Mugla Journal of Science and Technology

A PRACTICAL ESTIMATION OF THE REQUIRED SAMPLE SIZE IN FMRI STUDIES

Cemre CANDEMİR*, International Computer Institute, Ege University, Izmir,
cemre.candemir@ege.edu.tr
(https://orcid.org/0000-0001-9850-137X)

Received: 13.04.2023, Accepted: 24.08.2023
*Corresponding author
Research Article
DOI: 10.22531/muglajsci.1282492

Abstract

In functional Magnetic Resonance Imaging (fMRI) studies, the variability in fMRI data, the complexity of the analysis, and the need to correct for multiple comparisons make determining the appropriate sample size challenging. Hence, power analysis becomes an important tool to use for determining the appropriate sample size needed to achieve reliable and statistically significant results. In this context, this study aims to represent the process of conducting a power analysis and estimating the sample size for an fMRI study. To do this, three functional, affective, behavioral, and cognitive, data sets having different experimental task designs are used. This study provides a step-by-step guide on how to conduct a power analysis and estimate the sample size for various fMRI studies.

Keywords: Experimental task design, fMRI, number of participants, power analysis, sample size, statistical significance

FMRI ÇALIŞMALARINDA GEREKLİ ÖRNEK BÜYÜKLÜĞÜNÜN PRATİK BİR TAHMİNİ

Özet

Fonksiyonel Manyetik Rezonans Görüntüleme (fMRI) çalışmalarında, fMRI verilerindeki değişkenlik, analizin karmaşıklığı ve çoklu karşılaştırmalar için düzeltme ihtiyacı, uygun örneklem büyüklüğünü belirlemeyi zorlaştırır. Bu nedenle, güç analizi, güvenilir ve istatistiksel olarak anlamlı sonuçlar elde etmek için gereken uygun örneklem büyüklüğünü belirlemek için kullanılan önemli bir araç haline gelir. Bu bağlamda, bu çalışma, bir fMRI çalışması için güç analizi yapma ve örneklem büyüklüğünü tahmin etme sürecini temsil etmeyi amaçlamaktadır. Bunu yapmak için, farklı deneysel görev tasarımlarına sahip işlevsel, duyuşsal, davranışsal ve bilişsel üç veri seti kullanılır. Bu çalışma, güç analizinin nasıl yürütüleceğine ve çeşitli fMRI çalışmaları için örneklem boyutunun nasıl tahmin edileceğine ilişkin adım adım bir kılavuz sağlar.

Anahtar Kelimeler: Deneysel ödev tasarımı, fMRI, katılımcı sayısı, güç analizi, örneklem büyüklüğü, istatistiksel anlamlılık

Cite

Candemir, C., (2023). "A practical estimation of the required sample size in fMRI studies", Mugla Journal of Science and Technology, 9(2), 56-63.

1. Introduction

With the rapid development of brain imaging tools in the last decades, there have been great developments in brain studies. The most effective imaging method is magnetic resonance imaging (MRI), which is a non-invasive method, does not contain radiation, and has high resolution. Thanks to MRI, a detailed structural examination of the brain is possible. Thus, it provides important information on issues such as making critical decisions before surgery, understanding the effects of diseases and drugs on the structure of the brain, and monitoring the healing process [1]. Moreover, functional MRI (fMRI), which is a modulation of MRI, can understand the working mechanism of the brain. Questions such as which stimuli create activation in which part of the brain, which areas work together, and which part is responsible for which function have been answered by fMRI studies. Researches based on such questions can light the neuroscience of the brain [2].

However, performing fMRI studies is challenging in many ways, including technical difficulties, data complexity, participant recruitment, ethics and safety, cost, and data interpretation [3]. The high cost of an fMRI scan can limit the number of participants included in the study, as well as cannot to be repeated the same scan many times. Thus, estimating the sample size (i.e. the number of participants) becomes an important step in planning such research [4]. Another challenge is the complexity of the fMRI data, which can make it difficult to estimate the effect size and variability. Additionally, the heterogeneity of the fMRI data can make it challenging to determine the appropriate statistical methods to use for sample size estimation. In this context, the sample size of the study should be determined optimally so that it should be

statistically significant and large enough to generalize to a population, but not so large as to be prohibitively expensive or time-consuming. To overcome this issue, there are methods available, such as power analysis, to compute the sample sizes for an fMRI experiment, estimating the mixture distribution of null and active peaks [5], [6].

The power analysis in the research process helps to determine the appropriate sample size needed to detect a statistically significant effect with sufficient power. The level of power required depends on factors such as the effect size, alpha level (i.e., significance level), and type of statistical test. In this manner, [7] indicates the importance of power analysis and take attention to the well-established, reliable neuroscientific methodologies. In parallel, [7] and [8] points out the high correlations despite the low statistical power of the studies. The results presented with low power are more unlikely to be reflect the true effect and not reproducible. Thus, researchers strictly recommend to conduct the power analysis even before the fMRI data acquisition and report the parameters and results of power analysis in their study [9]–[11]. To summarize, the use of power analysis for estimating sample size in fMRI studies is an important step for ensuring reliable and statistically significant results.

In this study, the required sample size for fMRI data sets with different characteristics was examined. Statistically significant and generalizable numbers of participants, suitable for the characteristics of each stimulus and experiment, were investigated. Thus, it is aimed to be a guiding analysis of the number of participants to be included in group fMRI studies.

2. Conducting Power Analysis

2.1. Challenges and Advantages of Power Analysis

Conducting fMRI research can be difficult due to various reasons. *(i) Technical:* One of the main challenges is the limitations of the tool itself, since it requires sophisticated equipment, such as the MRI scanner, and specialized software and hardware for data acquisition and analysis. This can make it challenging to set up and conduct studies, particularly in low-resource settings.

(ii) Data complexity: fMRI data is complex and multi-dimensional, and requires specialized expertise in neuroimaging analysis to interpret and analyze. This can be time-consuming and require extensive training and experience. Additionally, fMRI studies are becoming harder to reproduce due to the development of more complex paradigms and the variety of analysis techniques available [12].

(iii) Participant recruitment: Recruiting participants for fMRI studies can be challenging, particularly for studies that require specific clinical or demographic characteristics. This can limit the generalizability of the study results.

(iv) Ethics and safety: fMRI studies involve exposing participants to the strong magnetic fields and radio waves of the MRI scanner, which can cause discomfort or even harm in rare cases. Ensuring participant safety and obtaining informed consent can be challenging.

(v) Cost: fMRI studies can be expensive to conduct, particularly in high-resource settings, which can limit the sample size.

(vi) Data interpretation: The interpretation of fMRI data requires expertise in both neuroimaging and the specific research domain being investigated. This can be challenging, particularly for studies that involve complex cognitive or behavioral tasks.

(vi) Reproducibility: fMRI studies can be difficult to reproduce due to the complexity of the data and the specialized equipment and expertise required. This can limit the generalizability of the study results and make it difficult to build on previous research.

Apart from the abovementioned challenges, another challenge is the lack of reporting of sample size calculations in fMRI studies, which can limit researchers running power analyses for new studies [13], [14]. Additionally, the estimation of effect size and variance components is required for sample size calculations, but these estimates are often not reported in fMRI studies [13].

In this context, power analysis is crucial for the validity of the fMRI results. It prevents the issue of low statistical power, which decreases the chance of detecting a true effect. In other words, the least replicable findings will have the lowest power. On the other hand, conducting a power analysis and estimating the adequate sample size as an a priori step in an fMRI study has several other benefits: *(i) More accurate results:* By estimating the required sample size before conducting the study, researchers can ensure that the study is adequately powered to detect the effects of interest. This can improve the accuracy and reliability of the results.

(ii) Efficient resource allocation: can help researchers allocate their resources more efficiently, the duration of the study, and the required equipment and personnel. *(iii) Reduced risk of false positives or negatives:* Adequately powered studies reduce the risk of false positives or negatives, as they can detect small but meaningful effects while avoiding over-detection of effects that may not be clinically or scientifically relevant. *(iv) Increased transparency:* Conducting a power analysis and reporting the estimated sample size in the study protocol increases the transparency of the study design and helps to promote scientific rigor and reproducibility. *(v)Better study planning:* researchers can plan their study more effectively, such as determining the appropriate statistical analysis methods and identifying potential limitations and sources of error.

2.2. Power Analysis Steps

Statistical power in fMRI studies refers to the ability to detect an effect when present. It is the probability of correctly rejecting the null hypothesis, which means concluding that an effect exists when it truly does. The

traditional null-hypothesis statistical testing (NHST) framework defines statistical power as the probability of correctly rejecting the null hypothesis. The aim is to achieve a power of 80% or higher, which means that if the study were repeated 100 times, it would be detected in 80 of the studies [5], [6], [15].

The power analysis for fMRI data involves some sequential steps. The flow chart of the process steps is given in Figure 1.

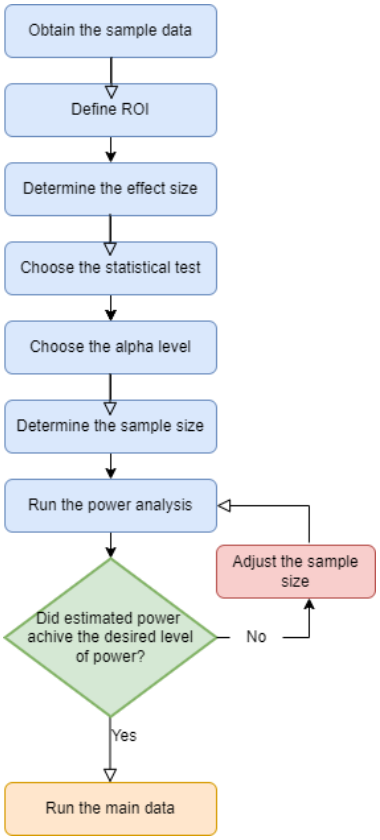


Figure 1. The flow chart of the power analysis process

i. *Obtaining the sample data:* The sample data can supply in several ways: Pilot study, expert opinion, simulation studies (i.e. synthetic data), meta-analysis or with searching prior studies in the same research area.

ii. *Defining a region of interest (ROI):* ROI selection specify the brain area and leads the accurate results.

iii. *Determine the effect size:* It is a measure of the strength of the relationship between the independent variable and the dependent variable in the study. In fMRI studies, the effect size can be estimated based on prior studies, pilot data, or literature review.

iv. *Choose a statistical test:* Different statistical tests have different assumptions and power characteristics. In fMRI studies, common statistical tests include t-tests, ANOVA, and regression analysis. The choice of the statistical test depends on the research question and experimental design.

v. *Choose the alpha level:* This is the significance level, which is the probability of rejecting the null hypothesis when it is true. The commonly used alpha level is 0.05, which means that there is a 5% chance of rejecting the null hypothesis.

vi. *Determine the sample size:* The sample size needed to reach a certain level of power depends on the effect size, alpha level, and the statistical test. A power analysis software package, such as NeuroPower, G*Power, FSL's randomize, or AFNI's 3dClustSim can be used to determine the adequate sample size to reach a desired level of power.

vii. *Run the power analysis:* Once the effect size, statistical test, alpha level, and sample size are determined, a power analysis can be run to estimate the power of the study.

viii. *Adjust the sample size:* If the estimated power is too low, the sample size can be adjusted to achieve a desired level of power. Alternatively, the effect size or alpha level can be adjusted to achieve a desired level of power. When the power is at the desired level, then the main study can be conducted with the determined sample size.

It should be also keep in mind that the other factors also affect power, such as variability and study design and these issues should be considered in the analysis [16].

3. Method

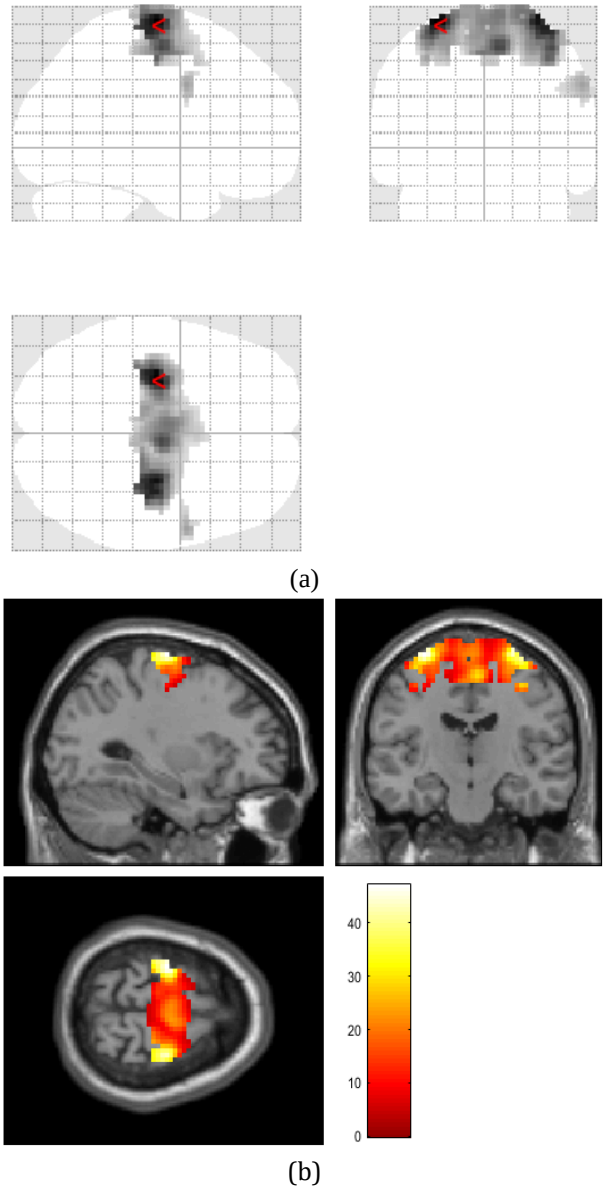
3.1. fMRI Data

This study, it is aimed to explore the sample size for different fMRI tasks. In this context, the analyses are run with using three diverse tasks: cognitive, behavioral, and affective functional data [17]. All fMRI data were acquired from healthy participants. All fMRI procedures were approved by the ethical committee of Ege University. Here, the cognitive task is stimulated by a memory-related task, the behavioral task is stimulated by a motor-related task, and the affective task is stimulated by an emotion-based task. Thus, each task has its own characteristics and may require different numbers of participants for group analysis. The details of each tasks and functional data structures can be reached through [17].

Before conducting the power analysis, the preprocessing of each functional data was done with the Statistical Parametric Maps (SPM) tool [18], which runs on the MatLab platform. A standard preprocessing procedure was followed, including the steps of realignment, slice timing, co-registration, segmentation, normalization, and smoothing. Finally, according to the research hypothesis of each task, the related ROIs are determined by the specialists, and ROI masks were generated with the WFU Pick Atlas.

3.2. Running Power Analysis

First, for each pilot data set, the statistical parametric maps were generated for a given sample (pilot) size by carrying out the first-level analysis on SPM. These maps represent the effect of individual analysis for each participant in the pilot study. After that, a sub-sample of 10 subjects was taken for each fMRI task to constitute the pilot group, and the second-level (group-wise) analyses were also performed. These analyses resulted in the statistical T-maps. With these maps, the power of the population was calculated and looked for the percentage of the peaks, and p-values, whether they show any activation in predetermined ROIs or not. An example of a statistical T-map is presented in Figure 2.



(a)

(b)

Figure 2. Active regions on the statistical T-map, which is acquired after the first and second level analysis using SPM (a) glass-brain display (b) canonical display

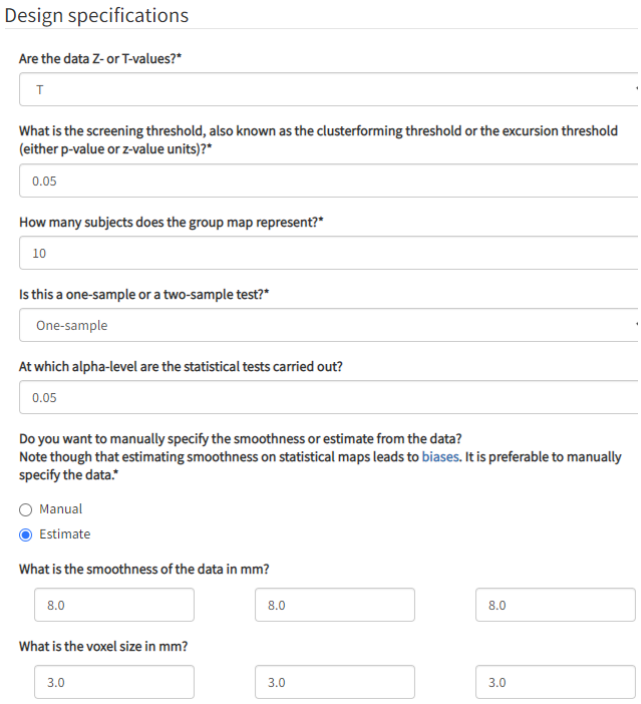
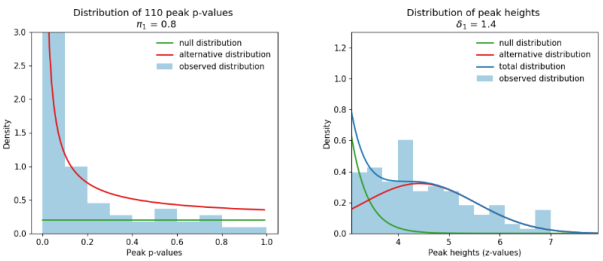


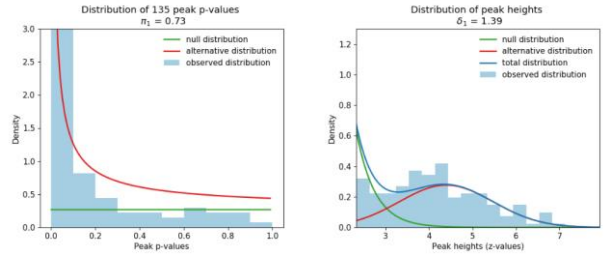
Figure 3. Importing the design specifications of the fMRI task

The power analyses were run using NeuroPower [19], which is an open toolbox used to perform this kind of calculation for fMRI tasks. The first step starts with the loading of the calculated statistical T-maps. In the next step, the generated ROI masks also can be selected, optionally. It may be beneficial to focus on the a-priori ROI rather than calculating whole-brain analysis. The next step is to insert the design specifications, which are unique for each task. The design specifications of a sample fMRI data set are given in Figure 3.

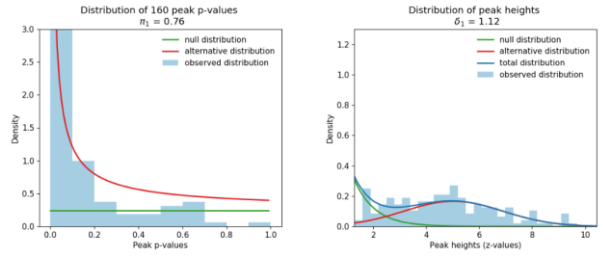
Here, the screening threshold can be determined as $p<0.001$ (uncorrected), $p<0.05$ for family-wise error (FEW) (corrected), or any desired value. The value for the subject number that the group map represent indicates the size of the pilot group (n=10). Afterward, the statistical test should be determined due to the different statistical tests have different assumptions and power characteristics. In this study, all data sets require a one-sample t-test at this step. Then, the significance level (alpha level) to be obtained from the results must be entered. Typically, it is used as 0.05. Finally, the special information for the functional data should be determined, including the smoothness level (the width of the Gaussian kernel) and the voxel size of the functional image in mm. In this study, all functional data were smoothed with an 8-mm Gaussian kernel and each image has a 3x3x3 mm voxel size.



(a)



(b)



(c)

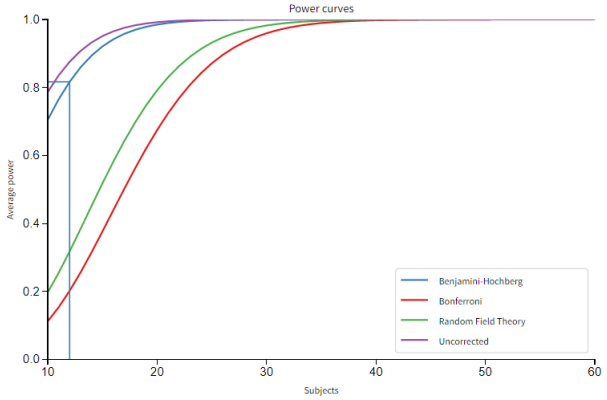
Figure 4. The distribution of the active peaks in each functional data set (a) Affective (b) Cognitive (c) Behavioral task design

4. Results

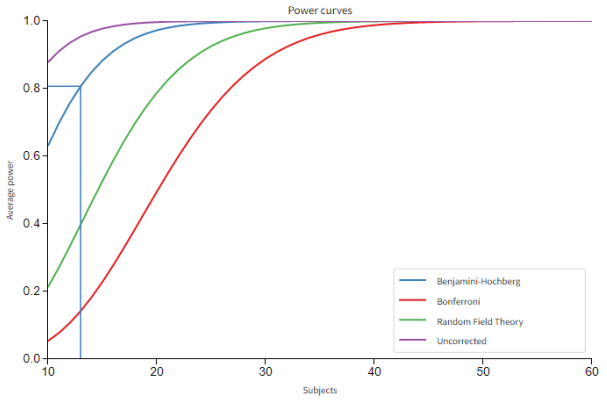
After importing the design specifications, the datasets were examined according to the local maxima. It gives the distribution of null and active peaks on functional data. Let's denote the active peaks as Z^u, which indicates the peaks above the determined monitoring threshold u. For the null hypothesis, i.e. non-active voxels ($Z^{u'}$), has an exponential distribution with $avg(\frac{u+1}{u})$ for the threshold u, which is given in Equation (1) [20].

$$f\left(z^{u'}\middle|H_o, Z^{u'} \geq u\right) = u * e^{(-u\left(z^{u'}-u\right))} \quad (1)$$

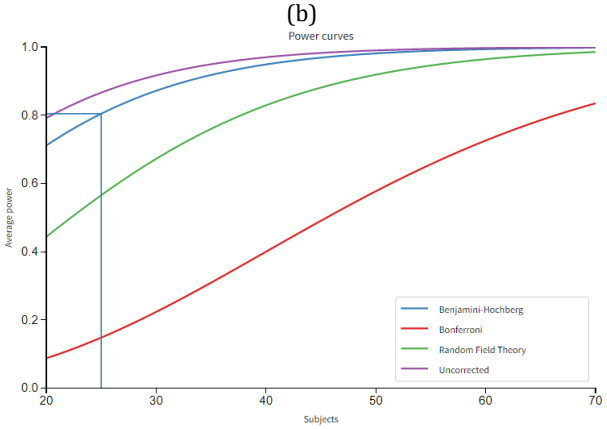
Furthermore, the total distribution of peak heights can be denoted with the mixture distribution, given in Equation (2) below.



(a)



(b)



(c)

Figure 5. The estimated power analysis with the multiple comparison procedure (a)Affective (b)Cognitive (c) Behavioral data

$$f\left(z^{u'}\middle|\pi_o, \mu_1, \sigma_1, Z^{u'} \geq u\right) = (1-\pi_1) f\left(z^{u'}\middle|H_0, Z^{u'} \geq u\right) + \pi_1 f\left(z^{u'}\middle|(H_1, \mu_1, \sigma_1, Z^{u'} \geq u\right) \quad (2)$$

Here, π_1 denotes the proportion of true positive peaks among all peaks above u, and μ_1 and σ_1 denotes the mean and standard deviation, sequentially. The resulting graphics for the distribution of active peaks for

behavioral, cognitive and affective data sets are given in Figure 4.

According to the distribution of the peaks, the results of the power analysis were exhibited with multiple comparison procedures (MCP). The results were carried with Bonferroni, Benjamini-Hochberg, Random Field Theory (RFT) statistical procedures, and uncorrected (default output) to control the False Discovery Rate (FDR) in multiple hypothesis testing [21], [22]. The Bonferroni correction is based on the Family Wise Error Rate (FWER), which is the probability of making at least one false discovery among all the tests. The Benjamini-Hochberg method is a modified Bonferroni correction, which is the expected proportion of false discoveries among all the rejected hypotheses. The Benjamini-Hochberg method has more power to detect true positives than the Bonferroni correction. On the other hand, RFT is a method used to correct for multiple comparisons in neuroimaging data by taking into account the spatial correlation between voxels. It assumes that the data are a discretization of a continuous random field with a certain smoothness and uses the Euler characteristic to threshold the image at a certain level to identify clusters of significant voxels [23].

For each data set, the parameters were set as threshold ($u = 0.01$), alpha-level $\alpha = 0.05$, the number of pilot subgroup $n = 10$, $8x8x8\ mm$ for smoothing kernel and $3x3x3\ mm$ for voxel size. The results of the power analysis according to each methodology are given in Figure 5. The intersection of the average power and power curves gives the required number of participants to reach the desired power in the main study. According to Figure 5, if we aim to reach 80% power, the affective fMRI task requires 12 subjects, cognitive data requires 13 and behavioral data requires 25 subjects for Benjamini-Hochberg error rate control. On the other hand, if we have a certain number of participants, it is also possible to know the power rate predictions with these graphics, as well. Furthermore, the detailed table for all possible sample sizes for the various correction methods is given in Table 1 for affective and cognitive functional data, and in Table 2 for the behavioral functional data.

When evaluating Table 1 and Table 2, it is also worth mentioning the low power with the small sample size. The low statistical power indicates a reduced likelihood of identifying effects that are true in reality. The studies with low power tend to yield a high number of false negatives compared to the studies with high power. For example, if a study has a design with a power of 10%, it signifies that the study is expected to find only 10 true effects among the existing 100 non-null effects. In addition to this, it is also reported that the low power is related to the other additional biases, such as publication bias. Thus, these factors result in the low reliability of the acquired results in studies with low power [7].

Table 1. Sample sizes with the estimated power analysis for affective and cognitive data sets.

Sample Size	Bonferonni		Benjamini-Hochberg		Random Field Theory		Uncorrected	
	Affective	Cognitive	Affective	Cognitive	Affective	Cognitive	Affective	Cognitive
10	.11	.05	.70	.63	.20	.21	.79	.88
11	.15	.08	.77	.70	.25	.27	.84	.91
12	.20	.11	.82	.76	.32	.33	.88	.94
13	.26	.14	.86	.81	.38	.40	.91	.95
14	.31	.18	.89	.85	.45	.46	.93	.97
15	.38	.23	.92	.88	.52	.53	.95	.98
16	.44	.28	.94	.91	.58	.59	.97	.99
17	.50	.33	.96	.93	.64	.65	.98	.99
18	.56	.38	.97	.95	.70	.70	.98	1.0
19	.62	.44	.98	.96	.75	.75	.99	1.0
20	.67	.50	.99	.97	.79	.79	.99	1.0
21	.72	.55	.99	.98	.83	.82	.99	1.0
22	.77	.60	.99	.99	.86	.86	1.0	1.0
23	.81	.65	1.0	.99	.89	.88	1.0	1.0
24	.84	.69	1.0	.99	.91	.91	1.0	1.0
25	.87	.74	1.0	.99	.93	.93	1.0	1.0
26	.90	.77	1.0	1.0	.95	.94	1.0	1.0
27	.92	.81	1.0	1.0	.96	.95	1.0	1.0
28	.93	.84	1.0	1.0	.97	.96	1.0	1.0
29	.95	.87	1.0	1.0	.98	.98	1.0	1.0
30	.96	.89	1.0	1.0	.98	.98	1.0	1.0

Table 2. Sample sizes with the estimated power analysis for behavioral data set.

Sample Size	Bonferonni	Benjamini-Hochberg	Random Field Theory	Uncorrected
	Behavioral	Behavioral	Behavioral	Behavioral
20	.09	.71	.44	.79
21	.10	.73	.47	.81
22	.11	.75	.49	.82
23	.12	.77	.52	.84
24	.13	.79	.54	.85
25	.15	.80	.57	.87
26	.16	.82	.59	.88
27	.18	.83	.61	.89
28	.19	.86	.63	.90
29	.21	.86	.65	.91
30	.22	.87	.67	.92
31	.24	.88	.69	.92
32	.26	.89	.71	.93
33	.27	.90	.73	.94
34	.29	.91	.74	.94
35	.31	.92	.76	.95
36	.33	.93	.78	.95
37	.35	.94	.79	.96
38	.36	.94	.80	.96
39	.38	.95	.82	.97
40	.40	.95	.83	.97

5. Conclusion

Power analysis is an important step in fMRI studies since it helps to constitute the appropriate number of participants to detect a significant effect with a desired level of statistical power. This is crucial to validate the study has enough statistical power and thus can avoid to deal negatives. Power analysis can also help to optimize the experimental functional task design besides the wasting resources. However, in a neuroimaging research study put forward that only 3% of the studies conducted the power calculations [14]. They reported that it results with missing information in the fMRI trials, underpowered results, and false negative outcomes.

In this context, in this study it was aimed to explore the sample size for different fMRI tasks. In this context, the analyses were run with using three diverse tasks: cognitive, behavioral, and affective functional data. The results indicated that on average, a sample size of 20 participants is needed to achieve a power of 80%, with an alpha level of 0.05, however it may change according to the nature of the functional task design. It is also correlated with the findings in a study, which is analyzed among popular fMRI studies. The results of it indicated that highly cited fMRI studies had a median sample size of 12 participants, most of the fMRI studies has a sample size of 14.5 participants on average, and these numbers are increasing at a rate of 0.74 participant per year [14].

Here, an important point of note is the significant absence of utilization of fMRI-tailored methods for calculating statistical power. Despite the availability of these procedures for a while [5], [6], [24], it seems that their application might be perceived as highly complex. Researchers might not have been introduced to these techniques, or perhaps, the incorporation of these methods might not have been deemed a sufficiently crucial aspect to warrant dedicated effort. The calculations of statistical power primarily center around singular t-tests and correlation examinations. It's probable that these calculations overly estimate statistical power due to their failure to account for any form of multiple-testing correction. Another possible approach may be using meta-analyses. Although the meta-analyses give a good estimation about the effect size, on the other hand, the challenges of conducting such an analysis limit its use.

As summary, it is obvious that power analysis is a critical step in planning and fMRI study. In the light of the findings, it is strongly suggested that researchers should conduct a power analysis, consider the effect size, statistical test, and alpha level while determining the appropriate sample size for an fMRI study.

6. Acknowledgment

The author thanks to all Socat team and Prof. Dr. Ali Saffet Gönül for his encourage and supervision.

7. References

[1] H. Aerts *et al.*, “Pre- and post-surgery brain tumor multimodal magnetic resonance imaging data optimized for large scale computational modelling,” *Sci. Data*, vol. 9, no. 1, Art. no. 1, Nov. 2022, doi: 10.1038/s41597-022-01806-4.

[2] S. Ogawa, T. M. Lee, A. R. Kay, and D. W. Tank, "Brain magnetic resonance imaging with contrast dependent on blood oxygenation.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 87, no. 24, pp. 9868–9872, Dec. 1990.

[3] R. T. Constable, "Challenges in fMRI and Its Limitations," in *Functional MRI: Basic Principles and Clinical Applications*, S. H. Faro and F. B. Mohamed, Eds., New York, NY: Springer, 2006, pp. 75–98. doi: 10.1007/0-387-34665-1_4.

[4] P. Lafaye de Micheaux, B. Liquet, S. Marque, and J. Riou, "Power and sample size determination in clinical trials with multiple primary continuous correlated endpoints," *J. Biopharm. Stat.*, vol. 24, no. 2, pp. 378–397, 2014, doi: 10.1080/10543406.2013.860156.

[5] J. Durnez *et al.*, "Power and sample size calculations for fMRI studies based on the prevalence of active peaks." bioRxiv, p. 049429, Apr. 20, 2016. doi: 10.1101/049429.

[6] J. A. Mumford, "A power calculation guide for fMRI studies," *Soc. Cogn. Affect. Neurosci.*, vol. 7, no. 6, pp. 738–742, Aug. 2012, doi: 10.1093/scan/nss059.

[7] K. S. Button *et al.*, "Power failure: why small sample size undermines the reliability of neuroscience," *Nat. Rev. Neurosci.*, vol. 14, no. 5, Art. no. 5, May 2013, doi: 10.1038/nrn3475.

[8] E. Vul and H. Pashler, "Voodoo and circularity errors," *NeuroImage*, vol. 62, no. 2, pp. 945–948, Aug. 2012, doi: 10.1016/j.neuroimage.2012.01.027.

[9] T. Yarkoni, "Big Correlations in Little Studies: Inflated fMRI Correlations Reflect Low Statistical Power-Commentary on Vul et al. (2009)," *Perspect. Psychol. Sci. J. Assoc. Psychol. Sci.*, vol. 4, no. 3, pp. 294–298, May 2009, doi: 10.1111/j.1745-6924.2009.01127.x.

[10] R. A. Poldrack, P. C. Fletcher, R. N. Henson, K. J. Worsley, M. Brett, and T. E. Nichols, "Guidelines for reporting an fMRI study," *NeuroImage*, vol. 40, no. 2, pp. 409–414, Apr. 2008, doi: 10.1016/j.neuroimage.2007.11.048.

[11] B. Thirion, P. Pinel, S. Mériaux, A. Roche, S. Dehaene, and J.-B. Poline, "Analysis of a large fMRI cohort: Statistical and methodological issues for group analyses," *NeuroImage*, vol. 35, no. 1, pp. 105–120, Mar. 2007, doi: 10.1016/j.neuroimage.2006.11.054.

[12] C. Pernet and J.-B. Poline, "Improving functional magnetic resonance imaging reproducibility," *GigaScience*, vol. 4, p. 15, Dec. 2015, doi: 10.1186/s13742-015-0055-8.

[13] Q. Guo, L. Thabane, G. Hall, M. McKinnon, R. Goeree, and E. Pullenayegum, "A systematic review of the reporting of sample size calculations and corresponding data components in observational functional magnetic resonance imaging studies," *NeuroImage*, vol. 86, pp. 172–181, Feb. 2014, doi: 10.1016/j.neuroimage.2013.08.012.

[14] D. Szucs and J. PA. Ioannidis, "Sample size evolution in neuroimaging research: An evaluation of highly-cited studies (1990–2012) and of latest practices (2017–2018) in high-impact journals," *NeuroImage*, vol. 221, p. 117164, Nov. 2020, doi: 10.1016/j.neuroimage.2020.117164.

[15] H. R. Cremers, T. D. Wager, and T. Yarkoni, "The relation between statistical power and inference in fMRI," *PLOS ONE*, vol. 12, no. 11, p. e0184923, Nov. 2017, doi: 10.1371/journal.pone.0184923.

[16] "Stata Programs for Data Analysis." https://stats.oarc.ucla.edu/stata/ado/analysis/ (accessed Apr. 11, 2023).

[17] O. T. Bişkin, C. Candemir, A. S. Gonul, and M. A. Selver, "Diverse Task Classification from Activation Patterns of Functional Neuro-Images Using Feature Fusion Module," *Sensors*, vol. 23, no. 7, Art. no. 7, Jan. 2023, doi: 10.3390/s23073382.

[18] J. Ashburner, "SPM: A history," *NeuroImage*, vol. 62, no. 2, pp. 791–800, Aug. 2012, doi: 10.1016/j.neuroimage.2011.10.025.

[19] "NeuroPower | NeuroPowerTools." http://neuropowertools.org/neuropower/neuropowerstart/ (accessed Apr. 11, 2023).

[20] K. J. Worsley, "Developments in random field theory," in *Human Brain Function*, R. S. J. Frackowiak, K. J. Friston, C. Frith, R. Dolan, K. J. Friston, C. J. Price, S. Zeki, J. Ashburner, and W. D. Penny, Eds., 2nd ed.Academic Press, 2003.

[21] W. Haynes, "Benjamini–Hochberg Method," in *Encyclopedia of Systems Biology*, W. Dubitzky, O. Wolkenhauer, K.-H. Cho, and H. Yokota, Eds., New York, NY: Springer, 2013, pp. 78–78. doi: 10.1007/978-1-4419-9863-7_1215.

[22] B. Walsh, "Multiple comparisons," Cornell University, 2004.

[23] H. Naouma and T. C. Pataky, "A comparison of random-field-theory and false-discovery-rate inference results in the analysis of registered one-dimensional biomechanical datasets," *PeerJ*, vol. 7, p. e8189, Dec. 2019, doi: 10.7717/peerj.8189.

[24] R. A. Poldrack *et al.*, "Scanning the horizon: towards transparent and reproducible neuroimaging research," *Nat. Rev. Neurosci.*, vol. 18, no. 2, pp. 115–126, Feb. 2017, doi: 10.1038/nrn.2016.167.