The complex association of pulmonary function with panic disorder: a rejoinder to Ley (1998)

Philip Spinhoven\textsuperscript{a,}\textsuperscript{*}, Peter J. Sterk\textsuperscript{b}, Leo van der Kamp\textsuperscript{c}, Evert J. Onstein\textsuperscript{d}

\textsuperscript{a}Leiden University, Faculty of Social and Behavioral Sciences, Division of Clinical and Health Psychology, P.O. Box 9555, 2300 RB Leiden, The Netherlands
\textsuperscript{b}Department of Pulmonology, Leiden University Medical Center, The Netherlands
\textsuperscript{c}Department of Psychology, Leiden University, The Netherlands
\textsuperscript{d}Psychiatric Hospital Endegeest, Oegstgeest, The Netherlands

Abstract

In a commentary on our paper entitled “Pulmonary function in panic disorder: evidence against the dyspnea-fear theory”, Ley (1998) provides a critical analysis of our study. He concludes that our failed attempt to replicate a relationship between pulmonary function and the severity of panic-related symptoms in panic disorder patients may have been a consequence of a lack of comparability between studies, a statistical anomaly, and experimenter-demand effects. After discussing his comments (with most of them we do not agree) in depth, we maintain our conclusion that: (a) pulmonary impairment is not directly associated with panic symptoms; and (b) that the existence of a distinct subgroup of panic disorder patients with signs of actual airways obstruction leading to uncontrollable dyspnea and fear of suffocation remains questionable. © 2000 Elsevier Science Ltd. All rights reserved.

1. Introduction

In a previous article (Spinhoven, Onstein & Sterk, 1995) we investigated whether evidence could be found for a subgroup of panic patients with a non-pathological pulmonary obstructive component inducing dyspnea, dyspneic fear and, ultimately,
panic, consistent with the dyspnea-fear theory of panic. In 38 consecutive patients who met DSM-III-R criteria for panic disorder (PD), pulmonary function was assessed and various measurements for panic symptoms and concomitant psychopathology were collected. In comparison to patients with a high forced expiratory flow at 50% (FEF 50%) of forced vital capacity (FVC), low FEF 50% patients demonstrated significantly lower levels of forced expiratory volume (first second) (FEV1) and peak expiratory flow (PEF) and significantly lower FEV1/FVC ratios. None of the differences on psychological measures for symptom severity between low and high FEF 50% patients proved to be significant. Moreover, FEF 50% scores and FEV1/FVC ratios were not correlated to any of the measurements of panic or associated psychopathology. It was concluded that the existence of a distinct subgroup of panic patients with signs of actual airways obstruction leading to uncontrollable dyspnea and fear of suffocation remained questionable.

Our paper elicited a critical reaction from Ley (1998). He compared our study with the study of Asmundson and Stein (1994), in which an association between forced expiratory flow rate and the severity of panic-related symptoms was reported in 15 panic disorder patients. In the opinion of Ley our failure to replicate their findings may have been due to the following points, which can be summarized as follows:

(a) There is a lack of comparability between the subjects in both studies: more specifically, since the subjects of our study have significantly higher FEV1/FVC scores than those of Asmundson and Stein, the severity of their dyspnea-related complaints would be expected to be relatively less.

(b) The differences between low and high FEF 50% subjects on psychological measures for symptom severity are disproportionately small and exceed predictions based on sampling error. Moreover, in a reanalysis of our data Ley showed that assignment of subjects to groups with or without regard to pulmonary measures yielded means and corresponding F’s as predicted on the basis of sampling error.

(c) The fact that Asmundson and Stein obtained significant effects with a relatively small sample of only 15 subjects shows that the effect was quite strong and a failure to replicate this effect with a larger sample of 38 subjects suggests that the replication may have been flawed. In this context the possibility of experimenter-demand effects as a strong source of conveying expectations of the outcome of pulmonary function in subjects is mentioned.

These three points of criticism are discussed below. Of course, we are grateful for Ley’s careful and detailed (re-)analysis. However, our arguments for disagreeing with most of Ley’s comments will be presented. Some of his criticism leads to an interesting discussion of problems and pitfalls in investigating the association of pulmonary function and panic disorder.

2. Lack of comparability of subjects between studies

With regard to the comparability of pulmonary function, Ley is touching upon an important point. We had purposely chosen to use the same criterion of airways
obstruction as Asmundson and Stein, namely the maximal expiratory flow at 50% of the forced vital capacity (FEF 50%). One may criticize such a choice, since FEF 50% is not very well reproducible (McDonald & Cole, 1980). Nevertheless, when using FEF 50% as a criterion the groups do not match appropriately for FEV₁/FVC between the two studies, our patients indeed having relatively higher FEV₁/FVC ratios as compared to Asmundson and Stein. However, when using the complete information on lung function as obtained from the Tables 1 of both studies, it appears that the studies are extremely well matched with regard to the values of FEV₁ itself in each of the two study groups. Moreover, as already mentioned in our 1995 article, the peak flow (PEF) values in our study are more distinctive between the groups, and appear to be even substantially lower in our “low FEF 50% group” as compared to the same group in the study of Asmundson and Stein.

The question then arises whether FEV₁/FVC is the “crucial” pulmonary function test, as Ley is suggesting. We believe it is not. Contrary to FEV₁, which is the gold standard of airways obstruction for many years (Crapo, 1994), the ratio of FEV₁/FVC is more subject to measurement error due to difficulties in recording a full, forced expiration (Quanjer, Tammeling, Cotes, Pedersen & Peslin, 1993; ATS, 1995). In addition, the ratio of FEV₁/FVC (albeit often used) is easily misinterpreted regarding the presence and severity of airways obstruction, because of the fact that both the numerator and the denominator of it decrease to a variable extent in case of airway narrowing (Quanjer et al., 1993; ATS, 1995). In fact, these are the reasons that the present European guidelines on pulmonary function testing have abandoned the FEV₁/FVC ratio in favor of the FEV₁/VC ratio, which is based on a slow inspiratory vital capacity manoeuvre (Quanjer et al., 1993). Hence, when using an adequate and complete set of spirometric data, we consider the groups to be adequately matched for the level of airways obstruction between the two publications.

Moreover, the assignment of patients into low FEF 50% and high 50% subjects taking a median split of FEF 50% predicted values is arbitrary since none of the PD patients in both studies displayed significant impairments in pulmonary function. This division into two groups entails a reduction of the level of measurement from at least an ordinal level to a nominal level and a subsequent loss of information. Therefore, in contrast to Asmundson and Stein we also calculated Pearson product–moment correlation coefficients between FEF 50% scores and FEV₁/FVC ratios on the one hand and measurements for panic and concomitant psychopathology on the other. None of these correlation coefficients (range 0.01–0.22) approached the level of significance. These results (not discussed by Ley in his reaction) also suggest that measures for pulmonary function and panic-related symptoms do not co-vary.

Finally, as also discussed in our 1995 paper, differences between study results may have been related to differences in the assessment of panic-related symptoms. The only significant differences reported by Asmundson and Stein were found on the unpublished anxiety symptom questionnaire (ASQ), a 20-item questionnaire that retrospectively assesses symptoms and symptom severity during periods of anxiety and panic and not on the Sheehan patient rated anxiety scale (SPRAS), a 35-item questionnaire that assesses a number of problems that an individual may have experienced in the
past week (of which past-week respiratory difficulties were analyzed by Asmundson and Stein). In our study, measures of panic-related symptoms included both a retrospective assessment of the most severe panic attack of the last two weeks (PQ) and a daily record of panic attacks (PAD) during the last two weeks. Consequently, the association of pulmonary function and panic-related symptoms in both studies has only been found on one out of four scales suggesting the necessity of investigating this relationship into further detail with the use of various standardized measures.

Unfortunately, due to these differences in measures between studies, a direct comparison of the measurements of panic-related symptoms is impossible. However, the mean frequency of past-week panic seems rather similar (see Tables 2 of both studies). As we used the present gold standard for assessing panic attacks (i.e. daily recording of attacks in a diary and assessment of panic attacks using DSM-III-R criteria), we do not agree with Ley’s presupposition that our sample probably manifested a lower level of (dyspnea-related) complaints than the sample of Asmundson and Stein.

2.1. A statistical anomaly

In a reanalysis of our data Ley finds disproportionally small differences between the means of each of our psychological measures that exceed predictions based upon sampling error. Apart from the unorthodox way of statistical reasoning (why not stick to the generally accepted classical procedure of null-hypothesis testing and leave it at that), Ley takes refuge to a debatable procedure to analyze these differences. Possible violations of initial ANOVA assumptions aside, the eleven F-distributions that constitute the basis of his conclusion of a statistical anomaly are highly likely not to be independent, due to the multicolliniarity of the eleven measures used in our study. This leaves us with all but flawless arguments for the so-called statistical anomaly of our findings.

In this context it is also relevant to refer to the growing body of research literature into the relationship of chronic pulmonary obstruction and PD in patients with asthma (Smoller, Pollack, Otto, Rosenbaum & Kradin, 1996). In our 1995 paper we already reviewed several of these studies suggesting that pulmonary function is not related to panic disorder in subjects with asthma. In a more recent study by Carr, Lehrer, Hochron and Jackson (1996) of physiological and airway reactivity to psychologically stressful laboratory tasks, it was even found that both asthmatics and non-asthmatics diagnosed with DSM-III-R PD had significantly lower airway impedance (i.e. better pulmonary function) than non-panickers.

As these results suggest that panic disorder may even be associated with better, not worse, pulmonary function Ley’s argument of a statistical anomaly may be invalid. His argument presupposes that sampling error alone is responsible for differences between the means on psychological measures between low and high FEF 50% subjects. However, this argument of random effects is invalid in case of better pulmonary function in PD patients.
3. Sample size and bias

It is a well-known fact that increasing sample size (i.e. increasing power) reduces probability of a Type II error but has no effect on Type I error rate. Asmundson and Stein obtained significant effects with a relatively small sample while we failed to replicate this effect within a larger sample. Clearly, our analysis has more statistical power than the one by Asmundson and Stein. Consequently, the credibility of our findings is higher than in the study by Asmundson and Stein. Why should the replication have been flawed? Traditional statistical theory invalidates Ley’s assertion that a failure to replicate the rejection of a null hypothesis in a large sample in contrast to a small sample necessarily implies that the methodology of the study with the larger sample must have been flawed. To be sure, applying ANOVA in studies of our kind has its limitations due to small sample size and the tenability of assumptions. In situations with small samples, results are sensitive to outliers. These outliers can produce either Type I or Type II error, with no clue in the analysis as to which is occurring. Normality has to be assumed and also homogeneity of variances. Are %-values normally distributed? The answer would be “no”. And to give an example of the violation of the homogeneity assumption, take Asmundson and Stein’s measure “% endorsed measure”: the violation of variance homogeneity is striking. The latter fact invalidates the use of the F-test. Should one use Bonferroni’s F-values (e.g. as reported in Maxwell & Delaney, 1990), Asmundson and Stein’s reported F-values must exceed the Bonferroni critical F-value 10.33 to be significant at a familywise alpha of < 0.05. Moreover, bias in the selection of subjects may introduce bias resulting in rejection of the null hypothesis. In this respect a larger sample size is favorable, because data from larger samples allow more reliable inferences about the target population involved.

Here we will not scrutinize any further the methodology of the Asmundson and Stein study. We only want to point out possible alternative reasons for the fact that in contrast to Asmundson and Stein we were unable to reject the null hypothesis in a larger sample. However, we do not think that in our study experimenter-demand effects have conveyed experimenters’ expectations of the outcomes of the pulmonary function tests to the subjects. Ley’s allegation seems to be far-fetched and unwarranted. It goes without saying that in our experiment the lung function technicians were not informed on the hypotheses of our study and were also fully blinded to the psychological measurements. Actually, these measures were collected by other research assistants at Psychiatric Hospital Endegeest before patients visited the Lung Function Laboratory of Leiden University Medical Center. Therefore, we hope that Ley will appreciate our difficulties in understanding the entitlement of “flawed methodology” in this respect.

4. Conclusion

As could be expected the results of our study, inconsistent with the dyspnea-fear theory of panic, has led Ley to defend his dyspnea-fear theory. As described above we
do not agree with his commentary on our study. After discussing his criticism of a lack of comparability between studies, a statistical anomaly, and a so-called “flawed” methodology, we maintain that: (a) pulmonary impairment is not directly associated with panic symptoms; and (b) that the existence of a distinct subgroup of panic disorder patients with signs of actual airways obstruction leading to uncontrollable dyspnea and fear of suffocation remains questionable. Evidently, more systematic studies are needed to disentangle the relationship between pulmonary function and panic, which may be more complex than it at first appears (Carr, 1998).

References


