



## Reliability of supraspinal correlates to lower urinary tract stimulation in healthy participants – A fMRI study



Matthias Walter<sup>a,b,1</sup>, Lorenz Leitner<sup>a,1</sup>, Lars Michels<sup>c,1</sup>, Martina D. Liechti<sup>a</sup>, Patrick Freund<sup>d,e,f,g</sup>, Thomas M. Kessler<sup>a</sup>, Spyros Kollias<sup>c</sup>, Ulrich Mehnert<sup>a,\*</sup>

<sup>a</sup> Department of Neuro-Urology, Balgrist University Hospital, University of Zürich, Zürich, Switzerland

<sup>b</sup> International Collaboration on Repair Discoveries (ICORD), University of British Columbia, Vancouver, Canada

<sup>c</sup> Institute of Neuroradiology, University Hospital Zürich, University of Zürich, Zürich, Switzerland

<sup>d</sup> Spinal Cord Injury Center, Balgrist University Hospital, University of Zürich, Zürich, Switzerland

<sup>e</sup> Department of Neurophysics, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

<sup>f</sup> Department of Brain Repair and Rehabilitation, UCL Institute of Neurology, London, UK

<sup>g</sup> Wellcome Trust Centre for Neuroimaging, UCL Institute of Neurology, London, UK

### ARTICLE INFO

#### Keywords:

Functional magnetic resonance imaging  
Interoception  
Lower urinary tract  
Reliability  
Strong desire to void  
Supraspinal control

### ABSTRACT

Previous functional neuroimaging studies provided evidence for a specific supraspinal network involved in lower urinary tract (LUT) control. However, data on the reliability of blood oxygenation level-dependent (BOLD) signal changes during LUT task-related functional magnetic resonance imaging (fMRI) across separate measurements are lacking. Proof of the latter is crucial to evaluate whether fMRI can be used to assess supraspinal responses to LUT treatments.

Therefore, we prospectively assessed task-specific supraspinal responses from 20 healthy participants undergoing two fMRI measurements (test-retest) within 5–8 weeks. The fMRI measurements, conducted in a 3T magnetic resonance (MR) scanner, comprised a block design of repetitive bladder filling and drainage using an automated MR-compatible and MR-synchronized infusion-drainage device. Following transurethral catheterization and bladder pre-filling with body warm saline until participants perceived a persistent desire to void (START condition), fMRI was recorded during repetitive blocks (each 15 s) of INFUSION and WITHDRAWAL of 100 mL body warm saline into respectively from the bladder. BOLD signal changes were calculated for INFUSION minus START. In addition to whole brain analysis, we assessed BOLD signal changes within multiple ‘a priori’ region of interest (ROI), i.e. brain areas known to be involved in the LUT control from previous literature. To evaluate reliability of the fMRI results between visits, we applied different types of analyses: coefficient of variation (CV), intraclass correlation coefficient (ICC), Sørensen-Dice index, Bland-Altman method, and block-wise BOLD signal comparison.

All participants completed the study without adverse events. The desire to void was rated significantly higher for INFUSION compared to START or WITHDRAWAL at both measurements without any effect of visit. At whole brain level, significant ( $p < 0.05$ , cluster corrected,  $k \geq 41$  voxels) BOLD signal changes were found for the contrast INFUSION compared to START in several brain areas. Overlap of activation maps from both measurements were observed in the orbitofrontal cortex, insula, ventrolateral prefrontal cortex (VLPFC), and inferior parietal lobe. The two highest ICCs, based on a ROI's mean beta weight, were 0.55 (right insular cortex) and 0.47 (VLPFC). Spatial congruency (Sørensen-Dice index) of all voxels within each ROI between measurements was highest in the insular cortex (left 0.55, right 0.44). In addition, the mean beta weight of the right insula and right VLPFC demonstrated the lowest CV and narrowest Bland and Altman 95% limits of agreement.

In conclusion, the right insula and right VLPFC were revealed as the two most reliable task-specific ROIs using our automated, MR-synchronized protocol. Achieving high reliability using a viscerosensory/interoceptive task such as repetitive bladder filling remains challenging and further endeavour is highly warranted to better understand which factors influence fMRI outcomes and finally to assess LUT treatment effects on the supraspinal level.

\* Corresponding author. Department of Neuro-Urology, Balgrist University Hospital, University of Zürich, Forchstrasse 340, 8008, Zürich, Switzerland.

E-mail address: [ulrich.mehnert@balgrist.ch](mailto:ulrich.mehnert@balgrist.ch) (U. Mehnert).

<sup>1</sup> Equal contribution to manuscript, i.e. shared first authorship.

<https://doi.org/10.1016/j.neuroimage.2019.02.031>

Received 4 December 2018; Received in revised form 2 February 2019; Accepted 13 February 2019

Available online 15 February 2019

1053-8119/© 2019 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The human lower urinary tract (LUT) has two important homeostatic functions: 1) low pressure continent and symptom-free storage of urine at adequate volumes, and 2) periodical, self-determined and complete release of the stored urine. To accomplish both, the LUT relies on a complex interplay of the central, peripheral and autonomic nervous system that is under supraspinal control. The latter, in particular, is relevant for the voluntary regulation of LUT function, which in turn is important for the appropriate adaption of LUT function to external circumstances, i.e., choosing safe and acceptable time point to empty the bladder.

Previous neuroimaging studies using positron emission tomography (PET) (Kavia et al., 2005; Griffiths and Tadic, 2008) and functional magnetic resonance imaging (fMRI) (Griffiths and Tadic, 2008; Fowler and Griffiths, 2010; Arya et al., 2017) have provided evidence of a supraspinal network (SSN) that is involved in LUT control. The SSN comprises the anterior cingulate cortex (ACC), basal ganglia (caudate nucleus, putamen, and globus pallidus), cerebellum, hypothalamus, insula, periaqueductal grey (PAG), parts of the prefrontal cortex (PFC), pons, supplementary motor area (SMA), and thalamus (Kavia et al., 2005; Fowler et al., 2008; Griffiths and Tadic, 2008; Fowler and Griffiths, 2010; Arya et al., 2017). Certainly, detection of these brain regions varied among studies depending on the investigational protocol and applied task.

Nowadays, fMRI has become the main neuroimaging modality to investigate supraspinal LUT control including post-treatment evaluation (Dasgupta et al., 2005; Blok et al., 2006; Kavia et al., 2010; Pontari et al., 2010). Despite previous studies investigating the reproducibility of fMRI findings using auditory (Wei et al., 2004; Caceres et al., 2009), memory (Caceres et al., 2009; Plichta et al., 2012), motor (Gountouna et al., 2010; Quiton et al., 2014) or pain tasks (Quiton et al., 2014), there is a paucity of reliability data in the context of interoceptive tasks, particularly in regard to the LUT. Yet, there is a single study providing data on this topic but only for same session repeat investigations (Clarkson et al., 2017).

However, information on separate session reliability is important for longitudinal studies and the evaluation of LUT treatment effects on supraspinal level. Therefore, our aim was to evaluate the reliability of supraspinal responses to bladder filling in healthy participants using our automated fMRI protocol (Jarrahi et al., 2016; Leitner et al., 2017). Considering high accuracy in timing and volume delivery using a standardized, automatic fMRI protocol, we hypothesized that our approach will generate reliable activations of literature-based supraspinal areas involved in LUT control.

## 2. Participants and methods

### 2.1. Study design

This is a prospective single center neuroimaging study conducted at the University of Zürich, Zürich, Switzerland. The study comprises two MRI measurements at two separate visits with an interval of 5–8 weeks (Fig. 1).

### 2.2. Ethics

This study has been approved by the local ethics committee (Kantonale Ethikkommission Zürich, KEK-ZH-Nr. 2011-0346) and was performed in accordance to the World Medical Association Declaration of Helsinki (World Medical Association, 1964) and the guidelines of the Swiss Academy of Medical Sciences (2009). Furthermore, handling of all personal data strictly complied with the federal law of data protection in Switzerland (The Federal Authorities of the Swiss Confederation, 1992, Swiss Academy of Medical Sciences, 2009). This study has been registered at [clinicaltrials.gov](http://www.clinicaltrials.gov) (<http://www.clinicaltrials.gov/ct2/show/NC T01768910>).

### 2.3. Participants

We recruited a cohort of 20 healthy participants within an age range of 18–55 years. All participants provided written informed consent prior to study inclusion.

Exclusion criteria were LUT symptoms (LUTS), urological or neurological pathology, current or recurrent urinary tract infection (UTI), hematuria, pregnancy, previous surgery for urological or neurological reasons, regular medication intake (except contraceptives), and incompatibility with MRI safety rules (e.g. ferromagnetic or electromagnetic implants). This was assessed on the basis of a complete medical history, vital signs, physical and neurological examinations (including examination of urogenital sensation, bulbocavernosus reflex, anal reflex, anal sphincter tone, and anal squeeze response), urine analysis, urodynamic investigation, Mini Mental State Examination (MMSE) (Folstein et al., 1975), the Hospital Anxiety and Depression Scale (HADS) (Zigmond and Snaith, 1983), and a 3-day bladder diary using predefined cut-offs (Table 1). The bladder diaries were completed over three consecutive days, recording the time points and volumes (mL) of drinking and micturition, as well as the number of incontinence episodes, pad usage, and pain levels associated with urine storage and/or micturition (0–10). Additionally, standardized urological questionnaires such as the International Consultation on Incontinence Questionnaire (ICIQ) for LUT

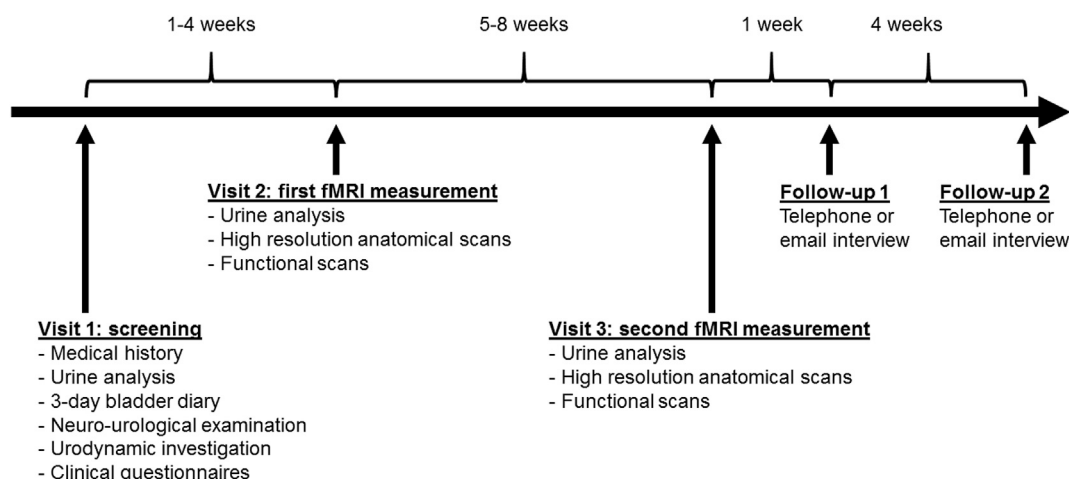


Fig. 1. Sequence of study visits and procedures.

**Table 1**

Baseline characteristics of healthy participants. All participants fulfilled pre-defined cut-offs for study inclusion: MMSE score  $\geq 24$ , HADS  $\leq 7$  each, 3-day bladder diaries:  $\frac{24h \text{ urinary frequency}}{\text{drinking volume [mL]}} \leq 0.0045$  with a maximum of  $1 \times$  nocturia, mean volume per void  $>150$  mL and absence of urinary incontinence or urgency.

	All participants (n = 20)	Women (n = 10)	Men (n = 10)
<b>Mean age <math>\pm</math> SD [years]</b>	39.2 $\pm$ 11.6	34.3 $\pm$ 10.2	44.1 $\pm$ 11.3
<b>3-day bladder diary</b>			
Mean micturition frequency per 24 h $\pm$ SD	6.0 $\pm$ 1.0	5.5 $\pm$ 1.0	6.5 $\pm$ 1.0
Mean micturition volume per micturition [mL] $\pm$ SD	337 $\pm$ 97	355 $\pm$ 108	318 $\pm$ 86
Mean fluid intake per 24 h [mL] $\pm$ SD	2246 $\pm$ 882	2013 $\pm$ 831	2478 $\pm$ 911
Mean incontinence episodes per 24 h	0	0	0
<b>Questionnaires (mean scores <math>\pm</math> SD)</b>			
ICIQ-FLUTS/MLUTS (score range 0–48/0–56)	–	2.5 $\pm$ 1.3	3.8 $\pm$ 3.6
<b>OAB-q SF</b>			
Symptoms (score range 6–36)	7.9 $\pm$ 1.5	7.6 $\pm$ 1.4	8.2 $\pm$ 1.7
QoL (score range 13–78)	14.1 $\pm$ 1.0	14.0 $\pm$ 0.8	14.1 $\pm$ 1.2
<b>HADS</b>			
Anxiety (score range 0–21, normal cut-off $\leq 7$ )	2.1 $\pm$ 1.7	2.9 $\pm$ 1.9	1.2 $\pm$ 1.0
Depression (score range 0–21, normal cut-off $\leq 7$ )	1.0 $\pm$ 1.5	0.9 $\pm$ 1.7	1.0 $\pm$ 1.3
MMSE (score range 0–30, normal cut-off $\geq 24$ )	29.5 $\pm$ 0.5	29.4 $\pm$ 0.5	29.5 $\pm$ 0.5
<b>Neuro-urological examination</b> (no. participants with intact or normal findings/no. participants with impaired findings)			
Urogenital sensation	20/0	10/0	10/0
Anal reflex	20/0	10/0	10/0
Anal sphincter tone	20/0	10/0	10/0
Voluntary anal sphincter contraction	20/0	10/0	10/0
Bulbocavernosus reflex	20/0	10/0	10/0

HADS Hospital Anxiety and Depression Scale, ICIQ-FLUTS International Consultation on Incontinence Questionnaire Female Lower Urinary Tract Symptoms, ICIQ-MLUTS International Consultation on Incontinence Questionnaire Male Lower Urinary Tract Symptoms, MMSE Mini Mental State Examination, OAB-q SF Overactive Bladder Questionnaire short form, QoL quality of life, SD standard deviation.

symptoms in females (ICIQ-FLUTS) and males (ICIQ-MLUTS) (Jackson et al., 1996; Donovan et al., 2000), and the Overactive Bladder Questionnaire short form (OAB-q SF) (Coyné et al., 2015) were completed.

#### 2.4. Pre- and post-measurement procedures

All participants were asked to abstain from any product containing caffeine or nicotine for at least six hours prior to each fMRI measurement visit. At the beginning of each fMRI measurement visit a urinalysis was performed to exclude urinary tract infection or pregnancy. Prior to entering the magnetic resonance (MR) scanner, all participants removed any ferromagnetic items and changed clothing to standardized clinical scrubs. All participants were trained to use a MR-compatible handheld response system (Jarrahi et al., 2013) in order to rate their desire to void on a numeric rating scale (NRS, ranging from 0 = no desire to void at all to 10 = strongest desire to void), which was presented on the video screen to every participant.

Prior to starting the fMRI paradigm, a 14 Fr Foley catheter (Uromed, Germany) was placed in the bladder. Then, the bladder was emptied and the balloon of the catheter was filled with 5 mL sodium chloride solution 0.9% (Braun, Germany) to avoid any unexpected dislocation. Subsequently, the bladder was manually prefilled until each participant reported a persistent desire to void, i.e. corresponding to a NRS score of 6. The individual prefilled bladder volume applied for the 1st fMRI

measurement was also used for the 2nd fMRI measurement to ensure comparable measurement conditions.

Before removing the Foley catheter at the end of each fMRI measurement, the bladder was emptied to record the corresponding urine volume. To screen for adverse events, each participant was interviewed via telephone or through email within one week and one month after the 2nd fMRI measurement visit (Fig. 1).

#### 2.5. Acquisition of neuro-imaging data

Structural and functional MRI data were recorded on a Philips Ingenia 3 Tesla scanner (Philips Medical Systems, Best, the Netherlands) using a 15-channel Philips Sense head coil.

##### 2.5.1. Anatomical imaging

Anatomical data were acquired with a high-resolution 3D T1-weighted turbo field gradient echo (TFE) pulse sequence covering the entire cerebrum as well as the cervical spinal cord. The imaging parameters were: isotropic  $1 \text{ mm}^3$  resolution, field of view  $256 \times 256 \times 180$  mm, matrix  $256 \times 256$  mm, 180 slices (scan time of 305 s), repetition time (TR) = 6.92 ms, echo time (TE) = 3.1 ms, slice thickness = 1 mm, no inter-slice gap, and flip angle =  $8^\circ$ .

##### 2.5.2. Functional MRI paradigm

The fMRI paradigm (Fig. 2) comprised eight blocks of repetitive filling (INFUSION condition) and draining (WITHDRAWAL condition) of 100 mL body warm saline. In order to precisely fill and drain the bladder, i.e. to provide exact volume delivery and timing, we used a custom-made automated MR-compatible and MR-synchronized infusion-drainage device (IDD) (Jarrahi et al., 2016; Leitner et al., 2017). The blocks of INFUSION (15 s) and WITHDRAWAL (15 s) were interspersed with PLATEAU (9s), REST (7–9s, mean 8s), and RATING (15s) periods (Fig. 2). REST periods served as interlude to allow the blood oxygenation level-dependent (BOLD) signal to return to baseline, hence avoiding any overlap from motor activity from the rating into a new block of filling. During the RATING periods that followed each INFUSION and WITHDRAWAL period, healthy participants rated their desire to void.

##### 2.5.3. Functional imaging

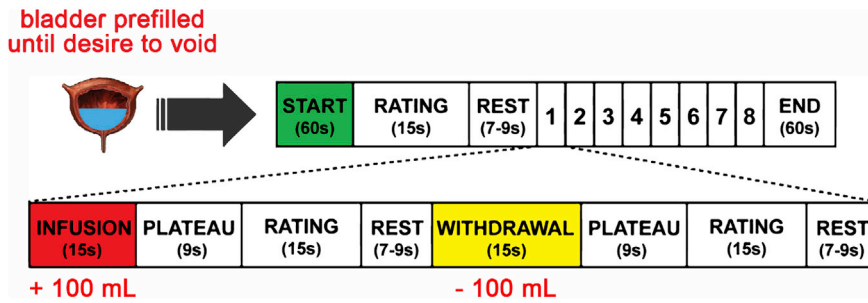
Functional time series were acquired with a field echo, i.e. gradient echo, echo planar imaging (FEEPI) pulse sequence. Thirty-four axial slices covered the entire cerebrum in ascending order. Other scan parameters were as follows: TR = 2000 ms, TE = 16 ms, slice thickness = 3 mm, inter-slice gap = 1 mm, flip angle =  $80^\circ$ , field of view =  $240 \times 135 \times 240$  mm, and reconstructed image matrix  $80 \times 80$  voxels, voxel resolution:  $3 \times 3$  mm. All images were obtained in an oblique axial orientation covering the entire brain including the cerebellum and rostral brainstem (including the pons). The task comprised 445 dynamic scans (scan duration of 902 s). The first four scans were dummy scans to reach steady-state magnetization.

#### 2.6. Statistical analyses of clinical and behavioural data

Statistical analyses of clinical and behavioural data were performed using IBM's Statistical Package for the Social Sciences (SPSS) version 25.0 (Armonk, New York, USA). Normal distribution of data and model residuals was tested using Shapiro-Wilk test and Q-Q plots. Depending on distribution, parametric or non-parametric tests were chosen accordingly. Bonferroni correction was used to account for multiple comparisons. The threshold for a significant difference was  $p < 0.05$ .

Descriptive statistics were used for clinical parameters. Results are presented as means and standard deviations (SD) for continuous, e.g. age, 3-day bladder diary, and questionnaires, and absolute numbers for binominal data, e.g. intact or impaired for results from the neuro-urological examination.

To test the effect of time and condition on the rating of the desire to



**Fig. 2.** Scan paradigm of the task-related functional magnetic resonance imaging (fMRI). Prior to the start of the task-related fMRI the bladder was manually pre-filled via a Foley catheter until a persistent desire to void was present. The actual functional paradigm begins with the baseline condition START (60s, no specific stimulus or task is performed), followed by a baseline rating of desire to void, and a short REST condition randomly shifting between 7 and 9 s (jitter). Subsequently, eight repetitive blocks were applied comprising the following conditions: (1) automated INFUSION of 100 mL warm saline, (2) PLATEAU condition (full bladder at stable volume, no further infusion), (3) RATING of desire to void, (4) REST condition randomly shifting between 7 and 9 s (jitter), (5) automated WITHDRAWAL of 100 mL, (6) PLATEAU condition (bladder at stable volume, no further withdrawal), (7) RATING of desire to void, and (8) REST condition randomly shifting between 7 and 9 s (jitter).

void, linear mixed models were employed. First, the interaction effect between the condition (i.e. INFUSION, WITHDRAWAL, and START) and time-point (i.e. 1st and 2nd fMRI measurement) on the rating of the desire to void was assessed. Rating was set as the dependent variable, while condition and time-point were set as independent variable. Random subject effects were included in this model. In case non-significant interaction effect, main effects of condition and time-point were tested in separate models.

Wilcoxon signed rank test was used for pair-wise comparison of post-measurement urine volume between both fMRI measurements.

## 2.7. Preprocessing of neuro-imaging data

Functional and anatomical data were pre-processed and analyzed in MATrix LABoratory (MATLAB) version 2016b (The MathWorks, Inc., Natick, Massachusetts, USA) using Statistical Parametric Mapping (SPM) version 8 (Wellcome Trust Center for Neuroimaging, University College London, London, UK). First, T1-weighted images were reoriented to set the image origin to the anterior commissure ( $x=0$ ,  $y=0$ , and  $z=0$ ) within the Montreal Neurological Institute (MNI) standard space. The images were initially realigned to the first scan, and then unwarped to control for movements and susceptibility-induced image distortions and temporally smoothed. Following co-registration of the anatomical and functional images, spatial normalisation of the functional images was performed (Friston et al., 1995). Using a unified segmentation approach, individual brains were normalized to the MNI standard space. Finally, spatial smoothing was conducted by applying an isotropic 8-mm full-width-at-half-maximum Gaussian kernel to reduce image noise. Temporal band pass filtering (128 s) was additionally applied to each data set. Prior to data analysis, fMRI data from each participant was visually screened for movement artefacts (comparing translation and rotation values over time).

## 2.8. Analysis of neuro-imaging data

A voxel-wise general linear model (GLM) (Friston et al., 1994) was used for the first-level analysis (i.e., within-participant) in order to calculate contrast images for each condition separately (i.e., INFUSION minus START) and for each fMRI measurement. Movement parameters were used as regressor of no interest in the subsequent first-level GLM. Significant changes in BOLD signal were identified using a repeated boxcar model convolved with a canonical form of the hemodynamic response function (HRF) (Friston et al., 1994). The second-level analysis was performed to identify specific pattern of activation for the entire group. Results were corrected for age, sex and total intracranial volume (included as covariates of no-interest in the GLM) to account for the between participant variability. One sample  $t$ -test was conducted to reveal BOLD signal changes of the contrast INFUSION minus START for

each time point (i.e., 1st and 2nd fMRI measurement). A paired  $t$ -test was used for pairwise comparison of each contrast between both time points (i.e., between both fMRI measurements). Results of whole brain analysis are shown at  $p < 0.05$  (cluster-corrected,  $k \geq 41$  voxels) (Slotnick et al., 2003).

### 2.8.1. Region of interest analysis

According to the neuroimaging literature on supraspinal LUT control (Fowler et al., 2008; Griffiths and Tadic, 2008; Fowler and Griffiths, 2010), we selected the most frequently reported areas for a region of interest (ROI) analysis: 1) dorsolateral PFC (DLPFC: BA 8, 9, and 46), 2) orbitofrontal/frontopolar PFC (OFC: BA 10 and 11), 3) ventrolateral PFC (VLPFC: BA 44, 45, and 47), 4) ACC, 5) insula, 6) basal ganglia, 7) hypothalamus, 8) thalamus, 9) PAG, 10) pons, 11) cerebellum, and 12) SMA. For each ROI, we applied a small volume correction. All ROIs were taken from the WFU PickAtlas (Advanced Neuroscience Imaging Research Laboratory, Wake Forest University School of Medicine, Winston-Salem, North Carolina, USA). The ROIs were included as a mask in order to restrict the voxel-by-voxel statistical analysis to pre-specified brain areas using a threshold of  $p < 0.05$  familywise error (FWE) corrected.

### 2.8.2. Reliability analysis

To evaluate reliability of the fMRI results between visits, we applied different types of analyses: coefficient of variation (CV) (Koopmans et al., 1964), intraclass correlation coefficient (ICC) (Shrout and Fleiss, 1979), Sørensen-Dice index (Dice, 1945), Bland and Altman 95% limits of agreement (Bland and Altman, 1986), and block-wise BOLD signal comparison.

Two different types of statistic reflect the scaling of the measurement error. One type is the CV, where the error variance is scaled by the magnitude of activation. The CV (in %) was calculated across repeated measurement sessions using the following formula for each individual:

$$CV = \left( \frac{\sqrt{\frac{\sum_{i=1}^k SD_i^2 * (n_i - 1)}{\sum_{i=1}^k n_i - k}}}{\frac{\sum_{i=1}^k mean_i}{k}} \right) * 100$$

where  $mean_i$  denotes the mean of measurement  $i$ ,  $SD_i$  denotes the standard deviation of measurement  $i$ ,  $n_i$  denotes the number of trials at measurement  $i$ , and  $k$  denotes the total number of measurements.

The other type of scaling of the measurement error is the ICC. First, we compared BOLD signal similarities between the 1st fMRI and 2nd fMRI across the whole brain at a voxel-to-voxel level. Then, we calculated reliability maps for the third ICC (ICC(3,1)) according to Shrout and Fleiss (1979):



$$ICC(3,1) = \frac{BMS - EMS}{BMS + (k-1)EMS}$$

This equation estimates the correlation of the participant signal intensities between sessions, modelled by a two-way ANOVA, with random participant effects and fixed session effects. In this model, the total sum of squares is split into participant (BMS), session (JMS) and error (EMS) sums of squares; and  $k$  is the number of repeated sessions. The main advantage of ICC(3,1) is that it assesses only the level of consistency between measurements (McGraw and Wong, 1996). ICC(3,1) is fully determined by the group effect ( $F=BMS/EMS$ ). Session effects ( $F = JMS/EMS$ ) and the group activation for the 1st session are considered as distinct factors. The latter, the group signal, is estimated by:

$$t = \frac{\text{mean}(X1) - \sqrt{((n-1)/SD(X1))}}$$

where with  $\times 1$  is the 1st session data, and  $n$  is number of participants. The assessment of a possible relationship between group activation and reliability was obtained from the joint probability distribution  $f(ICC, t)$  (Caceres et al., 2009).

ICC results range between 0 (no reliability) and 1 (maximum reliability) and are classified as poor ( $<0.4$ ), fair (0.41–0.59), good (0.6–0.74), and excellent (0.75–1) (Cicchetti, 1994).

Subsequently, ICC was calculated (in SPSS) at ROI level using a two mixed-model to investigate the agreement between parameter estimates (beta weights) averaged across all scan blocks for the contrast “INFUSION minus START” within each a priori ROI between the two scan time points  $t_1$  (1st fMRI) and  $t_2$  (2nd fMRI). The results are reported for the type absolute agreement [rater 1 (data from  $t_1$ ) “agrees” with rater 2 (data from  $t_2$ ), single measures] with a confidence interval of 95%.

In order to gain insight into the spatial congruence of activations at ROI level, the relative overlap of activations within each ROI between  $t_1$  and  $t_2$  was determined by calculating the Sørensen-Dice index for the 2nd-level group data from pairs of activation maps (from the two scan time points  $t_1$  and  $t_2$ ):

$$R_{\text{overlap}} = \frac{2 * V_{\text{overlap}}}{V_1 + V_2}$$

where  $V_{\text{overlap}}$  represents the number of voxels commonly activated in each ROI at  $t_1$  and  $t_2$ , and  $V_1 + V_2$  represent the number of voxels that were activated at  $t_1$  or  $t_2$  respectively. The ratio of commonly activated voxels and the sum of  $V_1 + V_2$  was calculated from activation maps that were limited to  $p \leq 0.001$  (uncorrected for multiple comparisons). The Sørensen-Dice index can range from 0 (no overlap) to 1 (perfect overlap) and is independent of the height of the  $t$ -values, once voxels have passed the threshold.

In addition to the ICC, Bland and Altman 95% limits of agreement were chosen to visualise the agreement between the parameter estimates (beta weights) averaged across all scan blocks for the contrast “INFUSION minus Start” within the selected ROIs (named  $S$ ) between the two scan time points  $t_1$  and  $t_2$ :

$$S(t_1, t_2) = \left( \frac{t_1 + t_2}{2}, t_1 - t_2 \right)$$

Subsequently, a linear regression analysis for each ROI was performed using the difference in activation between  $t_1$  and  $t_2$  as dependent variable and the mean activation of  $t_1$  and  $t_2$  as independent variable. Any significant effect (one-sample  $t$ -test) would indicate a proportional bias, i.e. that the number of data points (= mean activation for a participant for a given ROI) are not equally distributed around the mean activation across all participants.

Finally, we extracted for each participant the BOLD signal time courses of both scan time points for each ROI and filling block. For this analysis, we used the pre-processed fMRI data (including realignment, unwarping, normalisation, spatial smoothing, and temporal band pass filtering (128 s)). A repeated-measures ANOVA (using the mean activity across all time points within each block) with the factors time and block was applied in order to test for main effects of scan time point and block

as well as interaction effects. In case of a main effect of block, we used post-hoc paired  $t$ -tests comparing the differences of the fMRI signal between blocks within each scan time points. In case of a significant main effect of time or time  $\times$  block interaction, we would apply a paired  $t$ -test to compare changes of the pre-processed fMRI signal for each block between  $t_1$  and  $t_2$ .

### 3. Results

#### 3.1. Participants

Twenty right-handed healthy participants, 10 women and 10 men, mean age  $39 \pm 12$  years (range 22–54) were included in this study. As per our inclusion criteria, no pathologic findings with respect to 3-day bladder diary, questionnaires and neuro-urological examination were detected (Table 1). All participants completed the study. No adverse events were reported, neither during nor after the MR measurement itself or at follow-up.

#### 3.2. Desire to void and post fMRI bladder volume

The prefilled bladder volume prior to each fMRI measurement was  $453 \pm 153$  mL. The linear mixed model revealed a non-significant interaction effect (condition  $\times$  time-point) on the rating of desire to void ( $F = 2.9$ ,  $df: 2$ ,  $p = 0.061$ ). Moreover, a significant main effect of condition on the desire to void rating ( $F = 126.1$ ,  $df: 2$ ,  $p < 0.001$ ) was detected. In contrast, there was no significant effect of time-point (i.e. visits) on the desire to void rating ( $F = 1.4$ ,  $df: 1$ ,  $p = 0.237$ ). Significant differences between all conditions were found for each fMRI measurement. INFUSION induced the highest rating of desire to void, (1st fMRI:  $7.9 \pm 1.5$  and 2nd fMRI:  $7.8 \pm 1.1$ ) while WITHDRAWAL resulted in the lowest rating ( $4.5 \pm 1.6$  and  $3.4 \pm 1.8$ ). Baseline rating after bladder pre-filling (START) was  $6.3 \pm 0.6$  and  $6.2 \pm 0.5$ , respectively.

The post fMRI bladder volume was significantly higher ( $p = 0.043$ ) after the 2nd fMRI ( $511 \pm 174$  mL) compared to the 1st fMRI ( $491 \pm 170$  mL).

#### 3.3. Supraspinal responses to automated, repetitive LUT stimulation

Head motion (translation) was below 1.5 mm (i.e. half of the recorded voxel resolution) for each participant and fMRI run; hence, all participants were included in this study. At whole brain level, significant BOLD signal changes to bladder filling (= contrast INFUSION minus START) were detected in several brain areas during both fMRI measurements (Table 2). Results of the whole brain analysis are shown at  $p < 0.05$  (cluster-corrected) (Slotnick et al., 2003).

An illustration of the brain areas from the 1st and 2nd fMRI measurement as listed in Table 2 and their spatial overlap is demonstrated in Fig. 3.

#### 3.4. Reliability analysis

The whole brain ICC across both time points is shown in Fig. 4, some voxel locations showed an excellent ICC (for illustration purpose, we only report voxels with an  $ICC \geq 0.75$ ).

At ROI level, areas of significant ( $p = 0.05$ , FWE corrected) BOLD response within each of the 12 selected ROIs in response to bladder filling are summarized in Table 3. Significant BOLD responses were detected in 11 of 12 ROIs during the 1st fMRI and in 8 of 12 ROIs during the 2nd fMRI. The following eight ROIs demonstrated significant BOLD responses during both visits (Table 3): DLPFC, OFC, VLPFC, ACC, left insula, right insula, hypothalamus, and left basal ganglia. The ICC analysis, based on the mean beta weight of these eight ROIs, revealed values between  $-0.33$  (ACC) and  $0.55$  (right insula).

The relative overlap of activations between 1st and 2nd fMRI, calculated using the Sørensen-Dice index, showed the highest voxel

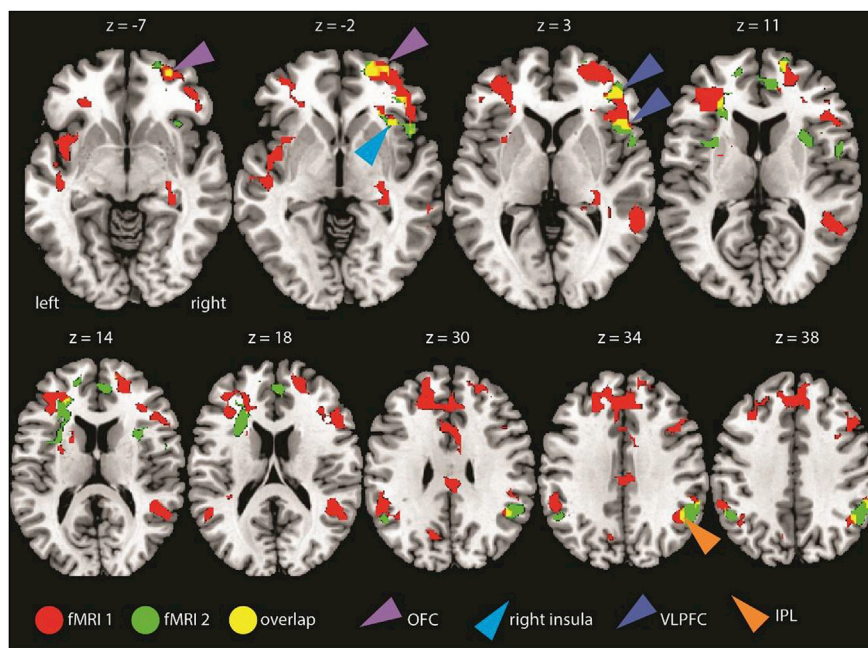
**Table 2**

Whole brain (voxel-wise) analysis of supraspinal areas with significant ( $p < 0.05$  (cluster-corrected,  $k \geq 41$  voxels)) BOLD signal changes during bladder filling (= contrast INFUSION minus START).

	Brain region	BA	Cluster size	T value	MNI peak coordinates		
					x	y	z
1 <sup>st</sup> fMRI	OFC*	10	3175	11.92	24	53	1
	Occipitotemporal cortex, fusiform gyrus	37	630	6.68	57	-49	4
	Inferior parietal lobe, angular gyrus*	39	441	7.39	-51	-46	31
	Posterior midcingulate gyrus	23	104	4.85	3	-25	31
	Dorsal posterior cingulate gyrus	31	88	4.96	-18	-61	25
2 <sup>nd</sup> fMRI	VLPPFC (pars triangularis)*	45	329	7.14	48	20	4
	OFC*	10	255	6.28	9	47	13
	Inferior parietal lobe, angular gyrus	39	212	8.87	51	-49	31
	VLPPFC (pars opercularis)*	44	190	6.49	-30	14	19
	Inferior parietal lobe, angular gyrus*	39	71	6.38	-51	-52	40

BA Brodmann area, BOLD blood oxygenation level-dependent, fMRI functional magnetic resonance imaging, MNI Montreal Neurological Institute, OFC orbitofrontal/ frontopolar prefrontal cortex, VLPPFC ventrolateral prefrontal cortex.

The asterisk (\*) indicates brain regions that show an overlap between both fMRI sessions (see Fig. 3). In this regard it has to be considered that the orbitofrontal cortex of the 1st fMRI is a large cluster of 3175 voxel that spreads widely beyond its peak activation, covering adjacent structures.



**Fig. 3.** Group-level maps (axial slices) of brain areas with significantly ( $p < 0.05$  (cluster-corrected)) increased BOLD signal during bladder filling (= contrast “INFUSION minus START”) from 1st fMRI (red) and 2nd fMRI (green). Overlap between identified brain areas is depicted in yellow. fMRI functional magnetic resonance imaging, IPL inferior parietal lobe; OFC orbitofrontal/frontopolar prefrontal cortex; VLPPFC ventrolateral prefrontal cortex.

overlap, i.e. the highest spatial congruence of activations, for the left ( $R_{\text{overlap}} = 0.55$ ) and right ( $R_{\text{overlap}} = 0.44$ ) insular cortex (Table 3).

The individual (per participant) CV between both fMRI measurements using the mean beta weights of each ROI that demonstrated significant BOLD responses in both fMRI measurements is displayed in Fig. 5. The CV ranged between -20.8% (basal ganglia left) and 26.3% (hypothalamus) (Table 4). The smallest CV spread in conjunction with a low median CV can be identified for the right insula and VLPPFC (Table 4).

The Bland and Altman 95% limits of agreement of the beta weights within each of those eight ROIs in relation to the scan time points are shown in Fig. 6. None of the eight ROIs showed a proportional bias (all  $p > 0.05$ ).

BOLD signal responses for each of the eight INFUSION blocks are shown in Fig. 7. Repeated-measures ANOVA did not reveal a main effect of block (at t1 or t2), time, or a time\*block interaction.

#### 4. Discussion

The results of this study demonstrate that using our automated, MR-synchronized bladder filling protocol, we were able to elicit significant supraspinal responses during two separate fMRI measurements. 8 of 12 predefined ROIs showed significant BOLD responses during both fMRI measurements: DLPFC, OFC, VLPPFC, ACC, left insula, right insula, hypothalamus, and left basal ganglia. The right insular cortex and the VLPPFC are the supraspinal ROIs that show the most reliable response to our automated bladder filling task.

The essential role the brain plays in the control of the LUT is beyond controversy and neuroimaging studies have largely helped to visualise and better understand the supraspinal correlates of LUT control in physiological but also pathophysiological context (Blok, 2002; Holstege, 2005; Fowler et al., 2008). Despite this gain in knowledge, clinical practice has not much changed in this regard yet. To increase the impact

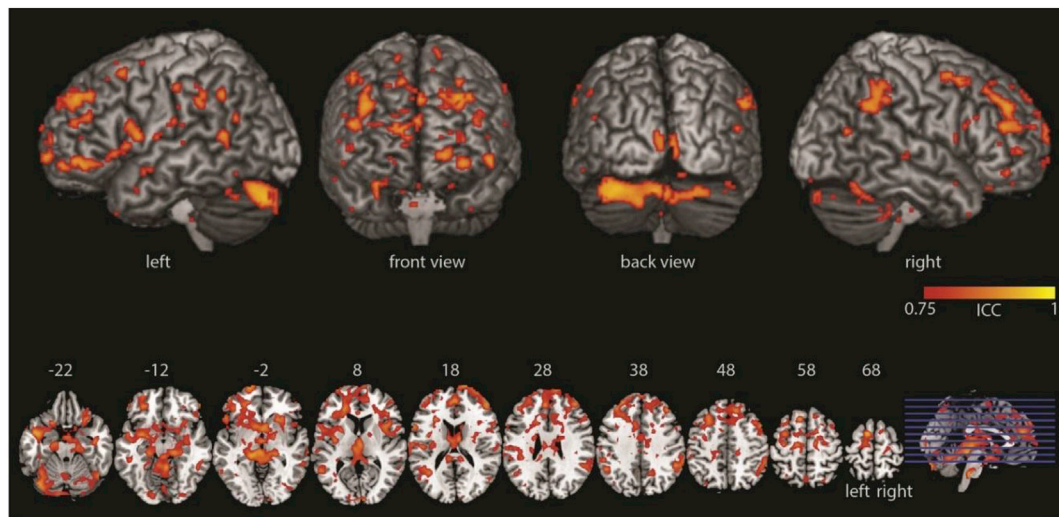


Fig. 4. Results of the whole brain intraclass correlation coefficient (ICC) analysis between the two scan time points for the contrast “INFUSION minus START”. Only regions with an ICC  $\geq 0.75$  are shown.

Table 3

ROI specific BOLD signal response during bladder filling (“INFUSION minus START”) for ROIs showing significant activation ( $p = 0.05$ , FWE corrected). The ICC represent a reliability measure (values ranging from 0 = no reliability to 1 = perfect reliability) of a ROI’s mean beta weight between 1st and 2nd fMRI. The relative voxel overlap (values ranging from 0 = no overlap to 1 = complete overlap) between 1st and 2nd fMRI, calculated using the Sørensen-Dice index, represents a measure of spatial congruency based on activation maps limited to  $p \leq 0.001$  (uncorrected for multiple comparisons).

ROI	Brain area	BA	1 <sup>st</sup> fMRI			2 <sup>nd</sup> fMRI			ICC of ROI mean beta weight (single measures)	Voxel overlap (Sørensen-Dice index)
			MNI peak coordinates			MNI peak coordinates				
			x	y	z	x	y	z		
1	DLPFC	8/9/46	3	38	37	3	47	19	-0.01	0.19
			45	23	40					
			-3	38	37					
			-21	35	25					
2	OFC	10	-9	35	31	45	47	1	0.33	0.27
			27	53	1					
			24	47	22					
3	VLPFC (pars triangularis, opercularis, and orbitalis)	44/45/47	-36	50	4	51	17	1	0.47	0.24
			51	23	1					
4	ACC	24/32	-27	32	-11	6	47	10	-0.33	0.08
			6	11	28					
			-3	14	28	-21	38	16		
5a	Insula (left)	13	-21	32	22				-0.05	0.55
5b	Insula (right)	13	-39	-4	-5	-27	20	-5	0.55	0.44
6	Hypothalamus		-9	-7	-8	-9	-4	-5	-0.20	0.50
7	Thalamus		15	-25	7	n/a	n/a	n/a	n/a	n/a
8	PAG		12	-16	-8	n/a	n/a	n/a	n/a	n/a
9	Pons		-9	-19	-32	n/a	n/a	n/a	n/a	n/a
10a	Cerebellum (left)		n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
10b	Cerebellum (right)		30	-52	-47	n/a	n/a	n/a	n/a	n/a
11a	Basal ganglia (left)		-27	-1	13	-27	8	7	0.03	0.21
			-15	2	-2	-24	-4	-8		
11b	Basal ganglia (right)		n/a	n/a	n/a	30	11	10	n/a	n/a
12	SMA		n/a	n/a	n/a	-9	14	61	n/a	n/a

ACC anterior cingulate cortex, BA Brodmann area, BOLD blood oxygenation level-dependent, DLPFC dorsolateral prefrontal cortex, fMRI functional magnetic resonance imaging, FWE family-wise error, ICC intraclass correlation coefficient, MNI Montreal Neurological Institute, OFC orbitofrontal/frontopolar prefrontal cortex, PAG periaqueductal grey, ROI region of interest, SMA supplementary motor area, VLPFC ventrolateral prefrontal cortex.

of functional neuroimaging on clinical practice, it is crucial to be able to reliably assess LUT treatment effects at supraspinal level to better understand the mechanism of action and to evaluate efficacy, including reason for failure. Several neuroimaging studies have investigated whether treatment-related improvement in LUT control correlate with supraspinal responses (Dasgupta et al., 2005; Blok et al., 2006; Kavia et al., 2010; Pontari et al., 2010; Griffiths et al., 2015). However, it remains questionable whether task-specific brain activities, acquired with

fMRI, are reliable and consequently to what extent a change in supraspinal response is treatment related.

To investigate for the first time the test-retest reliability of BOLD signal responses to a bladder filling task across two separate (5–8 weeks) fMRI measurements, we applied a block design and used an automated, MR-synchronized IDD (Jarrahi et al., 2016; Leitner et al., 2017) to achieve maximum consistency in regard to the task performance. The temporal distance between both fMRI measurements (visit 2 and 3), i.e. 5–8



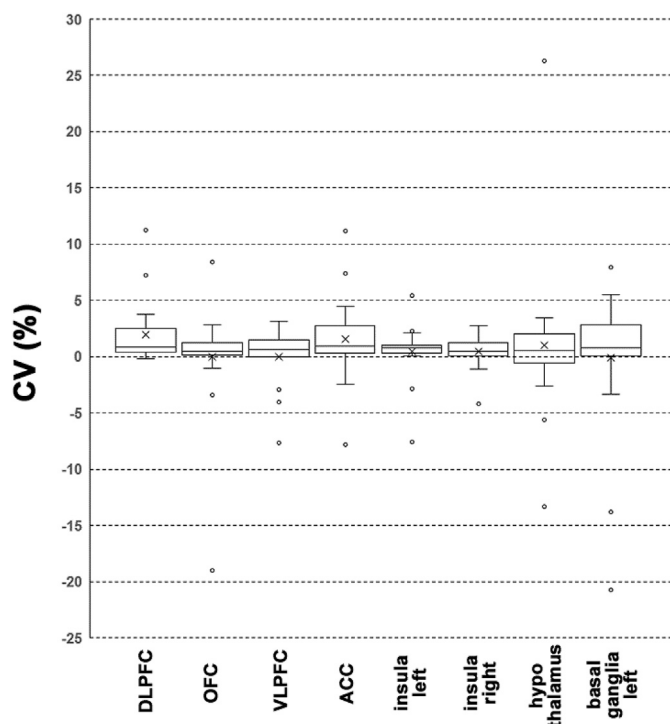


Fig. 5. Box plots (first quartile, median, third quartile, and mean marker (cross)) with whiskers (1.5× interquartile range of upper and lower quartile) and outlier of the individual coefficient of variation (CV) between visits of the mean beta weights of each ROI with significant BOLD response in both visits (Table 3). ACC anterior cingulate cortex, DLPFC dorsolateral prefrontal cortex, OFC orbitofrontal/frontopolar prefrontal cortex, VLPFC ventrolateral prefrontal cortex.

Table 4

Median, minimum, maximum, and range (= maximum – minimum) of the coefficient of variation (CV) of the mean beta weights of each region of interest (ROI).

ROI	Brain area	Median CV [%]	Minimum CV [%]	Maximum CV [%]	Spread (max-min) [%]
1	DLPFC	0.88	-0.21	11.26	11.47
2	OFC	0.42	-19.01	8.44	27.45
3	VLPFC	0.58	-7.70	3.14	10.84
4	ACC	0.96	-7.85	11.15	19.00
5a	Insula left	0.74	-7.62	5.45	13.07
5b	Insula right	0.45	-4.22	2.74	6.96
6	Hypothalamus	0.56	-13.36	26.30	39.66
11a	Basal ganglia left	0.79	-20.76	7.95	28.71

ACC anterior cingulate cortex, DLPFC dorsolateral prefrontal cortex, OFC orbitofrontal/frontopolar prefrontal cortex, VLPFC ventrolateral prefrontal cortex.

weeks, is in line with the time window to expect maximum treatment efficacy when treating individuals with LUT dysfunction (Siami et al., 2002; Karsenty et al., 2008) and to perform follow-up assessments in clinical trials (Maman et al., 2014). In addition, it helped to accommodate female participants to be investigated at the same time point of their menstrual cycle.

With this set-up, it was possible to demonstrate significant task-related BOLD signal responses in both fMRI visits at  $p < 0.05$  (corrected, using a cluster-threshold correction of  $k > 41$  voxel) (Slotnick et al., 2003). The intensity level in response to LUT stimulation either exceeded or was similar to what has been previously reported (Griffiths et al., 2005, 2007; Zhang et al., 2005; Di Gangi Herms et al., 2006; Tadic et al., 2008, 2010; Pontari et al., 2010).

Despite that MNI peak coordinates of each fMRI measurement (Table 2) did not highly correlate between measurements (except left inferior parietal lobe, BA 39), Fig. 3 demonstrates regional consistency and overlap of significant clusters located in the right OFC, right insula, right VLPFC, and right inferior parietal lobe (Fig. 3).

Using the voxel-based whole brain ICC analysis to quantify reliability of voxels between measurements (Fig. 4), a large number of voxels in brain areas attributed to the SSN of LUT control showed excellent reliability (ICC > 0.75). This would be an ideal result. However, this result is somewhat conflicting with our findings shown in Table 2. Moreover, one has to consider that voxel-based whole brain ICC analysis does not reflect the BOLD signal amplitude. Hence, as reported previously (Caceres et al., 2009; Plichta et al., 2012), there are voxels that demonstrate high reliability but low signal intensity (i.e. small t-values). Such voxels could be a task-related response but only indirectly or non-linearly to the stimulus (Caceres et al., 2009; Plichta et al., 2012).

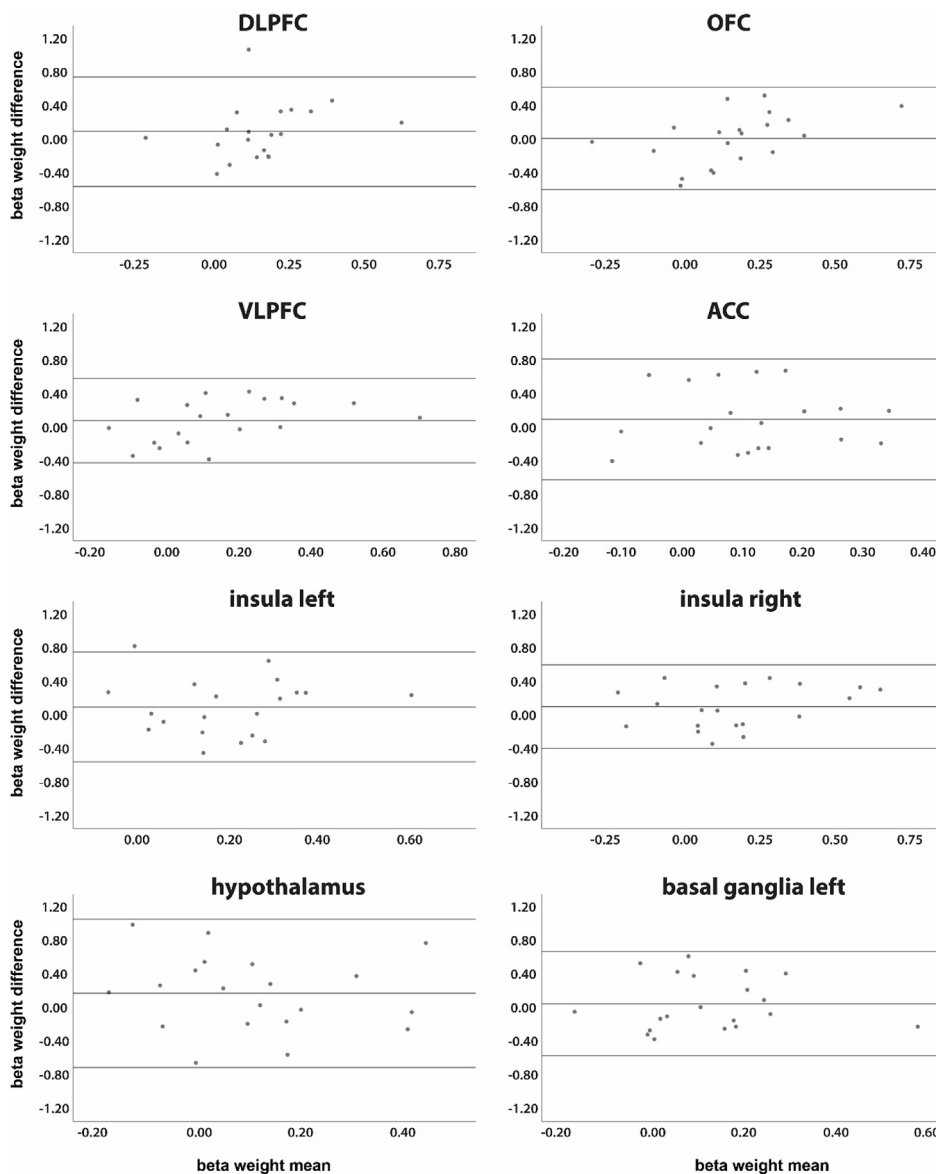
To assess both, reliability of task-related signal validity and spatial allocation, 12 ROIs were selected based on the most frequently reported supraspinal areas involved in LUT control (Kavia et al., 2005; Griffiths and Tadic, 2008; Fowler and Griffiths, 2010; Arya et al., 2017).

The current working models on supraspinal LUT control based on previous neuroimaging findings suggest that the ascending signals from the LUT are relayed through the PAG and thalamus to cortical areas, which further process the information and are involved in appropriate decision making and action taking processes (Fowler et al., 2008; Griffiths and Tadic, 2008; Fowler and Griffiths, 2010). These cortical areas specifically comprise the insula, the PFC, and the ACC. The insula as primary interoceptive cortex is supposed to be involved in a first-order mapping of visceral signals (e.g. bladder distention) with second-order regions such as the PFC and the ACC which are involved in the subjective awareness and feeling of the visceral signal (e.g. desire to void) (Craig, 2002, 2009; Fowler et al., 2008). The PFC is involved in planning complex cognitive and appropriate social behaviour (e.g. handling of strong desire to void in different circumstances) as well as attention and response-selection mechanisms (e.g. postponement of micturition) which are essential for the volitional control of the LUT (Pardo et al., 1991; Bechara et al., 2000; Kavia et al., 2005; Fowler and Griffiths, 2010). The PFC has strong and multiple connections to the cingulate cortex and hypothalamus which both are part of the limbic system and considered to provide emotional and motivational input to the prefrontal decision making process. The ACC for example is thought to modulate how much attention is provided to signals arising from the bladder and how one reacts to them (Fowler et al., 2008). The hypothalamus, which is also associated with autonomic control due to its projections to all autonomic preganglionic motor neurons in the spinal cord, including the sacral parasympathetic and sphincter motor nuclei, is one of the few regions that has direct afferent projections to the pontine micturition center (Fowler et al., 2008; Griffiths and Tadic, 2008). This connection has been interpreted as an additional “layer of control” that permits micturition in healthy subjects only if it is judged ‘safe’ to do so.

The SMA has been mainly observed in association with pelvic floor muscle contractions and may serve as additional continence mechanism in situations of strong desire to void or urgency to prevent premature leakage (Zhang et al., 2005; Griffiths et al., 2015).

Although previous neuroimaging studies have reported cerebellar and basal ganglia activity in response to different LUT conditions, their role in LUT control is not yet clear (Zhang et al., 2005; Seseke et al., 2006). Both structures are known to be involved in motor activity but also in other functions such as cognition, attention, emotion, and behaviour (Wolf et al., 2009; D’Angelo, 2018; Florio et al., 2018). Lesions or disorders of basal ganglia and cerebellum seem to frequently result in LUT dysfunction, predominantly detrusor overactivity and detrusor-sphincter-dyssynergia (Dietrichs and Haines, 2002; Sakakibara et al., 2004, 2008; Winge et al., 2006; Chou et al., 2013). Hence, both structures seem to contribute facilitatory but also inhibitory input to LUT control and may be involved in level setting and fine tuning processes of





**Fig. 6.** Bland-Altman plots demonstrating the 95% limits of agreement between 1st and 2nd fMRI. The plots display the beta weight differences of visits against the mean beta weight for each of the 8 ROIs that demonstrated significant BOLD signal increase during both visits (Table 3). The middle line represents the mean. The upper and lower line represent the upper and lower limits of agreement ( $\text{mean} \pm 1.96 \cdot \text{SD}$ ), respectively.

ACC anterior cingulate cortex, DLPFC dorsolateral prefrontal cortex, OFC orbitofrontal/frontopolar prefrontal cortex, VLPFC ventrolateral prefrontal cortex.

LUT control.

Despite that a high proportion of selected ROIs demonstrated significant BOLD signal changes to our current bladder filling task, it was not expected to find significant results in all 12 ROIs. The latter is aided by the fact that our ROI selection is based on combined findings from different studies, none of which reported significant responses in all ROIs within a single investigation. This is, at least partially, due to the high diversity of outcome measures, study design, bladder filling and applied scan protocol among functional neuroimaging studies investigating supraspinal LUT control.

In this study, we could demonstrate that 7 of 12 selected ROIs, respectively 8 of 15 (if bilateral ROIs are counted separately) selected ROIs showed significant BOLD signal changes in response to bladder filling during both fMRI visits. These 8 ROIs represent essential parts of the interoceptive network, i.e. hypothalamus, insula, ACC, and OFC (Craig, 2002). This fits well with the interoceptive character and transmission of the bladder filling sensation. Furthermore, it is in line with previous reports on the supraspinal correlate of bladder filling sensations in healthy participants at larger bladder volumes, i.e. at a strong desire to void (Griffiths et al., 2007; Mehnert et al., 2008). However, reliability of the mean beta weight of each ROI was “only” poor to fair with the two

highest ICC for right insula (0.55) and right VLPFC (0.47).

Recently, one study investigated the reliability of BOLD responses to bladder filling during same session repeat measurements representing the agreement between the first and second sequence of four blocks of total eight blocks (Clarkson et al., 2017). Reliability (ICC) was computed for three preselected ROIs (right insula, dorsal ACC, and medial PFC). Despite being recorded within the same session and without significant systematic change between first and second sequence of blocks, poor (0.19) to fair (0.44) reliability was achieved (Clarkson et al., 2017). These findings were interpreted as being related to random variability due to the complex and emotionally affected process of supraspinal LUT control (Clarkson et al., 2017). The authors concluded that there is need for improvements of methods used to provoke supraspinal responses related to bladder filling, i.e. standardizing the circumstances of the scanning, redesigning the stimulus protocol to yield larger brain responses (Clarkson et al., 2017).

Some of these suggested improvements have been already included in our study design, which may have contributed to the higher ICC values compared to Clarkson et al. even across a longer retest interval which usually tends to lower reliability (Bennett and Miller, 2010). However, direct comparison of our results to the study of Clarkson et al. has to be

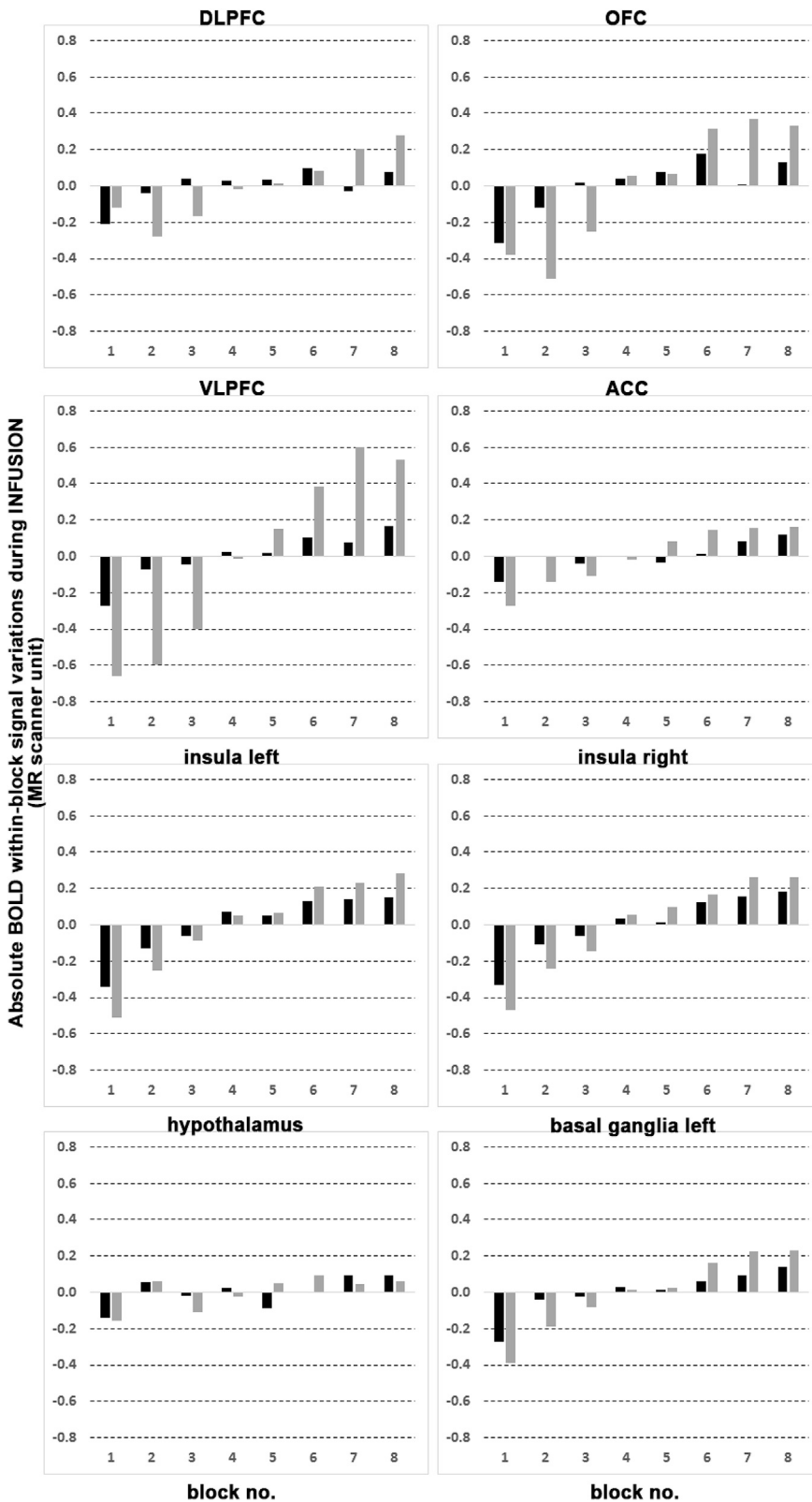


Fig. 7. Absolute INFUSION-related within-block blood oxygenation level-dependent (BOLD) signal variations relative to the mean BOLD activity across all blocks (= mean of all blocks – individual block) for each ROI and session (1st fMRI – black; 2nd fMRI – grey). The repeated measures ANOVA demonstrated no main effect of block or visit. ACC anterior cingulate cortex, DLPF dorsolateral prefrontal cortex, OFC orbitofrontal/frontopolar prefrontal cortex, VLPFC ventrolateral prefrontal cortex.

done with caution due to the different study population (women with urgency incontinence vs healthy participants) and the age difference ( $\geq 60$  years vs  $39.2 \pm 11.6$  years). Specifically, the latter can affect brain hemodynamics and consequently BOLD responses (D'Esposito et al., 2003). In addition, individuals with urinary incontinence may show a greater heterogeneity and variability of neuroimaging outcomes,

depending on various factors such as the aetiology of incontinence, concomitant medications and the presence of a potential underlying neurological disease. In the present study, using an automated, MR-synchronized IDD helped to provide consistency and accuracy of task performance throughout the measurements (Jarrahi et al., 2016; Leitner et al., 2017). The applied protocol (Fig. 2) incorporated a randomly

shifting time interval (jitter) of 7–9s, (mean 8s) during each REST condition to avoid conditioning/habituation. In contrast to previous studies, using a larger INFUSION/WITHDRAWAL volume (i.e. 100 mL) resulted in a strong and distinctive stimulus which is reflected by the participant's rating and the time course of the BOLD signal (Fig. 7), indicating a consistent stimulation and BOLD amplitudes within and between the two visits/fMRI measurements without any main effect of block or visit in the repeated-measures ANOVA.

Whether an ICC of 0.55 (as demonstrated for the right insula ROI) is sufficient enough to compare supraspinal LUT control before and after treatment has to be further elucidated. Considering the ICC classification of Cicchetti et al. (Cicchetti, 1994) our findings do not exceed the level of fair reliability which may be insufficient to differentiate true treatment effects from random variability. On the other side, our results are well within the range of previous investigations on fMRI reliability reporting ROI based ICC values from 0.16 to 0.88 with an overall mean of 0.52 (Bennett and Miller, 2010). In review of these studies, simpler motor and sensory tasks seems to be more reliable than more complex tasks involving interoception. This is not surprising, considering that instead of one distinct principle area such as primary motor, visual, or auditory cortex for simple motor, visual, or auditory tasks, respectively, more complex tasks require multiregional involvement, e.g. the SSN in case of LUT control (Fowler et al., 2008). Such network may undergo natural changes over time, be more variable between participants, and be more likely influenced/alterd by the actual motivational and affective condition of the participants during measurements (Stern et al., 2017). This may at least partly explain the poor reliability of ACC activity, despite significant activity in the ACC ROI during both fMRI measurements (Table 3). Although we did not expect or discover findings that suggest habituation to our task over the time period between both fMRI measurements, it has to be considered that despite all pre-scan instructions and accommodation time in the scanner, the 1st fMRI measurement was an emotionally “newer” experience than the 2nd fMRI measurement which may have affected our results. To better control for this effect it may be necessary to include a 3rd fMRI measurement at the beginning as accommodative scan which is not included in the reliability analysis. However, this may be extremely challenging in regard to participant/patient compliance, scanner time availability, and budget.

Assessment of supraspinal LUT control is certainly a more complex task including not only the ability to be aware of sensation but also how sensations are interpreted, regulated, and used to inform behaviour, with different dimensions relating to different aspects of health and disease (Stern et al., 2017). These processes involve several brain areas including areas of interoceptive, emotional, and cognitive processing which may also overlap in function (Fowler et al., 2008). Despite that there is an expectable variability within but also between participants (Fig. 5) in several areas of the supraspinal LUT control network, there are areas with a smaller variability than others, such as the right insula and the VLPFC (Table 4, Fig. 5). This corresponds well to the narrower limits of agreement in the Bland-Altman plots of those ROIs and to their ICC values, which were the highest among the selected ROIs. In addition, the right insula also demonstrated a considerable spatial congruency between both fMRI measurements (Sørensen-Dice index of 0.44 = 44% voxel overlap within ROI, Table 3). Other ROIs, such as the left insula or hypothalamus, demonstrated even higher values of spatial congruency, although their mean beta weight ICCs were even negative (Table 3). Again, this discrepancy is related to different thresholds and foci of both analyses (ICC vs Sørensen-Dice index). Voxels may have a task-related high spatial congruency but the agreement between the parameter estimates (beta weights) of these voxels may be poor due to their indirect or non-linear relation to the stimulus (Caceres et al., 2009; Plichta et al., 2012). Hence, of all investigated ROIs the right insular cortex appears to have the best congruency in both, BOLD signal location and validity.

In regard to the eminent role that has been attributed to the right insula in the function of interoception (Craig, 2002, 2009; Jarrahi et al., 2015), it may be only consequential that this area is most consistently

involved in a LUT stimulation task. Although we could demonstrate a fair reliability specifically for the right insula, there is room for further improvement and results need to be confirmed in individuals with LUTD. However, to improve fMRI protocols investigating task-specific responses of the SSN involved in LUT control, we need to better understand the factors that influence reliability in our current task designs. In general, increasing the repetition time (i.e., to increase the sampling size) and/or field strength (to achieve a better signal-to-noise ratio) could help to improve the reliability but only with an adequate scanner maintenance/quality control and task design (Bennett and Miller, 2010).

In conclusion, achieving high reliability using a viscerosensory/interoceptive task such as repetitive bladder filling is challenging. Applying an automated, MR-synchronized task with a stronger stimulus ( $\pm 100$  mL INFUSION/WITHDRAWAL) than previously used, we found in a cohort of healthy participants that the right insula and right VLPFC are the most reliable brain regions in relation to our protocol. This finding suggests that these two areas should be specifically considered and evaluated when using fMRI to investigate treatment efficacy. Additional endeavour is highly warranted to better understand involved supraspinal and potentially also spinal processes.

### Competing interest

None of the authors has a competing interest.

### Acknowledgement

This study was funded by the Swiss National Science Foundation (www.snf.ch; Grant no. 135774) and supported by the Swiss Continence Foundation (www.swisscontinencefoundation.ch).

### References

- Arya, N.G., Weissbart, S.J., Xu, S., Rao, H., 2017. Brain activation in response to bladder filling in healthy adults: an activation likelihood estimation meta-analysis of neuroimaging studies. *NeuroUrol. Urodyn.* 36, 960–965.
- Bechara, A., Damasio, H., Damasio, A.R., 2000. Emotion, decision making and the orbitofrontal cortex. *Cerebr. Cortex* 10, 295–307.
- Bennett, C.M., Miller, M.B., 2010. How reliable are the results from functional magnetic resonance imaging? *Ann. N. Y. Acad. Sci.* 1191, 133–155.
- Bland, J.M., Altman, D.G., 1986. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1, 307–310.
- Blok, B.F., 2002. Brain control of the lower urinary tract. *Scand. J. Urol. Nephrol. Suppl.* 11–15.
- Blok, B.F.M., Groen, J., Bosch, J.L.H.R., Veltman, D.J., Lammertsma, A.A., 2006. Different brain effects during chronic and acute sacral neuromodulation in urge incontinent patients with implanted neurostimulators. *BJU Int.* 98, 1238–1243.
- Caceres, A., Hall, D.L., Zelaya, F.O., Williams, S.C., Mehta, M.A., 2009. Measuring fMRI reliability with the intra-class correlation coefficient. *Neuroimage* 45, 758–768.
- Chou, Y.C., Jiang, Y.H., Harmod, T., Kuo, H.C., 2013. Characteristics of neurogenic voiding dysfunction in cerebellar stroke: a cross-sectional, retrospective video urodynamic study. *Cerebellum* 12, 601–606.
- Cicchetti, D.V., 1994. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychol. Assess.* 6, 284–290.
- Clarkson, B.D., Tyagi, S., Griffiths, D.J., Resnick, N.M., 2017. Test-retest repeatability of patterns of brain activation provoked by bladder filling. *NeuroUrol. Urodyn.* 36, 1472–1478.
- Coyne, K.S., Thompson, C.L., Lai, J.S., Sexton, C.C., 2015. An overactive bladder symptom and health-related quality of life short-form: validation of the OAB-q SF. *NeuroUrol. Urodyn.* 34, 255–263.
- Craig, A.D., 2002. How do you feel? Interoception: the sense of the physiological condition of the body. *Nat. Rev. Neurosci.* 3, 655–666.
- Craig, A.D., 2009. How do you feel—now? The anterior insula and human awareness. *Nat. Rev. Neurosci.* 10, 59–70.
- D'Angelo, E., 2018. Physiology of the cerebellum. *Handb. Clin. Neurol.* 154, 85–108.
- D'Esposito, M., Deouell, L.Y., Gazzaley, A., 2003. Alterations in the BOLD fMRI signal with ageing and disease: a challenge for neuroimaging. *Nat. Rev. Neurosci.* 4, 863–872.
- Dasgupta, R., Critchley, H.D., Dolan, R.J., Fowler, C.J., 2005. Changes in brain activity following sacral neuromodulation for urinary retention. *J. Urol.* 174, 2268–2272.
- Di Gangi Herms, A.M.R., Veit, R., Reisenauer, C., Herms, A., Grodd, W., Enck, P., Stenzl, A., Birbaumer, N., 2006. Functional imaging of stress urinary incontinence. *Neuroimage* 29, 267–275.
- Dice, L.R., 1945. Measures of the amount of ecologic association between species. *Ecology* 26, 297–302.
- Dietrichs, E., Haines, D.E., 2002. Possible pathways for cerebellar modulation of autonomic responses: micturition. *Scand. J. Urol. Nephrol. Suppl.* 16–20.

- Donovan, J.L., Peters, T.J., Abrams, P., Brookes, S.T., de aa Rosette, J.J., Schafer, W., 2000. Scoring the short form ICSmaleSF questionnaire. *International Continence Society. J. Urol.* 164, 1948–1955.
- Florio, T.M., Scarnati, E., Rosa, I., Di Censo, D., Ranieri, B., Cimini, A., Galante, A., Alecci, M., 2018. The Basal Ganglia: more than just a switching device. *CNS Neurosci. Ther.* 24, 677–684.
- Folstein, M.F., Folstein, S.E., McHugh, P.R., 1975. Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J. Psychiatr. Res.* 12, 189–198.
- Fowler, C.J., Griffiths, D., de Groat, W.C., 2008. The neural control of micturition. *Nat. Rev. Neurosci.* 9, 453–466.
- Fowler, C.J., Griffiths, D.J., 2010. A decade of functional brain imaging applied to bladder control. *NeuroUrol. Urodyn.* 29, 49–55.
- Friston, K., Ashburner, J., Frith, C.D., Poline, J.B., Heather, J.D., Frackowiak, R.S., 1995. Spatial registration and normalization of images. *Hum. Brain Mapp.* 3, 165–189.
- Friston, K.J., Holmes, A.P., Worsley, K.J., Poline, J.P., Frith, C.D., Frackowiak, R.S., 1994. Statistical parametric maps in functional imaging: a general linear approach. *Hum. Brain Mapp.* 2, 189–210.
- Gountouna, V.E., Job, D.E., McIntosh, A.M., Moorhead, T.W., Lymer, G.K., Whalley, H.C., Hall, J., Waiter, G.D., Brennan, D., McGonigle, D.J., Ahearn, T.S., Cavanagh, J., Condon, B., Hadley, D.M., Marshall, I., Murray, A.D., Steele, J.D., Wardlaw, J.M., Lawrie, S.M., 2010. Functional Magnetic Resonance Imaging (fMRI) reproducibility and variance components across visits and scanning sites with a finger tapping task. *Neuroimage* 49, 552–560.
- Griffiths, D., Clarkson, B., Tadic, S.D., Resnick, N.M., 2015. Brain mechanisms underlying urge incontinence and its response to pelvic floor muscle training. *J. Urol.* 194, 708–715.
- Griffiths, D., Derbyshire, S., Stenger, A., Resnick, N., 2005. Brain control of normal and overactive bladder. *J. Urol.* 174, 1862–1867.
- Griffiths, D., Tadic, S.D., 2008. Bladder control, urgency, and urge incontinence: evidence from functional brain imaging. *NeuroUrol. Urodyn.* 27, 466–474.
- Griffiths, D., Tadic, S.D., Schaefer, W., Resnick, N.M., 2007. Cerebral control of the bladder in normal and urge-incontinent women. *Neuroimage* 37, 1–7.
- Holstege, G., 2005. Micturition and the soul. *J. Comp. Neurol.* 493, 15–20.
- Jackson, S., Donovan, J., Brookes, S., Eckford, S., Swithinbank, L., Abrams, P., 1996. The Bristol female lower urinary tract symptoms questionnaire: development and psychometric testing. *Br. J. Urol.* 77, 805–812.
- Jarrahi, B., Gassert, R., Wanek, J., Michels, L., Mehnert, U., Kollias, S.S., 2016. Design and application of a new automated fluidic visceral stimulation device for human fMRI studies of interoception. *IEEE J Transl Eng Health Med* 4, 2000108.
- Jarrahi, B., Mantini, D., Balsters, J.H., Michels, L., Kessler, T.M., Mehnert, U., Kollias, S.S., 2015. Differential functional brain network connectivity during visceral interoception as revealed by independent component analysis of fMRI TIME-series. *Hum. Brain Mapp.* 36, 4438–4468.
- Jarrahi, B., Wanek, J., Mehnert, U., Kollias, S., 2013. An fMRI-compatible multi-configurable handheld response system using an intensity-modulated fiber-optic sensor. *Conf Proc IEEE Eng Med Biol Soc* 6349–6352, 2013.
- Karsenty, G., Denys, P., Amarenco, G., De Seze, M., Game, X., Haab, F., Kerdraon, J., Perrouin-Verbe, B., Ruffion, A., Saussine, C., Soler, J.M., Schurch, B., Chartier-Kastler, E., 2008. Botulinum toxin A (Botox) intradetrusor injections in adults with neurogenic detrusor overactivity/neurogenic overactive bladder: a systematic literature review. *Eur. Urol.* 53, 275–287.
- Kavia, R., DasGupta, R., Critchley, H., Fowler, C., Griffiths, D., 2010. A functional magnetic resonance imaging study of the effect of sacral neuromodulation on brain responses in women with Fowler's syndrome. *BJU Int.* 105, 366–372.
- Kavia, R.B., Dasgupta, R., Fowler, C.J., 2005. Functional imaging and the central control of the bladder. *J. Comp. Neurol.* 493, 27–32.
- Koopmans, L.H., Owen, D.B., Rosenblatt, J., 1964. Confidence intervals for the coefficient of variation for the normal and log normal distributions. *Biometrika* 51, 25–32.
- Leitner, L., Walter, M., Jarrahi, B., Wanek, J., Diefenbacher, J., Michels, L., Liechti, M.D., Kollias, S.S., Kessler, T.M., Mehnert, U., 2017. A novel infusion-drainage device to assess lower urinary tract function in neuro-imaging. *BJU Int.* 119, 305–316.
- Maman, K., Aballea, S., Nazir, J., Desroziars, K., Neine, M.E., Siddiqui, E., Odeyemi, I., Hakimi, Z., 2014. Comparative efficacy and safety of medical treatments for the management of overactive bladder: a systematic literature review and mixed treatment comparison. *Eur. Urol.* 65, 755–765.
- McGraw, K.O., Wong, S.P., 1996. Forming inferences about some intraclass correlation coefficients. *Psychol. Methods* 1, 30–46.
- Mehnert, U., Boy, S., Svensson, J., Michels, L., Reitz, A., Candia, V., Kleiser, R., Kollias, S., Schurch, B., 2008. Brain activation in response to bladder filling and simultaneous stimulation of the dorsal clitoral nerve-An fMRI study in healthy women. *Neuroimage* 41, 682–689.
- Pardo, J.V., Fox, P.T., Raichle, M.E., 1991. Localization of a human system for sustained attention by positron emission tomography. *Nature* 349, 61–64.
- Plichta, M.M., Schwarz, A.J., Grimm, O., Morgen, K., Mier, D., Haddad, L., Gerdes, A.B., Sauer, C., Tost, H., Esslinger, C., Colman, P., Wilson, F., Kirsch, P., Meyer-Lindenberg, A., 2012. Test-retest reliability of evoked BOLD signals from a cognitive-emotive fMRI test battery. *Neuroimage* 60, 1746–1758.
- Pontari, M.A., Mohamed, F.B., Lebovitch, S., Moonan, S., Lebed, B., Ruggieri, M.R., Faro, S.H., 2010. Central nervous system findings on functional magnetic resonance imaging in patients before and after treatment with anticholinergic medication. *J. Urol.* 183, 1899–1905.
- Quiton, R.L., Keaser, M.L., Zhuo, J., Gullapalli, R.P., Greenspan, J.D., 2014. Intersession reliability of fMRI activation for heat pain and motor tasks. *Neuroimage Clin* 5, 309–321.
- Sakakibara, R., Uchida, Y., Uchiyama, T., Yamanishi, T., Hattori, T., 2004. Reduced cerebellar vermis activation during urinary storage and micturition in multiple system atrophy: 99mTc-labelled ECD SPECT study. *Eur. J. Neurol.* 11, 705–708.
- Sakakibara, R., Uchiyama, T., Yamanishi, T., Shirai, K., Hattori, T., 2008. Bladder and bowel dysfunction in Parkinson's disease. *J. Neural. Transm.* 115, 443–460.
- Seseke, S., Baudewig, J., Kallenberg, K., Ringert, R.H., Seseke, F., Dechent, P., 2006. Voluntary pelvic floor muscle control—an fMRI study. *Neuroimage* 31, 1399–1407.
- Shrout, P.E., Fleiss, J.L., 1979. Intraclass correlations: uses in assessing rater reliability. *Psychol. Bull.* 86, 420–428.
- Siami, P., Seidman, L.S., Lama, D., 2002. A multicenter, prospective, open-label study of tolterodine extended-release 4 mg for overactive bladder: the speed of onset of therapeutic assessment trial (STAT). *Clin. Ther.* 24, 616–628.
- Slotnick, S.D., Moo, L.R., Segal, J.B., Hart Jr., J., 2003. Distinct prefrontal cortex activity associated with item memory and source memory for visual shapes. *Brain Res. Cogn. Brain Res* 17, 75–82.
- Stern, E.R., Grimaldi, S.J., Muratore, A., Murrrough, J., Leib, E., Fleysler, L., Goodman, W.K., Burdick, K.E., 2017. Neural correlates of interoception: effects of interoceptive focus and relationship to dimensional measures of body awareness. *Hum. Brain Mapp.* 38, 6068–6082.
- Swiss Academy of Medical Sciences, 2009. Guideline - Concerning Scientific Research Involving Human Beings.
- Tadic, S.D., Griffiths, D., Schaefer, W., Cheng, C.I., Resnick, N.M., 2010. Brain activity measured by functional magnetic resonance imaging is related to patient reported urgency urinary incontinence severity. *J. Urol.* 183, 221–228.
- Tadic, S.D., Griffiths, D., Schaefer, W., Resnick, N.M., 2008. Abnormal connections in the supraspinal bladder control network in women with urge urinary incontinence. *Neuroimage* 39, 1647–1653.
- The Federal Authorities of the Swiss Confederation, 1992. Bundesgesetz über den Datenschutz (DSG) vom 19. Juni 1992, Stand. 01.01.2011.
- Wei, X., Yoo, S.S., Dickey, C.C., Zou, K.H., Guttman, C.R., Panych, L.P., 2004. Functional MRI of auditory verbal working memory: long-term reproducibility analysis. *Neuroimage* 21, 1000–1008.
- Winge, K., Skau, A.M., Stimpel, H., Nielsen, K.K., Werdelin, L., 2006. Prevalence of bladder dysfunction in Parkinson's disease. *NeuroUrol. Urodyn.* 25, 116–122.
- Wolf, U., Rapoport, M.J., Schweizer, T.A., 2009. Evaluating the affective component of the cerebellar cognitive affective syndrome. *J. Neuropsychiatry Clin. Neurosci.* 21, 245–253.
- World Medical Association, 1964. Declaration of Helsinki - Ethical Principles for Medical Research Involving Human Subjects.
- Zhang, H., Reitz, A., Kollias, S., Summers, P., Curt, A., Schurch, B., 2005. An fMRI study of the role of suprapontine brain structures in the voluntary voiding control induced by pelvic floor contraction. *Neuroimage* 24, 174–180.
- Zigmond, A.S., Snaith, R.P., 1983. The hospital anxiety and depression scale. *Acta Psychiatr. Scand.* 67, 361–370.