for each participant by incorporating each participant's anatomical data into the tractography algorithm. The software outputs white matter integrity (FA) measures for each voxel inside the 18 tracts with a mean of about 500 voxels per tract. We used mean values by tract in our calculations.

Neuropsychological Outcome Measures

Twelve measures were selected from a battery of neuropsychological tests to assess cognitive functioning in 4 cognitive domains: episodic memory (MEM), reasoning ability (FLUID), perceptual speed (SPEED), and vocabulary (VOCAB). There were some missing data for the neuropsychological measures, but we decided to be as inclusive as possible. We calculated average z-scores within domain over the all of the 3 measures that were available; a missing value for the domain z-score was assigned only when all 3 measures were missing. All measures were adjusted such that a larger value indicated *better* performance, that is, completion times were flipped in sign. The measures that made up the domain z-scores all showed high correlation, lending good support for internal consistency, as can been below.

Three **memory** (=MEM) measures were based on sub-scores of the Selective Reminding Task (SRT; Buschke and Fuld 1974). Participants in this task were initially read a list of 12 words and asked to recall as many as they could. For the following 5 trials they were reminded of the words that they did not report and were asked to again recall all of the words in the list. Words are considered to enter long-term storage from the point when they are recalled twice in a row without reminders. The long-term storage sub-score (SRT_LTS) is the sum over all words of the number of trials when each word was in long-term storage. Continuous long-term retrieval (SRT_CLRT) is the sum over all words of the number of trials for which the word was continuously recalled. The third memory measure was the number of words recalled on the last trial (SRTLast). Of 368 participants in total, 351 participants all memory measure

Table 1 Participant demographics broken down by age and decade. Older participants were significantly better educated and possessed higher verbal intelligence. DRS refers to the Mattis-Dementia-Rating scale score (Mattis 1988). NART IQ refers to the American National Adult Reading Test (Grober and Sliwinski 1991)

	Age 20–29	Age 30–39	Age 40–49	Age 50–59	Age 60–69	Age 70–79
N	87	54	38	40	103	44
Sex	32 M, 55 F	16 M, 38 F	23 M, 15 F	20 M, 20 F	47 M, 56 F	22 M, 22 F
Education (year)	15.6 ± 2.1	16.4 ± 2.5	16.0 ± 2.7	16.1 ± 2.1	16.0 ± 2.7	17.5 ± 2.6
DRS total	140.4 ± 2.5	139.7 ± 2.5	139.4 ± 2.8	140.5 ± 3.1	140.2 ± 2.6	139.6 ± 2.9
NART IQ	113.2 ± 8.0	111.4 ± 8.5	114.3 ± 8.7	116.3 ± 8.2	117.8 ± 9.4	119.4 ± 10.7

recorded. The other 17 participants missed all 3 measures. The two pairwise correlations between all 3 measures yielded a minimum value of R = 0.79, p = 1 e −75.

Reasoning ability (=FLUID) was assessed with scores on 3 different tests. One test was the WAIS III (Wechsler 1997) Block Design test, in which participants are asked to reproduce a series of increasingly complex geometrical shapes using 4 or 9 identical blocks with red, white, or split red and white sides. A second test was the WAIS III (Wechsler 1997) Letter-Number Sequencing test in which participants are asked to recall progressively longer lists of intermixed letters and numbers in alphabetical and then numerical order. The third reasoning test was the Matrix Reasoning subtest from WAIS III (Wechsler 1997) in which participants are asked to select which pattern in a set of eight possible patterns best completes a missing cell in a matrix. Of 368 participants in total, 352 participants had at least one fluid-reasoning measure recorded, and 337 had all 3 measures recorded. The two pairwise correlations between all 3 measures yielded a minimum value of R = 0.28, p = 1 e −5.

Three measures were selected to assess **perceptual speed** (=SPEED). One was the score on the Digit Symbol subtest from the Wechsler Adult Intelligence Scale (WAIS III; (Wechsler 1997)). Participants in this test were instructed to write the symbol corresponding to specific numbers as quickly as possible based on a key specifying the appropriate symbol for each digit. The score is the number of correctly produced symbols in 90 s. A second measure was the score on Part A of the Trail Making Test (Reitan and Wolfson 1987), in which participants are instructed to connect circles numbered from 1 to 24 as rapidly as possible and performance is assessed as the time to connect all 24 circles. The third speed measure was the number of colored ink patches named in 45 s in the Stroop Color Naming test. Of 368 participants in total, 352 participants had at least one speed measure recorded, and 346 had all 3 measures recorded. The two pairwise correlations between all 3 measures yielded a minimum value of R = 0.52, p = 1 e −15.

Vocabulary (=VOCAB) was assessed with scores on the Vocabulary subtest from the WAIS III (Wechsler 1997), the Wechsler Test of Adult Reading (WTAR; Wechsler 2001) and the error score of the American National Adult Reading Test (NART; Grober and Sliwinski 1991). The Vocabulary subtest asks participants to provide definitions for a series of increasingly advanced words, and the WTAR and NART both involve participants correctly pronouncing irregularly spelled English words. Of 368 participants in total, 354 participants had at least one vocabulary measure recorded, and 332 had all 3 measures recorded. The two pairwise correlations between all 3 measures yielded a minimum value of R = 0.76, p = 1 e −61.

Computation of Brain Maintenance Measure

We now describe the algorithm for our operationalization of our measure for Brain Maintenance (BM). Because of possible differences in the rate of aging across the life span, we decided to forgo a simple regression approach, which presumes a linear relationship between age and brain-structural measures. Instead of using structural measures as independent variables to predict chronological age, we compute BM as the state of somebody's brain compared to a normative sample of age-matched peers. Education and verbal intelligence, while plausible causes or consequences of good BM, were left out of the brain-maintenance computation and serve for testing construct validity: BM should ideally be positively correlated with verbal

intelligence and education. Lack of a significant positive relationship is less desirable, while a significant *negative* relationship would cast severe doubt on the validity of the derived measure.

Our approach can be described as a nearest-neighbor version of regressing age on brain structure.

For any participant with age k, the normative reference sample are all participants with ages [k-bin, k + bin], where *bin* is the window size of the age window. Mean tract FA, mean thickness, and mean cortical volume values are computed for the reference sample. BM is now computed as the relative excess or shortfall of the participant's mean tract FA, cortical thickness, and cortical volume in regard to the mean of the age-matched sample of peers. A further advantage of this approach beyond allowing non-linear age-structure relationships is that our derived brain-maintenance measure is not correlated with age. In contrast, residual regression-based approaches necessarily have to endure a, possibly unwanted, level of positive correlation between the dependent variable and residual (see the text in Discussion section and Fig. 4).

We list the algorithm below and illustrate it with mock data in Figure 1.

1. Pick participant j with age k.
2. Identify reference sample as all participants {R} with ages in the range [k-bin,k + bin].
3. Compute mean tract FA, cortical thickness, and volume for reference sample {R} as FA, THX, VOL, across regions and subjects.
4. Compute across-region mean FA, thickness and volume for participant j as fa(j), thx(j), vol(j).
5. Compute BM as the sum of the relative shortfall or excess with respect to reference means in all 3 modalities:

$$bm(j) := \text{Brain Maintenance } (j) = fa(j) / FA - 1 + thx(j) / THX - 1 + vol(j) / VOL - 1 .$$
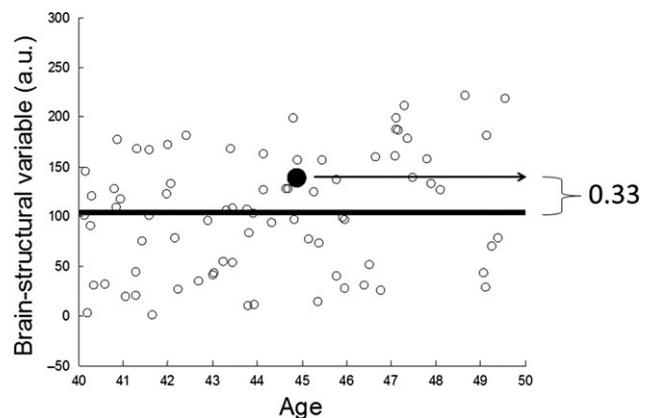

Figure 1. Illustration of the algorithm for computing BM, which is a simple non-parametric nearest-neighbor version of age regression, explained on mock data for one participant, marked with a large bolded dot. The x-axis described life years, while y-axis describes a generic brain-structural variable (brain volume, mean thickness, mean tract FA) in arbitrary units. The participant is 45 years old and her brain-structural variable is 1.33 of the mean of her age-matched peers in the 10-year age window [40,50]; her BM value thus is positive with BM = 0.33, that is, her brain structure compares favorably to her age-matched peers. In our use of the BM computation, the computation depicted in this Figure is performed for 3 modalities (cortical thickness, cortical volume, tract integrity) and the resulting BM values are averaged across modalities to result in one overall BM value.

The window size *bin* remains a critical dependency: ideally *bin* is small enough to allow the finest possible age resolution, while providing large-enough reference samples for every participant. We decided to optimize *bin* across a range of possible values such as to maximize the *F*-value of a regression model in which the independent variables verbal intelligence and education predicted the optimize brain-maintenance measure as a dependent variable. Our optimization yielded the best results for *bin* = 5, as shown in the Table 2 below.

All tested window sizes yielded highly significant relationships ($p < 0.0001$), but *bin* = 5 produced the largest association.

Computation of CR Measures

For CR we chose a linear-regression based approach: CR was computed for each cognitive domain (MEM, FLUID, SPEED, and VOCAB) as the residual, ε, when regressing the cognitive outcome measures on the brain structural measures of cortical volume, cortical thickness and tract FA. We thus willingly tolerated that collinearity between CR and the cognitive outcome measure under consideration. (Ultimately, beyond the scope of the current investigation, cognitive performance and the corresponding CR measure should – together and in spite of any collinearity- provide the optimal account for predicting somebody's future cognitive status.)

The individual-modality data, that is, 68 cortical ROIs for both volume and thickness, and FA for 18 tracts, were subjected to Principal Components Analysis prior to the 4 brain-behavioral regressions. All Principal Components surpassing the criterion of Eigen value >1 were retained, and their subject scores were used as independent variables in the regression to derive the CR measures. For the derivation of the CR measures in the full data set, the number of PCs retained for cortical volume, thickness, and tract FA was 9, 13, and 4, respectively.

We can explicitly write the brain-cognition regression as the following:

$$\mathbf{cog} = \mathbf{VOL}\,\beta_{vol} + \mathbf{THX}\,\beta_{Thx} + \mathbf{TRACT}\,\beta_{Tract} + \varepsilon$$

where **cog** is our cognitive measure of interest (=MEM, FLUID, SPEED, or VOCAB), **VOL** is a 368×9 matrix of component scores derived from the volume data with an associated 9×1 vector of regression weights β_{vol}. **THX** and **TRACT** are 368×13 and 368×4 matrices of component scores, respectively, with regression-weight vectors β_{Thx} (=13×1) and β_{Tract} (=4×1).

Age, verbal intelligence, and education were not utilized in the computation of the CR measures; we confined our derivation to the residual of cognitive performance *unexplained by brain structure*. This allows us to observe the relationship of the derived CR measures to these demographics for validation purposes, rather than making assumptions about their contribution to cognitive performance. The relationship to verbal intelligence and education should be positive, whereas the relationship to age has no prior constraints and might differ by cognitive domain.

Test Whether the Four CR Measures can be Described by one Underlying Construct

To assess whether the CR measures from the different cognitive domains described one underlying construct (and thus can be combined) we performed a Principal Component Analysis on the 4 CR measures with 10 000 permutations where the participant assignments for the 4 variables were randomized. For each of 10 000 permutations, we computed the 4 Eigen values, that is, we generated null-histograms for all Eigen values. If the point estimate of the first Eigen value lies in the right tail of the null distribution, while the other 3 Eigen values lie in the left tails of the null distribution, we are allowed to consider the 4 CR measures manifestations of a single underlying construct, and then average all scores to come up with a single generalized CR score.

To unburden the later description of the main results in the Results section, we already report here that the 4 derived CR-measures indeed could be described by one underlying construct. The point estimate of the first Eigen value lay in the right tail of the null distribution ($p < 0.001$), while all others lay in the left tail ($p = 1$). The loadings of the first Eigen vector were ([0.53, 0.49, 0.37, 0.59]) with the ordering MEM, FLUID, SPEED, and VOCAB. We thus felt that averaging all four CR scores to yield a single score is justified; this single score was terms "general CR".

Split-Sample Robustness Estimation

For ascertainment of statistical robustness of our general-CR and BM measures, we performed split-half simulations, where the pool of 368 subjects was randomly divided into a derivation sample and a replication sample of equal size (=184), with a subsequent derivation of general-CR measures and BM according to the protocol outlined in Computation of brain maintenance measure section and in the Computation of CR measures section. For general CR this means regression models are estimated in the derivation sample, and then applied without any re-estimation of regression weights to the replication sample. For BM, the algorithm in the replication sample employs the age-matched peer groups from the derivation sample only, with a slightly widened bin size for the age window of 8 years (instead of 5 years for the full-sample computation).

Correlation of CR and BM with NART and Education

Concerning the tracked performance metrics: for BM, we compute the association with education and NART in both derivation and replication samples. For the general-CR measure, we tracked how well the primary regression model (using brain structure and age as independent variables and cognitive performance as the dependent variable) predicted cognitive performance in the validation sample. Further, we computed associations with NART and education in both derivation and replication sample. We used 10 000 iterations in our split-half simulations and reported the proportion of iterations yielding significant correlations/predictions at $p < 0.01$.

Prediction of Cognitive Performance

We also used this opportunity to determine how well our derived BM and CR measures predict cognitive performance out of sample as compared to verbal intelligence and education. Age was used as a covariate for both prediction models, that is, Model 2 used NART, education and age as independent variables, whereas Model 1 used BM, CR and age as independent

Table 2 Brain behavioral regression using BM obtained from different age-window sizes (=bin) as the dependent variable, with NART and Education as independent variables

Bin size	3	4	**5**	6	7	8	9	10
F statistic	15.06	16.00	**16.45**	16.31	15.77	15.64	15.64	15.22

variables. The predicted sum of squares (PRESS) statistic in the replication sample was compared iteration-wise between the 2 sets of independent variables for all iterations. A larger PRESS value implies worse prediction.

Results

Derivation of CR and BM, and Relationship to NART and Education

We derived the Brain-Maintenance and 4 CR measures, as described above. In the derivation of the CR measures, the initial regression of the cognitive measures on the selected sets of component scores from cortical volume, thickness and tract integrity yielded the following fractions of explained variance in the cognitive outcomes: MEM – 20.4%, FLUID – 31.9%, SPEED – 38.8%, and VOCAB – 22.3 %. Relationships between CR and BM and other demographics are displayed in Supplementary Table 1. One can discern the positive relationships between BM and CR on the one hand and verbal intelligence and education on the other hand.

Test for a Single Underlying Construct for Domain-specific CR Measures

All CR measures are highly positively correlated, and as alluded in the Test whether the four CR measures can be described by one underlying construct section, our Monte-Carlo simulation indicated that they can be interpreted as manifestations of the one underlying construct. We therefore added all four CR measures to form a general CR measure.

As shown in Table 3, general CR is uncorrelated with BM and age, but some of the individual CRs did show significant negative as well as positive relationships to BM and age (positive: VOCAB; negative: MEM and SPEED, see Supplementary Table 1). For simplicity, in the remainder of the Results section, we will refer to our domain-summarized general CR measure just as "CR".

Association With Cognitive Performance and Split-half Validation Studies

To assess statistical robustness more thoroughly than single associations across the full subject sample, we employed the split-half simulations explained in the Split-sample robustness estimation section with 10 000 iterations. The fraction of significant correlations of positive sign with verbal intelligence and education were tallied in both derivation and replication samples. Since NART and education were not used in the derivation of either CR or BM, we expect the fraction of significant relationships to be similar in both samples.

Table 3 Correlation between general CR measure, BM and demographics

	Gen CR	BM	NART	Edu	Age
Gen CR	NA				
BM	0.01	NA			
NART	0.60**	0.29***	NA		
Edu	0.35**	0.19***	0.53***	NA	
Age	−0.01	−0.01	0.28***	0.14**	NA

**$p < 0.01$.
***$p < 0.001$.

Figure 2 shows horizontal scatter grams of the $\log_{10}(p)$ values of the correlations between BM and CR on the one hand and NART and education on the other hand. One can appreciate a monotonic relationship between the full-sample correlations and the replication success. As expected, there was no difference between derivation and replication sample in the fraction of significant correlations. The correlation between BM and education is modest, albeit significant, for the full sample, and replication success at $p < 0.01$ happens for 44% of iterations. All other associations had replication success of at least 95% of iterations at $p < 0.01$.

We next assessed whether CR and BM could explain the cognitive outcomes from the 4 cognitive domains. As Table 4 shows, CR and BM explained cognitive outcomes above and beyond age with significant contributions. The single exception was found for BM and its contribution to MEM, which was not significant.

Again, we substantiated these findings with split-half simulations. In particular, we were interested whether CR and BM (=model 1) were more powerful predictors of cognitive performance than NART and education (=model 2). Split-half simulations were used to fit both models in the derivation sample, and predict cognitive performance in the replication sample. Predictions in the replication were always significant at $p < 0.01$, that is, 100% of iterations gave a significant prediction. The PRESS statistic was thus compared between model 1 and model 2.

Figure 3 shows the results and plots the iteration-wise differences of the PRESS statistic between model 1 and model 2 for all 4 cognitive outcomes. For MEM, FLUID and SPEED, model 1 showed superior performance with PRESS values which were lower than for model 2 in more than 99% of iterations. For VOCAB, the situation was reversed: PRESS for model 2 was lower than for model 1 in 100% of iterations, probably reflecting the close relationships between NART and the NART error score that is being used as one of the measures to compute VOCAB.
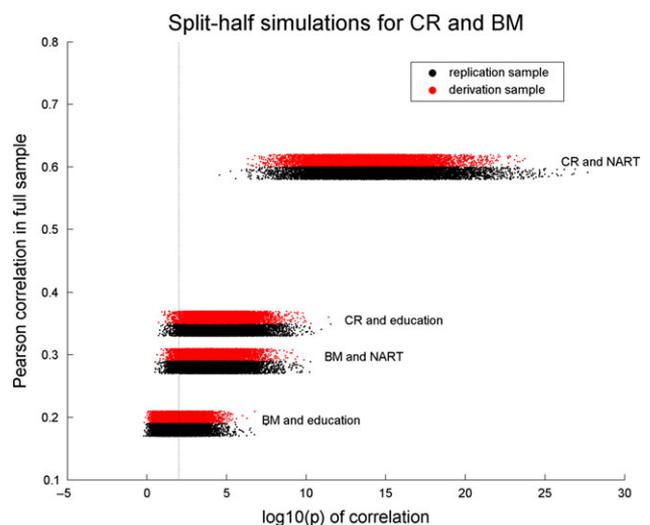

Figure 2. Split-half simulations with 10 000 iterations for the association of BM and CR with NART and education. Plotted on the x-axis is the log10(p) value of the correlation between the derived measures and NART/education; plotted on the y-axis is the Pearson correlation value in the full sample. One can discern a monotonic relationship between the correlation in the full sample and the correlations in the split-half simulations. The thin vertical line marks $p < 0.01$.

Table 4 T-statistics from 4 linear regressions of neuropsychological performance measures (MEM, FLUID, SPEED, VOCAB) against respective CR measures, BM and age, which were simultaneously entered as independent variables. All but one relationship are significant at nominal $p < 0.0001$, indicating that there are independent contributions of CR, BM and age to cognitive performance

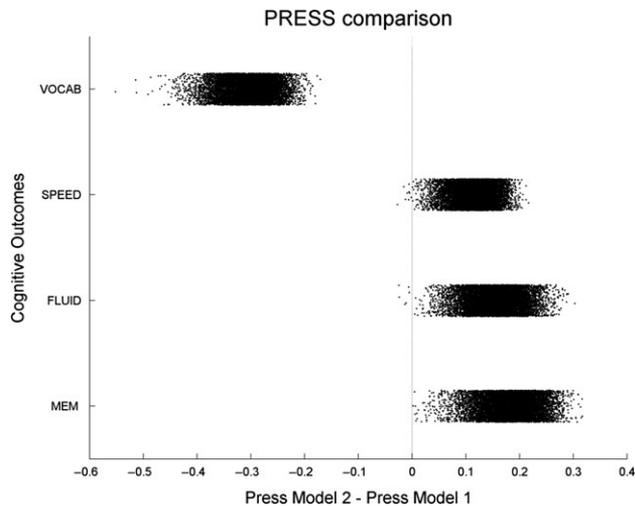| Cognitive outcome	CR	BM	Age
MEM	$T = 16.12, p < 0.0001$	$T = 0.66, p = 0.51$	$T = -11.42, p < 0.0001$
FLUID	$T = 18.93, p < 0.0001$	$T = 11.06, p < 0.0001$	$T = -10.02, p < 0.0001$
SPEED	$T = 14.01, p < 0.0001$	$T = 6.19, p < 0.0001$	$T = -15.27, p < 0.0001$
VOCAB	$T = 17.81, p < 0.0001$	$T = 9.81, p < 0.0001$	$T = 8.48, p < 0.0001$

Figure 3. Split-sample simulation with 10 000 iterations for the comparison of 2 models' predictive utility. Model 1 involves age, BM and general CR to predict cognitive outcomes, model 2 involves age, verbal intelligence and education as predictors. The scatter plot shows the difference of the PRESS statistic between Model 1 and Model 2. Model 1 has better (=lower) PRESS for the MEM, FLUID and SPEED outcomes, while Model 2 has lower PRESS for VOCAB.

Discussion

We operationalized 2 measures from structure–function relationships in 368 healthy adults: (1) a measure of BM, conceptualized as the excess or shortfall of chronological age over its age norm, and (2) general CR, conceptualized as the excess or shortfall of neuropsychological test performance in 4 cognitive domains over its brain-based prediction. Both measures were found to be orthogonal, which implies that CR can be seen as operating independently of the state of brain health of the participant, hinting at pure "software" aspects of brain functioning, while BM captures the quality of the "hardware." Both BM and CR correlated positively with years of education and verbal intelligence. Since education at least refers to a period in the participants' past, we assume that it is a formative influence for better brain aging (=BM) and more efficient use of given structural capacities (=CR).

Speculation about the formative role of education for BM and CR, while natural, has to be tempered somewhat: in several analyses run off-line and not included in this manuscript, the association of BM and CR with NART was shown to be significantly stronger than education (probed through permutation tests); further, when predicting BM and CR in a linear regressions with both NART and education as predictors, education did not contribute any effect above and beyond NART. However, the univariate associations of education with BM and CR are highly significant. Education might act through more

proximal mediators rooted in healthier life styles and reduced overall disease burden to endow participants with better brain health and higher NART too. Identifying the mechanisms underlying this mediation is an important project in its own right, which is however beyond the scope of this current report.

BM was derived from a simple nearest-neighbor version of regressing chronological age onto brain structure. Given that aging rates of structural measures might be different for different age ranges and the possibility of non-linear contributions, we decided against a least-squares based approach. A further advantage of our approach is that BM is not correlated with age from the outset (which would happen if BM was derived from a regression of age onto brain structure, since in linear regression, dependent variable and residual are forced to have a positive correlation. For clarification, Figure 4 reminds the reader of the inevitable positive correlation between dependent variable and residuals in a linear regression, which spells trouble for the regression approach in the derivation of BM.

In contrast to BM, CR was derived as the residual from domain-specific regressions that used brain structure as the independent variable; thus, by design, it is independent of brain structure. We found that the 4 different CR measures (memory, fluid reasoning, perceptual speed, and vocab) could be interpreted as manifestations of one underlying single CR construct, and subsequently added the CR scores to form one general CR score.

The residual approach was initially proposed for quantifying CR by Reed and Mungas (Reed et al. 2010). In their studies, they decomposed the variance in memory associated with demographic measures and some brain measures. The residual variance was treated as a quantified measure of CR, with validation by examining its correlation with traditional measures of CR (such as the NART), and investigating its ability to predict differential rates of cognitive decline or incident dementia. This residual model of CR was also applied to another large epidemiologic study, and tested with the same analytic approaches (Zahodne et al. 2013). With regard to BM, a relatively recent review (Nyberg et al. 2012) detailed numerous mechanisms, including genetics, lifestyle, exercise, et cetera, as credible factors in delaying brain aging successfully and establishing structure–function correlations. To our knowledge however, only a few groups have directly operationalized BM from the relationship between brain structure and chronological age for healthy as well as pathological aging (Franke et al. 2010; Bunge and Whitaker 2012; Cole et al. 2015; Wachinger et al. 2015; Steffener et al. 2016).

While, at first appearance, the residual model is an attractive approach towards estimating CR, it has some clear limitations. Conceptually, it provides a *negative* definition of CR, which is ultimately unsatisfactory. With regards to the regression model on which the residual computation is based, the model makes the assumption that the identified brain measures, such as gray matter volume, cortical thickness, white
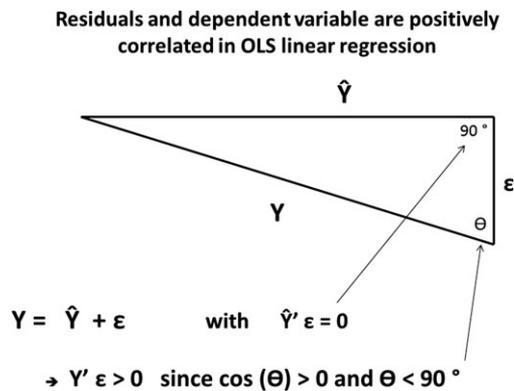
Residuals and dependent variable are positively correlated in OLS linear regression

$$Y = \hat{Y} + \varepsilon \quad \text{with} \quad \hat{Y}' \varepsilon = 0$$

$$\rightarrow Y' \varepsilon > 0 \quad \text{since } \cos(\Theta) > 0 \text{ and } \Theta < 90°$$

Figure 4. Geometric proof why residuals and dependent variable are positively correlated in ordinary-least-squares linear regression, arguing against an approach that derives BM as the residual of predicting age from brain-structural independent variables. The dependent variable Y, the model estimate $\hat{Y}$, and the residuals ε are linearly dependent and form a triangle in R^N. Since $\hat{Y}$ and ε are orthogonal, the subtending angle θ between ε and Y has to be smaller than 90°, implying that the correlation between ε and Y is positive.

matter hyperintensity burden, tract integrity, et cetera, give a structural account of substantial proportion of the variance in cognition. Ideally, for the residual to capture a true "reserve" against the effects of these brain changes there should only be minimal variance left that could be accounted for by structural brain measures, while avoiding overfitting the structural–behavioral model. In healthy aging, the maximally accounted-for proportion of variance in cognition is unclear, even when using multiple brain measures. For example, Hedden et al. (2016) used 7 separate measures to attempt to predict measures of cognition in elders age 65 to 95 and accounted for only 20% of the total variance, but 70–80% of the estimated age-related variance; however, the structural measures were pre-defined and not especially tailored towards the purpose of a maximal age- or cognition-variance account. Our structural measures were selected according to simple Eigen-value >1 criterion. They account for 20.4–38.3% of the variance in the cognitive measures. Our regression models for the derivation of CR did not subset by, or control for, age, but our structural measures accounted for 64.4% of the age-related variance in cognition. We thus feel that we pushed the residual approach to the maximum allowable variance explanation, while avoiding overfitting. The approach would further benefit from using additional brain markers to account for a hitherto unexplained amount of variance in cognition. Further, greater variation in the cognitive measure itself might be beneficial to elicit structure–function and CR-function relationships alike. Prodromal or early disease states, where the underlying pathology is beginning to cause significant changes in cognition, might thus offer the best testing ground for empirical validation of this CR measure. For example, in early Alzheimer's disease, it is conceivable that a much greater proportion of the variance in cognition would be accounted for by brain measures, including measures of amyloid and tau, and that the residual would more truly reflect CR.

Further, recalling the motivation for a re-formulation of our BM approach to avoid a necessary correlation between residual and dependent variable, we have to remind the reader that the regression-based approach *forces* a positive correlation between cognitive outcome measure and associated CR. Low performers are thus less likely to exhibit high CR than good performers, or:

high performers' residual performance unexplained by brain structure is higher than low performer's residual. We do not find this collinearity too troubling, since for it to be useful in practice, CR would have to prove its predictive utility above and beyond cognitive performance in any case, rather than being used as predictor by itself. Rather, CR and BM offer convenient summaries of structural health and performance beyond structure, respectively.

In summary, we derived 2 mutually independent measures for BM and CR which show anticipated relationships with verbal intelligence and education, while being independent of age. Further, the measures show statistical validity, as indicated by split-half replication of relationships with verbal intelligence, education and cognitive performance. Future applications of this research might aim at the prediction of the BM and CR from neuroimaging task-activation data. This might be relevant when high-quality measures of neuropsychological performance or brain structure are not available.

Both measures should be further refined by additional brain-structural variables such as White-Matter-Hyperintensities, beta-amyloid, et cetera. Particularly for BM, such multi-modal refinement on large normative reference samples would enable a quick estimation how well somebody's brain has fared in the aging process for any participant undergoing structural brain imaging, since no information other than the participant's age is needed. BM could further be augmented by blood-based biomarkers and general-health indicators. For research purposes in longitudinal cohort studies, the prognostic potential of BM could be evaluated rigorously, and younger and middle-aged individuals with low BM values could be tracked in a watchful manner. Longitudinal data could test the ecological validity of BM in midlife for the prognosis of cognitive decline in later life.

Our results in the current manuscript were purely based on cross-sectional analyses and results. Several longitudinal refinements could be considered: as mentioned before, the simplest analysis would be to track BM by itself. The BM computation at subsequent time points would be done in reference to the same normative distribution, the participant would just have a different age, (and would also provide another data point for the normative distribution itself). Worsening BM values over time might signal problems early, and initiate a "watchful waiting" regime. The next application we clearly envision would not be truly longitudinal: it would just take values of CR, BM and cognition at time point 1 and predict cognition at time point 2, which would be a straightforward regression model.

A genuine longitudinal analysis would require a full mixed-effects model, with computation of change scores for CR and BM. For CR, the values at *subsequent* time points should probably be computed as residuals to the regression equation estimated at the *first* time point to allow quantification on a single-individual basis, without having to refit the equation on which the residual computation is based. Assuming 2 time points for BM and CR, we might try to predict cognition at a third equally spaced time point with the model:

$$cog(t3) = cog(t2) + \Delta cog + BM(t2) + \Delta BM + CR(t2) + \Delta CR + age(t2)$$

where we assume that each term has its own regression weight (left out here for better legibility), and Δ indicates the change from time point 1 to time point 2. Whether *changes* in BM and CR would contribute above and beyond BM and CR, would have to be empirically tested and require a sufficient sample size.

Lastly, our measures of BM and CR offer ideal phenotypes which could be mapped in genome-wide association studies or studies of candidate genes. Genetic profiles might act through BM and CR as mediators to affect cognitive functioning. High-quality longitudinal data with complete genetics coverage will provide an exciting and fertile ground for exploration of BM and CR, their implication for future cognitive functioning and their potential for modification by interventions.

Supplementary Material

Supplementary material can be found at: http://www.cercor.oxfordjournals.org/

Notes

References

Akbaraly TN, Portet F, Fustinoni S, Dartigues JF, Artero S, Rouaud O, Touchon J, Ritchie K, Berr C. 2009. Leisure activities and the risk of dementia in the elderly: results from the Three-City Study. Neurology. 73:854–861.

Barulli D, Stern Y. 2013. Efficiency, capacity, compensation, maintenance, plasticity: emerging concepts in cognitive reserve. Trends Cogn Sci. 17:502–509.

Bunge SA, Whitaker KJ. 2012. Brain imaging: your brain scan doesn't lie about your age. Curr Biol. 22:R800–801.

Buschke H, Fuld PA. 1974. Evaluating storage, retention, and retrieval in disordered memory and learning. Neurology. 24:1019–1025.

Cole JH, Leech R, Sharp DJ, Alzheimer's Disease Neuroimaging I. 2015. Prediction of brain age suggests accelerated atrophy after traumatic brain injury. Ann Neurol. 77:571–581.

Crowe M, Andel R, Pedersen NL, Johansson B, Gatz M. 2003. Does participation in leisure activities lead to reduced risk of Alzheimer's disease? A prospective study of Swedish twins. J Gerontol B Psychol Sci Soc Sci. 58:249–255.

Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C, van der Kouwe A, Killiany R, Kennedy D, Klaveness S, et al. 2002. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron. 33:341–355.

Fischl B, van der Kouwe A, Destrieux C, Halgren E, Segonne F, Salat DH, Busa E, Seidman LJ, Goldstein J, Kennedy D, et al. 2004. Automatically Parcellating the Human Cerebral Cortex. Cereb Cortex. 14:11–22.

Franke K, Ziegler G, Kloppel S, Gaser C, Alzheimer's Disease Neuroimaging I. 2010. Estimating the age of healthy subjects from T1-weighted MRI scans using kernel methods: exploring the influence of various parameters. Neuroimage. 50:883–892.

Garibotto V, Borroni B, Kalbe E, Herholz K, Salmon E, Holtoff V, Sorbi S, Cappa SF, Padovani A, Fazio F, et al. 2008. Education and occupation as proxies for reserve in aMCI converters and AD: FDG-PET evidence. Neurology. 71:1342–1349.

Garibotto V, Borroni B, Sorbi S, Cappa SF, Padovani A, Perani D. 2012. Education and occupation provide reserve in both ApoE epsilon4 carrier and noncarrier patients with probable Alzheimer's disease. Neurol Sci. 33:1037–1042.

Grober E, Sliwinski M. 1991. Development and validation of a model for estimating premorbid verbal intelligence in the elderly. J Clin Exp Neuropsychol. 13:933–949.

Hedden T, Schultz AP, Rieckmann A, Mormino EC, Johnson KA, Sperling RA, Buckner RL. 2016. Multiple brain markers are linked to age-related variation in cognition. Cereb Cortex. 26:1388–1400.

Jones RN, Manly J, Glymour MM, Rentz DM, Jefferson AL, Stern Y. 2011. Conceptual and measurement challenges in research on cognitive reserve. J Int Neuropsychol Soc. 17:593–601.

Mattis S. 1988. Dementia Rating Scale (DRS). Odessa, FL: Psychological Assessment Resources.

Nyberg L, Lovden M, Riklund K, Lindenberger U, Backman L. 2012. Memory aging and brain maintenance. Trends Cogn Sci. 16:292–305.

Reed BR, Mungas D, Farias ST, Harvey D, Beckett L, Widaman K, Hinton L, DeCarli C. 2010. Measuring cognitive reserve based on the decomposition of episodic memory variance. Brain. 133:2196–2209.

Reitan RM, Wolfson D. 1987. The Halstead–Reitan neuropsychologican test battery. Tuscan, AZ: Neuropsychological Press.

Scarmeas N, Levy G, Tang MX, Manly J, Stern Y. 2001. Influence of leisure activity on the incidence of Alzheimer's disease. Neurology. 57:2236–2242.

Steffener J, Habeck C, O'Shea D, Razlighi Q, Bherer L, Stern Y. 2016. Differences between chronological and brain age are related to education and self-reported physical activity. Neurobiol Aging. 40:138–144.

Stern Y. 2002. What is cognitive reserve? Theory and research application of the reserve concept. J Int Neuropsychol Soc. 8:448–460.

Stern Y. 2007. Imaging cognitive reserve. In. Cognitive reserve: theory and applications. New York: Taylor & Francis. p. 251–264.

Stern Y, Albert S, Tang MX, Tsai WY. 1999. Rate of memory decline in AD is related to education and occupation: Cognitive reserve? Neurology. 53:1942–1957.

Stern Y, Alexander GE, Prohovnik I, Stricks L, Link B, Lennon MC, Mayeux R. 1995. Relationship between lifetime occupation and parietal flow: implications for a reserve against Alzheimer's disease pathology. Neurology. 45:55–60.

Stern Y, Tang MX, Denaro J, Mayeux R. 1995. Increased risk of mortality in Alzheimer's disease patients with more advanced educational and occupational attainment. Ann Neurol. 37:590–595.

Wachinger C, Golland P, Kremen W, Fischl B, Reuter M, Alzheimer's Disease Neuroimaging Initiative 2015. BrainPrint: a discriminative characterization of brain morphology. Neuroimage. 109:232–248.

Wechsler D. 1997. Wechsler adult intelligence scale – III. San Antonio, TX: Psychological Coorporation.

Wechsler D. 2001. Wechsler test of adult reading. San Antonio, TX: The Psychological Corporation.

Yendiki A, Panneck P, Srinivasan P, Stevens A, Zollei L, Augustinack J, Wang R, Salat D, Ehrlich S, Behrens T, et al. 2011. Automated probabilistic reconstruction of white-matter pathways in health and disease using an atlas of the underlying anatomy. Front Neuroinform. 5:23.

Zahodne LB, Manly JJ, Brickman AM, Siedlecki KL, Decarli C, Stern Y. 2013. Quantifying cognitive reserve in older adults by decomposing episodic memory variance: replication and extension. J Int Neuropsychol Soc. 19:854–862.