Sample size requirement for comparison of decompression outcomes using ultrasonically detected venous gas emboli (VGE): power calculations using Monte Carlo resampling from real data

David J Doolette, Keith A Gault and Christian R Gutvik

Abstract

(Doolette DJ, Gault KA, Gutvik CR. Sample size requirement for comparison of decompression outcomes using ultrasonically detected venous gas emboli (VGE): power calculations using Monte Carlo resampling from real data. *Diving and Hyperbaric Medicine*. 2014 March;44(1):14-19.)

Introduction: In studies of decompression procedures, ultrasonically detected venous gas emboli (VGE) are commonly used as a surrogate outcome if decompression sickness (DCS) is unlikely to be observed. There is substantial variability in observed VGE grades, and studies should be designed with sufficient power to detect an important effect.

Methods: Data for estimating sample size requirements for studies using VGE as an outcome is provided by a comparison of two decompression schedules that found corresponding differences in DCS incidence (3/192 [DCS/dives] vs. 10/198) and median maximum VGE grade (2 vs. 3, $P < 0.0001$, Wilcoxon test). Sixty-two subjects dived each schedule at least once, accounting for 183 and 180 man-dives on each schedule. From these data, the frequency with which 10,000 randomly resampled, paired samples of maximum VGE grade were significantly different (paired Wilcoxon test, one-sided $P \leq 0.05$ or 0.025) in the same direction as the VGE grades of the full data set were counted (estimated power). Resampling was also used to estimate power of a Bayesian method that ranks two samples based on DCS risks estimated from the VGE grades.

Results: Paired sample sizes of 50 subjects yielded about 80% power, but the power dropped to less than 50% with fewer than 30 subjects.

Conclusions: Comparisons of VGE grades that fail to find a difference between paired sample sizes of 30 or fewer must be interpreted cautiously. Studies can be considered well powered if the sample size is 50 even if only a one-grade difference in median VGE grade is of interest.

Key words

Decompression, diving, echocardiography, venous gas emboli, decompression sickness, statistics, research

Introduction

Decompression sickness (DCS) is thought to be caused by intracorporeal bubble formation. Venous bubbles (venous gas emboli, VGE) are sometimes used as an outcome in studies of decompression procedures because they can be easily detected by ultrasonic methods and graded, and because VGE grades have a general correlation with the incidence of DCS in large compilations of data.[1,2] This correlation may arise in part because VGE can cause some manifestations of DCS, but an increase in detectable VGE is also presumed to be correlated with an increase risk of bubble formation at other DCS sites. VGE grades are used to augment DCS incidence data or as a surrogate outcome if DCS is unlikely to be observed, for instance in anesthetized animals, or in studies of low-risk human procedures.

VGE occur commonly without DCS (which is rare); therefore, VGE data are potentially more information-rich than low-incidence DCS data. This additional information is counterbalanced by the facts that, owing to poor specificity, VGE grades have poor diagnostic value for DCS, and there is substantial inter- and intra-individual variability in VGE grades observed following identical exposures.[3–6] These latter facts impose a lower limit on sample size for studies of low-risk human procedures that use VGE as a surrogate outcome measure.

A common design of such studies is for two different procedures to be performed on separate occasions by the same subjects, and to test for a difference in VGE outcome using a paired statistical test such as the Wilcoxon signed-rank test. The power of a statistical test to detect a particular effect size at a particular statistical significance criterion (α) depends on the sample size, so power calculations may be used when designing an experiment to select an appropriate sample size. This study provides estimates of power for various sample sizes for human studies that use paired comparisons of VGE grades following decompression.

Methods

Monte Carlo experiments analyze outcomes in multiple computer-generated random samples. For instance, the probability of an outcome is estimated by the proportion of samples in which the outcome occurs. Monte Carlo experiments can be used to examine the properties of statistical hypothesis tests, for instance, the probability of rejecting a false null hypothesis (power) for a test procedure which produces a *P*-value and then rejects the null hypothesis if the *P*-value is less than or equal to a particular α-level. Monte Carlo estimation of the power involves computing the proportion of rejections in many random samples. Typically the random samples would be simulations generated from

parametric distributions and, in the case of a two-sample test, hypothetical effect sizes. However, in this report, samples were generated by resampling subsets of real data.

DATA

A recently published, large-scale comparison of two air decompression schedules provides unique data for estimating sample size requirements, finding corresponding statistically significant differences in DCS incidence and median peak VGE grade.[7] Eighty-one US Navy divers participated in a total of 390 man-dives, performing work during 30 minutes' bottom time at 622 kPa absolute (170 feet of sea water gauge, fsw). They were at rest and cold during either of two decompression schedules that differed only in the distribution of 174 minutes' total decompression stop time among stop depths: a shallow stop (A1) schedule and a deep stop (A2) schedule. The study reached an early stopping criterion at midpoint analysis, which found a lower incidence of DCS on the A1 than the A2 schedule at one-sided $\alpha = 0.05$ (an early 'opposite tail' finding relative to a final result that would have motivated changing US Navy procedures). DCS was diagnosed by the duty diving medical officer and full descriptions are given in the original report. During re-evaluation of the cases according to the criteria described in Temple et al,[8] one case with symptom onset 27 hours after surfacing from the A2 schedule was re-classified as not DCS. This resulted in 3/192 (DCS/dives) and 10/198 ($P = 0.0489$, one-sided Fisher's exact test), on the A1 and A2 schedules, respectively.

As a secondary outcome measure, subjects were monitored for VGE with trans-thoracic cardiac 2-D echo imaging at 30 minutes and two hours post dive. While the subjects reclined with left side down, the four heart chambers were imaged with the subject at rest and then, in turn, while they flexed each elbow and knee. VGE were graded according to the Table 1 scale, adapted from Eftedal and Brubbak.[9] The same ultrasound technician conducted all the examinations and all observed VGE grades are documented elsewhere.[7] However, in this report, only the maximum VGE grades observed at any time (rest or limb flexion, any examination) after each dive were used and will be referred to as 'VGE grade' without qualification. The median VGE grades were 2 and 3 (two-sided $P < 0.0001$, Wilcoxon rank sum test), on the A1 and A2 schedules, respectively. VGE data were missing for three man-dives: two subjects were recompressed to treat DCS before VGE examination, and results for a subject without symptoms were inadvertently not recorded. In each case, the same subject undertook the same schedule (for which data was missing) and had VGE recorded, on at least one other occasion.

The original study was not designed as a paired comparison, but of the 81 subjects who participated in the original trial, 62 dived each schedule at least once. The VGE outcome of all dives undertaken by these 62 subjects was designated the paired data set and was used to generate random samples of paired data (VGE grade after A1 and A2 schedules in the same subject). The paired data set contained 363 records, each representing one man-dive, and each comprised of a subject identifier, a schedule identifier, and the VGE grade. The distribution of VGE grades in the paired data set is given in Table 2. Median VGE grade was 2 (interquartile range [IQR] 1–3) following the A1 schedule and 3 (IQR 2–4) follow the A2 schedule. These VGE grades were significantly different (Wilcoxon rank sum test, two-sided $P < 0.0001$), and A1 less than A2 will be considered as the true outcome for power estimation. Many subjects dived the A1 and A2 schedules more than once. The mean number of dives per subject on the A1 schedule was 3 (range 1–9) accounting for a total of 183 man dives. The mean number of dives per subject on the A2 schedule was 3 (range 1–8) accounting for a total of 180 man-dives. There was no requirement in the original study for subjects to dive A1 and A2 schedules an equal number of times; however, the differences between the number of A1 and A2 schedules undertaken by each subject were relatively symmetrically distributed around zero with the absolute value of the difference/number of subjects: 0/25; 1/20; 2/12; 3/3; 4/2. Subjects refrained from any hyperbaric or hypobaric exposure for three days prior to any of the dives in the paired data set and the most common interval between these dives was seven days.

Table 1
Venous gas embolism grading (modified from reference 9)

Grade	Description
0	No bubble seen
1	Rare (< 1/s) bubble seen
2	Several discrete bubbles visible per image
3	Multiple bubbles visible per image but not obscuring image
4	Bubbles dominate image, may blur chamber outlines

RESAMPLING

For each of a range of paired sample sizes (n = 10 to 60 subjects), Monte Carlo resampling and testing of paired VGE grades was performed in the following manner. First, a subset of n subjects was randomly selected without replacement from a vector containing the 62 subject identifiers. Second, for each subject in this subset, one VGE grade was randomly selected from among the A1 schedules and one from among the A2 schedules that subject had completed. The resulting subset contained an A1-A2 pair of VGE grades for n different subjects. VGE grades from different subjects were considered independent and the resampling scheme took advantage of subjects who dived a schedule more than once by allowing different A1-A2 pairs for that subject in different subsets (there are more than 10^{41} possible such combinations in the paired data set for each value of n). Finally, for each

Table 2
VGE grades, paired data set (taken from reference 7)

	0	1	2	3	4
A1	27	36	36	53	31
A2	9	25	29	45	72

subset, the *P*-value of a paired Wilcoxon signed-rank test, with alternative hypothesis A1 less than A2 (in accord with the true outcome) was recorded. This three-step procedure was repeated 10,000 times for each value of *n*. The frequency with which *P*-values from the 10,000 subsets were less than or equal to a particular α-level provides an estimate of the probability of an α-level test on sample size of *n* subjects detecting the true one-grade difference in VGE in the paired data set (power). Power estimates are given for one sided $\alpha = 0.05$ because this level was an early stopping criterion for difference in DCS incidence in the original study that generated the data set, and for one-sided $\alpha = 0.025$ because this level is equivalent to two-sided $\alpha = 0.05$ that would commonly be used for comparisons where there is no justification for a one-sided test.

Within-subject variability in VGE grade for the same schedule was considered to be random since dives were sufficiently spaced so as not to influence each other either in terms of residual nitrogen or acclimatization. This assumption was not a requirement of the nonparametric statistical analysis. Some variability may result from measurement precision and, in particular, VGE measurements in the original study were infrequent (30 and 120 min post dive) and may not have consistently captured the peak VGE grade that occurred after each dive. To examine the consequence of possible frequent failure to record the peak VGE grade, a modified data set was drawn from the paired data set. The modified data set comprised only the maximum VGE grade observed among each repetition of the A1 schedule and each repetition of the A2 schedule for each of the 62 subjects (no intra-individual variability). The modified data set had median VGE grades of 3 (IQR 2.25–4) following the A1 schedule and 4 (IQR 3–4) following the A2 schedule (paired Wilcoxon signed-rank test, two-sided $P = 0.0056$). For each of a range of paired sample sizes ($n = 10$ to 50), a subset of *n* A1-A2 pairs of VGE grades was randomly selected without replacement from the 62 in the modified data set and tested with a paired Wilcoxon signed-rank test, with alternative hypothesis A1 less than A2. This resampling procedure was repeated 10,000 times and the power estimated as described for the paired data set. There are more than 10^{12} combinations of 50 from 62 subjects, but only 1,891 combinations of 60 from 62 subjects, so estimating power for $n = 60$ subjects by resampling from the modified data set was not considered meaningful.

Recently, a Bayesian method has been proposed to estimate the probability of DCS of a decompression procedure from maximum observed VGE grades and test for a difference in risk between two procedures.[10] We estimated the power of this latter test for comparison with the Wilcoxon test. Briefly, the method constructs posterior distributions of the probability of DCS given VGE grade (for instance based on the data given by Sawatzky[1]) and the probability of VGE grade given the test procedure, and then the total probability of DCS of a procedure is estimated by Monte Carlo simulation from these posteriors. Two procedures are tested for a difference in DCS risk by counting the frequency with which one procedure is estimated as riskier than the other (estimated confidence of the difference) in parallel Monte Carlo simulations. Using the same prior distributions as originally described[10] to produce posterior distributions from the present paired data set resulted in an estimated 99.98% confidence that the A2 schedule was riskier than the A1 schedule. Again using the same prior distributions, posterior distributions were produced from resampled subsets of the present paired data set. For each resampled subset, the confidence that the A2 schedule was riskier than the A1 schedule (in accord with the true outcome of both the Bayesian and Wilcoxon tests) was estimated. The frequency with which this confidence was greater than 95% in resampled subsets is comparable (but not identical) to the power estimate for the Wilcoxon rank sum test at one-sided $\alpha = 0.05$. Only sample sizes $n = 20$ and $n = 50$ were examined, and resampled 500 times, because the Bayesian method itself requires Monte Carlo simulations and is highly computing intensive.

Table 3
Power estimated from frequency of observed *P*-values of Wilcoxon test, paired data set

	Number of subjects					
	10	**20**	**30**	**40**	**50**	**60**
Power						
one-sided $P \leq 0.05$	0.27	0.48	0.65	0.78	0.88	0.94
one-sided $P \leq 0.025$	0.15	0.34	0.50	0.66	0.78	0.87

Data analysis was performed using R version 2.14.2 (Vienna, Austria: R Development Core Team; 2012) and MATLAB version 7.8.0.347 (R2009a) (Natwick, MA: The MathWorks Inc; 2009).

Results

Table 3 shows the power for various sample sizes for the Wilcoxon rank sum test, estimated by resampling from the paired data. These values are the probabilities of a significant test ($P \leq 0.05$ and $P \leq 0.025$) in accord with the true outcome. The fraction of results not in accord with the true outcome were usually failure to find a difference between A1 and A2 VGE grades (type II error) – the opposite tail finding of higher VGE grades on A1 than A2 was extremely rare, the highest frequency of this result was 0.0016 for $n = 10$ and $P \leq 0.05$, and otherwise zero. The choice of power depends on the consequences of making a type II error, but

Table 4
Power estimated from frequency of observed *P*-values of Wilcoxon test, modified data set

	Number of subjects				
	10	**20**	**30**	**40**	**50**
Power					
one-sided $P \leq 0.05$	0.22	0.39	0.56	0.75	0.95
one-sided $P \leq 0.025$	0.11	0.24	0.37	0.55	0.78

one convention is to design experiments with two-sided $\alpha = 0.05$ and 80% power. From the one-sided $P \leq 0.025$ row (equivalent to two-sided $\alpha = 0.05$) in Table 3, it can be seen that VGE grades from a paired sample size of about $n = 50$ subjects would have 80% power to detect a difference of one VGE grade. Power dropped quickly with sample size so that at $n = 30$ subjects ($P \leq 0.025$) there was equal probability of a true answer and a type II error.

Table 4 shows the power for various sample sizes for the Wilcoxon rank sum test, estimated by resampling from the modified data comprising only the highest VGE scores from repeated dives on the same schedules. Although there are some differences from the results of the paired data set, a sample size of about $n = 50$ is required for 80% power at two-sided $\alpha = 0.05$.

Power estimates for the Bayesian test were similar to those of the Wilcoxon rank sum test. The frequency of predicting the A2 schedule to be riskier than A1 schedule with 95% confidence was 0.40 for $n = 20$ resampled subsets and 0.80 for $n = 50$ resampled subsets. These power estimates are comparable to the values for these sample sizes in the $P \leq 0.05$ row of Table 3. The opposite tail finding (A1 riskier than A2 with 95% confidence) never occurred.

Discussion

Statistical power (or sensitivity) is the probability of rejecting a false null hypothesis (not making a type II error). In the current context, this is the probability of finding a difference (rejecting the null hypothesis of no difference) between paired samples of VGE grades for each schedule given that the VGE grades are different for each schedule in the population. The power of a statistical test depends on the magnitude of the effect to be detected, the α-value of the test, and the sample size. Power calculations are used to select appropriate sample sizes when designing experiments and Table 3 provides guidelines for designing paired comparisons using VGE as an outcome. For instance, a paired sample size of about 50 subjects is required for 80% power to detect a one-grade difference in median VGE at one-sided $\alpha = 0.025$ (equivalent to two-sided $\alpha = 0.05$) in this relatively homogenous group of subjects diving under rigidly controlled conditions.

The present results are only relevant to a one-grade difference in VGE. For instance, analysis of a simulated data set with a two-grade difference in median VGE (not shown) found a paired sample size of about 20 was required for 80% power to detect the difference at two-sided $\alpha = 0.05$. Nevertheless, the present guidelines are broadly applicable for two reasons: one VGE grade is the precision that is common across the most frequently used grading systems and many published studies report one-grade or less difference in VGE. With respect to grading precision, the present VGE grading system was a modification of the Eftedal-Brubakk system for grading VGE in 2D echocardiographic images, and the Eftedal-Brubakk grading system is broadly similar to the Spenser and Kisman-Masurel systems for aural grading of VGE detected by ultrasonic Doppler shift, in that they all grade human VGE data on an approximately equivalent zero to four ordinal scale (although the Kisman-Masurel system reports "+" and "−" intergrades and the Eftedal-Brubakk system has a grade 5 which has not been reported in humans).[2,9,11] Sample size guidelines based on the minimum measurable difference in peak VGE grade (e.g., Table 3) are useful if there is no reason to expect or require a greater difference.

The estimated power to detect a one-grade difference in median VGE is relevant to many published studies. A Medline search for the 10 years up to 2012 identified 23 publications that were paired comparisons of VGE following diving (68% of all publications found concerning VGE and diving in humans in this period). Of these, 16 reported the individual or summary statistics of the observed VGE grades (Eftedal-Brubakk, Spencer or Kisman-Masurel systems).[12–27] Only three of these 16 papers reported more than a one-grade difference in median VGE.[17,24,27] Sample sizes in these studies ranged from 6 to 28 subjects and only four of these papers reported a significant difference in VGE grades. Four papers reported no significant difference in VGE grades, and eight reported significant difference in transformations of the data. The most common transformations were to bubble count·cm^{-2} and to the Kisman-Masurel integrated severity score.[2,5] Bubble count·cm^{-2}, if a transformation from peak VGE grades (i.e., not measured directly), is subject to the same power constraints as the underlying VGE grades. The current power calculations are not applicable to the Kisman-Masurel integrated severity score which includes additional time-course information. If the Kisman-Masurel integrated severity score were demonstrated to have a stronger correlation with DCS incidence than has maximum VGE grades, sample size guidelines would be useful, but the present data did not include sufficiently frequent VGE measurements to calculate a meaningful score.

Power estimates are dependent on the precision of measurement. A limitation of the present estimates is that the paired data set may have unnecessary variance because infrequent measurements of VGE may not have always captured the true peak VGE grade. Any such aliasing may

not have been severe because the two VGE examinations (at 30 and 120 minutes) span the period during which peak VGE are typically recorded following bounce dives and VGE grades were similar at these two examinations.[28] There was no difference in VGE grades between examination times following the A1 schedule; however, there was a significant difference in VGE grades between examinations following the A2 schedule (Wilcoxon rank sum test two-sided, $P = 0.0006$) but the estimated location shift was only one-half a VGE grade. Also, the modified data set, which had no intra-individual variability in VGE scores, produced similar power estimates to those extracted from the paired data set.

The concordance between VGE grades and DCS incidence in the present data is of interest since VGE grades are often used as a surrogate for DCS (although not in the original study). The dives in the present data set were relatively risky air decompression dives; for instance, in the US Navy Diving Manual, an air dive to 170 fsw for 30 minute bottom time requires the use of oxygen decompression, and the two air schedules had a measurable difference in DCS incidences.[29] The original study planned 375 man-dives on each schedule, which would have had approximately 80% power to detect the actually observed difference in DCS incidences (a difference which was larger than expected) at two-sided $\alpha = 0.05$. This is compared with a paired sample size of about $n = 50$ subjects to detect the observed one-grade difference in median VGE at the same power and significance. While this comparison is interesting in hindsight, the objective of the original comparison of decompression procedures was to discern any practical difference in the DCS incidence, not VGE grades per se.

The concordance of differences in VGE grades and differences in DCS risk (estimated from observed DCS incidence) in the present data will not necessarily hold for all experiments. In the largest compilation of VGE and DCS incidence following diving, there was no DCS associated with Kisman-Masurel grade 0 (0 DCS/819 dives) and DCS incidence was indistinguishable between grades I (3 DCS/287 dives) and II (2 DCS/183 dives) or between grades III (27 DCS/365 dives) and IV (9 DCS/72 dives), although the DCS incidence does differ between these low and high VGE grades.[1] Therefore, an experiment that demonstrates a statistically significant difference between, for instance, median VGE grades I and II using a Wilcoxon signed rank test, may not reflect a demonstrable difference in DCS risk. Misinterpretation is less likely with the Bayesian method of Eftedal and colleagues.[10] This Bayesian method compares estimates of the probability of DCS derived from information about the distribution of DCS incidence with VGE grades, in this case a prior distribution from the data compilation noted above.[1] Because the Bayesian method incorporates this prior, it is unlikely to find a difference between a sample dominated by VGE grade I and a sample dominated by VGE grade II, unless there is also substantial difference in the distribution of other VGE grades between the samples. Conversely, any analysis of VGE may fail to identify a true difference in DCS risk between two samples dominated by grade IV VGE, since this is the highest grade observable, irrespective of DCS risk. The similarity of power and sample size estimates between the Wilcoxon and Bayesian test on the present data arises because the median VGE grades on the A1 and A2 schedule were 2 and 3 (equivalent to Kisman-Masurel grades II and III), respectively, and there is a significant difference in DCS incidence between these grades in the prior distribution.

Conclusions

Comparisons of two decompression procedures using only VGE as an endpoint that fail to find a difference between paired sample sizes of 30 or fewer must be interpreted cautiously. Studies can be considered well powered if the sample size is above 50 even if only a one-grade difference in median VGE is of interest. Maximum VGE grades can provide more power than DCS incidence to distinguish between two decompression procedures; however, a difference in VGE grades does not necessarily reflect a difference in DCS risk. If the purpose of the study is to infer a difference in DCS risk from VGE grades alone, VGE data must be interpreted cautiously, and the Bayesian method incorporating appropriate prior information about the distribution of DCS incidence with VGE grades is preferred over simple statistical tests such as the Wilcoxon signed-rank test.

References

1 Sawatzky KD. *The relationship between intravascular Doppler-detected gas bubbles and decompression sickness after bounce diving in humans* [MSc Thesis]. Toronto, ON: (Canada): York University; 1991.

2 Nishi RY, Brubakk AO, Eftedal OS. Bubble detection. In: Brubakk AO, Neuman,TS, editors. *Bennett and Elliott's physiology and medicine of diving*. 5th ed. Edinburgh: Saunders; 2003. p. 501-29.

3 Kumar VK, Billica RD, Waligora JM. Utility of Doppler-detectable microbubbles in the diagnosis and treatment of decompression sickness. *Aviat Space Environ Med*. 1997;68:151-8.

4 Gerth WA, Ruterbusch VL, Long ET. *The influence of thermal exposure on diver susceptibility to decompression sickness.* Technical Report. Panama City, FL: Navy Experimental Diving Unit; 2007 Nov. Report No.: NEDU TR 06-07. Available at http://archive.rubicon-foundation.org/xmlui/handle/123456789/5063.

5 Nishi RY, Kisman KE, Eatock BC, Buckingham IP, Masurel G. Assessment of decompression profiles and divers by Doppler ultrasonic monitoring. In: Bachrach AJ, Matzen MM, editors. *Underwater physiology VII. Proceedings of the 7th Symposium on Underwater Physiology*. Bethesda, MD: Undersea Medical Society; 1981. p. 717-27.

6 Eckenhoff RG, Hughes JS. Acclimatization to decompression stress. In: Bachrach AJ, Matzen MM, editors. *Underwater physiology VIII. Proceedings of the 8th Symposium on Underwater Physiology*. Bethesda, MD: Undersea Medical Society; 1984. p. 93-100.

7 Doolette DJ, Gerth WA, Gault KA. *Redistribution of decompression stop time from shallow to deep stops increases incidence of decompression sickness in air decompression dives.* Technical Report. Panama City, FL: Navy Experimental Diving Unit; 2011 Jul. Report No.: NEDU TR 11-06. Available at http://archive.rubicon-foundation.org/xmlui/handle/123456789/10269.
8 Temple DJ, Ball R, Weathersby PK, Parker EC, Survanshi SS. *The dive profiles and manifestations of decompression sickness cases after air and nitrogen-oxygen dives*. Technical Report. Bethesda, MD: Naval Medical Research Center; 1999. Vol 1. Report No.: 99-02. Available at http://archive.rubicon-foundation.org/xmlui/handle/123456789/4975.
9 Eftedal O, Brubakk AO. Agreement between trained and untrained observers in grading intravascular bubble signals in ultrasonic images. *Undersea Hyperb Med*. 1997;24:293-9.
10 Eftedal OS, Tjelmeland H, Brubakk AO. Validation of decompression procedures based on detection of venous gas bubbles: a Bayesian approach. *Aviat Space Environ Med*. 2007;78:94-9.
11 Brubakk AO, Eftedal O. Comparison of three different ultrasonic methods for quantification of intravascular gas bubbles. *Undersea Hyperb Med*. 2001;28:131-6.
12 Dujic Z, Duplancic D, Marinovic-Terzic I, Bakovic D, Ivancev V, Valic Z, et al. Aerobic exercise before diving reduces venous gas bubble formation in humans. *J Physiol*. 2004;555:637-42.
13 Marroni A, Bennett PB, Cronje FJ, Cali-Corleo R, Germonpre P, Pieri M, et al. A deep stop during decompression from 82 fsw (25 m) significantly reduces bubbles and fast tissue gas tensions. *Undersea Hyperb Med*. 2004;31:233-43.
14 Blatteau JE, Gempp E, Galland FM, Pontier JM, Sainty JM, Robinet C. Aerobic exercise 2 hours before a dive to 30 msw decreases bubble formation after decompression. *Aviat Space Environ Med*. 2005;76:666-9.
15 Dujic Z, Palada I, Valic Z, Duplancic D, Obad A, Wisloff U, et al. Exogenous nitric oxide and bubble formation in divers. *Med Sci Sports Exerc*. 2006;38:1432-5.
16 Blatteau JE, Boussuges A, Gempp E, Pontier JM, Castagna O, Robinet C, et al. Haemodynamic changes induced by submaximal exercise before a dive and its consequences on bubble formation. *Br J Sports Med*. 2007;41:375-9.
17 Blatteau JE, Pontier JM. Effect of in-water recompression with oxygen to 6 msw versus normobaric oxygen breathing on bubble formation in divers. *Eur J Appl Physiol*. 2009;106:691-5.
18 Bosco G, Yang ZJ, Di Tano G, Camporesi EM, Faralli F, Savini F, et al. Effect of in-water oxygen prebreathing at different depths on decompression-induced bubble formation and platelet activation. *J Appl Physiol*. 2010;108:1077-83.
19 Jurd KM, Thacker JC, Seddon FM, Gennser M, Loveman GA. The effect of pre-dive exercise timing, intensity and mode on post-decompression venous gas emboli. *Diving Hyperb Med*. 2011;41:183-8.
20 Blatteau JE, Hugon J, Gempp E, Pény C, Vallée N. Oxygen breathing or recompression during decompression from nitrox dives with a rebreather: effects on intravascular bubble burden and ramifications for decompression profiles. *Eur J Appl Physiol*. 2012;112:2257-65.
21 Schellart NA, Sterk W. Venous gas embolism after an open-water air dive and identical repetitive dive. *Undersea Hyperb Med.* 2012;39:577-87.
22 Gennser M, Jurd KM, Blogg SL. Pre-dive exercise and post-dive evolution of venous gas emboli. *Aviat Space Environ Med*. 2012;83:30-4.
23 Risberg J, Englund M, Aanderud L, Eftedal O, Flook V, Thorsen E. Venous gas embolism in chamber attendants after hyperbaric exposure. *Undersea Hyperb Med*. 2004;31:417-29.
24 Marinovic J, Ljubkovic M, Breskovic T, Gunjaca G, Obad A, Modun D, et al. Effects of successive air and nitrox dives on human vascular function. *Eur J Appl Physiol*. 2012;112:2131-7.
25 Castagna O, Brisswalter J, Vallee N, Blatteau JE. Endurance exercise immediately before sea diving reduces bubble formation in scuba divers. *Eur J Appl Physiol*. 2011;111:1047-54.
26 Dujic Z, Palada I, Obad A, Duplancic D, Bakovic D, Valic Z. Exercise during a 3-min decompression stop reduces postdive venous gas bubbles. *Med Sci Sports Exerc*. 2005;37:1319-23.
27 Møllerløkken A, Breskovic T, Palada I, Valic Z, Dujic Z, Brubakk AO. Observation of increased venous gas emboli after wet dives compared to dry dives. *Diving Hyperb Med*. 2011;41:124-8.
28 Blogg SL, Gennser M. The need for optimisation of post-dive ultrasound monitoring to properly evaluate the evolution of venous gas emboli. *Diving Hyperb Med*. 2011;41:139-46.
29 Naval Sea Systems Command. *US Navy Diving Manual*. Revision 6, NAVSEA 0910-LP-106-0957/SS521-AG-PRO-010. Arlington, VA: Naval Sea Systems Command; 2008.

Conflicts of interest: None

Acknowledgements

Aspects of this work were supported by the Naval Sea Systems Command Deep Submergence Biomedical Research Program. We thank WA Gerth for lively discussions on analysis of VGE data. The Bayesian analysis was based on the original MATLAB code authored by OS Eftedal.

Submitted: 06 August 2013
Accepted: 01 January 2014

David J Doolette[1], Keith A Gault[1] and Christian R Gutvik[2]

[1] Navy Experimental Diving Unit, Panama City, Florida, USA
[2] Technology Transfer Office, Norwegian University of Science and Technology (NTNU)

Address for correspondence:
David Doolette, PhD
Navy Experimental Diving Unit
321 Bullfinch Road
Panama City, FL 32407, USA
Phone: *+1-(0)850-230-3100*
E-mail: *<david.doolette.as@navy.mil>*

Decompression illness in divers treated in Auckland, New Zealand, 1996–2012

Rachel M Haas, Jacqueline A Hannam, Christopher Sames, Robert Schmidt, Andrew Tyson, Marion Francombe, Drew Richardson and Simon J Mitchell

Abstract

(Haas RM, Hannam JA, Sames C, Schmidt R, Tyson A, Francombe M, Richardson D, Mitchell SJ. Decompression illness in divers treated in Auckland, New Zealand, 1996–2012. *Diving and Hyperbaric Medicine*. 2014 March;44(1):20-25.)

Introduction: The treatment of divers for decompression illness (DCI) in Auckland, New Zealand, has not been described since 1996, and subsequent trends in patient numbers and demographics are unmeasured.

Methods: This was a retrospective audit of DCI cases requiring recompression in Auckland between 01 January 1996 and 31 December 2012. Data describing patient demographics, dive characteristics, presentation of DCI and outcomes were extracted from case notes and facility databases. Trends in annual case numbers were evaluated using Spearman's correlation coefficients (ρ) and compared with trends in entry-level diver certifications. Trends in patient demographics and delay between diving and recompression were evaluated using regression analyses.

Results: There were 520 DCI cases. Annual caseload decreased over the study period ($\rho = 0.813$, $P < 0.0001$) as did entry-level diving certifications in New Zealand ($\rho = 0.962$, $P < 0.0001$). Mean diver age was 33.6 (95% confidence limits (CI) 32.7 to 34.5) years and age increased ($P < 0.0001$) over the study period. Median (range) delay to recompression was 2.06 (95% CI 0.02 to 23.6) days, and delay declined over the study period ($P = 0.005$).

Conclusions: Numbers of DCI cases recompressed in Auckland have declined significantly over the last 17 years. The most plausible explanation is declining diving activity but improvements in diving safety cannot be excluded. The delay between diving and recompression has reduced.

Key words

Diving, embolism, decompression illness, hyperbaric oxygenation therapy, air/diagnosis/etiology/therapy, decompression sickness/diagnosis/epidemiology/etiology/physiopathology/therapy

Introduction

Decompression illness (DCI) may occur following compressed gas dives if intra-corporeal bubbles form from dissolved inert gas, or if air is introduced to the arterial circulation by pulmonary barotrauma. The definitive treatment of DCI involves recompression and oxygen administration in a hyperbaric chamber.[1] Recompression facilities in New Zealand are located in Auckland and Christchurch and these have, in general, served divers from the North and South Islands respectively, although lower North Island divers are sometimes evacuated to Christchurch for recompression. The recompression facility (the Slark Hyperbaric Unit, SHU) in Auckland has been based at the Royal New Zealand Navy Hospital (RNZNH). Another unit, operated by Hyperbaric Health (a private company), has offered treatment for DCI since 2006. The caseload of the SHU was last reported for the 1996 calendar year.[2]

We undertook this study to describe the numbers and characteristics of DCI cases treated in Auckland from 1996 to the present time. In particular, we set out to document any trends in case numbers, and in relevant parameters such as patient demographics, type of diving, and latency between the incident dive and recompression.

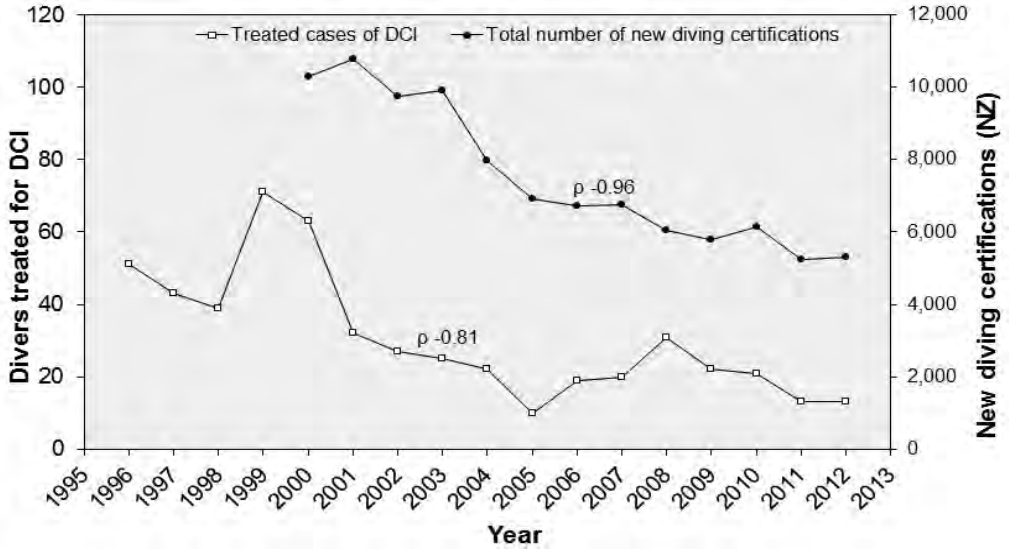
Methods

The study was approved by the University of Auckland Human Participants Ethics Committee (Reference: 9287). Locality approval was given by the Royal New Zealand Navy and Hyperbaric Health Limited. This was a retrospective, longitudinal audit of DCI cases treated in Auckland between 01 January 1996 and 31 December 2012. We chose 1996 as the start point because from this year forward the Christchurch unit was in continuous operation and patient numbers were not influenced by the need for evacuations from the South Island. A small number of cases were treated at the Hyperbaric Health Unit from 2006 and so these were also included in the audit.

Scuba divers who were recompressed and given a discharge diagnosis of DCI, probable DCI, or possible DCI were included. Cases considered 'unlikely' to have DCI or given alternative discharge diagnoses were excluded. At the SHU, case data were accessed by the principal author from two sources. The primary source was a Microsoft® Access 2 database maintained by the hyperbaric technicians and updated with each new patient's data during or soon after their admission. Where available, original case notes were also accessed for comparison against the database and extraction of missing or additional parameters. Data for cases treated at the Hyperbaric Health unit were extracted directly

Figure 1

Temporal trends in the number of divers treated for decompression illness in Auckland between 1996 and 2012 ($P < 0.0001$); the temporal trend in the number of newly certified divers (all New Zealand) is also displayed for 2000 to 2012 ($P < 0.0001$)



from case notes by a Hyperbaric Health clinician. Data from both units were combined into a single spreadsheet. Each case was given a unique study identifier. No data were collected that could identify patients or hospital staff.

The following data were extracted for each case: age; gender; height; weight; diver certification level; number of previous dives; maximum depth of incident dive or dive series; method of assessing decompression status during incident dive (dive table, dive computer, nil); breathing gas used (air or nitrox/mixed gases); equipment used (open-circuit scuba, surface supply or rebreather); latency between last dive and symptoms; nature of first-aid treatment; latency between last dive and recompression; qualitative nature of the symptoms; the presence or absence of objective signs on examination; putative risk factors for DCI; initial recompression treatment table; number of follow up recompressions and the recovery status at discharge (categorically divided as complete or incomplete recovery).

In an attempt to compare trends in annual case numbers against an indirect index of diving activity, annual numbers of new diver certifications in New Zealand over the years 2000–2012 were obtained by courtesy of a major global and national provider of diver training, the Professional Association of Diving Instructors (PADI).

This was a descriptive study rather than an investigation of hypotheses. Nevertheless, we identified the measurement of any trend in annual cases recompressed between 1996 and 2012 as the primary endpoint. Secondary endpoints were the trends over time in: maximum depth of the incident dive or dive series; breathing gases used; latency between the incident dive or dive series and recompression and in diver demographics such as age, body mass index (BMI) and gender.

STATISTICAL ANALYSIS

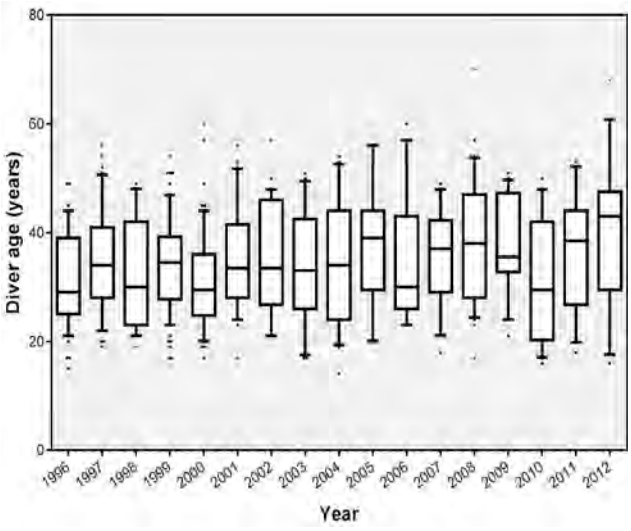
Trends in annual case numbers were investigated using Spearman's *rho* (ρ) correlation coefficients. Trends in secondary outcomes were investigated using regression analyses with year as a covariate. Linear regressions were conducted using normal distributions where appropriate, and Poisson distributions for count data. Binary data were investigated using logistic regression. A P value of < 0.05 is usually considered to indicate statistical significance; however, a total of eight analyses were conducted in this study and therefore the predefined criterion for statistical significance was adjusted using a Bonferroni correction (to $P < 0.00625$). All analyses were conducted using SPSS Statistics V. 19.

Results

NUMBER OF CASES

A total of 522 divers recompressed for DCI were identified. Two cases were excluded (one was an erroneous entry in the RNZNH database, and a second case was eventually diagnosed as feigned or 'factitious' DCI[3]), leaving a total cohort of 520 cases of which 506 were treated at the SHU and 14 at Hyperbaric Health. The annual DCI case load has trended downward over this period (Spearman's $\rho = 0.813$, $P < 0.0001$). Similarly, new diving certifications issued in New Zealand by PADI have also trended downward over a similar period (2000–2012) (Spearman's $\rho = 0.962$, $P < 0.0001$) (Figure 1).

Figure 2
Age of divers recompressed over the study period; the box plot shows the median (horizontal line inside boxes), interquartile range (boxes), and 10th-90th percentile (vertical lines). Outlier data are indicated by black dots. A significant upward trend in age is shown ($P < 0.0001$).



DIVER AGE, GENDER AND BMI

Demographics and diving characteristics of the recompressed divers are summarised in Table 1. Mean age (95% confidence limits, CI) across the cohort was 33.6 (32.7 to 34.5) years and age increased over the study period ($P < 0.0001$) (Figure 2). No significant trends were identified for gender or BMI over time.

DIVER EXPERIENCE

Certification levels among recreational divers covered a spectrum from no formal certification to instructor. There were also a number of so-called recreational 'technical divers' and occasional divers from professional groups such as commercial and military divers. Fifty-four per cent of divers for whom the previous number of dives was recorded had completed fewer than 100 dives at the time of injury, and 19% had undertaken more than 500 (Table 1).

NATURE OF DIVING

The vast majority of cases of DCI occurred in divers using standard open-circuit scuba equipment (95%), with six (~1%) using rebreathers, and six (~1%) using surface-supply equipment. In 13 cases (~2%) the equipment was not recorded. Of the 496 cases where the diving activity was explicitly recorded, 460 (93%) were diving recreationally, with only three involved in military diving and 33 (~6%) in occupational diving. Note, this distribution of activity does not intuitively match the certification data because some occupational diving (such as diving instruction) is undertaken by divers with recreational qualifications.

Table 1
Demographics and diving experience of 520 divers treated for decompression illness in Auckland between 1996 and 2012

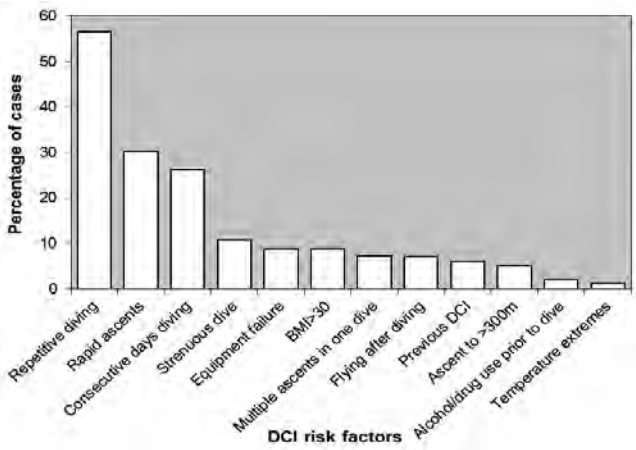
Diver demographics	***n***	**mean**	**range**
Age (yr)	512	34	14–70
Male:Female	419:101		
Weight (kg)	384	83	45–198
Height (cm)	379	178	140–210
BMI (kg m^{-2})	377	26.2	16.5–59.1
Certification level	***n***	**%**	
No certification	24	4.6	
Training	14	2.7	
Basic open water	274	52.7	
Advanced/Rescue/ Dive-master	98	18.8	
Instructor	49	9.4	
Commercial/Military/ Technical	4	0.8	
Unknown	57	11.0	
Experience level	***n***	**%**	
No previous dives	8	1.9	
≤ 5	25	5.8	
≤ 10	29	6.8	
≤ 100	169	39.4	
≤ 500	118	27.5	
> 500	80	18.6	
Breathing gas used	***n***	***%***	
Air diving	488	96.6	
Mixed-gas diving	17	3.4	

The depth of incident dives (or dive series) ranged from 1.8 to 80 metres, with a mean (95% CI) of 25.8 (24.74 to 26.92) m. There were no significant trends over time for depth of incident dive or use of air versus nitrox and mixed gas. There was an apparent increase over time in the proportion of recompressed divers who used a dive computer as the primary method of depth and time control. For example, in 1996 45% of incident dives were controlled according to a table plan whereas 18% were controlled by a computer (37% of divers either used nothing or the planning tool was not recorded). By 2012, this situation had reversed and 46% of divers were controlled by a computer, and 15% according to a table. Unfortunately over the period of the study, many data were missing in relation to this parameter, and we did not attempt to analyse the trend.

RISK FACTORS

In addition to provocative depth/time profiles, a number of putative risk factors for DCI have been proposed. The most prevalent of these among cases in this study was repetitive diving (57%). Rapid ascents (30%), consecutive days' diving

Figure 3
Putative risk factors for DCI among divers recompressed over the study period; data are the percentage of the total cases in which the risk factor was reported



(26%) and subjectively 'strenuous' diving (11%) were also features in many cases (Figure 3).

PRESENTATION OF DCI

The latency of symptom onset ranged from "*present on surfacing*" to 168 hours after diving, with a median time of 1.5 hours. The most frequently reported symptom was musculoskeletal pain (65% of cases), followed by cutaneous tingling (45%), headache (35%), fatigue (32%), weakness (31%), numbness (28%) and dizziness (22%). Objective signs were seen in 180 (36%) of the 499 divers for whom symptoms and objective signs were explicitly recorded. Objective signs included an abnormal sharpened Romberg test.[4] The percentage of cases in which each reported symptom occurred is given in Figure 4.

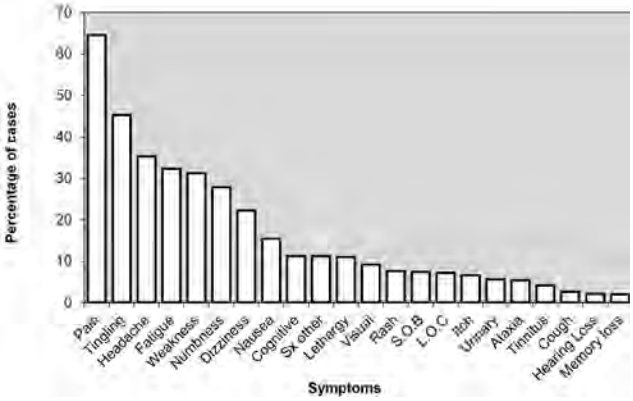
FIRST AID, REFERRAL AND TREATMENT

In 60% of cases, whether first-aid oxygen was administered was not recorded. Of the 210 (40%) recorded cases, only 87 (41%) received oxygen prior to recompression. Divers were referred mainly by their local doctor (31%), a hospital (30%), or were self-referred (28%). For the entire cohort, the median (range) latency between the incident dive and the time to recompression was 2.06 (0.02–23.6) days. This declined to a small but significant extent over the 17 years audited ($P = 0.005$).

RECOMPRESSION PROTOCOL

In accordance with widely accepted practice, divers underwent an initial recompression prescribed by a protocol chosen according to perceived DCI severity and physician preference. Most commonly this was the US Navy Treatment Table 6, used in 338 (65%) of cases. A 4 ATA (405 kPa) table utilising 50:50 oxygen-helium breathing, the so-called

Figure 4
Presenting symptoms of the divers treated over the study period (percentage of the total cases). S.O.B – short of breath. L.O.C – loss of consciousness; 'Cognitive' refers to complaints of dysexecutive problems such as poor memory and difficulty concentrating; 'Weakness' refers to subjective perceptions of weakness (frequently associated with pain but not always associated with objective signs of weakness)



'RNZN 1A', was prescribed in a further 109 (21%) cases which were generally of a more serious nature. Divers with residual symptoms after the first recompression underwent further once-daily recompressions until there was either full recovery or no sustained improvement over two consecutive days. These follow-up treatments were conducted according to a shorter protocol specifying oxygen breathing at either 284 or 203 kPa for 60 or 90 minutes respectively. The mean number of re-treatments was 1.27.

DIVER OUTCOMES

At discharge, 438 divers (84%) were either without sequelae or with an expectation that minor subjective symptoms would resolve within one month. Though this was usually confirmed by follow-up, these latter cases were deemed to have experienced a complete recovery. Sixty-one (12%) patients were considered to have had an incomplete recovery. Outcome data were not recorded for 21 (4%) divers.

Discussion

We have described the caseload of DCI patients treated in Auckland between 1996 and 2012. The most striking feature of these data is the significant decline in annual case numbers that has occurred over the 17-year period. The mid- to late-1990s was characterised by high numbers of DCI cases treated in Auckland. Indeed, 100 cases were treated in 1995 though this cohort included patients from the South Island because the Christchurch chamber was not operational.[5] Whereas annual numbers above 50 were typical in the 1990s, these have dwindled to fewer than 30 in recent years. There are few published accounts of comparable data from other centres but it is notable that a similar decline in the numbers of divers recompressed for DCI in Australia also occurred

between 1995 and 2007.[6] The number of calls from within Australia to the Australian Diver Emergency Service hotline also declined between 1991 and 2007.[7] Thus, the decline in the number of DCI cases treated at Auckland is confluent with the Australian experience. The cause of this decline is unknown.

One potential explanation is that it reflects a regional decrease in diving activity, but the latter has not been measured and it would be difficult to do so.[6] We have reported annual numbers of entry-level certifications issued in New Zealand by the predominant diver training organisation as one plausible index of diving activity over an approximately corresponding period. There has been a significant decline in certification numbers (Figure 1). Similar data were provided by PADI to estimate the incidence rate of scuba diving fatalities for a previous New Zealand study.[8] Although this lends some strength to the hypothesis that the decline in DCI is owing (at least in part) to a decline in diving activity, the observation deserves cautious interpretation. There are other training organisations operating in New Zealand whose training numbers were not obtained, and the number of new certifications cannot be assumed to directly equate with diving activity because this could also be influenced by fluctuations in the activity of previously trained divers, due to factors such as changing economic conditions, or by changes in diving tourism activity.

Another potential explanation is that diving has become safer. The imposition of regulation and safety strategies can produce dramatic declines in DCI cases in high-risk populations, but it is debateable whether there were any pivotal positive influences on diver safety in New Zealand over the reference period.[9] One possibility might be that an increasing proportion of divers adopted the use of dive computers instead of tables for planning and controlling of their depth/time profiles. Computers have several potential advantages such as: ensuring the diver at least uses some means of controlling depth and time; the monitoring of ascent rates and provision of alarms when safe rates are exceeded and avoidance of the errors that divers frequently make when performing dive table calculations.[10] It is known that computer use has increased markedly over the last 20 years to such an extent that, whereas dive table instruction was previously mandatory, the PADI entry level course now offers the option of only learning to use a computer. Little can be deduced from our finding of a trend to increasing computer use among DCI patients without accurate data describing the relative use of computers and tables in the community. The apparent trend in our data probably reflects increasing use of dive computers in the community, and it is possible that dive computer users are actually under-represented in our cohort. Other plausible contributors to improved decompression safety over the audit period include the progressive inculcation of divers in the use of a 'safety stop' for 3 minutes at 5 metres' depth as a routine on every dive. Similarly, relevant educational initiatives such as the SAFE Dive (Slowly Ascend From Every Dive) campaign have become ubiquitous in the instructional and training pedagogy.

A third potential explanation is that fewer divers suffering symptoms of DCI are choosing to present for recompression treatment. This would seem unlikely in the face of serious manifestations, but divers with mild symptoms might invoke the findings of the 2004 remote DCI workshop to justify such a choice.[11] Specifically, a workshop consensus statement reads: "*The workshop acknowledges that some patients with mild symptoms and signs after diving can be treated adequately without recompression. For those with DCI, recovery may be slower in the absence of recompression.*"[11] We doubt this has had significant, if any impact on divers' choices in respect of seeking recompression in New Zealand. Awareness of the workshop's findings among divers is not widespread. In addition, the SHU policy of offering recompression to all local divers diagnosed with DCI has not changed. Moreover, this explanation is not consistent with our data which show that the trend to declining numbers was well established prior to publication of the workshop proceedings in 2005. Finally, if declining case numbers reflected an increasingly frequent choice not to present for mild symptoms, we would expect to see serious cases making up a greater proportion of the total. In fact, the proportion of cases (36%) with objective signs (which tend to be seen in the serious neurological events) is less in this series than the 45% recorded for the 100 cases treated in 1995.[5]

There were several other significant trends revealed by our data. First, the average age of divers treated for DCI increased over the study period. The most plausible explanation for this is that it simply reflects the demographic of the diving population. It is certainly possible that if fewer new divers are being trained (as the PADI data indicate) then a greater proportion of the total diving is being conducted by an aging population of established divers. Second, the median latency between incident dive and recompression also declined over the study period. It is more difficult to generate a plausible hypothesis to explain this trend. The most obvious (but entirely speculative) explanation would be that divers are becoming better educated, such that the diagnosis of DCI has become 'de-stigmatised' and, combined with better understanding on the potential benefits of timely treatment, this has resulted in earlier reporting of symptoms. In respect to evacuation for treatment and since first-aid oxygen can improve the early response to recompression, it was disappointing that in those cases where first-aid strategies were recorded, less than half received first-aid oxygen.

The clinical aspects of the cases in this series were confluent with those reported from a 1995 cohort treated at the SHU.[5] The most common symptoms were pain and patchy paraesthesiae, with objective signs in only 36% of cases. The choice of a higher pressure oxygen/helium table for

cases of greater perceived severity is consistent with practice among hyperbaric units in Australia.[12] Most cases (84%) were considered to have made a complete recovery. This was higher than for the 1995 SHU cohort (70%), but the difference is probably explained by changes in the definition of complete recovery at the point of discharge rather than a change in actual outcomes.[5] Over the period considered in the present study mild residual symptoms thought likely to resolve within a month were not considered in determining categorisation as 'incomplete' recovery.

This study has several weaknesses that should be acknowledged. The retrospective method placed reliance on the accuracy and completeness of data recorded in the patient notes and SHU patient database. In some cases, the notes were not available for reconciliation with the database, mandating total reliance on the database. Not surprisingly, there were some missing data. The retrospective design also precluded the accurate application of potentially useful severity scoring systems to individual cases which would have helped inform some of the preceding discussion.[13] Finally, many symptoms of DCI are non-specific and there is an undeniable potential for cautious practitioners to over-apply the diagnosis resulting in contamination of our dataset by non-DCI cases. Such conservative practice is widespread. The 'marginal' cases included in our dataset were recompressed and discharged with the diagnosis of DCI, and by definition they constitute part of the case load. They are, therefore, included in our report. Despite these limitations, our study describes one of the larger single-centre cohorts of DCI patients reported to date in Australia and New Zealand. In addition, the longitudinal design has facilitated identification of several interesting and potentially important trends in the number and nature of cases.

We conclude that the annual number of cases of DCI recompressed at Auckland has declined significantly over the past 17 years. A decrease in diving activity is the most plausible explanation, but other factors cannot be excluded.

References

1 Vann RD, Butler FK, Mitchell SJ, Moon RE. Decompression illness. *Lancet*. 2011;377:153-64.

2 Richardson K, Mitchell SJ, Davis M, Richards M. Decompression illness in New Zealand divers: the 1996 experience. *SPUMS Journal*. 1998;28:50-5.

3 Kenedi C, Sames C, Paice R. A systematic review of factitious decompression sickness. *Undersea Hyperb Med*. 2013;40:267-74.

4 Fitzgerald B. A review of the sharpened Romberg test in diving medicine. *SPUMS Journal*. 1996; 26:142-6.

5 Gardner M, Forbes C, Mitchell SJ. One hundred divers with DCI treated in New Zealand during 1995. *SPUMS Journal*. 1996;26:222-6.

6 Lippmann J. Review of scuba diving fatalities and decompression illness in Australia. *Diving Hyperb Med*. 2008;38:71-8.

7 Wilkinson D, Goble S. A review of 17 years of telephone calls to the Australian Diver Emergency Service (DES). *Diving Hyperb Med*. 2012;42:137-45.

8 Davis FM, Warner M, Ward B. Snorkelling and scuba deaths in New Zealand, 1980–2000. *SPUMS Journal*. 2002;32:70-80.

9 Xu W, Liu W, Huang G, Zou Z, Cai Z, Xu W. Decompression illness: clinical aspects of 5278 consecutive cases treated in a single hyperbaric unit. *PloS one*. 2012;7:e50079.

10 Wilks J, O'Hagan V. Queensland scuba divers and their tables. *SPUMS Journal*. 1991;21:12.

11 Mitchell SJ, Doolette DJ, Wacholz CJ, Vann RD, editors. *Management of mild or marginal decompression illness in remote locations*. Workshop Proceedings. Divers Alert Network. 2005: Sydney, Australia. p. 240.

12 Bennett MH, Mitchell SJ, Young D, King D. The use of deep tables in the treatment of decompression illness: the Hyperbaric Technicians and Nurses Association 2011 Workshop. *Diving Hyperb Med*. 2012;42:171-80.

13 Mitchell SJ. Severity scoring in decompression illness. *SPUMS Journal*. 2005;35:199-205.

Conflicts of interest: None

Acknowledgements

The authors would like to gratefully acknowledge Dr Matthew Pawley for providing guidance on statistical processes.

Submitted: 02 October 2013
Accepted: 03 January 2014

Haas Rachel M[1], Hannam Jacqueline A[2], Sames Christopher[3], Schmidt Robert[3], Tyson Andrew[4], Francombe Marion[3], Richardson Drew[5], Mitchell Simon J[2,3,6]

[1] School of Medicine, University of Auckland, New Zealand
[2] Department of Anaesthesiology, Faculty of Medical and Health Sciences, University of Auckland, Auckland, New Zealand
[3] Slark Hyperbaric Unit, Waitemata District Health Board, Auckland, New Zealand
[4] Hyperbaric Health, Quay Park, Auckland, New Zealand
[5] Professional Association of Diving Instructors, Rancho Santa Margarita, CA, USA
[6] Department of Anaesthesia, Auckland City Hospital, 2 Park Rd, Grafton, Auckland, New Zealand

Address for correspondence:
Associate Professor Simon Mitchell
Department of Anaesthesiology,
Faculty of Medical and Health Sciences,
University of Auckland,
Auckland, New Zealand
E-mail: *<sj.mitchell@auckland.ac.nz>*
Phone: *+64-(0)9-373-7599*
Fax: *+64-(0)9-373-7970*

Neuron-specific enolase and S100B protein levels in recreational scuba divers with neurological decompression sickness

Emmanuel Gempp, Pierre Louge, Sébastien De Maistre, Loïc Emile and Jean-Eric Blatteau

Abstract

(Gempp E, Louge P, De Maistre S, Emile L, Blatteau J-E. Neuron-specific enolase and S100B protein levels in recreational scuba divers with neurological decompression sickness. *Diving and Hyperbaric Medicine*. 2014 March;44(1):26-29.)

Introduction: Neuron-specific enolase (NSE) and S100B protein are brain-origin proteins commonly described to assess the presence and severity of neurological injury. To date, there are limited data examining the influence of scuba diving on these biomarkers, particularly when symptoms of decompression sickness (DCS) occur. The purpose of this controlled study was to determine whether these serum neurochemical markers could be used as 1) indicators of neurological DCS and 2) predictors of incomplete recovery.

Methods: Fifty-nine divers with neurological DCS and 37 asymptomatic divers admitted for inadequate decompression, serving as controls, were consecutively enrolled between 2010 and 2012. Blood samples were collected at initial presentation up to 6 hours after dive completion (controls) or onset of symptoms (DCS divers). Biomarkers were quantified in non-haemolysed samples only. Clinical outcome was assessed at 6 months post-injury.

Results: The two groups did not differ regarding the variables examined, except for the total dive time which was slightly shorter in the control group. NSE, but not S100B protein, was higher in the DCS group than in controls ($P < 0.0001$). An NSE level > 15.9 µg L^{-1} determined by ROC analysis predicted DCS development with a specificity of 100% (95% confidence interval (CI) 90 to 100) and a sensitivity of 24% (95% CI 14 to 36). There was a trend towards a higher likelihood of residual neurological deficits above this cut-off value ($P = 0.08$).

Conclusions: Early determination of NSE was found to be useful for the diagnosis of neurological DCS with a high specificity. However, its clinical applicability in decision making for determining treatment as well as its prognostic value remains to be established. Reliability of S100B protein was not demonstrated in the present study.

Key words

Decompression sickness, central nervous system, brain injury, proteins, severity, outcome, diving research

Introduction

Neurological decompression sickness (DCS) in scuba divers is a rare event with an incidence estimated between 0.02 and 0.03% per dive.[1] This disorder is the leading cause of morbidity with potential residual deficits of around 30% reported in the literature.[2,3] While there have been a number of clinical scoring systems devised for acute neurological DCS that have proved reliable for the prediction of incomplete recovery, little in the way of research into biological markers in humans has been conducted to test their value in diagnosing DCS and assessing prognosis.[4–6] Numerous studies have documented a variety of haematological and biochemical changes associated with decompression stress or the occurrence of DCS, but their utility as diagnostic tools has not yet been proven.[7–10] Particular attention has focused on the measurement of haematocrit, which has been noted to rise in severe cases of neurological DCS.[11] However, normal values have also been observed commonly in patients with a poor outcome, limiting the prognostic performance of this test in routine clinical use. Recent work also showed that elevated plasma D-dimer levels during the acute phase of neurological DCS was associated with the occurrence of sequelae at three months but the sensitivity of the test is still rather low.[12]

Neuron-specific enolase (NSE), a glycolytic enzyme predominantly localized in the cytoplasm of neurons and cells with neuro-endocrine differentiation, and S100B, a calcium-binding protein found in abundance in astroglial and Schwann cells, are commonly elevated during the acute phase of neurological damage after global cerebral ischaemia, stroke and traumatic brain injury.[13–15] The value of these neurochemical biomarkers in spinal cord injury is still unknown with very few investigations conducted in this field of study.[16]

To date, there are limited data examining the influence of scuba diving on these biomarkers, particularly when symptoms of DCS occur.[17–19] The purpose of this retrospective observational study in a large cohort of divers was to determine whether serum NSE and S100B protein levels could be used as 1) supplementary indicators to a clinical diagnosis of neurological DCS and 2) predictors of incomplete recovery.

Methods

The ethics committee of Saint Anne's Military Hospital approved the study, and all patients gave their informed consent. Between January 2010 and February 2012, 94 recreational divers with clinical signs of neurological DCS and 38 asymptomatic divers referred for inadequate decompression (i.e., fast ascent, omitted decompression

Figure 1
Flow diagram describing the selection of DCS divers

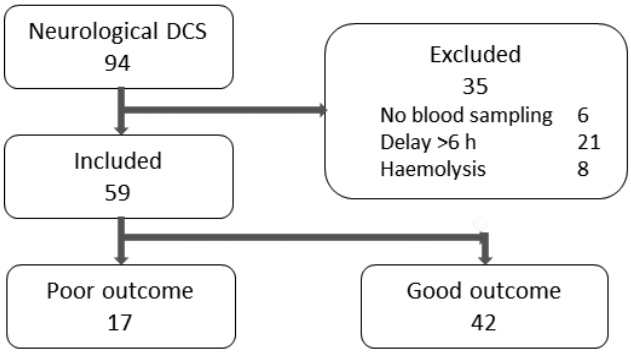


Table 1
Characteristics of DCS divers and control divers; mean (SD); * means $P < 0.05$

Characteristics	**DCS divers**	**Controls**	***P*-value**
	n = 59	*n* = 37	
Age (years)	46 (10)	49 (12)	0.16
Gender (M/F)	43/16	27/10	0.82
Mean depth (msw)	40.5 (10.5)	41.5 (12.5)	0.65
Mean dive time (min)	35 (10)	30 (14)	0.02*
Repetitive dive	12/59	7/3	0.9
Delay for blood collection (min)	170 (70)	156 (34)	0.27

stops), serving as controls, were admitted to our hyperbaric facility (Toulon, France). Cases suspected of cerebral arterial gas embolism, patients with incomplete data or those who presented more than 6 hours after onset of symptoms (DCS divers) or more than 6 hours after surfacing (controls) were excluded. Demographics, diving parameters and delay between blood collection and dive completion were recorded in each group. Clinical outcome was classified as poor (presence of residual neurological manifestations defined as persistent objective sensory, motor or urinary disorders) or good (full recovery) after clinical evaluation at six months post injury.

Venous blood samples were collected from all divers on initial presentation and drawn in dry and EDTA tubes (8 ml). Serum NSE and S100B were obtained by centrifugation (5000 rpm for 10 min at 4^{O}C) and stored at -80^{O}C until measurement of both biomarkers with commercially available electrochemicoluminescence immunoassay kits (Elecsys, Roche Diagnostics). All samples with visible haemolysis were discarded from analysis to avoid any falsely elevated values for NSE.

STATISTICAL ANALYSIS

Data were expressed as mean ± SD or median with range for nonparametric variables. Differences between groups were compared using the unpaired Student's t-test or the Mann-Whitney U test where appropriate. Correlations between continuous variables were evaluated by calculating Spearman's coefficient (ρ). Associations between categorical variables were measured by the Fisher exact test. A receiver operating characteristics (ROC) curve was performed to discriminate the highest measurement of NSE levels in predicting DCS development while specificity (Sp) and sensitivity (Se) were obtained with the use of predefined thresholds. Odds ratios with 95% confidence intervals (CI) were calculated when needed and *P*-values lower than 0.05 were considered significant. Statistical calculations were performed with Graphpad Prism 5.00 (GraphPad Software, San Diego, CA).

Results

Fifty-nine DCS divers and 37 controls (after exclusion of 1 diver with haemolysis) were eligible for this study (Figure 1). Both groups were similar regarding the variables examined, except for total dive time which was shorter in the control group compared with DCS divers (35 ± 10 min vs. 30 ± 14 min, $P = 0.02$, Table 1). There was no significant difference in the mean delay to collection of blood between the groups (170 ± 70 min vs. 156 ± 34 min for DCS divers and controls, respectively).

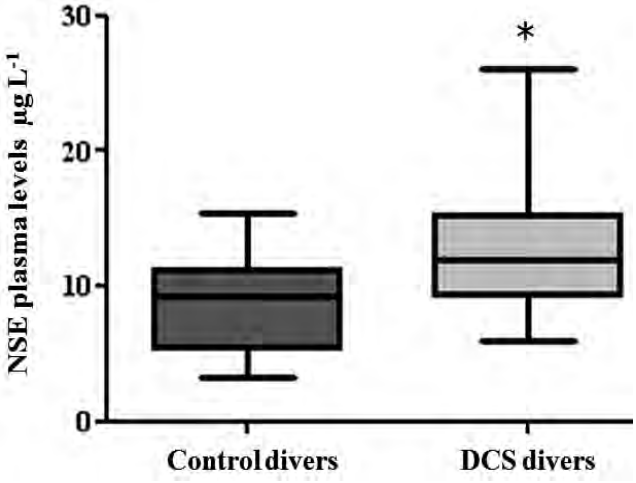
Of the 59 injured divers, 17 were found to have incomplete recovery after follow-up evaluation and were considered the severe group. Among them, four had disabling sequelae including urinary or bowel disturbance, ataxia due to sensory spinal myelopathy and mild degrees of limb spasticity. The remaining 42 (of 59) DCS divers did not exhibit neurological residual symptoms, and thus belonged to the benign group.

Serum NSE was higher in the DCS group than in controls (12.5 ± 4.3 µg L^{-1} vs. 8.8 ± 3.2 µg L^{-1}, $P < 0.0001$) (Figure 2). The level with the highest specificity and sensitivity was 12.1 µg L^{-1} (Sp = 89%, 95% CI 75 to 97; Se = 44%, 95% CI 32 to 58). A cut-off value of 15.9 µg L^{-1} predicted DCS development with a specificity of 100% (95% CI 90 to 100) and a sensitivity of 24% (95% CI 14 to 36).

The mean NSE level was significantly higher among patients in the severe group than those with a good outcome (14.5 ± 5.2 µg L^{-1} vs. 11.7 ± 3.6 µg L^{-1}; $P = 0.02$). However, association between NSE ≥ 15.9 µg L^{-1} and DCS severity did not reach statistical significance although there was a trend towards a poorer outcome above this cut-off value (OR = 3.5, 95% CI 0.99 to 2.3, $P = 0.08$).

There was no difference in the median S100B levels between injured divers and controls (0.087 ng L^{-1}, 95% CI 0.010 to 0.270 vs. 0.083 ng L^{-1}, 95% CI 0.045 to 0.260, respectively) or between severe DCS divers and those with

Figure 2
NSE concentrations in the serum of diver subgroups;
* $P < 0.0001$



benign evolution (0.081 ng L^{-1}, 95% CI 0.036 to 0.227 vs. 0.087 ng L^{-1}, 95% CI 0.010 to 0.272, respectively). In addition, there was no statistically significant correlation ($\rho = 0.08$) between NSE and S100B levels.

Discussion

To our knowledge, this is the first study investigating the concomitant use of NSE and S100B in divers with neurological DCS and comparing them to a control population. Our findings indicate that NSE, but not S100B, is elevated in the serum of divers presenting with neurological decompression sickness as compared to asymptomatic divers who had performed a dive with inadequate decompression. It appears that NSE is a specific biomarker which allows ruling in the diagnosis of neurological DCS with a very good reliability when the values exceed 12.1 µg L^{-1}. We also identified a cut-off value for NSE (> 15.9 µg L^{-1}) predicting the development of DCS with no false positives. However, the clinical usefulness of this test is hampered by its low sensitivity, meaning that a negative result does not necessarily rule out the occurrence of neurological DCS. In addition, the assay procedures make the clinical applicability difficult for the acute evaluation of the severity of DCS and consequently, for the choice of hyperbaric treatment regimen.

To our knowledge, there are only two reports assessing the influence of scuba diving on these two humoral indicators of neuronal damage.[17,18] Although no cumulative effect of repetitive dives on serum S100B levels was found in either study, there were small but significant post-dive increases in S100B concentrations in one study.[18] However, the concomitant rise in creatine kinase activity following each dive led the authors to suggest a skeletal muscle origin for this protein, as already observed after swimming.[19] In addition to S100B, NSE release did not seem be affected by four days of consecutive diving despite detection of significant amounts of vascular bubbles post dive.[18] These findings may indicate that uneventful no-decompression scuba dives do not cause discernable neuronal damage. On the other hand, a previous study in rats demonstrated a rise in serum S100B following simulated dives, with a strong correlation between S100B expression, bubble formation and/or the extent of hyperbaric exposure, suggesting a potential influence of decompression stress severity on alterations of the blood brain barrier.[20]

In a study of divers with neurological DCS, S100B also did not appear to be of clinical use in diagnosis as this marker did not increase over the next few days following the onset of symptoms.[21] Our data are in agreement with these findings, although the blood samples were drawn at different times, with an average time of less than 3 hours in the present study. To date, no clinical study has focused on the analysis of NSE concentration in DCS divers, hence making direct comparison with our findings difficult. Further research is warranted to evaluate the potential role of this biomarker in predicting outcome, in particular, after serial measurements over time since it has been reported that the release of NSE may reach a peak value at 48 to 72 h following acute ischaemic stroke.[22]

Conclusion

The present study reveals that plasma NSE concentrations in divers with neurological DCS exceed the levels found in control subjects who had performed dives with inadequate decompression. Our findings suggest that an increase of NSE level above a cut-off value of 15.9 µg L^{-1} measured early on admission appears to have a specificity of 100% but a low sensitivity for neurological DCS. The clinical relevance of this test in the acute assessment of divers with suspected neurological DCS remains to be established, considering the relatively long time needed to perform the biomarker analysis. The combined measurement of S100B with NSE does not add diagnostic or prognostic information, suggesting that damage of neurones is more involved in neurological DCS than glial alterations.

References

1 Vann RD, Butler FK, Mitchell SJ, Moon RE. Decompression illness. *Lancet.* 2011;377:154-64.
2 Gempp E, Blatteau JE. Risk factors and treatment outcome in scuba divers with spinal cord decompression sickness. *J Crit Care*. 2010;25:236-42.
3 Vann R, Freiberger JJ, Caruso JL, Denoble P, Pollock NW, Uguccioni DM, et al. *DAN report on decompression illness, diving fatalities and project dive exploration*. Durham, NC: Divers Alert Network; 2005. p. 63-5.
4 Blatteau JE, Gempp E, Simon O, Coulange M, Delafosse B, Souday V, et al. Prognostic factors of spinal cord decompression sickness in recreational diving: retrospective and multicentric analysis of 279 cases. *Neurocrit Care.* 2011;15:120-7.

5 Boussuges A, Thirion X, Blanc P, Molenat F, Sainty JM. Neurologic decompression illness: a gravity score. *Undersea Hyperb Med*. 1996;23:151-5.
6 Dick APK, Massey EW. Neurologic presentation of decompression sickness and air embolism in sport divers. *Neurology.* 1985;35:667-71.
7 Nyquist P, Ball R, Sheridan MJ. Complement levels before and after dives with a high risk of DCS. *Undersea Hyperb Med.* 2007;34:191-7.
8 Ersson A, Walles M, Ohlsson K, Hekholm A. Chronic hyperbaric exposure actives proinflammatory mediators in humans. *J Appl Physiol.* 2002;92:2375-80.
9 Pontier JM, Gempp E, Ignatescu M. Blood platelet-derived microparticule release and bubble formation after an open-sea air dive. *Appl Physiol Nut Metab*. 2012;37:888-92.
10 Philp RB. A review of blood changes associated with compression decompression: relationship to decompression sickness. *Undersea Biomedical Research*. 1974;1:117-50.
11 Boussuges A, Blanc P, Molenat F, Bergmann E, Sainty JM. Haemoconcentration in neurological decompression illness. *Int J Sports Med.* 1996;17:351-5.
12 Gempp E, Morin J, Louge P, Blatteau JE. Reliability of plasma D-dimers for predicting severe neurological decompression sickness in scuba divers. *Aviat Space Environ Med*. 2012;83:771-5.
13 Martens P, Raabe A, Johnsson P. Serum S100 and neuron-specific enolase for prediction of regaining consciousness after global cerebral ischemia. *Stroke*. 1998;29:2363-6.
14 Bloomfield SM, McKinney J, Smith L, Brisman J. Reliability of S100B in predicting severity of central nervous system injury. *Neurocrit Care*. 2007;6:121-38.
15 Anand N, Stead LG. Neuron specific enolase as a marker for acute ischemic stroke: a systematic review. *Cerebrovasc Disord*. 2005;20:213-9.
16 Pouw MH, Hosman AJF, Van Middendorp JJ, Verbeek MM, Vos PE, Van de Meent H. Biomarkers in spinal cord injury. *Spinal Cord*. 2009;47:519-25.
17 Stavrinou LC, Kalamatianos T, Stavrinou P, Papasilekas T, Koutsarnakis C, Psachoulia C, et al. Serum levels of S100B after recreational scuba diving. *Int J Sports Med*. 2011;32:912-5.
18 Bilopavlovic N, Marinovic J, Ljubkovic M, Obad A, Zanchi J, Pollock NW, et al. Effect of repetitive SCUBA diving on humoral markers of endothelial and cerebral nervous system integrity. *Eur J Appl Physiol.* 2013;113:1737-43.
19 Dietrich MO, Tort AB, Schaf DV, Farina M, Goncalves CA, Sousa DO, et al. Increase in serum S100B protein level after a swimming race. *Can J Appl Physiol*. 2003;28:710-6.
20 Havnes MB, Hjelde A, Brubbak AO, Møllerløkken A. S100B and its relation to intravascular bubbles following decompression. *Diving Hyperb Med*. 2010;40:210-2.
21 Poff DJ, Wong R, Bulsara M. Acute decompression illness and serum S100B levels: a prospective observational pilot study. *Undersea Hyperb Med*. 2007;34:359-67.
22 Brea D, Sobrino T, Blanco M, Cristobo I, Rodriguez-Gonzalez R, Rodriguez-Yanez M, et al. Temporal profile and clinical significance of serum neuron-specific enolase and S100 in ischemic and hemorrhagic stroke. *Clin Chem Lab Med*. 2009;47:1513-8.

Conflict of interest: None

Acknowledgments

The authors thank Mihaela Ignatescu, London Hyperbaric Centre, Whipps Cross Hospital, London, for reviewing the manuscript.

Submitted: 09 October 2013
Accepted: 20 December 2013

Emmanuel Gempp[1], Pierre Louge[1], Sébastien De Maistre[1], Loïc Emile[2], Jean-Eric Blatteau[3]

[1] Department of Diving and Hyperbaric Medicine, Ste Anne's Military Hospital, Toulon, France
[2] Department of Biochemistry, Ste Anne's Military Hospital, Toulon
[3] ERRSO, Institute of Biomedical Research of the Armed Forces Health Service, Toulon

Address for correspondence:
Emmanuel Gempp, MD
Department of Diving and Hyperbaric Medicine
Ste Anne's Military Hospital
83800 Toulon cedex 9, France.
Phone: *+33-(0)4-8316-2320*
E-mail: *<gempp@voila.fr>*