2013; 35: 472–480 MEDICAL TEACHER

Effect of stress on clinical reasoning during simulated ambulatory consultations

P. POTTIER[1], T. DEJOIE[1], J. B. HARDOUIN[1], A. G. LE LOUPP[1], B. PLANCHON[1], A. BONNAUD[1] & V. R. LEBLANC[2]

[1]CHU Nantes, France, [2]University of Toronto, Canada

Abstract

Background: The goal of this study was to examine the impact of subjective and physiological stress responses on medical students' diagnostic reasoning and communication skills.

Method: A prospective randomized quantitative study was undertaken, looking at ambulatory consultations in internal medicine. On the first day (baseline day), volunteer year 6 students ($n = 41$) participated in a simulated ambulatory consultation with standardized patients (SPs). On the second day (study day), one week later, they were randomly assigned to two groups: a low stress ($n = 20$) and a high stress ($n = 21$) simulated ambulatory consultation. Stress was measured using validated questionnaires and salivary cortisol. The SPs assessed the students' reasoning and communication. The students completed assessments of their clinical reasoning after the consultations.

Results: Although stress measures were all significantly higher in the high-stress condition (all $p < 0.05$), no differences were found in diagnostic accuracy and justification scores. However, correlational analyses revealed a negative correlation between multiple-stress measures and the students' ability to generate arguments for differential diagnoses.

Conclusion: Stress was associated with impairments in clinical reasoning, of a nature typically suggestive of premature closure.

Introduction

According to Lazarus and Folkman (1984), a situation is perceived as stressful when the demands of the situation are appraised as exceeding the available resources, and thus endangering well-being or the attainment of an important goal. This cognitive appraisal of a situation occurs in two steps. The primary appraisal results in the perception of the demands induced by the situation. The secondary appraisal results in the perceptions of the resources or abilities to cope with the demands of the situation. A situation is perceived as a threat, and thus stressful, when the demands are appraised as exceeding the resources. In the opposite case, the situation is perceived as a challenge, and thus not stressful.

Medical training has been recognized as a highly stressful experience for medical students (Toews et al. 1993; Dyrbye et al. 2005). Numerous studies show that the stress of medical training has an impact on the physical and mental health of medical students, and many academic institutions have implemented broad interventions aimed at providing students with the skills and resources to cope with the chronic stressors of medical training (Campo et al. 2008). In contrast, there has been relatively little work exploring the occurrence of acute stress or its effect on the clinical performance of medical students. In the medical setting, students' acute stress response varies according to the type of professional exercise, with higher subjective and physiological levels of stress in ambulatory consultations compared to in-hospital consultations (Pottier et al. 2011).

Practice points

- No published studies examine students' clinical reasoning under realistic stressful conditions in a context of ambulatory consultation.
- Stress levels observed in real ambulatory consultations had been recreated in a reliable way using standardized patients.
- High-stress conditions were associated with some alterations in diagnostic reasoning.
- Students who exhibited greater levels of stress generated fewer arguments for differential diagnoses.

Some of the effects of stress on clinical performance have already been described. Previous studies have assessed the effect of acute subjective stress on academic performance of medical trainees (Pamphlett & Farnill 1989; Reteguiz 2006; LeBlanc & Bandiera 2007; LeBlanc et al. 2008) and dental students (Sanders & Lushington 2002) with conflicting results. In contrast, other studies have looked at the effects of physiological stress (elevated cortisol levels) on clinical performance. With a few exceptions showing no clear relationship between stress and performance (Van Dulmen et al. 2007; Wetzel et al. 2010), clinical performance is impaired in situations that lead to elevations in cortisol levels (LeBlanc et al. 2005, 2012; Arora et al. 2010; Harvey et al. 2012). As such, the cortisol response in stressful situations appears to be a key determinant of impaired clinical performance.

Correspondence: Dr Pierre Pottier, Department of Internal Medicine, Faculty of Medicine, CHU Nantes, Place Alexis Ricordeau, Nantes 44093, France. Tel: 00 33 240 083805; fax: 00 33 240 083379; email: pierre.pottier@univ-nantes.fr

ISSN 0142–159X print/ISSN 1466–187X online/13/060472–9 © 2013 Informa UK Ltd.
DOI: 10.3109/0142159X.2013.774336

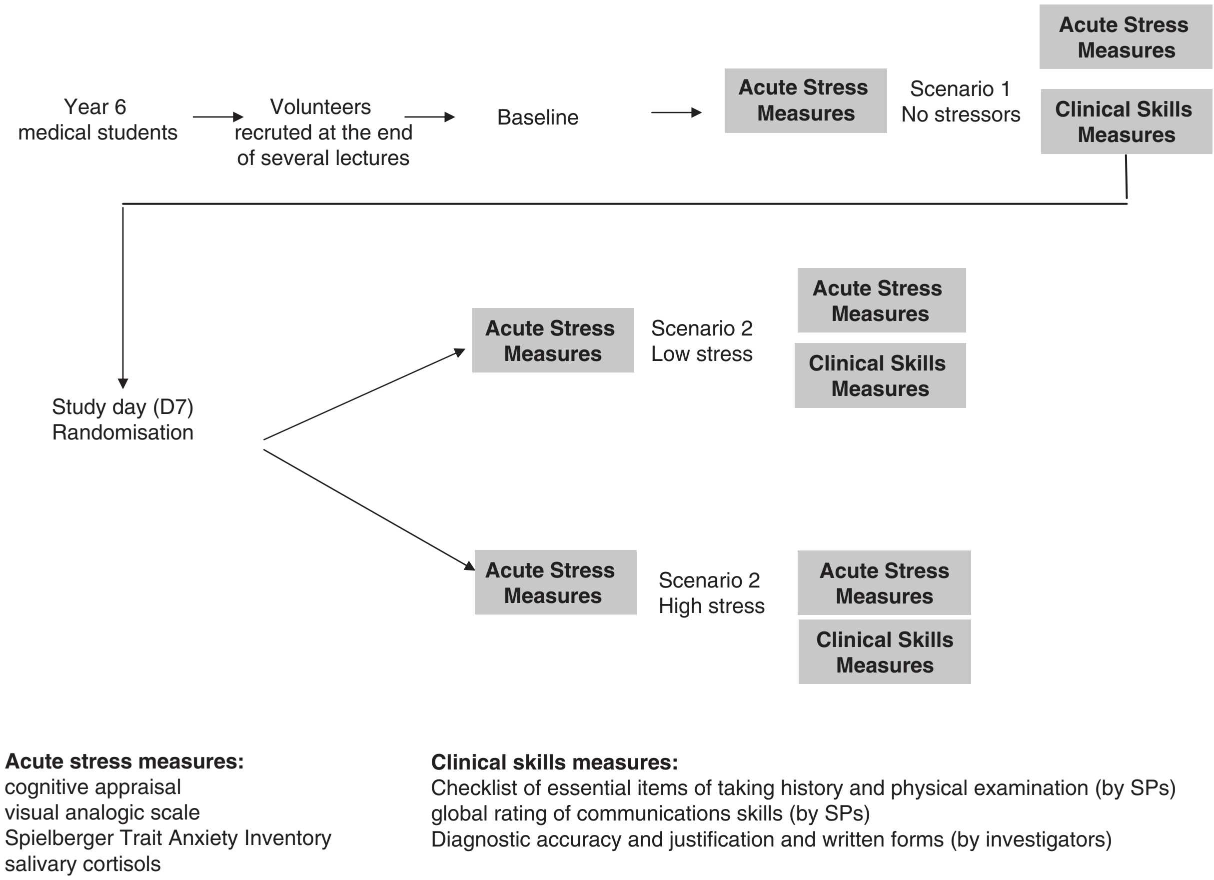

Figure 1. Design of the study.

Despite the growing literature of the effects of subjective and physiological stress on clinical performance in general, there are currently no published studies examining medical students' clinical reasoning and decision-making under realistic stressful conditions. Clinical reasoning is a broad concept that includes the ability to collect relevant clinical data and to reach a clinical diagnosis, the ability to generate alternative (or differential) diagnoses, to order diagnostic tests, to initiate first treatments, and to organize an appropriate follow-up (Yates & Tschirthart 2006). Communication is a key competency for health professionals and is a part of clinical reasoning (Ajjawi & Higgs 2012), so better communication with the patient is suspected to improve the gathering of the clinical data.

The objective of the study was to assess the impact of subjective and physiological stress on the decision making and communication skills of medical students in a context of ambulatory consultation. Our primary study hypothesis was that medical clinical reasoning would be impaired under high-stress conditions in a simulated ambulatory setting.

Methods

Design of the study

The study protocol was approved by an external ethics committee (Hospital and University ethics committee of Louvain Medicine School, Louvain Catholic University, Belgium, record number B403201110916) (Figure 1).

A randomized prospective study was undertaken, with medical students conducting medical consultations with simulated ambulatory patients complaining of a symptom requiring a clinical diagnosis. To ensure that the task difficulty was appropriate for the students' knowledge and skill levels, frequent and already-taught medical problems were chosen (osteoporotic rib fracture and pneumothorax).

At baseline day, all students conducted a simulated ambulatory consultation without any added stressful components (osteoporotic rib fracture) in order to control for stress resulting from the novelty of the situation, as few students had previously conducted such ambulatory consultations. Moreover, it was their first experience with standardized patients (SPs).

On the study day, one week after the first consultation, the students were randomly assigned to two groups: a low-stress (LS) group and a high-stress (HS) group with the same problem in each group (thoracic pain due to pneumothorax). Students were not aware of their group assignment (LS or HS) before this consultation.

Subjective and physiological stress responses were assessed 10 minutes before and 10 minutes after the simulated consultations. Clinical reasoning was assessed at different times: (1) at the end of each consultation, the SP and the standardized family member (who was also an actor) independently completed a pre-determined checklist described in the "Assessment of clinical skills" section and (2) immediately after the session, the students completed a clinical assessment form, also described in the same section.

Student recruitment

Final year (year 6) medical students were recruited through oral presentations made at the end of several, large group lectures. Written informed consent was obtained.

Simulation of ambulatory setting

The consultations were performed in an authentic environment (i.e., the consultation rooms of the Internal Medicine Department), which included a desk, seats, telephone, computer, and examination devices such as stethoscope, blood pressure cuff, reflex hammer, and so on.

All the simulated sessions included the following characteristics usually found in ambulatory consultations: (1) a time limit (15 minutes), (2) a clinical complaint expressed by the patient which needed to be solved in real time by the student, (3) an accompanying member of the patient's family (patient's son on day 1, patient's girlfriend on day 2), and (4) a synthesis given to the patient by the student at the end of the consultation, which included complementary tests and treatment.

Simulation of low- and high-stress conditions

Low-stress scenario. The patient and the family members were cooperative, confident and in a pleasant mood. There were no distracting elements.

High-stress scenario. Several stressful components occurred during the consultation:

(1) On arrival, just before stress measurements, the trainee was informed that he/she was 90 minutes late in his/her schedule.
(2) At the beginning of the consultation, the medical student met an upset and aggressive family member who questioned/challenged his or her competency in performing the consultation.
(3) During the consultation, the family member commented aloud to herself about the actions/questions of the student.
(4) At the end of the consultation, the family member and the patient asked explanations from the medical student in a skeptical tone, and challenged the student's explanations.
(5) The patient was noncooperative (mostly nonverbal communication) but answered the questions in the same way and with the same precision as in the LS scenario.
(6) A disagreement between the patient and the family member developed progressively during the consultation.
(7) At the end of the consultation, the patient presented a short vagal syncope (due to the thoracic pain) from which he recovered quickly.

Standardized patients training

Ten professional actors were hired from a local theater company to portray patients and family members. During a two-hour training session, the actors were given a detailed description of the patient's recent symptoms and past medical history, and a character description for both the patient and the family member including name, age, behavior, affect, mannerisms, questions and prompts, social history, and family history. The actors were also trained regarding their demeanor for both the LS and the HS scenarios. Each item of the clinical skills checklist they were to complete was carefully explained in order to allow them to distinguish between a good and a bad performance.

Assessment of clinical skills

Five dimensions of clinical skills were assessed. Two components of decision-making were assessed by the SPs and family members:

(1) At the end of each consultation, the SPs and family members independently completed a checklist regarding the student's performance in obtaining the clinical history and in performing the physical examination (clinical abilities scores). The checklist contained 20 items, and each item was scored 0, 0.5 or 1. Scores on the clinical examination checklist could range from 0 to 20. The items of the checklist were established using a Delphi method (including a panel of six experts in clinical reasoning).
(2) The students' communication skills were evaluated by the SPs and family members on a Likert-type four-item scale that assessed empathy, discourse coherence, verbal, and nonverbal communication (Hodges 2003). A single communication score was generated using the mean of the four items for both the patient and the family member scores.

For these two latter dimensions (abilities and communication scores), the reliability between patient and family member scores was assessed using Pearson's correlation test. In the case of good reliability (determined by a Pearson's coefficient over 0.6), the two scores were then averaged into a global score (Fleiss 1981). In the case of low reliability (Pearson's coefficient under 0.6), the two scores were considered separately for statistical analysis.

As no external clinical assessor was present in the consultation rooms in order to closely mimic naturalistic conditions, to ensure the reliability of the actors scorings (SP and family member), 28 consultations (18 at baseline day and 10 at study day) were randomly video-recorded and scored by a medical teacher, included in the panel of experts previously cited, using the same checklist. Reliability of the actors' scoring was assessed by the calculation of Pearson's coefficient between the medical teacher and the actors.

Three components of clinical skills were assessed using written assessment forms completed by the student immediately following the consultation. These were scored by two independent raters, blinded to the group assignment, based on the method described by LeBlanc et al. (2002). The reliability between raters was tested using a Pearson's correlation test. In the case of good reliability (Pearson's coefficient over 0.6), the two measures were then converted into one, by the calculation of the mean. Thus, the following scores were created:

(1) Diagnostic accuracy score: the students were asked to indicate their clinical diagnosis for the patient. For both scenarios, accuracy was scored on a three-point Likert scale (false = 0, partially correct = 1, correct = 2).
(2) Positive arguments score: the students were asked to report the relevant clinical information that supported their main diagnosis.
(3) Differential arguments score: the students were asked to report all relevant clinical information that supported their differential diagnosis.

For the calculation of these two latter scores, the following procedure was applied. For each scenario, the first author generated an exhaustive list of possible symptoms. Using a modified Delphi process, six raters (experienced internal medicine physicians) then scored each of the items for its relevance to the scenario, on a scale of 1–7. Those items that received a mean rating of less than 3 were eliminated. Thus, scenario at baseline day and scenario at study day were given 7 and 6 items for the positive argument score and 17 and 19 items for the differential argument score, respectively. In order to make comparisons between the scenarios, these two scores were transformed into a percentage of possible total scores.

Stress measures

Several measures were used as manipulation check, to ensure that the different stress conditions played by the actors did induce different acute states of stress in students. Acute stress measures were obtained 10 minutes before and after each consultation. Subjective acute stress was assessed using three measures:

(1) A cognitive appraisal (threat/challenge) score, based on Tomaka's framework (Tomaka et al. 1993), was assessed before and after each consultation by calculating the ratio of primary appraisal (perceived demands) to secondary appraisal (perceived resources) for each student. Primary and secondary appraisals were respectively evaluated by the following questions (translated in French): "How demanding do you expect the upcoming task to be?" and "How are you able to cope with this task?" Responses were recorded with a seven-point Likert-type scale. Threat appraisal was defined by a ratio of demands to resources above 1. Challenge appraisal was defined by a ratio of demands to resources lower than or equal to 1. In laboratory studies, threat appraisal has been associated with greater subjective and physiological stress responses than challenge appraisals (Dickerson & Kemeny 2004; Denson et al. 2009; Harvey et al. 2010).
(2) A visual analog scale (VAS) asking "quantify your stress from 0 (not stressed) to 100 (very stressed) on the following visual analog scale" was completed by the students before and after each consultation. Such scales have been used as a marker of subjective stress in previous research (Reteguiz 2006; Pottier et al. 2011).
(3) The French version of the State-Trait Anxiety Inventory (STAI), validated by Bruchon-Schweitzer and Paulhan (1993), measures anxiety experienced at a given moment. It includes 20 items scored from 1 to 4 on a Likert scale, with a total score ranging from 20 to 80. A high internal consistency of the STAI (α coefficient: 0.92) has been demonstrated (Spielberger et al. 1983).

Physiological stress was measured using salivary cortisol levels. Salivary cortisol sampling is a noninvasive test (students chew on a swab for one minute, which is then frozen until analysis). It has been shown to increase in stressful situations (Pottier et al. 2011) and to correlate with some personality traits (Pruessner et al. 1997). Salivary cortisol levels were measured in duplicate by radioimmunoassay (Diasorin Inc 1951, Stillwater, MN, ref CA-1549E RIA) from frozen centrifuged salivary sampling. Saliva samples were collected 10 minutes before the start and 10 minutes after the end of the consultations, using Salivettes from Sartstedt D-5188 (Nümbrecht, Germany). All the samples were taken in the afternoon between 1:30 and 6:30 PM, when diurnal variations in cortisol are minimal (Hindmarsh et al. 1989).

At day 2, students in the HS group had just been informed they were 90 minutes late in their schedule when the stress measures were performed.

Statistical analysis

In educational interventions, the studied effects are likely to be diminished when mixed with all of the other factors in the learning environment that can affect learning and performance. As such, when studying education factors of interest in highly controlled settings, we are interested in fairly large effect sizes. For effect sizes of 1 and setting alpha (p value) at 0.05, power (1-beta) at 0.8, and a two-tailed analysis, 16 participants per group are needed to detect a statistically significant difference. Data analysis was performed using STATA 10 and SPSS softwares.

Clinical skills, as defined by the five dimensions described above, were compared between the two groups (HS and LS) using mixed-design analyses of variance, with Session (Session 1, Session 2) as the repeated measure and Group (HS, LS) as the between subject variable.

The subjective stress responses (STAI, VAS, Tomaka) were analyzed with mixed-design analyses of variance, with Session (Session 1, Session 2) and Time (pre-scenario, post-scenario) as the repeated measures and Group (HS, LS) as the between subject variable. The cortisol responses, calculated as the post-scenario levels minus the pre-scenario levels, were analyzed with a mixed-design analysis of variance, with Session (Session 1, Session 2) as the repeated measure and Group (HS, LS) as the between subject variable. Effects of gender, SP group, and group on the five dimensions of clinical skills were tested by multivariate ANOVAs.

Correlations between stress measures and clinical skill measures were assessed using Pearson's correlation coefficients. For all the analyses, the significance was fixed at $p = 0.05$.

Results

Population characteristics

The students' age ranged from 23 to 30, with a majority of students aged from 23 to 24 (32/41). The students' mean academic level (respective mean of the year 1 and year 5 rankings of each participant among their class) was moderate: 101/220 at year 1 and 71/171 at year 5, with no differences between the HS and the LS groups. The percentage of women was 55% in group 1 and 43% in group 2.

Stress responses

In multivariate analysis, no effect of SP team was found on the different stress measures indicating that the stress conditions were portrayed similarly by the different SP team (Table 1).

Visual analogical scale. There was a main effect of day on the scores of the visual analogical scale (VAS) ($F = 7.2$, $p < 0.01$) with overall scores being higher on the second day than on the baseline day. A significant day by group interaction ($F = 8.8$, $p < 0.01$) revealed that while the two groups' stress levels did not differ on the baseline day, the HS group showed significantly higher pre- and post-scenario scores on the study day than did the LS group.

Spielberger trait anxiety inventory. The analyses revealed an effect of day ($F = 16.9$, $p < 0.01$), an effect of group ($F = 13.4$, $p < 0.01$), and a significant day by group interaction ($F = 7.3$, $p < 0.01$). Both groups showed similar STAI scores on the baseline day, and these scores did not increase in response to the scenario. On the study day, the HS group reported greater pre- and post-scenario scores than did the LS group.

Cognitive appraisal. The analyses revealed a main effect of day ($F = 15.0$, $p < 0.01$) and a significant day by group interaction ($F = 5.6$, $p < 0.05$). The students in both groups assessed the demands on the baseline day to be greater than their resources, thus interpreting the scenarios as a threat. However, the two groups did not differ from each other on the baseline day. On the study day, the students in the HS group interpreted the scenarios as more threatening than did the LS students, both at pre- and post-scenario.

Salivary cortisol. The analyses revealed a significant group by day interaction ($F = 4.4$, $p < 0.05$). The cortisol responses of the two groups did not differ on the baseline day. On the study day, however, the students in the HS group showed significantly greater cortisol responses to the scenario than did the students in the LS group.

Effect of stress on clinical skills

Diagnosis accuracy. The inter-rater correlation was excellent on the baseline day ($R = 0.92$, $p < 0.0001$) and on the study day ($R = 1$, $p < 0.0001$) for the diagnostic accuracy score (Table 2).

The analysis of the diagnostic accuracy revealed no main effect of day ($F = 0.1$, $p = 0.77$), no main effect of group ($F = 0.36$, $p = 0.55$), and no significant group by day interaction ($F = 0.72$, $p = 0.4$), indicating that there were no effects on diagnostic accuracy as a result of our experimental manipulations.

Positive and differential argumentation. The inter-rater correlation coefficients were 0.9 ($p < 0.0001$) and 0.67 ($p < 0.0001$) on the baseline day and on the study day for the positive argumentation score and 0.8 ($p < 0.0001$) and 0.89 ($p < 0.0001$) for the differential argumentation score,

Table 1. Acute stress measures before and after ambulatory consultations.

	Baseline		Study day		
	Pre-scenario	Post-scenario	Pre-scenario	Post-scenario	Effect of day and group
VAS-LS group	45.95 (5.01)	40.60 (6.05)	46.2 (4.90)	38.70 (5.59)	Day effect: $F = 7.2$, $p < 0.01$
VAS- HS group	43.33 (4.89)	47.38 (5.90)	61.61 (4.78)	61.54 (5.46)	Interaction day-group: $F = 8.8$, $p < 0.01$
STAI-LS group	39.0 (1.82)	41.45 (2.06)	39.9 (2.13)	41.55 (2.32)	Day effect: $F = 16.9$, $p < 0.01$ Group effect: $F = 13.4$, $p < 0.01$
STAI-HS group	42.28 (1.77)	42.14 (2.01)	50.33 (2.08)	51.76 (2.26)	Interaction day-group: $F = 7.3$, $p < 0.01$
CA-LS group	1.05 (0.14)	1.32 (0.22)	1.29 (0.28)	1.45 (0.30)	Day effect: $F = 15.0$, $p < 0.01$
CA-HS group	1.26 (0.14)	1.34 (0.22)	1.88 (0.27)	2.26 (0.29)	Interaction day-group: $F = 5.6$, $p < 0.05$
	Day 1: change from pre-scenario		Day 2: change from pre-scenario		
Cortisol response: LS group	1.75 (1.01)		0.10 (0.73)		Interaction day-group: $F = 4.4$, $p < 0.05$
Cortisol response: HS group	1.51 (0.99)		3.63 (0.71)		

Note: LS, low stress; HS, high stress; VAS, visual analogic scale; STAI, Spielberger trait anxiety inventory; CA, cognitive appraisal means followed with standard deviations between brackets only significant effects are reported.

Table 2. Effect of time and stress on clinical skills.

	Baseline	Study day	Effect of time and group
Diagnostic accuracy* – LS Group	0.55 (0.38)	0.65 (0.37)	NS
Diagnostic accuracy* – HS Group	0.57 (0.39)	0.52 (0.43)	
Positive arguments score – LS group	30.4% (23.5)	28.4% (24.2)	NS
Positive arguments score – HS group	33.4% (32.1)	21.8% (23.9)	
Differential arguments score – LS group	24.2% (13.5)	38.6% (17.2)	Time effect: $F = 30.3$, $p < 0.0001$
Differential arguments score – HS group	18.6% (10.2%)	37.8% (13.7)	
Global communication score – LS group	64.9% (21.7)	69.7% (12.9)	Time effect: $F = 5.2$, $p < 0.05$
Global communication score – HS group	55.7% (18.4)	68.2% (14.1)	
Global clinical abilities score – LS group	62.8% (15.0)	64.8% (10.8)	NS
Global clinical abilities score – HS group	58.0% (11.5)	60.9% (10.9)	

Note: LS, low stress; HS, high stress; NS, non significant means and standard deviations between brackets below only significant effects are reported.
*Diagnostic accuracy is expressed in terms of mean of a three-point likert scale (0/1/2).

respectively. The analysis of the positive argument score revealed no main effect of day ($F = 1.4$, $p = 0.24$), no main effect of group ($F = 0.1$, $p = 0.71$), and no day by group interaction ($F = 0.7$, $p = 0.41$).

The analysis of the differential score revealed a main effect of day ($F = 30.3$, $p < 0.0001$), with overall scores increasing from the baseline day to the study day. There was no significant main effect of group ($F = 1.12$, $p = 0.29$) and no significant group by day interaction ($F = 0.62$, $p = 0.43$). Paired sample *t*-tests revealed that the increase in differential diagnosis scores were significant for both the LS group ($p < 0.01$) and the HS group ($p < 0.0001$).

Clinical abilities scores. The scores given by the patients and by the family members were significantly correlated with each other ($r = 0.60$, $p < 0.01$ on the baseline day; $r = 0.78$, $p < 0.01$ on the study day).

The scores given by the medical teacher from video-recordings and the global clinical abilities score given by the actors were significantly correlated with each other ($r = 0.81$, $p < 0.01$).

There were no effect of group ($F = 2.6$, $p = 0.11$), no effect of day ($F = 0.9$, $p = 0.36$), and no significant day by group interaction ($F = 0.03$, $p = 0.87$) on the global clinical abilities score.

Communication scores. The communication scores given by the patients and family members were significantly correlated with each other ($r = 0.77$, $p < 0.01$ on the baseline day; $r = 0.60$, $p < 0.01$ on the study day).

The scores given by the medical teacher from video-recordings and the global communication score given by the actors were significantly correlated with each other ($r = 0.58$, $p < 0.01$).

Analyses revealed a significant main effect of day on the global communication score ($F = 5.2$, $p < 0.05$), due to an increase from the baseline day to the study day but no significant effect of group ($F = 2.0$, $p = 0.16$) and no day by group interaction ($F = 1.1$, $p = 0.30$). Paired samples *t*-tests showed that while the global communication score increased from the baseline day to the study day, they did not reach significance for either the LS group ($p = 0.08$) or for the HS group ($p = 0.14$).

Correlations between subjective and physiological stress and diagnostic and communications scores

There were individual differences in stress responses to the scenarios on the study day. Some of the students in the LS scenario exhibited stress responses, while some of the students in the HS scenario did not exhibit any stress responses. As such, we conducted correlational analyses between the stress responses and performance scores of the students collapsed across both conditions.

The differential diagnosis arguments scores were negatively correlated to the post-scenario VAS scores ($r = -0.28$, $p = 0.04$), pre-scenario STAI scores ($r = -0.27$, $p = 0.04$), post-scenario STAI scores ($r = -0.39$, $p = 0.01$), pre-scenario cognitive appraisals ($r = -0.31$, $p = 0.02$), and the pre-scenario salivary cortisol levels ($r = -0.32$, $p = 0.02$). The correlations between the remainder of the diagnostic performance scores were not significantly correlated with any of the pre- or post-scenario stress levels on the study day.

Discussion

A previous study by this research group showed that ambulatory consultations are appraised by students as stressful situations (Pottier et al. 2011). The goal of this study was to examine the relationship between medical students' stress responses and their clinical reasoning and communication skills in simulated ambulatory consultations.

This study shows that stress levels previously observed in real ambulatory consultations could be recreated in a reliable way using actors who were new to the role of SPs. A study with real ambulatory patients in internal medicine (Pottier et al. 2011) showed approximately the same scores using VAS (49.4 ± 20.9), Spielberger trait anxiety inventory (42.2 ± 9.8), and Tomaka's ratio (1.7 ± 0.7) and the same level of salivary cortisols (5.1 ng/mL). The students in the current study showed similar levels at baseline day and in the LS group on study day. The stress levels in the HS group on the study day were higher than the mean scores observed in the Pottier study with real patients. However, they were within one standard deviation of the scores in the Pottier study, suggesting that the HS scenarios created for this study capture the higher end of actual cases, so they are within the range of what is observed with real patients. The increase in subjective stress measures but not in cortisol in the HS group before the consultation is probably due to the announcement of the delay in the student schedule, which was certainly interpreted as a clue. Indeed, subjective stress measures were performed immediately after that information and knowing that salivary cortisols assess stress experienced 20 minutes before, it is not surprising that no differences were found between both groups since, at this time, students could not find out in which group they were.

This is one of the first comparisons of stress levels in real clinical settings with those elicited by simulated clinical scenarios. This shows that simulation scenarios can realistically be used to recreate the affective components of the clinical setting.

The second finding of this study is that the HS experienced by students was associated with subtle alterations in some aspects of their diagnostic reasoning. We did not observe any impairment in clinical reasoning or communication scores when comparing the means of the HS and LS groups. However, correlational analyses of the study day scores revealed that, regardless of the experimental condition, those students who exhibited greater subjective and physiological stress levels received lower scores on the differential diagnosis scores. This suggests that stress may be associated with the breadth of reasoning of students. When stressed, they did not appear to be impaired in their ability to reach a correct diagnosis or to identify the symptoms in support of this diagnosis. However, they appear less likely to report signs of differential diagnoses, suggesting that they may be engaging in premature closure; the tendency to stop considering other alternatives once a diagnosis is reached. This interpretation is consistent with previous research that suggests attentional narrowing under stress (Janis & Mann 1977; Johnston et al. 1997; Chajut & Algom 2003). Pattern recognition is an unconscious, non-analytical cognitive process leading to the generation of one single hypothesis when enough features are available to match with a memorized concrete case (called instance) or a prototype of the disease which itself represents an average of the situations already encountered (Eva 2004; Norman et al. 2007). Pattern recognition process is based on the recognition of positive features while hypothetico-deductive reasoning is also built from an argumentation relying on negative signs. Our study suggests that acute stress could have impaired the ability to develop a differential argumentation while the pattern recognition process (indirectly assessed here by the positive arguments score) seemed independent from stress state of the students. Since excessive reliance on pattern recognition and premature closure are shown to be two sources of diagnostic errors (Eva 2004; Graber et al. 2005), we can hypothesize that acute stress may represent a risk factor for diagnostic errors, which should be taken into consideration during the medical curriculum implementation.

Several limitations need to be highlighted in this study. First, having included volunteer students may have introduced a representativeness bias. Students who volunteered to be exposed to stress conditions may be less likely to feel stress than those who did not. Second, it is possible that the limited relationship between stress and clinical reasoning observed in this study may be due to a lack in sensitivity of our assessment tools. To our knowledge, there is no validated tool allowing an assessment of clinical reasoning and decision-making processes during an ambulatory consultation. Asking for the most likely hypothesis, the main arguments supporting this hypothesis and alternative diagnosis, our performance assessment tool was very close to the one used by Chamberland et al. (2011) in a recent quantitative study assessing the effect of self-explanations on diagnostic performance. In addition, some of the measures were based on LeBlanc's assessment of clinical reasoning (Leblanc 2002), which showed differences due to the level of training and due to various experimental manipulations. Third, the relative high level of difficulty for the diagnostic task, suggested by the low mean score of diagnostic accuracy in both groups (0.56 and 0.58) could have impaired the diagnostic justification and consequently biased the analysis of the positive and negative argument scores. Fourth, one of our assessment tools of clinical reasoning used written assessments performed *a posteriori*, approximately 10 minutes after the task, which may pose some problems of construct validity. Indeed, some students may have retrieved positive or negative arguments for some hypotheses while writing their thinking though they did not use them for their reasoning during the consultation.

This study highlights the potential deleterious effect of stress on clinical reasoning. An educational implication of this finding would be that students may need to develop mindful practice or metacognition as proposed by Borrell-Carrió and Epstein (2004), to be able to detect periods particularly at risk for making diagnostic errors. Teachers should help them through clinical feedbacks or academic lessons coping with their stress and more systematically developing differential argumentations under stressful contexts. Further researches are needed to evaluate the efficacy of such educational interventions in the curriculum on stress levels and diagnostic performance.

In conclusion, this study shows that an emotional clinical context can be realistically simulated using novice SPs during ambulatory consultations. As such, this approach represents a viable means in which to study the effects of emotions on clinical skills, something that is challenging to achieve with actual patients due to variation in patient presentations. The results of this study further show some subtle negative relationships between clinical reasoning and stress responses. Regardless of the experimental conditions, students who exhibited greater levels of subjective and physiological stress generated fewer arguments for differential diagnoses. This is consistent with previous research suggesting that stress may be associated with attentional narrowing and premature closure (Janis & Mann 1977; Johnston et al. 1997; Chajut & Algom 2003). Given the subtle relationship between stress and clinical reasoning observed in this study, further research is required to replicate these findings and to better understand the relationship between the complex constructs of stress and clinical reasoning.

Acknowledgments

Authors thank the following:

- Professor Jean-Michel Rogez, Dean of the Nantes University Medical School, for his moral and logistic support.
- The following researchers in health professional education for their valuable scientific revisions: Arija Birze, Laura Dempster, Dominique Piquette, Carilynne Yarascavitch, Sylvain Boet, and Walter Tavares.
- M.T. Biais, Mrs N. Coudrier, administrative directors of the Nantes Hospital and Professor J. Barrier, Professor M. Hamidou, medical directors of the Department of Internal Medicine who gave their approval for using the internal medicine consultation's area.
- The 41 year six medical students who kindly accepted to submit themselves to stress conditions.

Declaration of interest: The authors report no declarations of interest. The study protocol has been approved by an external ethic committee (from Belgium, record number B403201110916). This study was funded by ADREMSE (*Association pour le développement de la recherche en médecine et en sciences de l'éducation*).

Notes on contributors

PIERRE POTTIER, MD, PhD, is a specialist in internal medicine. He has achieved a master's degree in education research. He contributed to the conception and design of the study, the acquisition, analysis and interpretation of data, and the drafting and revision of the article.

THOMAS DEJOIE, MD, and ANNE-GAELLE LE LOUPP, MD, are specialists in biochemistry. They contributed to the acquisition of data, and the drafting and revision of the article.

BERNARD PLANCHON, MD, PhD, is a specialist in internal medicine and the Vice-Dean in charge of the educational program at the Nantes Medical School. He contributed to the conception and design of the study and the drafting and revision of the article.

JEAN-BENOIT HARDOUIN, PhD, is specialist in biostatistics. He contributed to the conception and design of the study, the analysis and interpretation of data, and the drafting and revision of the article.

VICKI LEBLANC, PhD, is a Researcher in Cognitive Psychology. She contributed to the conception and design of the study, the analysis and interpretation of data, and the drafting and revision of the article.

ANGELIQUE BONNAUD, PhD, is a Medical Psychologist. She contributed to the conception and design of the study, the interpretation of data, and the drafting and revision of the article.

All authors approved the final manuscript for publication.

References

Ajjawi R, Higgs J. 2012. Core components of communication of clinical reasoning: A qualitative study with experienced Australian physiotherapists. Adv Health Sci Educ Theory Pract 17(1):107–119.

Arora S, Sevdalis N, Aggarwal R, Sirimanna P, Darzi A, Kneebone R. 2010. Stress impairs psychomotor performance in novice laparoscopic surgeons. Surg Endosc 24: 2588–2593.

Borrell-Carrió F, Epstein RM. 2004. Preventing errors in clinical practice: A call for self –awareness. Ann Fam Med 2:310–316.

Bruchon-Schweitzer M. 1993. Paulhan I: Manuel français de l'échelle d'anxiété-trait. Anxiété état de CD, ECPA. Paris, France: Spielberger.

Buchanan TW, Tranel D. 2008. Stress and emotional memory retrieval: Effects of sex and cortisol response. Neurobiol Learn Mem 89:134–141.

Campo AE, Williams V, Williams RB, Segundo MA, Lydston D, Weiss SM. 2008. Effects of Life Skills training on medical students' performance in dealing with complex clinical cases. Acad Psychiatr 32(3):188–193.

Chajut E, Algom D. 2003. Selective attention improves under stress: Implications for theories of social cognition. J Pers Soc Psychol 85(2):231–248.

Chamberland M, St-Onge C, Setrakian J, Lanthier L, Bergeron L, Bourget A, Mamede S, Schmidt H, Rikers R. 2011. The influence of medical students' self-explanations on diagnostic performance. Med Educ 45:688–695.

Denson TF, Spanovic M, Miller N. 2009. Cognitive appraisals and emotions predict cortisol and immune responses: A meta-analysis of acute laboratory social stressors and emotion inductions. Psychol Bull 135(6):823–853.

Dickerson S, Kemeny ME. 2004. Acute stressors and cortisol responses: A theoretical integration and synthesis of laboratory research. Psychol Bull 130(3):355–391.

Domes G, Heinrichs M, Rimmele U, Reichwald U, Hautzinger M. 2004. Acute stress impairs recognition for positive words association with stress induced cortisol secretion. Stress 7:173–181.

Dyrbye LN, Thomas MR, Shanafelt TD. 2005. Medical student distress: Causes, consequences, and proposed solutions. Mayo Clin Proceed 80(12):1613–1622.

Elzinga BM, Roelofs K. 2005. Cortisol-induced impairments of working memory require acute sympathic activation. Behav Neur 119(1):98–103.

Eva K. 2004. What every teacher needs to know about clinical reasoning. Med Educ 39:98–106.

Fleiss JL. 1981. Statistical methods for rates and proportions. New York: Wiley.

Graber ML, Franklin N, Gordon R. 2005. Diagnostic error in internal medicine. Arch Intern Med 165(13):1493–1499.

Harvey A, Bandiera G, Nathens AB, LeBlanc VR. 2012. Impact of stress on resident performance in simulated trauma scenarios. J Trauma Acute Care Surg 72(2):497–503.

Harvey A, Nathens AB, Bandiera G, LeBlanc VR. 2010. Threat and challenge: Cognitive appraisal and stress responses in simulated trauma resuscitations. Med Educ 44:587–594.

Hindmarsh KW, Tan L, Sankaran K, Laxdal VA. 1989. Diurnal rhythms of cortisol, ACTH, and beta-endorphin levels in neonates and adults. West J Med 151(2):153–156.

Hodges B, Mc Ilroy JH. 2003. Analytic global OSCE ratings are sensitive to level of training. Med Educ 37:1012–1016.

Janis IL, Mann L. 1977. Decision making: A psychological analysis of conflict, choice, and commitment. New York: Free Press.

Johnston JH, Driskell JE, Salas E. 1997. Vigilant and hyper vigilant decision-making. J App Psychol 82(4):614–622.

Lazarus RS, Folkman S. 1984. Stress appraisal and coping. 1st ed. New York: Springer.

LeBlanc VR. 2009. The effects of acute stress on performance: Implications for health professions education. Acad Med 84(10 Suppl):S25–S33.

LeBlanc VR, Bandiera GW. 2007. The effect of examination stress on the performance of emergency medicine residents. Med Educ 41:556–564.

LeBlanc VR, Brooks LR, Norman GR. 2002. Believing is seeing: The influence of a diagnostic hypothesis on the interpretation of clinical features. Acad Med 77(10):S67–S69.

LeBlanc VR, MacDonald RD, McArthur B, King K, Lepine T. 2005. Paramedic performance in calculating drug dosages following stressful scenarios in a human patient simulator. Prehosp Emerg Care 9(4):439–44.

LeBlanc VR, Regehr C, Tavares W, Scott AK, MacDonald R, King K. 2012. The impact of stress on paramedic performance during simulated critical events. Prehosp Disaster Med 27(4):369–374.

LeBlanc VR, Woodrow SI, Sidhu R, Dubrowski A. 2008. Examination stress leads to improvements on fundamental technical skills for surgery. Am J Surg 196:114–119.

Norman G, Young M, Brooks L. 2007. Non-analytical models of clinical reasoning: The role of experience. Med Educ 41:1140–1145.

Pamphlett R, Farnill D. 1989. Effect of anxiety on performance in multiple choice examinations. Med Educ 23(5):416–421.

Pottier P, Hardouin JB, Desjoie T, Le Loupp AG, Bonnaud-Antignac A, Planchon B, LeBlanc VR. 2011. Stress responses in medical students in ambulatory and in-hospital patient consultations. Med Educ 45:678–687.

Pruessner JC, Gaab J, Hellhammer DH, Lintz D, Schommer N, Kirschbaum C. 1997. Increasing correlations between personality traits and cortisol stress responses obtained by data aggregation. Psychoneuroendocrino 22(8):615–625.

Reteguiz JA. 2006. Relationship between anxiety and standardized patient test performance in the medicine clerkship. J Gen Intern Med 21:415–418.

Sanders AE, Lushington K. 2002. Effect of perceived stress on student performance in dental school. J Dent Educ 66(1):75–81.

Spielberger CD, Gorsuch RL, Lushene R, Vagg PR, Jacobs GA. 1983. Manual for the State-Trait anxiety inventory. Palo Alto, CA: Consulting Psychologists Press.

Toews JA, Lockyer JM, Dobson DJG, Brownell AK. 1993. Stress among residents, medical students, and graduate science (MSc/PhD) students. Acad Med 68(10 suppl):S46–S48.

Tomaka J, Blaskovich J, Kelser RM, Leitten CL. 1993. Subjective, physiological and behavioural effects of threat and challenge appraisal. J Pers Soc Psychol 65(2):248–260.

Van Dulmen S, Tromp F, Grosfeld F, ten Cate O, Bensing J. 2007. The impact of assessing simulated bad news consultations on medical students' stress response and communication performance. Psychoneuroendocrino 32(8):943–950.

Wetzel CM, Black SA, Hanna GB, Athanasiou T, Kneebone RL, Nestel D, Wolfe JH, Woloshynowych M. 2010. The effects of stress and coping on surgical performance during simulations. Ann Surg 251:171–176.

Yates JF, Tschirthart MD. 2006. Chapter 24: Decision-making. In: Eltovich P, Hoffman P, editors. The Cambridge handbook of expertise and expert performance. Cambridge, New York: Cambridge University Press.