Breast Cancer Research and Treatment (2006) 95: 125–129
DOI 10.1007/s10549-005-9055-1

Clinical trial

The (mis)classification of chemo-fog – methodological inconsistencies in the investigation of cognitive impairment after chemotherapy

Val Shilling[1], Valerie Jenkins[1], and Ivonne Solis Trapala[2]
[1]*Cancer Research UK Psychosocial Oncology Group, Brighton and Sussex Medical School University of Sussex, Falmer East Sussex, UK;* [2]*Centro de Investigacion en Matematicas, A.C, Guanajuato Gto, Mexico*

Key words: breast cancer, chemotherapy, cognitive impairment

Abstract

Background. A growing number of studies report cognitive impairment after chemotherapy; indeed the phenomenon of chemo-fog is now almost universally accepted. We are concerned however that there is little if any consistency in the way in which patients are classified as showing cognitive impairment or not. We aim to demonstrate that different methods of analysis produce markedly different results, making the true extent of impairment unclear.

Methods. We analysed data from 92 breast cancer patients 4 weeks post-chemotherapy and from 42 healthy controls using 7 different methods, each taken from a different research paper in the area of cognitive impairment post-chemotherapy.

Findings. The extent of impairment was dependent on the method of analysis. Impairment ranged from 12 to 68.5% in the chemotherapy group and from 4.8 to 64.3% in the healthy control group.

Interpretation. This brief report highlights the contrasting degrees of cognitive impairment calculated by using legitimate statistical methods and demonstrates the need for a collaborative effort to standardise our methods that we might better understand the phenomenon of chemo-fog.

Introduction

Memory and attention problems after chemotherapy are frequently reported by patients [1] and have been objectively demonstrated in a number of studies [2–4]. We, like others, have previously commented on the problems associated with existing research – primarily its cross-sectional nature. Such studies run the risk of declaring a post-treatment difference that was in fact already present or of failing to show a difference when a high functioning group pre-treatment decline to meet a lower functioning group. Initially high functioning patients can experience significant cognitive loss but still score within the normal range.

We have also noted that comparisons between these studies are difficult for a number of reasons. The main hurdle is that many use a diverse range of measures assessing different aspects of cognitive function. No less serious are the variations in type and dose of chemotherapy and time since completion of chemotherapy.

We would like to draw attention to another disparity between the studies, which has the potential to change the results reported. Many studies classify participants as cognitively impaired or not impaired on the basis of their cognitive test scores. The proportion of each group reported as showing cognitive impairment can vary quite dramatically and ranges from 17% [5] to 75% [6]. The criteria for the classification of cognitive impairments differ quite substantially between studies however. This may result in a situation whereby a significant increase in the risk of cognitive impairment as reported in one published study might reveal a non-significant effect if it was analysed by different means – noted by Desai and colleagues in their own data [7]. To investigate the extent to which each method used determines the results of a study, we analysed data from our ongoing longitudinal study of cognitive function after chemotherapy for breast cancer, using different criteria for cognitive impairment as reported in other published work. This brief report presents the findings of these analyses and emphasises the need for consistency across research methods in order to establish a realistic appraisal of the nature of cognitive impairment after chemotherapy.

Method

The data used in these analyses are from an ongoing longitudinal study. Details on the methodology,

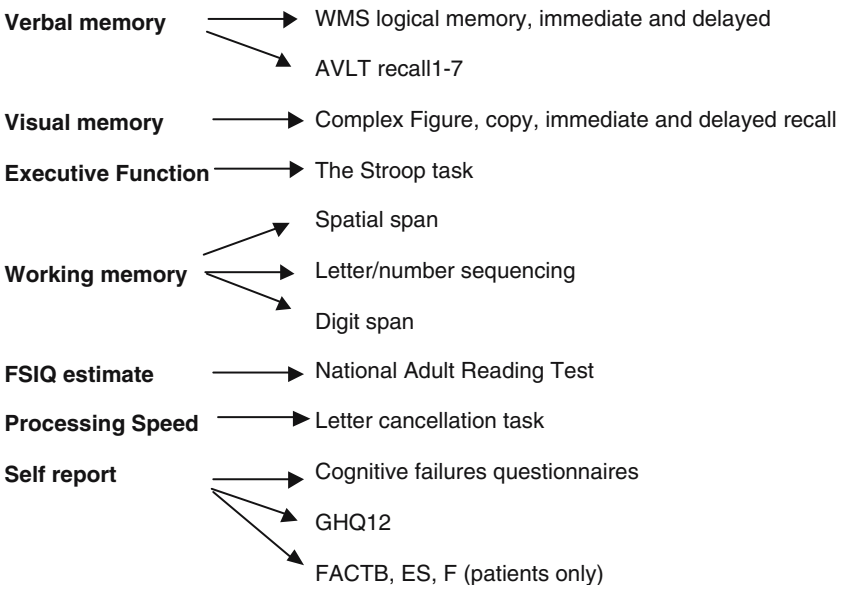

Figure 1. Cognitive test battery

participants and cognitive test battery are reported elsewhere [1] and are only summarised here.

Participants

Women with early stage breast cancer were recruited into the study prior to the start of adjuvant chemotherapy. The control group were a sample of convenience made up of friends and family of the patients and experimenters and from a local women's group. Participant groups did not differ significantly on age, estimated IQ or years of education. Cognitive assessments were made at baseline, four weeks after the final chemotherapy session (6 months in the control groups) and at 18 months. Data from the time 2 assessments of 92 chemotherapy patients and 42 healthy controls are discussed here.

Assessments

The cognitive test battery assesses several broad areas of cognitive function as outlined in Figure 1. All measures are standardized, validated and from published neuropsychological test batteries.

Statistical analyses

The data were analysed in seven different ways based on the analysis reported by other researchers working in the field. This is not an exhaustive list – we were unable to adopt the methods of analyses reported in all studies. In some cases individual participants were not classified as impaired or not impaired, rather, group comparisons were made on test scores. Other studies, whilst classifying participants failed to publish the criteria by which classification was made. The methods of analyses we used are listed below.

Method 1

Reliable change index [8]. Details of the method are reported elsewhere [1]. In summary, using the control group data as a reference, the amount of change between baseline and T2 is classified as showing reliable decline or not on each measure. Reliable decline on 2 or more (of 14) measures was classified as overall cognitive decline, based on the number of tests for mild cognitive impairment reported elsewhere [9]. As a strict criterion we also calculated the percentage of participants who showed reliable decline on two or more measures with no corresponding reliable improvement on any measure.

Method 2

Scores on each task were converted to z scores. If a participant had multiple test scores where $z \leq -1.5$, they were considered to show impairment. If impairment was seen on only 1 measure then in order for that participant to be classified as cognitively impaired, the score on that task must be in the lowest 2.5 percentile, corresponding to a z score of ≤ -2 [4].

Method 3

Performance in the lower quartile of each task is classified as impaired. Performance in the lower quartile on four or more tasks was classified as overall impairment [10]. In the reference paper domain scores averaged across a number of tasks were reported rather than individual test scores as is the case here and norms were used whereas we use control data as our reference for Z transformations.

Method 4

Failure on an individual test is a score of at least 2SD below the control mean. The number of tasks failed are

Table 1. Number and percentage of each participant group classified as impaired by each method and the relative risk

	Patient group	Control group	Odds ratio (95% CI)	χ^2 (p)
Method 1	26 (28.3%)	8 (19.0%)	1.67 (0.689–4.09)	1.29 (0.179)
Method 1 (strict criteria)	11 (12.0%)	2 (4.8%)	2.72 (0.57–12.84)	1.70 (0.16)
Method 2	42 (45.7%)	8 (19%)	3.57 (1.49–8.54)	8.73 (0.002)
Method 3	56 (60.9%)	20 (47.6%)	1.71 (0.82–3.57)	2.06 (0.106)
Method 4	35 (38%)	6 (9.5%)	3.68 (1.41–9.64)	7.67 (0.004)
Method 5[a]	63 (68.5%)	27 (64.3%)	1.21 (0.56–2.60)	0.23 (0.386)
Method 6	53 (57.6%)	13 (31%)	3.03 (1.40–6.57)	8.20 (0.004)
Method 7	49 (53.3%)	19 (45.2%)	1.38 (0.66–2.87)	0.74 0.250)

[a]30 patients were classified as moderately impaired and 33 as mildly impaired. Six controls were classified as moderately impaired and 21 as mildly impaired.

totalled. Overall cognitive impairment was then classified based on the 5th percentile of the control group – this corresponded to failure on three or more tests in the reference studies [3,5] (the fifth percentile cut off point for the control group actually included 12% of the control group due to the restricted range of scores). In our study the fifth percentile cut off point included 6 (9.5%) of the control group and corresponded to a score of at least 2SD below control mean on one or more tasks.

Method 5

Participants are divided into three classes. Not impaired corresponds to no more than one test score at less than or equal to 1 S.D. below control mean. Mildly impaired corresponds to performance less than or equal to 1SD below control mean on at least two measures. Finally, moderately impaired is less than or equal to 1SD below control mean on at least two measures and less than or equal to 2SD on at least one measure [6] (N.B. in the reference study population norms are used rather than a control group). For the purpose of this analysis, participants will be classified as cognitively impaired if they show mild or moderate impairment as described above [11].

Method 6

Raw scores are converted into z scores (the study cited here used norms whereas we use control data as our reference). Each test is then classified as impaired if performance was below the 10th percentile. Participants were classed as impaired if performance was below the 10th percentile on two or more tests [12].

Method 7

Overall performance was considered impaired if scores were at least 1 standard deviation below the control mean on at least three domains [13].

Results

Table 1 shows the percentage of the patient and control group that are classified as cognitively impaired by each method of analysis. The odds ratio and chi-square for impairment rate in healthy versus the patient group for each method is also shown.

The proportion of participants identified as showing cognitive impairment ranges from 12 to 68.5% in the patient group and from 4.8 to 64.3% in the control group. Method 1 (the Reliable Change Index) is the only method that takes into account change in performance rather than performance at a single time point and as such provides the most conservative estimation of impairment. Excluding this method the rate of impairment ranged from 38 to 68.5% in the patient group and 9.5–64.3% in the control group.

These rates of impairment in the chemotherapy group are largely in keeping with the rates reported in the reference studies from which the methodologies were taken. For example 38% in our study compared to 28% in the reference study [3]; 45.7% in our study, 35% in the reference study [4]; 68.5% in our study, 75% in the reference study [6]. The rates of impairment reported in our study differed from the reference study more substantially where the reference study had used domain scores and/or norms as compared to individual test scores and a healthy control group as used in our analysis [10,12].

It should be noted that cancer site varied in the reference studies. While we assessed only breast cancer patients (method 1) as did the reference studies for method 2 [4], method 4 [3,5] and method 5 [6], the reference study for method 3 included patients with breast cancer (55%) and lymphoma [10] and methods 6 and 7 used a mixed patient group comprising 35% [12] and 70% [13] breast cancer patients, respectively.

The absolute proportion of participants classified as impaired is to some extent irrelevant. What is important is the proportion of patients relative to controls/norms that are classified as impaired. As Table 1 shows, the risk of being classified as cognitively impaired ranges from 1.21 times as likely if you have had chemotherapy to 3.68 times as likely, depending on which method is used. Three of the seven methods show a significant relationship between classification and group at the conventional 5% significance level.

Discussion

We have demonstrated that methods of analyses used by different groups working in the area of chemotherapy and cognition do not produce the same results when applied to the same data set. It is important to note that we are not trying to demonstrate the superiority of one method over another, merely to draw attention to the fact that using different methods to analyse the same data can lead to very different results and obscure our understanding of the true rate of cognitive impairment.

We also do not suggest that the authors of the reference studies would necessarily have employed the same methodology to our data as they did to their own and acknowledge that the way our data is organised is not best suited to some of the methods of analysis. We use individual test scores while some of our reference studies used domain scores. We use a control group as the reference for z transformations while other studies used population norms. Undoubtedly this has led to some of the very high rates of impairment seen when some of the methods are applied to our data. In a sense however it is irrelevant what absolute proportion of participants is classified as impaired and not impaired as long as the relationship between the two groups is constant. The worrying thing about our investigation is the difference in odds ratios calculated using each method.

To illustrate this point, using the more lenient reliable change methodology we would state that patients were not significantly more likely to show cognitive impairment than the control group (OR 1.674, χ^2 1.29, p 0.179). Had we used a different method we might report that patients were significantly more likely than controls to be classified as cognitively impaired (OR 3.68, χ^2 7.665, p 0.004). While this may be an unfair comparison because the Reliable Change Indices used to analyse our data are not truly comparable with the other, single time point, methods of analysis, odds ratios for the remaining studies ranged from 1.21 (not significant) to 3.68 (highly significant).

These findings, commonly circulated to the national press, have the potential to reach a huge number of women who have had, are having or are about to start chemotherapy, as well as thousands of health care professionals yet a vastly different picture could be painted using seemingly valid methods of analysis. The impact of these findings on patient well-being and treatment choices is as yet unknown but should not be underestimated.

This investigation demonstrates the need for consensus in the way in which we classify cognitive impairment. In addition to establishing a more universal approach to the method of classification, the number of tasks a participant must fail to classify as impaired currently varies between studies and is to some extent arbitrary. Some researchers [14] have argued that two tests are not sufficient in a large battery and may artificially inflate the rate of impairment. In our study as criteria by which each task is classified are very strict, the low number of tasks required to 'fail' for classification is justified.

It may not be enough to establish a fixed number of tasks to be failed, as the rate of impairment will vary dramatically depending on how many neuropsychological tests are included in each study. As the number of tests in a battery increases so does the probability of being classified as impaired if a fixed number of tasks are used as a criterion. It may be more appropriate to establish a minimum percentage of tasks that must be impaired, rather than an absolute number.

A further factor is that some studies report single test scores [3], as do we. Others however report domain scores based on an average of several tasks measuring one cognitive domain, such as verbal memory [10]. This would undoubtedly be a more conservative assessment of impairment and may well be more valid. By dividing tasks into domains and calculating an overall score within that domain, classification of overall impairment is less likely to be influenced by performance on a single task.

We choose the reliable change index because it allows for analysis of cognitive change between time points rather than static comparison at one time point and it allows one to control for the practice effects inherent in repeated cognitive assessment. It should be noted however that the reliable change method, as all of the other methods described here, makes the assumption that the raw scores are normally distributed. This may go some way to explaining the discrepancy in the risk of being classified as cognitively impaired, as the different methods are sensitive to distributional and variance assumptions. The reliable change index, based on two time points, makes the further assumption that the variance is the same at both time points and that the patients' difference scores follow the same distribution as those from the control group. In fact it is likely that the variance of change scores in the patient group is larger than in the control group due to different individual responses to chemotherapy. We feel justified in using reliable change indices with correction for practice (see Heaton et al. for a comparison of RCI and other methods of detecting change [15]). While we feel justified in the method we have selected the fact remains that those employed by other researchers produce very different results. The differences between methods do not help to clarify a literature already somewhat muddied by other methodological differences.

Conclusions

This investigation demonstrates the clear need for longitudinal studies wherever possible as they allow us to examine real cognitive change. In our data, this resulted in a far lower rate of cognitive impairment than if only the post-chemotherapy data is considered. Where cross-sectional data are used we have demonstrated an

obvious need for consensus on the method and criteria by which impairment is classified. This will not only help researchers working in the field but would also help to provide a clearer take home message for clinicians and patients on the real risk of impairment.

Acknowledgements

We thank the women who participated in the study, all hospital clinic and administrative staff and Roberta Morris for conducting many of the patient assessments. We also thank all of the consultants (Mr Allan, Mr Bishop, Dr Bloomfield, Dr Deutsch, Dr Hodson, Dr Mitra, Dr Sadler, Miss Shah, Dr Stein, Mr Whitehead, Mr Winstanley) and research nurses (Rose Errington, Jayne Hughes, Sonya Mash, Helen Mitchell, Elaine Noon, Victoria Rawlins, Frances Scott), for helping recruit patients to the study. Cancer Research UK funded the study, Dr Shilling and Dr Jenkins. Conflict of Interest The authors declare no conflict of interest. Dr Shilling had full access to all the data in the study and had final responsibility for the decision to submit for publication.

References

1. Shilling V, Jenkins V, Morris R, Deutsch D, Bloomfield D: The effects of adjuvant chemotherapy on cognition in women with breast cancer – preliminary results of an observational longitudinal study. The Breast 14(2): 142–150, 2005
2. Ahles T, Saykin A: Cognitive effects of standard-dose chemotherapy in patients with cancer. Cancer Investigations 19: , 2001
3. Schagen SB, van Dam FS, Muller MJ, Boogerd W, Lindeboom J, Bruning PF: Cognitive deficits after postoperative adjuvant chemotherapy for breast carcinoma. Cancer 85(3): 640–50, 1999
4. Wefel JS, Lenzi R, Theriault RL, Davis RN, Meyers CA: The cognitive sequelae of standard-dose adjuvant chemotherapy in women with breast carcinoma: results of a prospective, randomized, longitudinal trial. Cancer 100(11): 2292–9, 2004
5. van Dam FS, Schagen SB, Muller MJ, Boogerd W, vd Wall E, et al. Impairment of cognitive function in women receiving adjuvant treatment for high-risk breast cancer: high-dose versus standard-dose chemotherapy. J Natl Cancer Inst 90(3): 210–8, 1998
6. Wieneke M, Dienst ER: Neuropsychological assessment of cognitive functioning following chemotherapy for breast cancer. Psycho-oncology 4: 61–66, 1995
7. Desai J, Wagner LI, Beaumont JL, Sweet JJ, Butt Z, Cella D: Cognitive functioning among breast cancer patients prior to chemotherapy. Psycho-oncology 14(1S): S32, 2005
8. Jacobson NS, Truax P: Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. J Consult Clin Psychol 59(1): 12, 1991
9. Heaton RK, Grant I, C.G. M. Comprehensive norms for an expanded Halstead-Reitan battery: Demographic corrections, research findings and clinical applications. Florida: Psychological Assessments Resources; 1991
10. Ahles T, Saykin A, Furstenberg C, et al. Neuropsychologic impact of standard-dose systemic chemotherapy in long-term survivors of breast cancer and lymphoma. J Clin Oncol 20: 485–493, 2002
11. Tager FA, McKinley PS: Cognitive functioning before chemotherapy treatment in women with breast cancer. J Clin Oncol 22(14S): 14, 2004
12. Poppelreuter M, Weis J, Kulz AK, Tucha O, Lange KW, Bartsch HH: Cognitive dysfunction and subjective complaints of cancer patients. a cross-sectional study in a cancer rehabilitation center. Eur J Cancer 40(1): 43–9, 2004
13. Wagner LI, Sweet JJ, Cella D, Havlin KA, Forrestal S, Parrish T, et al. Performance of cognitively impaired oncology patients and normal controls on functional MRI tasks. J Clin Oncol 22(14S): 734, 2004
14. Ingraham LJ, Aiken CB: An empirical approach to determining criteria for abnormality in test batteries with multiple measures. Neuropsychology 10(1): 120–124, 1996
15. Heaton RK, Temkin N, Dikmen S, Avitable N, Taylor MJ, Marcotte TD, et al. Detecting change: A comparison of three neuropsychological methods, using normal and clinical samples. Arch Clin Neuropsychol 16(1): 75–911, 2001

Address for offprints and correspondence: Val Shilling, Cancer Research UK Psychosocial Oncology Group, Brighton and Sussex Medical School University of Sussex, BN3 1PU, Falmer East Sussex, UK; *Tel.:* +01273-873019; *Fax:* +01273-873022; *E-mail:* v.m.shilling@sussex.ac.uk