The American Journal of Chinese Medicine, Vol. 36, No. 1, 55–70

Institute for Advanced Research in Asian Science and Medicine

Carryover Effects Alter fMRI Statistical Analysis in an Acupuncture Study

Tsung-Jung Ho,[*,†] Jeng-Ren Duann,[‡] Chun-Ming Chen,[§,¶] Jeon-Hor Chen,[§]
Wu-Chung Shen,[§] Tung-Wu Lu,[//] Jan-Ray Liao[¶] and Jaung-Geng Lin[*]
[*]*Graduate Institute of Chinese Medical Science*, *China Medical University*, *Taichung*, *Taiwan*
[†]*Chang-Hua Hospital*, *Department of Health*, *Executive Yuan*, *Taiwan*
[‡]*Institute for Neural Computation*, *University of California San Diego*, *La Jolla*, *CA*
[§]*Department of Radiology*, *China Medical University Hospital*, *Taichung*, *Taiwan*
[¶]*Department of Electric Engineering*, *National Chung-Shin University*, *Taichung*, *Taiwan*
[//]*Institute of Biomedical Engineering*, *National Taiwan University*, *Taipei*, *Taiwan*

Abstract: Carryover effects can contaminate ON/OFF BOLD contrasts designated in an fMRI experiment. Yet, the ON/OFF contrasts are essential to facilitate statistical analysis based on the significance of contrast levels. Here, we conducted an fMRI experiment with acupuncture stimulation applied on ST42 acupoint as well as with tactile stimulation on its skin surface. Experiment consisted of three two-block acupuncture and one two-block tactile fMRI runs. Each block started with 26-sec OFF period followed by either 26-sec needle manipulation in the acupuncture runs or by scratching skin surface with sand paper in the tactile. To test if carryover effects could alter the BOLD contrasts, we analyzed different portions of fMRI data using GLM method. Our results showed analyses on different portions of acupuncture fMRI data gave significantly different results. Statistical parametric maps of group random effects resulted from the analysis on the very first fMRI trial formed the broadest coverage of the active brain areas. BOLD model time course also best explained the adjusted raw time course at peak active voxel (coefficient of determination = 0.88). Analyses on other portions of fMRI data only selected subset of the active brain areas delineated by the analysis on the very first data trial and the BOLD model only mildly accounted for the adjusted raw time courses. In tactile runs, results were more consistent across analyses. Therefore, in fMRI experiments with strong carryover effects, a single-block experimental design with multiple repetitions, separated by long enough periods of time, should be more suitable to extract task BOLD effects.

Keywords: fMRI, Acupuncture; Carryover Effects; Contrast-Based Statistics; Single-Block; Tactile; BOLD; GLM.

Correspondence to: Dr. Jaung-Geng Lin, Graduate Institute of Chinese Medical Science, China Medical University, 91 Hsueh-Shih Road, Taichung 404, Taiwan. Tel: (+886) 4-2205-3366 (ext. 1018), E-mail: jglin@mail.cmu.edu.tw
Remark: T.-J. Ho and J.-R. Duann contributed equally to this work.

Introduction

The signal-to-noise ratio of the blood oxygenation-level dependent (BOLD) signals obtained in a functional magnetic resonance imaging (fMRI) study is often weak. Sophisticated data process and analysis are thus needed to help detect the tiny BOLD signals. Among many analysis methods, the general linear model (GLM) is the most frequently used. In the use of GLM method, BOLD contrast introduced by different treatment conditions (e.g., on- versus off-treatment) is essential to determine active brain regions in response to the treatment. In addition, multiple repetitions of the on- versus off-treatment "block" are also crucial to boost the signal-to-noise ratio given the weak BOLD signals. In doing so, researchers expect to find repeated brain activities to further lower the standard errors and, in turn, task-related BOLD effects can pass some rigorous statistical hypothesis tests.

However, during the alternations of on- and off-treatments, the effects of the treatment administered earlier in the experiment may persist so long that its effects are still present even when the subject is receiving treatments coming after the earlier ones. In such a case, we may believe that the participant's behavior is due to the treatment just delivered when, in reality, the behavior is due to the lingering effects of a treatment administered some time earlier (Koyama *et al*., 2003; Raja *et al.*, 1999; Wu *et al*., 2001). In turn, this so-called "carryover effect" (or treatment carryover effect) can become a problem in experimental design and in data analysis based on contrasts between treatment conditions.

Generally, the carryover effects can be ubiquitous in most psychophysical and cognitive experiments. For example, Wu *et al*. demonstrated by heating up the skin surface or intradermally injecting capsaicin to elevate the skin temperature of subjects' forearms from their baseline temperature, it could alter the temperature threshold of heat pain sensation (2001). In other words, in a heat pain study consisting of several suprathreshold heat pain stimuli, if the inter-stimulus interval is not long enough, the second or later stimuli might not be able to induce the same level of pain perception because the previous stimuli may already raise the threshold of heat pain sensation. Also, in a study conducted by Raja and colleagues, they showed that after the first heat pain stimulus, heat pain sensation was saturated (Raja *et al.*, 1999). This in turn suppressed the pain ratings of the following heat pain stimuli. On the other hand, Monsell (2003) showed that the carryover effects could profoundly exist in a cognitive task as well. In his study, the subjects were asked to perform several "task sets" and cued to switch between them (for example, read letters versus read colors out of the same character presentations). It could be clearly seen that the carryover effects of a task-set engagement could take some time for subjects to reconfigure processes during task switch, which could in turn substantially slower subjects' responses and cause more errors. This is without mention of the effects from habituation and fatigue; these state changes can also make prolonged influence on the following experimental trials since.

Not only can the carryover effects camouflage our cognitive and sensory processes, but they can also profoundly influence data analysis. For statistical analyses based on contrast of the BOLD effects between ON- and OFF-treatments, it is essential to have stable "baseline" BOLD signals associated with OFF-treatment so that the BOLD effects registered to ON-treatment can be easily detected. If the BOLD signal changes associated

with task-related brain activities are very weak, it is necessary to repeat ON-/OFF-treatment alternations several times to build up effective statistical power. Therefore, it becomes even more crucial to maintain the contrast structure across multiple ON-OFF blocks in a multiple-run fMRI experiment such that the task effects may have been accumulated. However, once the carryover effects exists, the ON-OFF contrast structure may be distorted and thus become less detectable by a contrast-based analysis method, such as the regular general linear model (GLM) analysis.

Hence, efforts have been made to alleviate the influence of carryover effects and in turn to preserve statistical power. The proposed methods include (1) prolonging the experimental sessions up to days or weeks so that the task effects introduced by the earlier treatments are completely gone before the following treatments (Aron *et al.*, 2005; Kolbitcsh *et al*., 2003) (2) using a counterbalanced design, (3) excluding the data portions contaminated by the carryover effects (Michelon *et al.*, 2005), and (4) modeling mathematically the carryover effects in data (Locascio *et al.*, 1997).

However, the above-mentioned methods might not be feasible in some cases. In order to remove or model carryover effects, we have to know in what forms the carryover effects may exist in data. In fact, it is possible for carryover effects to either augment (prolonged excitation) or suppress (prolonged inhibition) the baseline brain activities associated with the OFF-treatments in the upcoming experimental blocks. In addition, they can also co-exist with other artifectual processes, such as cardiac or respiratory rhythms. Therefore, it is extremely difficult, if not impossible, to effectively model carryover effects. On the other hand, given limited time and very slow fMRI acquisition protocol, 2 sec or more for each time point, an fMRI study in general can only have limited ON-/OFF-treatment alternations. It will thus be impractical to drop the fMRI trials contaminated by carryover effects, given that carryover effects can be everywhere. After all, the counter-balanced experimental design is still the most popular strategy to manage carryover effects.

In a previous study, Ho *et al*. showed that the needling sensation introduced by acupuncture needling could last for at least 2 min among a group of healthy subjects (Ho *et al.*, 2007). This profound carryover effects could by no means be controlled by a counterbalanced experimental design because the effects were so strong and long-lasting. Consequently, an acupuncture fMRI protocol lends itself as a good paradigm to test how carryover effects may contaminate BOLD contrast between ON- and OFF-treatments and in turn affect statistical results obtained by contrast-based analysis methods. We hypothesize that if there is no carryover effects involved in recorded data, the same contrast-based analysis, for example, we used the GLM method as implemented in SPM2 (http://fil.ion.ucl.ac.uk/spm2) (Friston *et al*., 1995) here, applied to different portions of fMRI data should give similar task-related BOLD effects. In contrast, if carryover effects happen to alter the baseline BOLD responses registered to the later experimental trials, the statistical power of the data analysis applied to more data blocks might decrease.

Here, we conducted a multiple-block acupuncture fMRI experiment with acupuncture stimulation applied to subjects' left Chong Yang (ST42) acupoints. The fMRI experiment consisted of 6 52-sec trials in 3 separate fMRI runs. Each trial started with a 26-sec OFF block followed by a 26-sec ON block. Between fMRI runs, there was only 15-sec

separation in time for operator to reload the imaging pulse sequence. This short down time between runs allowed us to investigate if the carryover effects can be less contaminating when a new fMRI run was resumed after a short break. We then conducted one additional fMRI run using the same experimental design as the acupuncture fMRI runs but with tactile stimulation on the skin surface of ST42 acupoint. This additional run worked as a control study because carryover effects associated with tactile stimulation might be less pronounced. Also, the same stimulation site should activate exactly the same brain area (primary sensory cortex, S1). In data analysis, we applied GLM analysis to (1) only the first data block from the first run, (2) the entire first fMRI run, (3) the first data blocks of all 3 runs, (4) the second data blocks of all 3 runs, (5) all 3 fMRI runs, and (6) the 3 runs except the very first data block. For tactile fMRI data, we only compared the analysis conditions of (1) and (2). At the end, the task effects as explained by the model BOLD signals used in GLM regression as well as the group random effects maps obtained with different analysis conditions were compared.

Methods

Participants

Eleven normal healthy right-handed subjects (5 females, aged 21.7 ± 2.7 years) participated in this study. None of these subjects has had pathological history of stomachache, head trauma, neurological diseases, and substance abuse. None of them reported painful or stressful responses to the experiment and its environment before and after experimental sessions. The Institutional Review Board (IRB) of China Medical University Hospital approved this experimental protocol. All the subjects provided their signed consent prior to experiment and understood that they could withdraw from the experiment any time they felt uncomfortable.

Experimental Paradigms

All participants were requested to come 20 min before fMRI sessions. After explaining to them their rights and experimental procedures, an experienced acupuncturist (TJH) performed acupuncture on the subjects' left ST42 acupoint using an aseptic acupuncture needle (0.25 mm in diameter and 25.4 mm in length). Needle was manipulated by twisting in both clockwise and counterclockwise directions at a rate of 2 Hz to produce needling sensation and to make sure the needle was inserted in the right position. After needling sensation (the feeling of sourness, numbness, distension, and pain surrounding the needling point) was produced, the subjects were asked to subjectively rate the strength of their needling sensation, from 0 to 10. The highest score (10) was defined as the maximum level of needling sensation the subjects could ever imagine (Beijing College of Traditional Chinese Medicine *et al*., 1993; Stux and Pomeranz, 1998).

Subjects were then positioned in the MR scanner after their needling sensation returned to its baseline. Before the first acupuncture fMRI run was performed, subjective needling

Figure 1. Experimental design. Each experiment consisted of 4 fMRI runs, 3 for acupuncture stimulation and 1 for tactile stimulation. Each fMRI run consisted of 2 stimulus blocks, each started with 26-sec OFF followed by 26-sec ON treatment.

sensation rating was recorded again to confirm that subjects' needling sensation had returned to baseline. Three fMRI runs with acupuncture stimulation followed. Each acupuncture fMRI run consisted of 2 alternations of 26-sec OFF and 26-sec ON blocks and started with OFF condition. During ON period, the acupuncturist manipulated the needle at the same pace as was used in the pre-screening phase; during OFF period, acupuncturist kept his hand off the needle and the subjects were asked to hold still. Subjects' needling sensation ratings were given after all 3 acupuncture fMRI runs were completed. Between acupuncture runs, minimum delay time (15 sec in average, only for prescribing and launching MR pulse sequence) was assured to make the fMRI runs as continuous as possible (see Fig. 1). The acupuncture needle was removed after acupuncture fMRI runs were finished.

Next, an anatomical scan was acquired. Until the needling sensation returned to the baseline again, another fMRI run with tactile stimulation was performed. Since needling sensation could somehow desensitize the skin surface of the acupoint, it was necessary to wait for the needling sensation to be completely gone before the tactile fMRI run could be performed. For tactile fMRI experiment, we used a matched paradigm, two alternations of 26-sec OFF and 26-sec ON blocks; however, only one run was conducted. The stimulation was delivered by slightly scratching the skin surface of the same left ST42 acupoint with a piece of sandpaper in 2-cm^2 area.

Image Acquisition

MR images were acquired using a 1.5-T MR scanner (GE Excite-2, Milwaukee, MI) installed in China Medical University Hospital, Taichung, Taiwan. For each subject, the four fMRI image data sets (three acupuncture and one tactile fMRI runs) were acquired using an echoplanar image sequence (EPI, TR/TE = 2000/60 ms, flip angle = 90°, image matrix = 64 × 64, FOV = 23 × 23 cm^2, resolving a 3.6 × 3.6 mm^2 in-plane resolution, slice thickness = 7 mm plus 0.7 mm gap). Eighteen axial slices parallel to the AC-PC line were acquired from brain base to the vertex. Each fMRI run consisted 52 image volumes. Dummy scans were excluded from saved image data during image acquisition. Anatomical image data set was obtained using a 3D Space SPGR sequence (TR/TE = 33/3 ms, flip angle = 35°, image matrix = 256 × 256, FOV 23 × 23 cm^2 resulting in a 0.9 × 0.9 mm^2 in-plane resolution).

Data Analysis

FMRI data were analyzed using SPM2 running under a Windows XP platform. All the functional imaging data were first slice timing adjusted to compensate the inhomogeneity introduced by acquiring different image slices at slightly different timing. The image volumes were then realigned to the middle (14th) image volume of the image sequence of each fMRI run. After realignment, the image data were spatially normalized to the standard MNI brain template and then spatially smoothed using an 8-mm FWHM Gaussian kernel. The preprocessed data were first subjected to GLM analysis to derive fixed effects at individual level. The model BOLD signal used in the GLM analysis was created by convolving a canonical hemodynamic response function (HRF) with a step function resembling the OFF/ON task design time course. Temporal derivatives were added to model different onsets of the task-related BOLD responses. The extracted fixed effects were then grouped using SPM2 second-level analysis.

In order to test if the statistical results were affected by the acupuncture carryover effects, we applied GLM analysis to different portions of fMRI data at individual level (Fig. 2):

(1) Single block: Using only the data from the first trial of the first acupuncture fMRI run (Fig. 2a, this analysis condition was also applied to tactile fMRI runs).
(2) Multiple blocks: Using the entire fMRI data from the first acupuncture fMRI run (Fig. 2b, this analysis condition was also applied to tactile fMRI runs).
(3) Odd number blocks: Using only the first trials of fMRI data from all 3 acupuncture fMRI runs (Fig. 2c).
(4) Even number blocks: Using only the second trials of fMRI data from all 3 acupuncture fMRI runs (Fig. 2d).
(5) All 6 blocks: Using the entire fMRI data from all 3 acupuncture fMRI runs (Fig. 2e). This is the analysis condition most commonly used.
(6) Last 5 blocks: Using the entire fMRI data, except the first trial from the first acupuncture fMRI run (Fig. 2f).

The contrast images of fixed effects obtained by different analysis conditions were all thresholded at corrected $p < 0.05$ (by false discovery rate, FDR). These thresholded contrast images from individual subjects were then spatially smoothed with an 8-mm FWHM kernel again to account for anatomical differences at individual level. Then, they were grouped using SPM2 second-level analysis with "one-sample t-test" option. The group random effects were also thresholded at corrected $p < 0.05$.

The influence of carryover effects was evaluated by comparing the random effects maps and the coefficient of determination (COD = r^2, square of correlation coefficient) between the adjusted time course at the peak activation voxel (max. Z score) and the model BOLD time course used in the GLM analysis. The COD measures were first derived at individual level and then averaged across subjects. If no carryover effects were involved, analyses on different data portions should give similar resultant CODs and random-effect maps. Only slight differences in t-values might be resolved due to different data variances.

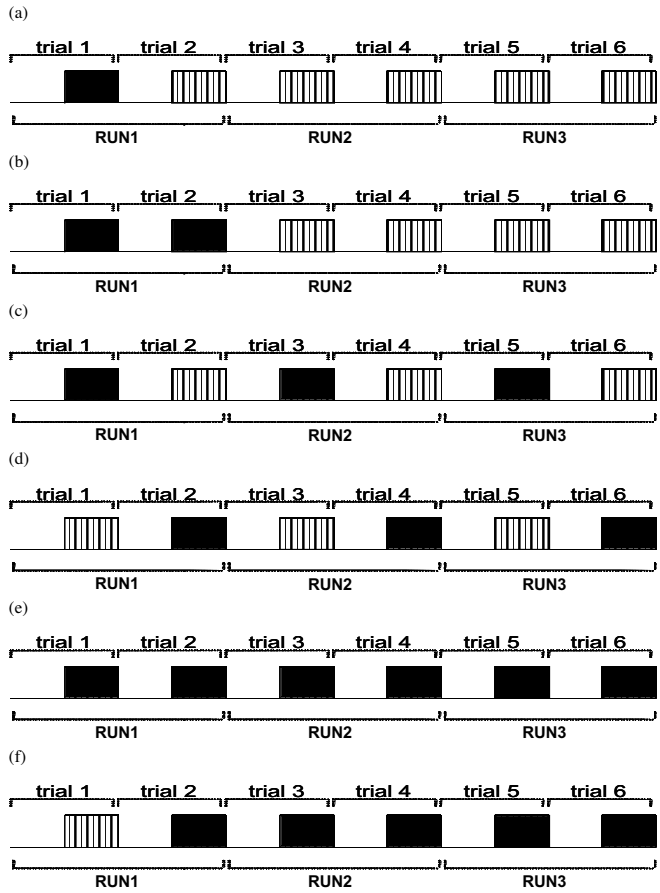

Figure 2. Different fMRI data portions were included in SPM2 GLM analysis. (a) Single block: only the data from the first trial of the first fMRI run were included in data analysis. (b) Multiple blocks: the entire data from the first fMRI run. (c) Odd number blocks: the first trials from all 3 fMRI runs. (d) Even number blocks: the second trials from all 3 fMRI runs. (e) All 6 blocks: the entire data. (f) Last 5 blocks: the entire data, except the first trial from the first run. The solid black blocks indicate the data portions that are included in the data analysis.

Results

Needling Sensation Ratings

Before fMRI experiment, needling sensation was produced by manipulating needle on subjects' left ST42 acupoint. On average, subjects reported needling sensation ratings of 8.1 ± 0.8. In order to allow the needling sensation to return to its baseline, we waited for at least 10 min before we placed the subjects into MR scanner. The first acupuncture fMRI run was not conducted until subject reported his/her needling sensation returned to baseline. The needling sensation scores were averagely below 2 before we started the first fMRI scanning (1.7 ± 1.2). After all 3 acupuncture fMRI runs were finished, the subjects

were asked to report their needling sensation ratings again. Averagely, the ratings were as high as the measures in the pre-screening phase (8.2 ± 1.1).

FMRI Results

Figure 3 shows the fMRI results obtained by SPM2 GLM analysis applied to different portions of fMRI data: single block (Fig. 3a), multiple blocks (3b), odd number blocks (3c), even number blocks (3d), all 6 blocks (3e), last 5 blocks (3f), and the single-block and multiple-block conditions of the tactile fMRI data (3g and 3h). The 3 glass-brain views on the left show the random effects grouped from all 11 subjects. The significance level of these random effects maps was all set to FDR-corrected $p < 0.05$. On the right, it shows the mean (solid line) and ± standard deviation (broken lines) of the adjusted raw time courses at the peak activation voxels averaged across all 11 subjects and the model time course (dotted line) of the model BOLD signal.

Single- versus multiple-block condition (on both tactile and acupuncture fMRI data): We hypothesized that if no carryover effects involved in an fMRI experiment, the results obtained by applying GLM analysis to different portions of fMRI data should give similar results. In our results, we showed that the GLM analysis applied to the first trial and the entire tactile fMRI data gave identical group random effects maps (Figs. 3g and 3h). However, subtle differences could be found that the analysis on the data of entire run rendered more compact and smoother active brain areas. This was mainly because more trials of data were included in the data analysis. For time courses, the average BOLD effects extracted by analyzing the first trial of the tactile fMRI data were well fit by the BOLD model ($r = 0.89$). In the same time, the BOLD model also well fit the averaged BOLD effects extracted by analyzing the data from entire tactile fMRI run ($r = 0.78$). The BOLD baseline of the OFF period of the second trial was also well maintained as expected. In other words, the BOLD effects induced by the tactile stimulation were nicely repeated to retain the contrasts between ON and OFF conditions across the entire experimental runs.

On the other hand, the group results obtained by analyzing the data from the first trial of and from the entire data of the first acupuncture fMRI run were shown in Figs. 3a and 3b. In the group random effects map derived by analyzing single-block data (Fig. 3a), it covered the broadest brain areas compared with the maps by the analyses on all other data portions. The random effects were peaking in the brain area of right primary sensory cortex (S1) representing the subjects' left foot area. However, when the analysis extended to include also the second half of the fMRI data from the first run, the random effects map dramatically shrank (Fig. 3b). It only selected a subset of what was found in the single block result. The maximum effects were found in secondary sensory cortex (S2) instead of S1.

For time courses, the extracted BOLD time course by the analysis on only the very first block (single block condition) was highly aligned with the BOLD model ($r = 0.94$, Fig. 3a). However, apparently, the BOLD time course extracted from multiple blocks of data failed to return to its baseline at the beginning (OFF period) of the second trial

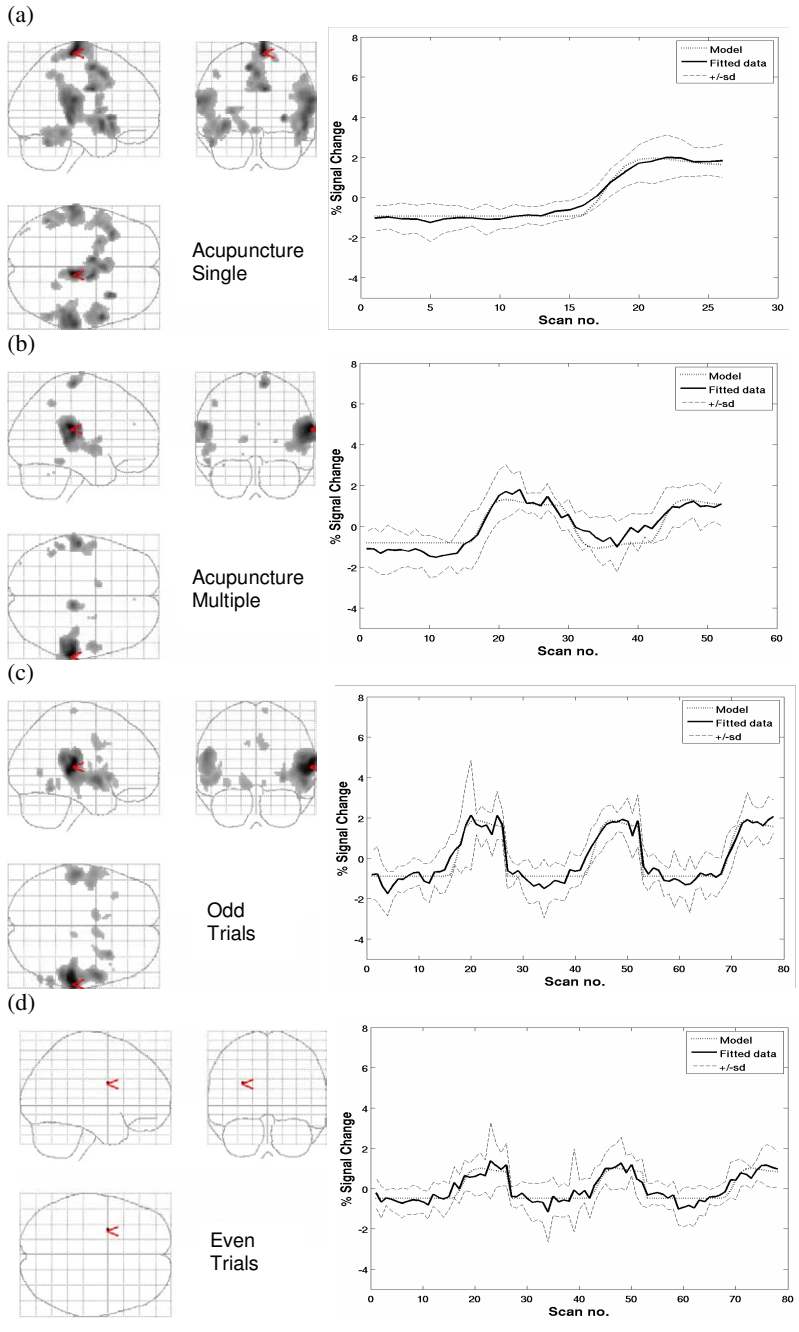

Figure 3. Results of applying SPM2 GLM analysis on different fMRI data portions. (a) Single block acupuncture fMRI data. (b) Multiple blocks of acupuncture fMRI data. (c) Odd number blocks of acupuncture fMRI data. (d) Even number blocks of acupuncture fMRI data. (e) All 6 blocks of acupuncture fMRI data. (f) Last 5 blocks of acupuncture fMRI data. (g) Single block of tactile fMRI data. (h) Multiple blocks of tactile fMRI data.

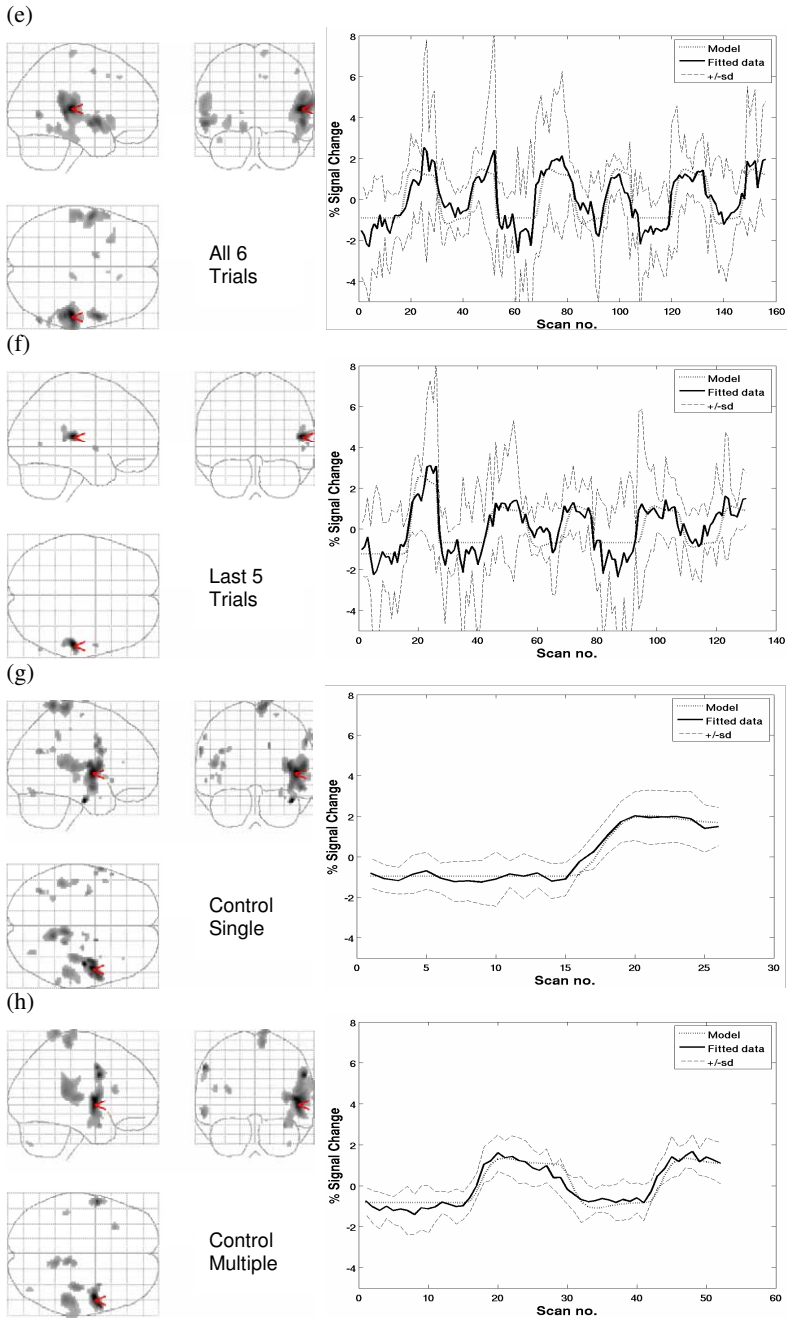

Figure 3. (*Continued*)

(Fig. 3b). As a result, the needling sensation introduced in the very first trial had lasted too long and thus elevated the baseline of the second block. Also, the BOLD contrast between ON and OFF treatments of the second block in the extracted time course was more distorted compared with the first one. This highly accounted for the lower correlation coefficient between the extracted BOLD time course and the model ($r = 0.74$).

Odd-number- versus even-number-block condition: In Figs. 3c and 3d, we compare the group random effects maps and the average BOLD time courses resulted from the analysis on the odd-number trials (3c) and on the even-number trials of acupuncture fMRI data. These 2 analyses also selected on subsets of the random effects map rendered by the analysis on the single-block data (3a). However, the analysis on the odd-number blocks selected much broader active brain areas than the analysis on the even-number blocks. Actually, the random effects maps resulted from the analysis on odd-number blocks best resembled the result of the analysis on single block. On the other hand, as shown in the figure, the analysis on even-number blocks gave the worst random effects map among all the analysis conditions. Almost nothing significant could be found under the threshold of corrected $p < 0.05$.

Regarding the detected BOLD time course at the peak active voxel, the ON/OFF contrast was much preserved in the time course obtained by the analysis on the odd-number blocks (Fig. 3c). Compared with the time course resulted from the analysis on the even-number blocks (Fig. 3d), it was almost 3 folds larger. This implied that the ON/OFF contrasts were strongly distorted by carryover effects — made it 3 times smaller — and made the task-related effects much harder to find in data. On the other hand, with a small separation between fMRI runs, it was helpful to mitigate the influence of the carryover effects. However, the carryover effects introduced by the acupuncture stimulation used here was too strong to be completely eradicated by such a small separation between fMRI runs (Ho *et al*., 2007). This was reflected by the differences between the random effects maps obtained by the analyses on odd-number blocks and single block only of the fMRI data.

All-6 versus last-5-block condition: Figs. 3e and 3f show the resulting group effects maps and the BOLD time courses at the peak active voxels from the analyses on the entire fMRI data and on the data without the very first block. Again, these 2 analysis conditions also selected subsets of the active brain areas rendered by the analysis on the very first block of data. Although the means of the effect time courses of the peak active voxels across subjects did resemble the ON/OFF task reference time course, their standard errors were much higher compared to the mean effect time courses obtained by other analysis conditions. This implies the effect time courses in the selected active brain areas are less consistent and much less voxels, whose effect time courses resembled the BOLD model, could be found in these 2 analyses.

Coefficient of determination: CODs were calculated as square of the correlation coefficient (r^2) between the BOLD time course at the peak active voxel and the BOLD model used in first-level GLM analysis. This value was first derived at individual level and then averaged across all 11 subjects. Figure 4 shows the mean ± standard deviation of the CODs obtained by different analysis conditions (e.g., “Single”, “Multiple”, etc.). The

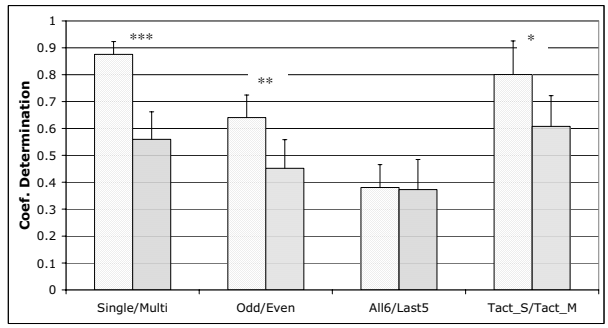

Figure 4. Paired t-test of the fitting coefficients between model and extracted BOLD time courses obtained by SPM2 GLM analysis on different fMRI data portions. "Single/Multi": single block versus multiple blocks of acupuncture fMRI data. "Odd/Even": odd number blocks versus even number blocks of acupuncture fMRI data. "All6/Last5": all 6 blocks versus last 5 blocks of acupuncture fMRI data. "Tact_S/Tact_M": single block versus multiple blocks of tactile fMRI data. ***indicates $p < 0.000001$, $^{**}p < 0.00005$, and $^{*}p < 0.001$.

resultant CODs were also compared using a paired t-test on condition pairs (e.g., "Single" versus "Multiple," etc.). In the figure, "Tact S" means the single-block analysis condition on tactile fMRI data and "Tact M" the multiple-block analysis condition.

We expected the same experimental effects should be extracted by the same type of analysis if no carryover effects were involved, although different portions of fMRI data were analyzed. As a result, the average CODs should be rather comparable across different analyses. However, in our results obtained from acupuncture fMRI data, we found significant differences between the CODs in the paired comparisons of "Single" versus "Multiple" ($p < 0.000001$) and "Odd" versus "Even" ($p < 0.00005$). No difference was found in the paired comparison of "All 6" versus "Last 5." This implied that the carryover effects of acupuncture needling sensation did profoundly affect the results of data analysis. Surprisingly, even in the tactile fMRI runs, the carryover effects were also mildly revealed. This tactile paradigm was otherwise expected to have no contamination from the carryover effects whatsoever. However, the average CODs did show slight difference ($p < 0.001$) if different portions of fMRI data were analyzed.

Active brain areas: The active brain area (as shown in Fig. 3a) rendered the contralateral (right) primary sensory cortex (S1), which represented the (left) foot area corresponding to the location of the ST42 acupoint and showed the highest t-value. In addition, the dorsal anterior cingulate cortex (dACC), supplementary motor area (SMA), contralateral primary motor cortex (M1), bilateral secondary sensory area (S2), bilateral anterior insula, bilateral inferior frontal cortex, bilateral inferior parietal lobule, thalamus, ipsilateral amygdala, and bilateral inferior temporal gyrus were also included. This result was highly in line with what generally defined as the pain matrix (Jones, 1998; Wager, 2005). However, much stronger activities could be found in the S2 areas. This might in turn render the acupuncture-specific brain responses to the needling on ST42 acupoint.

Among the resultant random effects maps obtained by other analysis conditions, the one resulted from the analysis on the odd-number blocks best resembled the maps of the single block result. However, the activations in the S1, dACC, and SMA were less pronounced

and the activations in the inferior parietal lobule were completely missing. Others picked even smaller subset of the single block result.

As a control paradigm, the tactile stimulation activated exactly the same primary sensory cortex, representing the left foot area, since the same site was stimulated (Figs. 3a, 3g and 3h). Other brain areas, such as contralateral M1, bilateral S2 and insula — more on the contralateral side — were also included in the significantly active brain areas. However, the activation in M1 was more pronounced in response to the tactile stimulation; in contrast, insula was more activated in response to the acupuncture stimulation. In addition, no activations were found in the dACC, amygdala, and thalamus registered to the tactile stimulation. As a result, the tactile stimulation might lend itself as a better control paradigm to the acupuncture stimulation to extract the acupuncture specific brain responses (Hui *et al*., 2006).

T-values of the peak active voxels: In Table 1, we list the t-values (corrected with FDR) of the peak active voxels in the fixed effects maps derived by SPM2 GLM analysis at individual level. Different columns show results obtained by analyzing different fMRI data portions; different rows give results from different subjects. The results are also summarized by their means and standard deviations across subjects for each of the analyses. To assess if any differences between the peak t-values obtained by different analysis conditions, the maximum t-values were first paired according to conditions of "Single versus Multiple,"

Table 1. T-*Values* of the Peak Activation Voxels Found in the Individual Fixed Effects Obtained by the First-Level Analysis of SPM2

	Single	Multiple	Odd	Even	All 6	Last 5	Tact S	Tact M
S01	15.26	8.54	14.36	9.01	11.91	11.23	15.40	11.24
S02	11.00	6.94	13.27	10.06	7.40	6.47	7.79	8.06
S03	13.07	7.63	8.68	5.66	8.51	7.30	13.14	10.13
S04	10.91	6.29	12.82	8.35	9.94	9.33	14.44	9.58
S05	13.91	10.02	11.37	10.29	12.41	10.77	20.36	12.39
S06	13.00	5.21	12.58	5.86	8.42	6.96	4.93	6.23
S07	23.77	9.01	10.74	5.90	9.14	6.97	9.73	7.00
S08	12.47	9.52	9.12	7.03	3.91	5.90	7.90	12.63
S09	15.01	10.37	13.00	7.99	11.92	9.07	11.91	7.66
S10	15.00	7.99	12.66	9.34	10.15	12.00	10.97	6.65
S11	9.19	9.43	9.79	6.95	9.31	9.99	9.50	9.19
Mean	13.87	8.27	11.67	7.86	9.37	8.73	11.46	9.16
Std. Dev.	3.80	1.63	1.86	1.69	2.42	2.11	4.27	2.26
p	<0.0006		<0.00002					
Max. Group t	12.20	11.73	12.10	6.49	11.39	7.17	7.89	9.75

Each column represents the resultant t-values of the data analysis applied to different portions of fMRI data (e.g. "Single", "Multiple", etc.); each row represents an individual subject ("S01" to "S11" plus mean and standard deviation). Significance level of the differences between the group results of "Single" versus "Multiple", "Odd" versus "Even", "All 6" versus "Last 5", and "Tact S" (single block of tactile fMRI data) versus "Tact M" (multiple block of tactile fMRI data) were also evaluated using paired t-tests. The row entitled "p" shows the comparisons of the group results (shaded) with significant differences: "Single" > "Multiple" ($p < 0.0006$) and "Odd" > "Even" ($p < 0.00002$). At the bottom, the t-values of the peak activation voxels as obtained in the group random effects are listed.

"Odd versus Even," etc. and then compared using paired t-test. The columns listed with significance level in Table 1 show the comparisons with significant differences. In summary, the analysis on the first block fMRI data yielded much higher t-values at the peak active voxels than the analysis on the data of the entire first fMRI run ($p < 0.0006$). Also, the peak t-values obtained by analyzing the odd-number blocks of data were significantly higher than those obtained by analyzing the even-number blocks ($p < 0.00002$). No difference was found in comparing the results of analyses on all 6 blocks and last 5 blocks of the acupuncture fMRI data as well as on the single block and multiple blocks of tactile fMRI data.

Discussion

Here, we demonstrated that a contrast-based statistical analysis could be strongly affected by the carryover effects introduced by needling sensation. In a psychophysical experiment done by Ho *et al.* (2007), the authors showed that once needling sensation was produced, it took at least 2 min for the needling sensation to return to its baseline. This strong and long-lasting effect did not seem easy to be removed by manipulating experimental designs or by modeling in data analysis. This might explain why the acupuncture fMRI studies with repetitions of one-minute or shorter ON/OFF block had drawn incongruent conclusions as addressed in the literature. The incongruent conclusions could be caused by the unexplained carryover effects of the acupuncture needling sensation. The evidence shown in this study revealed that the needling sensation introduced in the early trials in an acupuncture fMRI run was carried over to the upcoming ones and the BOLD baselines of the following trials were thus altered and the designated ON/OFF contrasts were distorted. As a result, if data analysis was performed on the image data without taking the carryover effects into account, the resulting brain BOLD activation could be largely underestimated or even missed completely.

In our results, we simply showed that by analyzing different blocks of fMRI data, the same GLM method resulted in very different fixed effects and in turn arrived at different random effects at group level. Basically, the GLM analysis applied to the very first image block of the acupuncture fMRI data gave the most promising estimation of the random effects map and the BOLD activities across all 11 subjects. The random effects map showed the broadest coverage of the active brain areas, peaking at the primary sensory cortex of the acupuncture site. The BOLD model could best explain the variance of the extracted BOLD activities at the peak active voxels. Once we tried to add into the analysis with more data blocks, the extent of rendered active brain areas as well as the statistical power started to drop dramatically. According to the maximum group statistics as listed in Table 1 (row "Max. Group t"), we can rank the statistical results as "Single" (max. t = 12.20) > "Odd" (12.10) > "Multiple" (11.73) > "All 6" (11.39) > "Last 5" (7.17) > "Even" (6.49). It is really interesting and counter-intuitive to see the analysis on the data of only one trial of data giving the best statistical power and rendering broader active brain regions compared with the ones on more trials. This counter-intuitive result implies that the main experimental effects were most pronounced in the very first trial of the first fMRI run and were getting worse and worse when the experiment moved on. On the other hand, a small

separation in between the fMRI runs helped to bring the experimental effects back but not completely. This was reflected by the analysis on the odd number blocks of data, which should have otherwise produced a better statistical result than the analysis on single block data, if the experiment effects could be completely brought back to the original strength. As a result, we ascribed this distinction between the statistical results to the carryover effects. Of course, one can argue that this distinction might be caused by adaptation or habituation to the acupuncture stimulation. However, if this is the case, we should also see gradual degradation of the extracted BOLD effects along the time axis in the result derived from the odd number blocks of data.

It is also worth noting that in comparing the statistical results derived from the very first data block (Figs. 3a and 5a) and the entire image data (Figs. 3e and 5e), the latter had only picked very limited subset of the former one. Unfortunately, the result shown in Figs. 3e and 5e are the most commonly seen in a regular block-designed fMRI study in which fMRI data are analyzed by a contrast-based analysis method. It implies that if we simply analyze an fMRI data set without knowing or caring about whether or not carryover effects are involved in the recorded BOLD signals, it is very likely that we will extremely underestimate the task-related BOLD effects. This is mainly because the designated BOLD contrasts in the later part of the experimental runs/sessions may be highly distorted by the carryover effects introduced in the earlier trials/sessions. Therefore, the carryover effects are really the key to why acupuncture fMRI studies have long been deriving inconsistent results and in turn drawing such incongruent conclusions in the literature.

Single-Block Experimental Design

Our findings were highly in line with the results found in Koyama *et al*. (2003), in which they studied the differences between the single-block and multiple-block experimental designs for an fMRI experiment. In the fMRI experiment, they delivered thermal pain stimulation to subjects and found that the stimulus-induced BOLD changes recorded in the fMRI runs with single-block design were easier to be found. Accordingly, they suggested that a single-block experimental design might be more preferable for a pain fMRI study.

However, unlike our findings, their result still showed increasing statistical power as the block number included in data analysis was increased. The reason might be that their data were less contaminated by carryover effects since they were using thermal pain stimulation and pain sensation dropped much faster after stimulus stopped compared with needling sensation induced by acupuncture stimulation. Another reason could be that their experiment had longer inter-run interval for the subjects to rate their pain sensation. In our case, we only allowed for less than 15 sec for reloading the imaging pulse sequence. In our results comparing the analyses applied to the odd number blocks and even number blocks of data, we did find that brief separation between fMRI runs did help to mitigate the influences of carryover effects. Therefore, in an fMRI study if carryover effects are profoundly involved and difficult to avoid, a multiple single-block experimental design with long inter-run interval may be more appropriate. Single-block design with long inter-run interval helps to alleviate the influences of carryover effects; in the same time, multiple repetitions provide sufficient statistical power for the data analysis.

Acknowledgments

This study was supported by National Science Council (NSC 94-2314-B-039), Taiwan. The authors would like to thank Mr. Chang-Hai Tsai, Chairman of China Medical University and Hospital, for his great support on this study.

References

Aron, A., H. Fisher, D.J. Machek, G. Strong, H. Li and L.L. Brown. Reward, motovation, and emotion systems associated with early-stage intense romantic love. *J. Neurophysiol.* 94: 327–337, 2005.

Beijing College of Traditional Chinese Medicine, Shanghai College of Traditional Chinese Medicine, Nanjing College of Traditional Chinese Medicine, and The Acupuncture Institute of the Academy of Traditional Chinese Medicine. *Essentials of Chinese Acupuncture*. Foreign Language Press, Beijing, 1993.

Friston, K.J., C.D. Frith, R. Turner and R.S.J. Frackowiak. Characterizing evoked hemodynamics with fMRI. *NeuroImage* 2: 157–165, 1995.

Ho, T.J., J.R. Duann, W.C. Shen and J.G. Lin. Needling sensation: explanation of incongruent conclusion drawn from acupuncture fMRI study. *J. Altern. Complement. Med.* 13: 13–14, 2007.

Hui, K.K.S., E. Nixon, M. Vangel, S.M. Hodge, J. Liu, O. Marina, V. Napadow, N. Makris and D.N. Kennedy. Comparing the psychophysical response to acupuncture and tactile stimulation. *The 36th Annual Meeting of Society for Neuroscience*, Atlanta, 14–18 October 2006.

Jones, A. The pain matrix and neuropathic pain. *Brain* 121: 783–784, 1998.

Kolbitcsh, C., M. Schocke, C. Hormann, I.H. Lorenz, C. Kremser, R. Ellinger, F. Zschiegner, S. Felber, F. Anchner and A. Benzer. Effects of hyperoxia and hypocapnia on regional venous oxygen saturation in the primary visual cortex in conscious humans. *Br. J. Anesth.* 83: 835–838, 2003.

Koyama, T., J.G. McHaffie, P.J. Laurienti and R.C. Coghill. The single-epoch fMRI design: validation of a simplified paradigm for the collection of subjective ratings. *NeuroImage* 19: 976–987, 2003.

Locascio, J.J., P.J. Jennings, C.I. Moore and S. Corkin. Time series analysis in the time domain and resampling methods for studies of functional magnetic resonance imaging. *Human Brain Mapp.* 5: 168–193, 1997.

Michelon, P.M., J.M. Vettel and J.M. Zacks. Lateral somototopic organization during imagined and prepared movements. *J. Neurophysiol.* 95: 811–822, 2005.

Monsell, S. Task switching. *Trends Cogn. Sci.* 7: 134–140, 2003.

Raja, S.N., R.A. Meyer, M. Ringkamp and J.N. Campbell. Peripheral neural mechanisms of nociception. In: P.D. Wall and R. Melzack (eds.) *Textbook of Pain*. Churchill Livingstone, Edinburgh, 1999, pp. 13–44.

Stux, G. and B. Pomeranz (eds.). *Basics of Acupuncture*. Springer-Verlag, Berlin, 1998.

Wager, T.D. The neural bases of placebo effects in anticipation and pain. *Semin. Pain Med.* 3: 22–30, 2005.

Wu, G., J.N. Campbell and R.A. Meyer. Effects of baseline skin temperature on pain ratings to suprathreshold temperature-controlled stimuli. *Pain* 90: 151–156, 2001.