IOP Publishing Physiol. Meas. 39 (2018) 125002 (13pp) https://doi.org/10.1088/1361-6579/aae9fd

Physiological Measurement

IPEM Institute of Physics and Engineering in Medicine

PAPER

Reproducibility of dynamic cerebral autoregulation parameters: a multi-centre, multi-method study

Marit L Sanders[1], Jurgen A H R Claassen[1], Marcel Aries[2], Edson Bor-Seng-Shu[3], Alexander Caicedo[4], Max Chacon[5], Erik D Gommer[6], Sabine Van Huffel[7,8], José L Jara[5], Kyriaki Kostoglou[9], Adam Mahdi[10], Vasilis Z Marmarelis[11], Georgios D Mitsis[12], Martin Müller[13], Dragana Nikolic[14], Ricardo C Nogueira[3], Stephen J Payne[10], Corina Puppo[15], Dae C Shin[11], David M Simpson[14], Takashi Tarumi[16], Bernardo Yelicich[15], Rong Zhang[16], Ronney B Panerai[17] and Jan Willem J Elting[18]

[1] Department of Geriatric Medicine, Radboudumc Alzheimer Centre and Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center, Nijmegen, The Netherlands
[2] Department of Intensive Care, University of Maastricht, Maastricht University Medical Center, Maastricht, The Netherlands
[3] Department of Neurology, Hospital das Clinicas University of Sao Paulo, Sao Paulo, Brazil
[4] Mathematics and Computer Science, Faculty of Natural Sciences and Mathematics, Universidad del Rosario, Bogotá, Colombia
[5] Department of Engineering Informatics, University of Santiago, Santiago, Chile
[6] Department of Clinical Neurophysiology, University Hospital Maastricht, Maastricht, The Netherlands
[7] Department of Electronic Engineering (ESAT), STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics, KU Leuven, Leuven, Belgium
[8] IMEC, Leuven, Belgium
[9] Department of Electrical, Computer and Software Engineering, McGill University, Montreal, Canada
[10] Department of Engineering Science, University of Oxford, Oxford, United Kingdom
[11] Department of Biomedical Engineering, University of Southern California, Los Angeles, CA, United States of America
[12] Department of Bioengineering, McGill University, Montreal, Canada
[13] Department of Neurology, Luzerner Kantonsspital, Luzern, Switzerland
[14] Institute of Sound and Vibration Research, University of Southampton, Southampton, United Kingdom
[15] Departamento de Emergencia, Hospital de Clínicas, Universidad de la República, Montevideo, Uruguay
[16] IEEM, Presbyterian Hospital Dallas, University of Texas Southwestern Medical Center, Dallas, United States of America
[17] Department of Cardiovascular Sciences and Leicester NIHR Biomedical Research Centre in Cardiovascular Sciences, Glenfield Hospital, Leicester, United Kingdom
[18] Department of Neurology, University Medical Center Groningen, Groningen, The Netherlands

E-mail: Jurgen.Claassen@radboudumc.nl

RECEIVED
5 June 2018

REVISED
1 October 2018

ACCEPTED FOR PUBLICATION
22 October 2018

PUBLISHED
7 December 2018

Keywords: cerebral autoregulation, method comparison, reproducibility, surrogate data

Supplementary material for this article is available online

Abstract

Objective: Different methods to calculate dynamic cerebral autoregulation (dCA) parameters are available. However, most of these methods demonstrate poor reproducibility that limit their reliability for clinical use. Inter-centre differences in study protocols, modelling approaches and default parameter settings have all led to a lack of standardisation and comparability between studies. We evaluated reproducibility of dCA parameters by assessing systematic errors in surrogate data resulting from different modelling techniques. *Approach*: Fourteen centres analysed 22 datasets consisting of two repeated physiological blood pressure measurements with surrogate cerebral blood flow velocity signals, generated using Tiecks curves (autoregulation index, ARI 0–9) and added noise. For reproducibility, dCA methods were grouped in three broad categories: 1. Transfer function analysis (TFA)-like output; 2. ARI-like output; 3. Correlation coefficient-like output. For all methods, reproducibility was determined by one-way intraclass correlation coefficient analysis (ICC). *Main results*: For TFA-like methods the mean (SD; [range]) ICC gain was 0.71 (0.10; [0.49–0.86]) and 0.80 (0.17; [0.36–0.94]) for VLF and LF ($p = 0.003$) respectively. For phase, ICC values were 0.53 (0.21; [0.09–0.80]) for VLF, and 0.92 (0.13; [0.44–1.00]) for LF ($p < 0.001$). Finally, ICC for ARI-like methods was equal to 0.84 (0.19; [0.41–0.94]), and for correlation-like methods, ICC was 0.21 (0.21; [0.056–0.35]). *Significance*: When applied to realistic surrogate data, free from the additional exogenous influences of physiological variability on cerebral blood flow, most methods of dCA modelling showed ICC values considerably higher than what has been reported

© 2018 Institute of Physics and Engineering in Medicine

for physiological data. This finding suggests that the poor reproducibility reported by previous studies may be mainly due to the inherent physiological variability of cerebral blood flow regulatory mechanisms rather than related to (stationary) random noise and the signal analysis methods.

Introduction

Cerebral autoregulation (CA) is an important mechanism for maintaining adequate cerebral perfusion despite changes in blood pressure (BP) (Panerai 1998). Dynamic CA (dCA) is measured as the CBF response to a transient, short-lasting change in BP (Aaslid *et al* 1989, Panerai 1998). Usually, CBF velocity (CBFV) measured with transcranial Doppler ultrasound, as an estimate of CBF, is analysed simultaneously with BP recordings (Panerai *et al* 1998b).

Different indices of dCA have been shown to reflect pathological conditions such as stroke, severe head injury, subarachnoid haemorrhage, carotid artery disease and others (Reinhard *et al* 2003b, 2004, 2012, Immink *et al* 2005, Panerai 2008, van Beek *et al* 2008, Czosnyka *et al* 2009). However, despite the potential to bring considerable benefits to early diagnosis, management and prognosis of patients with cerebrovascular disease, the reliability of these indices of dCA has not been fully validated. Not surprisingly, therefore, dCA measurements have not yet been incorporated into routine clinical practice. There are a number of reasons for this.

First, currently no gold standard test for the assessment of dCA exists. Many different methods to calculate dCA parameters are now available and with the growing number of possibilities to measure and analyse dCA, more information is needed about the diagnostic performance and reliability of different methods (Panerai 2008, van Beek *et al* 2008, Meel-van den Abeelen *et al* 2014b).

Second, the reproducibility of dCA indices is a major concern; a relatively small number of studies show that most techniques do not demonstrate reproducibility that would be acceptable for clinical use (Gommer *et al* 2010, van Beek *et al* 2010, Elting *et al* 2014a).

Third, a lack of metric convergence between different methods (Tzeng *et al* 2012), inter-centre differences in study protocols, modelling approaches and default parameter settings, for techniques such as transfer function analysis (TFA) have led to a lack of standardisation and comparability of studies, thus limiting the possibility of using the literature to overcome the problems above (Meel-van den Abeelen *et al* 2014b).

To address these issues, The Cerebral Autoregulation Research Network (CARNet, www.car-net.org) has embarked on several multi-centre studies aiming to improve standardisation and reliability of techniques for assessment of dCA. The first initiative was limited to one method—TFA—following the observation of considerable disparity in parameter settings and reporting of TFA coherence, gain and phase. In that study, a common dataset was analysed with TFA by multiple centres (Meel-van den Abeelen *et al* 2014a). The outcome of that study led to recommendations to improve the standardisation of TFA to assess dCA (Claassen *et al* 2016).

The present study is part of a more ambitious CARNet project, wherein we aim to address the question of reproducibility for a wider range of dCA methods, not only TFA, but also time-domain models and correlation-coefficient- based methods (table 1).

We considered that the reproducibility of dCA has two main components, namely (1) 'methodological noise': systematic errors that are inherent to the methods and modelling techniques, and (2) 'physiological noise': random errors due to physiological variability between repeated measurements or due to noise or artefacts in the recorded signals. In order to advance the field, it is important to study each component separately.

The main purpose of the present study was to address the first possible cause of poor dCA reproducibility: the systematic errors resulting from different modelling techniques proposed for dCA assessment. To achieve this, we have used surrogate CBF data to reduce and control physiologic variability in repeated dCA measurements. We report herein how reproducibility varies when a single dataset is analysed by different centres using various methods to assess dCA.

Specifically, this study assessed the repeatability of dCA measurements for spontaneous oscillations in BP and CBFV, under the ideal conditions where the expected dCA is known and invariant between repeated records, covering the full range of autoregulation from absent to very efficient. This was achieved through simulations of CBF, using physiological (truly recorded) repeated BP signals as input. dCA analysis methods were grouped in three broad categories: 1. TFA-like output; 2. ARI-like output; and 3. correlation coefficient-like outputs.

Methods

Subjects

A database was created from available datasets of cerebral hemodynamic measurements from three of the 14 participating centres listed in table 2. For the purpose of this study, only the BP signal was selected and the corresponding CBF signal was ignored and replaced by a generated signal (see below). Twenty-two healthy adults aged 66.3 $\pm$ 7.0 years were selected. Exclusion criteria were uncontrolled hypertension, smoking, cardiovascular

Table 1. Methods with corresponding output variables per centre.

Centre number	Method	Output variables	Method category	Method group	References
1	1.1 Transfer function analysis	Coherence, gain (cm/s/mmHg) and phase (rad) in VLF, LF	1	1	Zhang *et al* (1998)
	1.2 Autoregulation index	ARI	2	6	Panerai *et al* (1998b)
2	2.1 Laguerre expansion of 1st-order Volterra kernels, single input (BP)	Gain (cm/s/mmHg) and phase (rad) in VLF, LF	1	2	Marmarelis (2004), Marmarelis *et al* (2013, 2014a, 2014b)
	2.2 Laguerre expansion of 1st-order Volterra kernels, dual input (BP, CO_2)	Gain (cm/s/mmHg) and phase (rad) in VLF, LF	1	2	
3	3.1 Transfer function analysis	Coherence, gain (cm/s/mmHg), phase (rad) in VLF, LF	1	1	Zhang *et al* (1998)
	3.2 Transfer function analysis	Coherence, gain (%/%) in VLF, LF	1	1	
4	4.1 Autoregulation index (FFT)	ARI	2	6	Panerai *et al* (1998b) and (2003)
	4.2 Autoregulation index (moving average 1)	ARI	2	7	
	4.3 Autoregulation index (moving average 2)	ARI	2	7	
5	5.1 Transfer function analysis	Coherence, gain (cm/s/mmHg), phase (rad) in VLF, LF	1	1	Zhang *et al* (1998)
	5.2 Oblique and orthogonal subspace projections	Subspace ratio's	3	10	Caicedo *et al* (2016)
6	6.1 Transfer function analysis	Coherence, gain (cm/s/mmHg), phase (rad) in VLF, LF	1	1	Muller *et al* (2003), Muller and Osterreich (2014)
7	7.1 Transfer function analysis	Coherence, gain (cm/s/mmHg), phase (rad) in VLF, LF	1	1	Gommer *et al* (2010)
8	8.1 ARX	ARX coefficient (3rd)	2	7	Liu and Allen (2002), Liu *et al* (2003), Panerai *et al* (2003)
	8.2 Wavelet analysis	Synchronisation index, phase (rad) in VLF,LF	1	3	Peng *et al* (2010)

(*Continued*)

Table 1. (*Continued*)

Centre number	Method	Output variables	Method cate-gory	Method group	References
9	9.1 Transfer function analysis	Coherence, gain (cm/s/mmHg), phase (rad) in VLF, LF	1	1	van Beek *et al* (2010, 2012)
	9.2 Convergent cross mapping	CCM correlation coefficient	1	3	Heskamp *et al* (2014)
11	11.1 Transfer function analysis	Coherence, Gain (cm/s/mmHg), phase (rad) in VLF, LF	1	1	Panerai *et al* (1998b)
	11.2 Transfer function analysis	Coherence, gain (%/mmHg), phase (rad) in VLF, LF	1	1	
	11.3 Transfer function analysis	Coherence, gain (%/%) in VLF, LF	1	1	Panerai *et al* (2000) and Simpson *et al* (2001)
	11.4 Univariate transfer function analysis (parametric method)	Coherence, gain (%/%), phase (rad) in LF	1	4	
	11.5 Univariate impulse response (parametric method)	The second filter coefficient (h_1) of the estimated FIR	2	9	
	11.6 Multivariate transfer function analysis (parametric method)	Gain (%/%) and phase (rad) for LF band	1	4	
12	12.1 Transfer function analysis	Coherence, gain (cm/s/mmHg), phase (rad) in VLF, LF	1	1	Zhang *et al* (1998)
	12.2 Autoregulation index	ARI	2	6	Panerai *et al* (1998b)
	12.3 Wavelet coherence analysis	Gain (cm/s/mmHg) and phase (rad) in VLF, LF	1	3	Torrence and Webster (1999) and Grinsted *et al* (2004)
13	13.1 Transfer function analysis	Coherence, gain (cm/s/mmHg), phase (rad) in VLF, LF	1	1	Panerai *et al* (1998a, 1998b)
14	14.1 ARX models: 1 input	Gain (cm/s/mmHg), phase (rad) in VLF, LF	1	5	Mitsis *et al* (2002, 2009)
	14.2 ARX models: 2 inputs	Gain (cm/s/mmHg), phase (rad) in VLF, LF	1	5	
	14.3 Laguerre expansion FIR models, single input (BP)	Gain (cm/s/mmHg), phase (rad) in VLF, LF	1	2	Mitsis *et al* (2004) and Kostoglou *et al* (2014)
	14.4 Laguerre expansion FIR models, dual input (BP, CO_2)	Gain (cm/s/mmHg), phase (rad) in VLF, LF	1	2	
	14.5 Transfer function analysis	Coherence, gain (cm/s/mmHg), Phase (rad) in VLF, LF	1	1	Meel-van den Abeelen *et al* (2014a)

Method category: 1 = TFA-like methods, 2 = ARI-like methods, 3 = correlation-like methods; method group: 1 = TFA, 2 = Laguerre expansions, 3 = wavelets, 4 = IR-filter, 5 = ARX, 6 = ARI, 7 = ARMA-ARI/ARX, 9 = IR-filter, 10 = correlation coefficient; VLF: very low frequency; LF: low frequency; BP: blood pressure; FFT: fast Fourier transform; ARI: autoregulation index; ARX: autoregressive model with exogenous input; centre names are listed in table 2. Pre-processing settings are listed in table S7.

Table 2. Participating centres and their roles.

Name	Institution	Country	Role	Centre number
E Borg-Seng-Shu RC Nogueira	Department of Neurology, Hospital das Clinicas University of Sao Paulo	Brazil	Analysis	1
VZ Marmarelis DC Shin	Department of Biomedical Engineering University of Southern California, Los Angeles	USA	Analysis	2
R Zhang T Tarumi	IEEM, Presbyterian Hospital Dallas University of Texas Southwestern Medical Center	USA	Analysis Data provider	3
RB Panerai	Department of Cardiovascular Sciences University of Leicester	UK	Analysis Data provider Trial coordination	4
S van Huffel A Caicedo	Department of Electronic Engineering (ESAT), STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics, KU Leuven, Belgium; imec	Belgium	Analysis	5
M Müller	Department of Neurology Luzerner Kantonsspital	Switzerland	Analysis	6
ED Gommer	Department of Clinical Neurophysiology University Hospital Maastricht	Netherlands	Analysis Data provider	7
SJ Payne A Mahdi	Department of Engineering Science University of Oxford	UK	Analysis	8
JAHR Claassen ML Sanders	Department of Geriatric Medicine Radboud University Nijmegen	Netherlands	Analysis Data provider Trial coordination	9
DM Simpson D Nikolic	Institute of Sound and Vibration Research University of Southampton	UK	Analysis Data provider	11
JWJ Elting M Aries	Department of Neurology University Medical Center Groningen	Netherlands	Analysis Data provider Trial coordination	12
C PuppoB Yelicich	Departamento de Emergencia, Hospital de Clínicas Universidad de la República, Montevideo	Uruguay	Analysis	13
GD Mitsis K Kostoglou	Department of Bioengineering Department of Electrical, Computer and Software Engineering McGill University, Montreal	Canada	Analysis	14

Centre 10 withdrew their results from analysis because their method was superseded by recent developments that would disadvantage their original approach.

disease, diabetes, irregular heart rhythm, transient ischemic attack/stroke or significant pulmonary disease. The study has been carried out in accordance with The Code of Ethics of the World Medical Association (Declaration of Helsinki). Informed consent was obtained of all subjects.

Description of datasets

The three centres that provided the data had similar protocols expecting participants to refrain from nicotine, alcohol and caffeine from midnight prior to the study. Time between the two measurements varied between centres from minutes to a maximum of three months. These interval differences were not considered in the analysis. Datasets consisted of five minutes of beat-to-beat mean BP (digital artery volume clamping) and end-tidal CO_2 ($etCO_2$, capnography) measurements at rest.

The BP time-series were used to generate surrogate CBF data, which were expressed as CBFV to allow comparison with physiological data and with the literature in this field. For each BP signal in this set, the original Tiecks curves (ARI 0–9) (Panerai *et al* 1998b) were used to generate one surrogate CBFV signal from autoregulation index (ARI) values ranging from ARI = 0 (absence of autoregulation) to ARI = 9 (best CA that can be observed). The ARI value used to generate the CBFV signal will be referred to as ARI_{INPUT}. ARI 0–9 were all represented: ARI0 $n = 3$; ARI1 $n = 2$; ARI2 $n = 3$; ARI3 $= 2$; ARI4 $n = 2$; ARI5 $n = 3$; ARI6 $n = 1$; ARI7 $n = 2$; ARI8 $n = 1$; ARI9 $n = 3$ (total: $n = 22$). For the repeated measurements, identical ARI values were used to generate

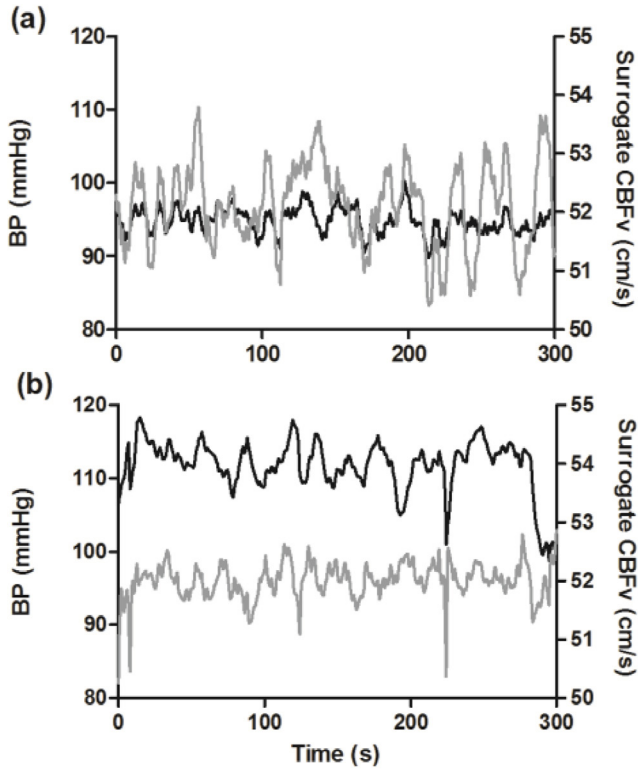

Figure 1. BP and CBFV signal examples. Representative mean blood pressure (BP, black) recordings and their corresponding surrogate cerebral blood flow velocity (CBFV, grey) signals, which were generated using Tiecks' model (Tiecks *et al* 1995). In panel (a), ARI 0 was used to generate the CBFV signal, in panel (b) ARI 9 was used. An ARI value of 9 corresponds to very efficient autoregulation which makes the relationship between BP and CBFV less pronounced, also considering there is noise added.

the repeated CBFV signals. Random Gaussian band-pass noise (0.02–0.1 Hz) was added to the generated CBFV signals, to produce outputs with a signal-to-noise ratio (SNR) of 6 dB in this frequency range, thus mimicking CBFV signals as they would be measured in 'real life'. (Katsogridakis *et al* 2011).

dCA analysis

Data analyses were performed by 14 participating centres on 44 datasets from 22 volunteers with two measurements each. The following dCA analysis methods were used: TFA (Panerai *et al* 1998a, Zhang *et al* 1998, Mitsis *et al* 2002, Muller *et al* 2003, Reinhard *et al* 2003a, Liu *et al* 2005, Gommer *et al* 2010, van Beek *et al* 2010, Meel-van den Abeelen *et al* 2014a, Muller and Osterreich 2014, Panerai 2014), Laguerre expansion of 1st-order Volterra kernels or finite impulse response models (Marmarelis 2004, Marmarelis *et al* 2013, 2014a, 2014b, Mitsis *et al* 2004, 2009), wavelet analysis (Torrence and Webster 1999, Grinsted *et al* 2004, Peng *et al* 2010), parametric finite-impulse response filter based methods (Panerai *et al* 2000, Simpson *et al* 2001), ARI anaysis (Panerai *et al* 1998b), autoregressive moving average (ARMA) based ARI methods and variant ARI methods (Panerai *et al* 2003), autoregressive with exogenous input (ARX) methods (Liu and Allen 2002, Liu *et al* 2003, Panerai *et al* 2003) and correlation coefficient-like indices (Heskamp *et al* 2014, Caicedo *et al* 2016). A summary of the methods and corresponding references are given in table 1.

Analysis of dCA outcome parameters

For the reproducibility and variability analysis of the dCA parameters, analysis methods were grouped into three broad categories: 1. TFA-like output; 2. ARI-like output; and 3. correlation coefficient-like outputs. For the TFA-like output methods, we provided suggested settings that were similar to the recent CARNet white paper (Claassen *et al* 2016). However, it should be noted that that paper had not yet been published when we performed this study. Because of this, and given the specific purpose of this study, adherence to these white paper settings was not strictly enforced. In summary, the suggested settings involved spectral estimates using the Welch method with multiple segments of data of at least 100 s, 50% superposition, and cosine windowing to reduce spectral leakage. Estimates of gain and phase were averaged for different frequency bands. All centres were free to use their own settings to cover the frequency range between 0–0.5 Hz. The ARI-like output methods consisted of time domain estimates of the impulse or step response, using the inverse Fourier transform of gain and phase, or ARMA models. Finally, the correlation coefficient-like outputs consist of a single parameter, obtained by linear regression or similar methods (table 1). These categories were created from the perspective of similar output parameters, not because of similarity on mathematical grounds.

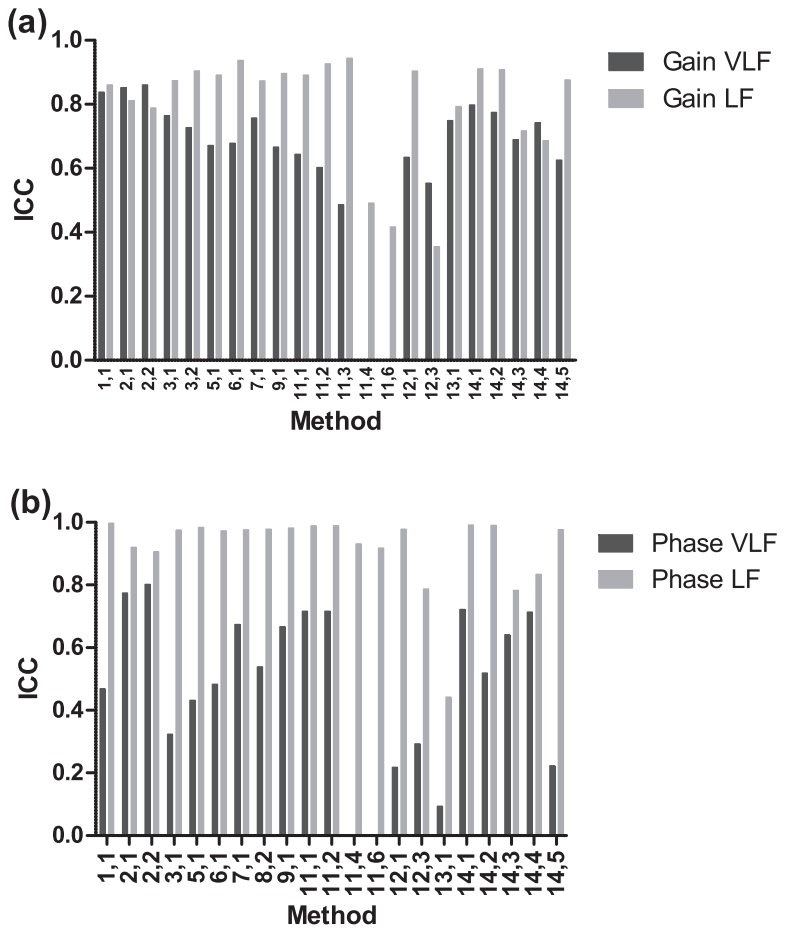

Figure 2. ICC values for methods using TFA or similar approaches ('TFA-like') with multiple outcome parameters. (a) Gain VLF and Gain LF; (b) phase VLF and Phase LF. Results are shown per method. See tables 1 and 2 for a list of centres and methods.

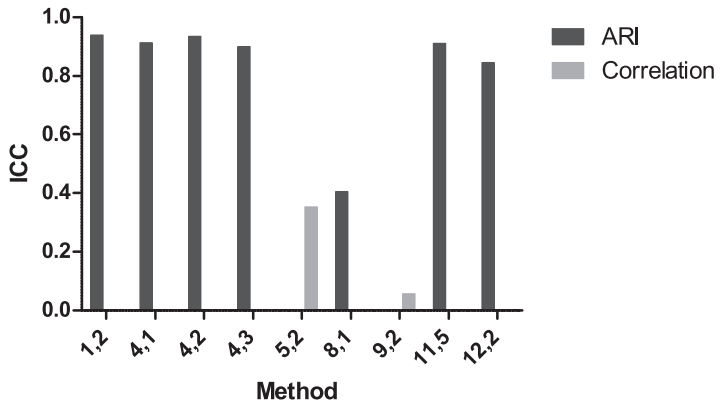

Figure 3. ICC values for methods with single outcome parameters: ARI (dark grey) and correlation based methods (light grey). Results are shown per method. See tables 1 and 2 for a list of centres and methods.

Statistical analysis

Reproducibility of the repeated measurements for all dCA analysis methods was determined by one-way intraclass correlation coefficient analysis (ICC). Because the ICC does not reflect the accuracy of the measurement, just the consistency, and because outliers could yield high ICC values, an additional evaluation of accuracy was performed. For this, the results of dCA analysis methods were compared to the reference ARI (ARI_{INPUT}) that was used to generate the surrogate CBFV signals (used here as the 'gold-standard' for the simulated signals), and to the reference gain and phase values that corresponded to the ARI_{INPUT}.

Differences between VLF and LF gain and phase values were tested with the paired Wilcoxon signed rank test, considering that most parameters, such as TFA estimates, are not normally distributed. A value of $p < 0.05$ was adopted to indicate statistical significance.

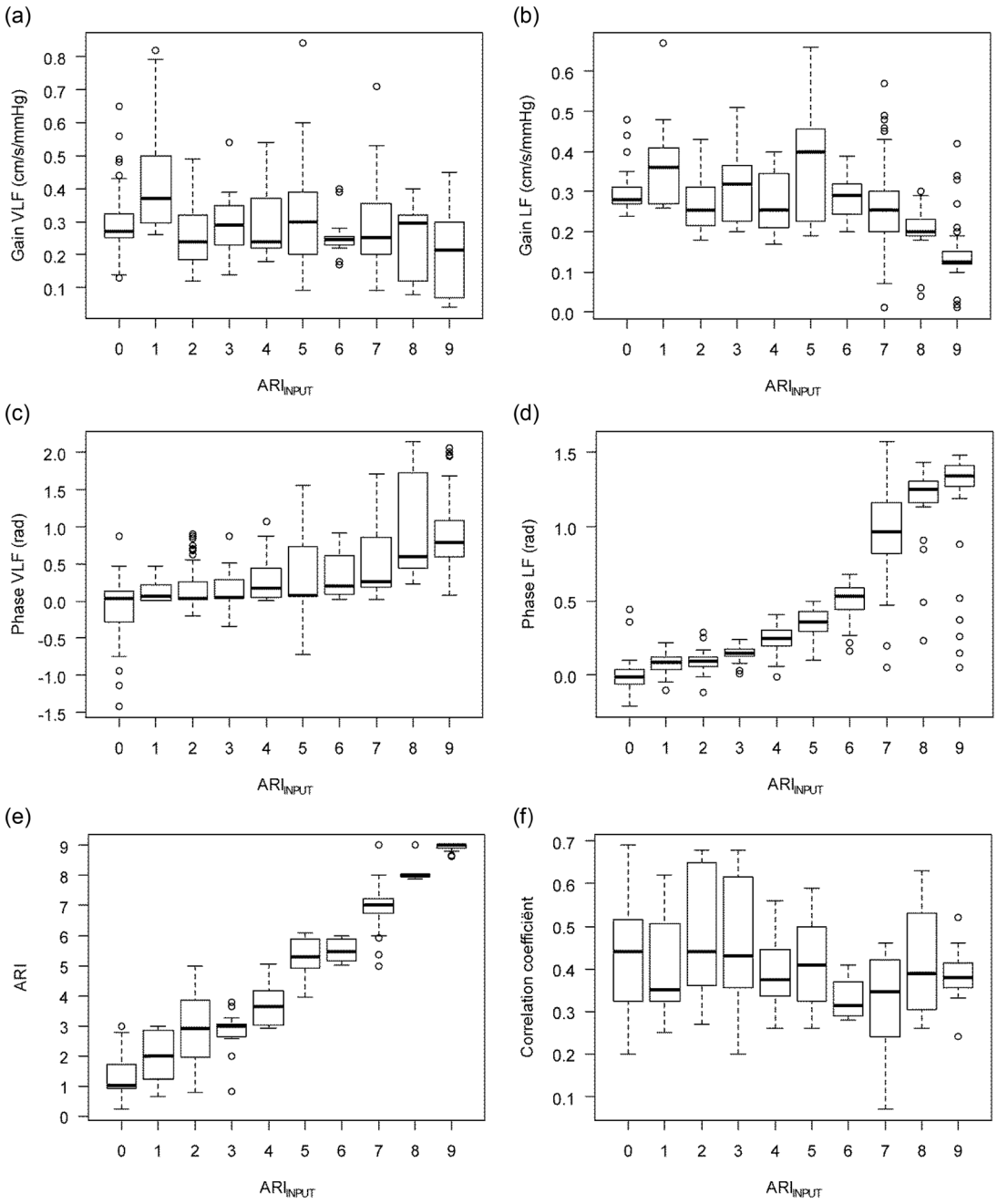

Figure 4. Combined dCA results per method category (table 1) compared to reference ARI values. TFA gain and phase values, ARI and correlation coefficient values as estimated and reported by centres pooled and plotted against the reference ARI_{INPUT} value that was used to generate the surrogate CBFV signal. (a): VLF gain; (b): LF gain; (c): VLF phase; (d): LF phase; (e): ARI; (f): correlation coefficient.

Results

An example showing two BP signals from two different subjects with the corresponding surrogate CBFV signals is presented in figure 1. This figure shows that the two BP signals differed due to expected physiological variation between repeated measurements, and also demonstrates the difference in the generated CBFV signals between ARI = 0 (poor autoregulation, high variability in CBFV) and ARI = 9 (very efficient autoregulation, limited variability in CBFV).

ICC

S2–S6 tables present an overview of the outcome parameters as reported by all centres and all the different dCA methods that were used for the repeated measurements (T1 and T2).

Figures 2 and 3 depict the results of the ICC analysis. For TFA-like methods (figure 2), ICC values for gain and phase in the low frequency band (LF: 0.07–0.2 Hz, light grey) are higher than in the very low frequency

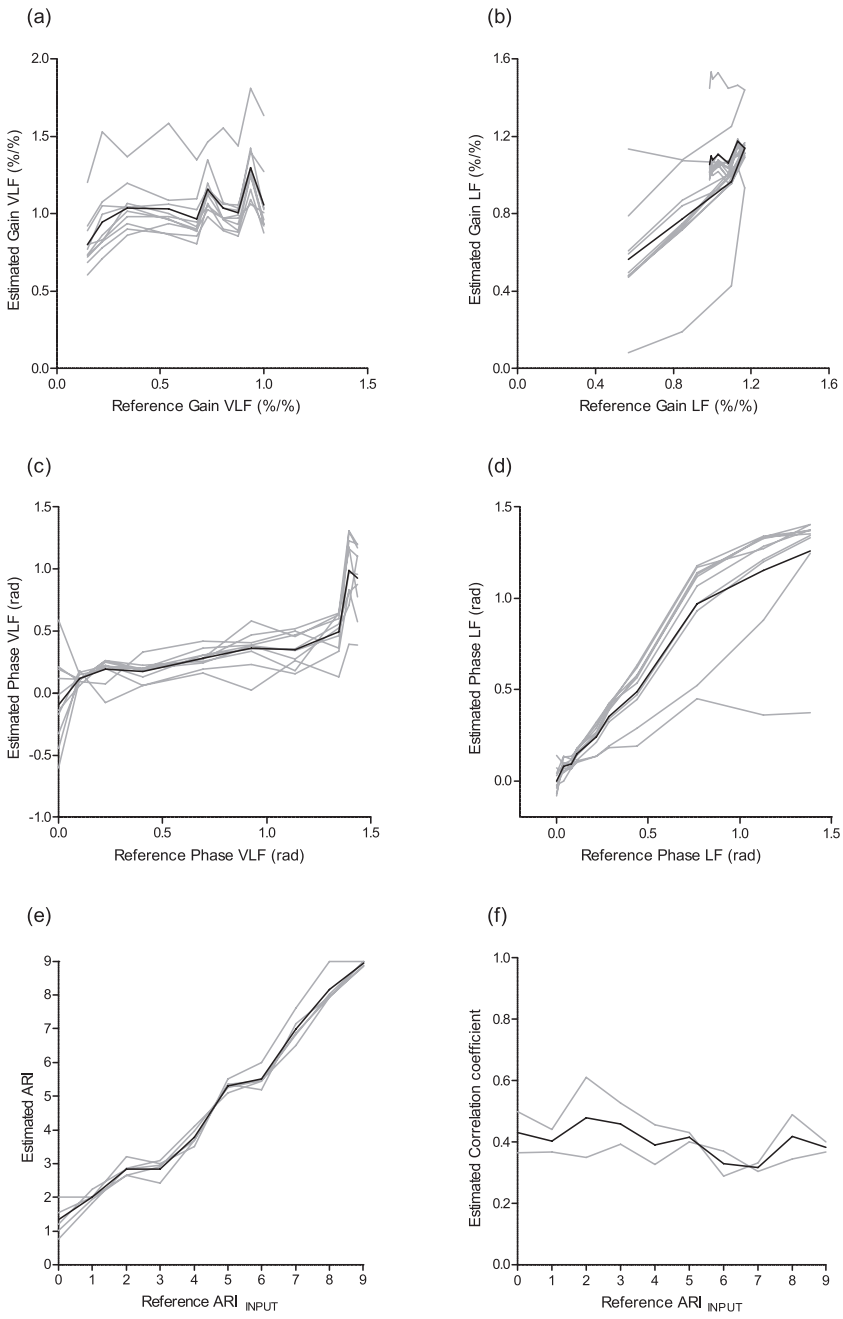

Figure 5. dCA results of individual methods. (a)–(d) Gain (normalised) and phase, values as estimated and reported by centres plotted against the reference gain and phase values, as depicted in figure S1 (stacks.iop.org/PM/39/125002/mmedia), that were determined with standardised CARNet settings (Claassen *et al* 2016) (grey lines) and the average of all methods (black line). Figures (e) and (f): Estimated ARI and correlation coefficient values from individual methods against the reference ARI_{INPUT} value that was used to generate the surrogate CBFV signal. (a): VLF gain; (b): LF gain; (c): VLF phase; (d): LF phase; (e): ARI; (f): correlation coefficient.

band (VLF: 0.02–0.07 Hz, dark grey) for almost all methods. Mean (SD; [range]) ICC for gain was 0.71 (0.10; [0.49–0.86]) and 0.80 (0.17; [0.36–0.94]), respectively for VLF and LF ($p = 0.003$). For phase, the corresponding ICC values were 0.53 (0.21; [0.09–0.80]), and 0.92 (0.13; [0.44–1.00]) respectively ($p < 0.001$). Figure 3 provides ICC values for ARI-based and correlation-based methods, which provide a single output parameter. Mean (SD; [range]) ICC for ARI-like methods was 0.84 (0.19; [0.41–0.94]). For correlation-like methods, the ICC was 0.21 (0.21; [0.056–0.35]).

Relationship between reference gain, phase, ARI_{INPUT} and estimated parameters

In figure 4, box plots were used to plot the individual outcomes of TFA gain and phase, ARI and correlation coefficient values. The outcomes were pooled and compared against the ARI_{INPUT} that was used to generate the surrogate CBFV signals. Overall, a higher ARI_{INPUT} was associated with lower gain and higher phase outcomes

(figures 4(a)–(d)). Variation between the gain parameters was larger, compared to phase, especially LF phase, which showed minimal variation between different methods.

Figure 4(e) depicts a strong relationship between input ARI_{INPUT} and measured ARI, with a stronger relationship for higher ARI values. On the other hand, there was a lack of association between input ARI_{INPUT} and measured correlation coefficient (figure 4(f)).

In figures 5(a)–(f) averages for each parameter (black lines) are presented in combination with the individual estimated outcomes from each method (grey lines) and are compared to corresponding reference gain, phase, and ARI_{INPUT} values. Different methods show similar patterns. In figures 5(a) and (b) one method reported systematically shifted gain VLF results, while the other methods yielded more comparable results. The phase VLF results (figure 5(c)) show more pronounced between-method variability for the lowest and highest phase results. Variability between phase LF measurements (figure 5(d)) increases with increasing phase. In figure 5(e) the estimated and reference ARI show a clear association, however in figure 5(f) any association between estimated correlation coefficient and the reference ARI_{INPUT} is lacking.

Discussion

The main aim of this study was to investigate to what extent shortcomings in methods could be responsible for the high variability (Meel-van den Abeelen *et al* 2014a) and therefore poor reproducibility that have been reported in the literature on dCA measurements using spontaneous oscillations (Birch *et al* 2002, Reinhard *et al* 2003a, Hu *et al* 2008, Brodie *et al* 2009, Gommer *et al* 2010, van Beek *et al* 2010). Since poor reproducibility may be explained by physiological variability, combined with methodological shortcomings, we removed the contribution of physiological variability by generating CBFV signals (surrogate output signals) based on true BP recordings as input signals.

Main findings

Overall, our main finding was that reproducibility, quantified by the ICC, was high for most TFA-like and ARI-like dCA methods for these realistic surrogate data, when compared to previous studies on reproducibility of dCA using physiological data (Birch *et al* 2002, Reinhard *et al* 2003a, Hu *et al* 2008, Brodie *et al* 2009, Gommer *et al* 2010, van Beek *et al* 2010). No single dCA analysis method performed clearly better than others. As a result of these findings, we can conclude that the largest contributors to the longitudinal variability of dCA parameters are more likely to be physiological factors, rather than inherent limitations of analytical methods

Assessment of dCA reproducibility

The choice of ICC as a method to assess reproducibility was based on the difficulty of comparing different methods with multiple outcome parameters. The benefit of using ICC is that it assesses the correlation between repeated measurements in a manner independent of the number or nature of the outcome parameter. On the other hand, one limitation of ICC is that it does not reflect whether these assessments are accurate, i.e. two measurements may both be highly inaccurate but still agree with each other. To account for this, we investigated the agreement of each method by comparing the outcome parameters with the reference ARI (ARI_{INPUT}) that had been used to generate the surrogate CBFV signals and the reference gain and phase derived from ARI_{INPUT} (figures 4 and 5). If a higher ARI (more efficient autoregulation) is used to generate CBFV signals, the estimated gain based on that signal is expected to be lower (Tiecks *et al* 1995). Trends in figures 4(a), (b) and 5(a), (b) show this correlation, however the variation is large for both VLF and LF gain. Similarly, phase is expected to be higher for signals generated with higher ARI (Tiecks *et al* 1995). Figures 4(c), (d) and 5(c) and (d) show this correlation, with a clearly smaller variation for LF phase. The results of phase and gain estimated by the different methods (figure 5) were compared to the reference gain and phase values corresponding to the reference ARI_{INPUT}. Supplemental figure S1 shows the distribution of phase and gain for each ARI_{INPUT}.

Differences between analytical methods

The results of the measured ARI values compared to the input ARI show the expected linear correlation (figure 4(e)) with hardly any deviation from the line of perfect agreement. For the lower ARI values, corresponding to a less efficient or absent autoregulation, the variation is larger. This is in agreement with previous studies describing the increased contribution of the VLF on the ARI estimation for lower ARI values (Elting *et al* 2014b). The added noise in the VLF range leads to an increase in variation. The stronger linear relationship between ARI estimates and ARI_{INPUT} should not in itself be taken as an endorsement of the use of ARI in the assessment of autoregulation in signals recorded from human volunteers. Such a linear relationship might have been expected

as ARI was used to generate the surrogate signals. More recent methods, including multivariate analysis, exhibited similar or poorer reproducibility compared to standard TFA and ARI methods. These methods have been partly proposed to overcome problems from time-varying behaviour or the confounding influence of additional inputs, such as CO_2 (Mitsis *et al* 2004, 2009, Marmarelis *et al* 2013, 2014a, 2014b, Kostoglou *et al* 2014). The simulations did not attempt to emulate these problems and therefore do not show the potential benefit of these methods. The correlation-like methods were underrepresented because only methods that could be applied to short data segments (5 min) were evaluated, but these clearly showed reduced reproducibility compared to the other categories (figures 2 and 3) under these conditions.The results obtained with the correlation index were not related to ARI_{INPUT}, suggesting that the poor reproducibility of these methods for the present dataset may be related to a mismatch in the underlying Tiecks model adopted to generate the surrogate data as explained above. LF phase reproducibility was higher than VLF reproducibility (figure 2(b)). This also applies, to a lesser extent, to the gain results (figure 2(a)). The VLF band is more susceptible to the occurrence of large negative values of phase, due to the phenomenon of 'wrap-around' (Claassen *et al* 2016). Unless these negative values are removed from the mean, estimates of mean phase for the VLF band will be considerably distorted. VLF gain (figure 5(a)) showed a difference between the reference value and the measured values for lower gain values, corresponding to higher ARI_{INPUT} values. This could be the result of inter-centre differences in pre-processing settings (table S7). For example, a lack of mean subtraction, a normalisation over the whole data segment or the use of a VLF band including frequencies below 0.02 Hz could have increased the VLF gain compared to the reference values. Comparing VLF gain and phase line plots (figures 5(a) and (c)) with corresponding values for LF (figures 5(b) and (d)), a smaller variability is observed in the LF band, especially for the lower ARIs and the agreement between measured output and reference gain and phase is improved in the LF band. These differences between the frequency bands are the result of the added noise that was stronger (in relative terms) in the VLF compared to the LF band. This is also reflected in the low coherence results for VLF in S5 table. A further contributor may also be a lower number of VLF oscillations in both BP and CBFV signals or the different distribution of the theoretical ARI curves for VLF compared to LF.

In summary, different analytical methods for assessment of dCA have different characteristics and peculiarities, but none of the methods included in this study showed distinctive superiority regarding the reproducibility of estimates based on surrogate data. As the next step of this ongoing study, we will investigate the performance of these methods when applied to real physiological measurements.

Limitations of the study

We used physiological BP data as input. An alternative could have been to first test purely surrogate data with also surrogate BP signals, say with a wide-band spectral power. That should produce the exact systems parameters regardless of the methods used if the surrogate data is generated by a linear system. However, in a future study we aim to compare reproducibility of physiological data with these surrogate data, and therefore the use of physiological BP data in this set was felt to allow better comparison than if we had used purely synthetic data for both BP and CBFV.

To generate realistic CBFV surrogate data, low-pass filtered random noise was added to the output of Tiecks model. The noise power adopted corresponded to a SNR level of 6 dB as suggested by previous studies (Katsogridakis *et al* 2011). Most results presented above are likely to change with different levels of SNR. On one hand, much noisier measurements will undoubtedly worsen the ICC and the scatter diagrams (figure 5), and on the other, high quality measurements could have much better ICC. Both situations deserve more detailed investigation, but, essentially, would not be expected to change the main conclusions of the study, suggesting that the main sources of poor reproducibility lie with the influences of multiple physiological mechanisms, and not with the assessment methods adopted for quantification of dynamic CA. Similar considerations apply to the choice of 0.1 Hz as cut-off frequency for low-pass filtering the noise added to the CBFV time-series generated with Tiecks model. This choice was based on the power spectral distribution of measured BP and CBFV beat-to-beat values (Zhang *et al* 1998, Mitsis *et al* 2004). Different cut-off frequencies are likely to change the results above, but would lead to less realistic CBFV surrogate signals, unless non-random sources of noise are considered, for example Doppler probes being disturbed in synchronism with respiratory frequency.

The time interval between repeated measurements that were available varied from same day to several weeks. For the purpose of this study it was not essential to have matching intervals -which would reduce interindividual variability- and therefore this limitation was accepted.

It is difficult to find suitable analysis methods for comparing reproducibility between multiple methods with different outcome values and outcome ranges. Not all criteria for using ICC were met (bivariate normal distributions, and equal variances) which may have influenced the results. However, this influence was reduced by using surrogate CBFV data which do not include external physiological influences except for BP and are free from sporadic artefacts that are common in data recorded from human subjects. For example, ICC is sensitive to outliers, but this would be more of an issue when using purely physiological data. The use of the Tiecks model for the

generation of CBFV signals directly influences the results of the nonlinear methods, since this model is a linear time-invariant system. Results may thus be biased towards higher reproducibility for linear analysis methods, since nonlinear (or time-varying or multivariate) models have additional degrees of freedom that allow more variability, in accordance with the general principle of parsimony in fitting models to data. Such methods might outperform linear ones when the assumptions inherent in the use of linear models are not justified in physiologically 'noisy' data.

The correlation-like methods should be extended with other types of correlation methods to make more extensive analysis feasible for this dCA analysis category. Well known correlation-like methods such as Mx (Czosnyka *et al* 1996) were not applicable for this study since longer data recording and ICP data are required.

This study was initiated before the recommendations of the white paper (Claassen *et al* 2016) were formulated. Therefore, the different centres could not be instructed to use standardized settings for TFA. For the purpose of this study, this was not a major limitation because it allowed us better to study the influence of methodological differences on reproducibility. Nonetheless, it also complicated comparisons. For example, there were differences between the frequency band used for the TFA-like analysis. Some centres used different settings than the current white paper recommendations of VLF: 0.02–0.07 Hz; LF: 0.07–0.2 Hz; HF: 0.2–0.5 Hz. This again emphasises the need to apply the white paper standardised settings in future publications on TFA, and to establish consensus on other commonly used methods.

Conclusion

When applied to realistic surrogate data, free from the influences of physiological variability on the BP-CBF relationship, most methods of dCA modelling yield parameters showing ICC values considerably higher than what has been reported for physiological data. This finding suggests that the poor reproducibility reported by previous studies may be mainly due to the inherent physiological variability of CBF regulatory mechanisms, rather than related to (stationary) random noise and the signal analysis methods. Further work is warranted to test this hypothesis, by comparing the performance of different methods using a common set of repeated recordings, aiming to identify methods that could optimise the reproducibility of dynamic CA parameters while at the same time clearly distinguishing between normal and impaired blood flow regulation.

References

Aaslid R, Lindegaard K F, Sorteberg W and Nornes H 1989 Cerebral autoregulation dynamics in humans *Stroke* **20** 45–52

Birch A A, Neil-Dwyer G and Murrills A J 2002 The repeatability of cerebral autoregulation assessment using sinusoidal lower body negative pressure *Physiol. Meas.* **23** 73–83

Brodie F G, Atkins E R, Robinson T G and Panerai R B 2009 Reliability of dynamic cerebral autoregulation measurement using spontaneous fluctuations in blood pressure *Clin. Sci.* **116** 513–20

Caicedo A, Varon C, Hunyadi B, Papademetriou M, Tachtsidis I and van Huffel S 2016 Decomposition of near-infrared spectroscopy signals using oblique subspace projections: applications in brain hemodynamic monitoring *Frontiers Physiol.* **7** 515

Claassen J A, Meel-van den Abeelen A S, Simpson D M, Panerai R B and International Cerebral Autoregulation Research Network 2016 Transfer function analysis of dynamic cerebral autoregulation: a white paper from the International Cerebral Autoregulation Research Network *J. Cereb. Blood Flow Metab.* **36** 665–80

Czosnyka M, Brady K, Reinhard M, Smielewski P and Steiner L A 2009 Monitoring of cerebrovascular autoregulation: facts, myths, and missing links *Neurocrit. Care* **10** 373–86

Czosnyka M, Smielewski P, Kirkpatrick P, Menon D K and Pickard J D 1996 Monitoring of cerebral autoregulation in head-injured patients *Stroke* **27** 1829–34

Elting J W, Aries M J, van der Hoeven J H, Vroomen P C and Maurits N M 2014a Reproducibility and variability of dynamic cerebral autoregulation during passive cyclic leg raising *Med. Eng. Phys.* **36** 585–91

Elting J W, Maurits N M and Aries M J 2014b Variability of the autoregulation index decreases after removing the effect of the very low frequency band *Med. Eng. Phys.* **36** 601–6

Gommer E D, Shijaku E, Mess W H and Reulen J P 2010 Dynamic cerebral autoregulation: different signal processing methods without influence on results and reproducibility *Med. Biol. Eng. Comput.* **48** 1243–50

Grinsted A, Moore J C and Jevrejeva S 2004 Application of the cross wavelet transform and wavelet coherence to geophysical time series *Nonlinear Process. Geophys.* **11** 561–6

Heskamp L, Meel-van den Abeelen A, Lagro J and Claassen J A 2014 Convergent cross mapping: a promising technique for future cerebral autoregulation estimation *IJCNMH.* **1** (Suppl. 1) S20

Hu K, Peng C K, Czosnyka M, Zhao P and Novak V 2008 Nonlinear assessment of cerebral autoregulation from spontaneous blood pressure and cerebral blood flow fluctuations *Cardiovascular Eng.* **8** 60–71

Immink R V, van Montfrans G A, Stam J, Karemaker J M, Diamant M and van Lieshout J J 2005 Dynamic cerebral autoregulation in acute lacunar and middle cerebral artery territory ischemic stroke *Stroke* **36** 2595–600

Katsogridakis E, Dineen N E, Brodie F G, Robinson T G and Panerai R B 2011 Signal-to-noise ratio of bilateral nonimaging transcranial Doppler recordings of the middle cerebral artery is not affected by age and sex *Ultrasound Med. Biol.* **37** 530–8

Kostoglou K, Debert C T, Poulin M J and Mitsis G D 2014 Nonstationary multivariate modeling of cerebral autoregulation during hypercapnia *Med. Eng. Phys.* **36** 592–600

Liu J, Simpson D M and Allen R 2005 High spontaneous fluctuation in arterial blood pressure improves the assessment of cerebral autoregulation *Physiol. Meas.* **26** 725–41

Liu Y and Allen R 2002 Analysis of dynamic cerebral autoregulation using an ARX model based on arterial blood pressure and middle cerebral artery velocity simulation *Med. Biol. Eng. Comput.* **40** 600–5
Liu Y, Birch A A and Allen R 2003 Dynamic cerebral autoregulation assessment using an ARX model: comparative study using step response and phase shift analysis *Med. Eng. Phys.* **25** 647–53
Marmarelis V Z 2004 *Nonlinear Dynamic Modeling of Physiological Systems* (New York: Wiley)
Marmarelis V Z, Shin D C, Orme M and Rong Z 2014a Time-varying modeling of cerebral hemodynamics *IEEE Trans. Biomed. Eng.* **61** 694–704
Marmarelis V Z, Shin D C, Orme M E and Zhang R 2013 Model-based quantification of cerebral hemodynamics as a physiomarker for Alzheimer's disease? *Ann. Biomed. Eng.* **41** 2296–317
Marmarelis V Z, Shin D C, Orme M E and Zhang R 2014b Model-based physiomarkers of cerebral hemodynamics in patients with mild cognitive impairment *Med. Eng. Phys.* **36** 628–37
Meel-van den Abeelen A S *et al* 2014a Between-centre variability in transfer function analysis, a widely used method for linear quantification of the dynamic pressure-flow relation: the CARNet study *Med. Eng. Phys.* **36** 620–7
Meel-van den Abeelen A S, van Beek A H, Slump C H, Panerai R B and Claassen J A 2014b Transfer function analysis for the assessment of cerebral autoregulation using spontaneous oscillations in blood pressure and cerebral blood flow *Med. Eng. Phys.* **36** 563–75
Mitsis G D, Poulin M J, Robbins P A and Marmarelis V Z 2004 Nonlinear modeling of the dynamic effects of arterial pressure and CO_2 variations on cerebral blood flow in healthy humans *IEEE Trans. Biomed. Eng.* **51** 1932–43
Mitsis G D, Zhang R, Levine B D and Marmarelis V Z 2002 Modeling of nonlinear physiological systems with fast and slow dynamics. II. Application to cerebral autoregulation *Ann. Biomed. Eng.* **30** 555–65
Mitsis G D, Zhang R, Levine B D, Tzanalaridou E, Katritsis D G and Marmarelis V Z 2009 Autonomic neural control of cerebral hemodynamics *IEEE Eng. Med. Biol. Mag.* **28** 54–62
Muller M, Bianchi O, Erulku S, Stock C, Schwerdtfeger K and Homburg Traumatic Brain Injury Group 2003 Changes in linear dynamics of cerebrovascular system after severe traumatic brain injury *Stroke* **34** 1197–202
Muller M W and Osterreich M 2014 A comparison of dynamic cerebral autoregulation across changes in cerebral blood flow velocity for 200 s *Frontiers Physiol.* **5** 327
Panerai R B 1998 Assessment of cerebral pressure autoregulation in humans—a review of measurement methods *Physiol. Meas.* **19** 305–38
Panerai R B 2008 Cerebral autoregulation: from models to clinical applications *Cardiovascular Eng.* **8** 42–59
Panerai R B 2014 Nonstationarity of dynamic cerebral autoregulation *Med. Eng. Phys.* **36** 576–84
Panerai R B, Eames P J and Potter J F 2003 Variability of time-domain indices of dynamic cerebral autoregulation *Physiol. Meas.* **24** 367–81
Panerai R B, Rennie J M, Kelsall A W and Evans D H 1998a Frequency-domain analysis of cerebral autoregulation from spontaneous fluctuations in arterial blood pressure *Med. Biol. Eng. Comput.* **36** 315–22
Panerai R B, Simpson D M, Deverson S T, Mahony P, Hayes P and Evans D H 2000 Multivariate dynamic analysis of cerebral blood flow regulation in humans *IEEE Trans. Biomed. Eng.* **47** 419–23
Panerai R B, White R P, Markus H S and Evans D H 1998b Grading of cerebral dynamic autoregulation from spontaneous fluctuations in arterial blood pressure *Stroke* **29** 2341–6
Peng T, Rowley A B, Ainslie P N, Poulin M J and Payne S J 2010 Wavelet phase synchronization analysis of cerebral blood flow autoregulation *IEEE Trans. Biomed. Eng.* **57** 960–8
Reinhard M, Muller T, Guschlbauer B, Timmer J and Hetzel A 2003a Transfer function analysis for clinical evaluation of dynamic cerebral autoregulation—a comparison between spontaneous and respiratory-induced oscillations *Physiol. Meas.* **24** 27–43
Reinhard M, Roth M, Muller T, Czosnyka M, Timmer J and Hetzel A 2003b Cerebral autoregulation in carotid artery occlusive disease assessed from spontaneous blood pressure fluctuations by the correlation coefficient index *Stroke* **34** 2138–44
Reinhard M, Roth M, Muller T, Guschlbauer B, Timmer J, Czosnyka M and Hetzel A 2004 Effect of carotid endarterectomy or stenting on impairment of dynamic cerebral autoregulation *Stroke* **35** 1381–7
Reinhard M, Rutsch S, Lambeck J, Wihler C, Czosnyka M, Weiller C and Hetzel A 2012 Dynamic cerebral autoregulation associates with infarct size and outcome after ischemic stroke *Acta Neurol. Scand.* **125** 156–62
Simpson D M, Panerai R B, Evans D H and Naylor A R 2001 A parametric approach to measuring cerebral blood flow autoregulation from spontaneous variations in blood pressure *Ann. Biomed. Eng.* **29** 18–25
Tiecks F P, Lam A M, Aaslid R and Newell D W 1995 Comparison of static and dynamic cerebral autoregulation measurements *Stroke* **26** 1014–9
Torrence C and Webster P J 1999 Interdecadal changes in the ENSO-monsoon system *J. Clim.* **12** 2679–90
Tzeng Y C, Ainslie P N, Cooke W H, Peebles K C, Willie C K, Macrae B A, Smirl J D, Horsman H M and Rickards C A 2012 Assessment of cerebral autoregulation: the quandary of quantification *Am. J. Physiol.* **303** H658–71
van Beek A H, Claassen J A, Rikkert M G and Jansen R W 2008 Cerebral autoregulation: an overview of current concepts and methodology with special focus on the elderly *J. Cereb. Blood Flow Metab.* **28** 1071–85
van Beek A H, Lagro J, Olde-Rikkert M G, Zhang R and Claassen J A 2012 Oscillations in cerebral blood flow and cortical oxygenation in Alzheimer's disease *Neurobiol. Aging* **33** 428
van Beek A H, Olde Rikkert M G, Pasman J W, Hopman M T and Claassen J A 2010 Dynamic cerebral autoregulation in the old using a repeated sit-stand maneuver *Ultrasound Med. Biol.* **36** 192–201
Zhang R, Zuckerman J H, Giller C A and Levine B D 1998 Transfer function analysis of dynamic cerebral autoregulation in humans *Am. J. Physiol.* **274** H233–41