Original articles

Venous gas emboli detected by two-dimensional echocardiography are an imperfect surrogate endpoint for decompression sickness

David J Doolette

Abstract

(Doolette DJ. Venous gas emboli detected by two-dimensional echocardiography are an imperfect surrogate endpoint for decompression sickness. *Diving and Hyperbaric Medicine*. 2016 March;46(1):4-10.)

Introduction: In studies of decompression procedures, ultrasonically detected venous gas emboli (VGE) are commonly used as a surrogate endpoint for decompression sickness (DCS). However, VGE have not been rigorously validated as a surrogate endpoint for DCS.

Methods: A data set for validation of VGE as a surrogate endpoint for DCS was retrospectively assembled comprising maximum VGE grades measured using two-dimensional echocardiography and DCS outcome following 868 laboratory man-dives. Dives were conducted according to only ten different experimental interventions such that the ten cumulative incidences of DCS (0–22%) provide relatively precise point estimates of the probability of DCS, *P(DCS)*. Logistic models relating the *P(DCS)* to VGE grade and intervention were fitted to these validation data. Assessment of the models was used to evaluate the Prentice criteria for validating a surrogate endpoint.

Results: The *P(DCS)*) increased with increasing VGE grade. However, the difference in the *P(DCS)* between interventions was larger than explained by differences in VGE grades. Therefore, VGE grades did not largely capture the intervention effect on the true endpoint (DCS) in accord with the Prentice definition of a surrogate endpoint.

Conclusions: VGE can be used for comparisons of decompression procedures in samples of subjects but must be interpreted cautiously. A significant difference in VGE grade probably indicates a difference in the *P(DCS)*. However, failure to find a significant difference in VGE grades does not necessarily indicate no difference in *P(DCS)*.

Key words

Venous gas emboli; echocardiography; decompression sickness; decompression; diving; research; statistics

---

Introduction

A reduction in ambient pressure (decompression) can result in decompression sickness (DCS). The conventional approach to evaluating the efficacy of a new decompression procedure aimed at reducing the risk of DCS is to conduct a trial of the procedure with DCS as the endpoint. The incidence of DCS is necessarily kept low to protect subjects and so that the tested procedure is operationally relevant, and such trials require many man-dives. As with any clinical trial in which the true endpoint is rare, replacement of the true endpoint with a more frequently occurring surrogate endpoint has the potential to reduce the trial sample size.

DCS is thought to be caused by intracorporeal bubble formation.[1] Venous bubbles (venous gas emboli, VGE) can be detected by ultrasonic methods after dives, whether the dive results in DCS or not. The number of VGE is usually represented by an ordinal grade. The VGE grade in the mixed venous blood is presumed to be correlated with the risk of bubbles forming at, or impacting, sites where they will cause DCS.[1] The cumulative incidence of DCS does increase with increasing VGE grade in large compilations of data from decompression trials.[1,2] On the bases of these presumed and actual correlations, VGE grades are sometimes used as a surrogate endpoint for DCS in studies of decompression procedures.

Using inappropriate surrogate endpoints can lead to misleading results and prescription of improper interventions. Consequently, criteria have been developed for validating surrogate endpoints for clinical trials.[3,4] VGE have not been rigorously validated as a surrogate endpoint for DCS. This paper reviews the operational definition of a surrogate endpoint and examines whether VGE meet the criteria for a surrogate endpoint for DCS.

Methods

OPERATIONAL DEFINITION OF A SURROGATE ENDPOINT

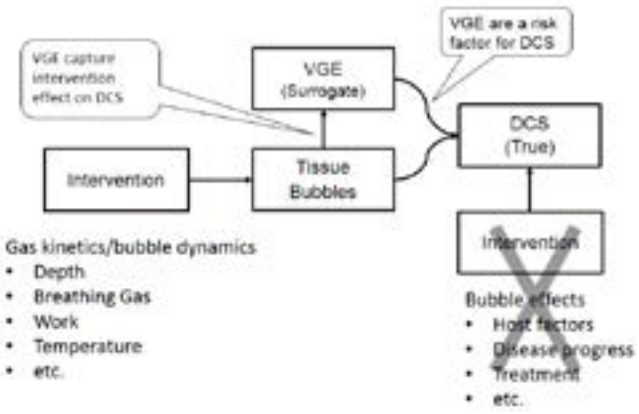
Prentice defined surrogate endpoints with respect to the effect of a particular intervention on the surrogate and true endpoints: for a specified intervention, the test of a null hypothesis on a surrogate endpoint is a valid test of the corresponding null hypothesis on the true endpoint.[3] This operational definition requires that a surrogate endpoint meet the following 'Prentice' criteria: 1) the surrogate endpoint captures the intervention effect on the true endpoint; and 2) the surrogate endpoint is prognostic for the true endpoint. These two Prentice criteria are expressed formally as:

$$P(DCS|VGE_i, X_j) = P(DCS|VGE_i) \quad [1]$$

and

$$P(DCS|VGE_i) \neq P(DCS) \quad [2]$$

Figure 1
Model of intervention, VGE as a surrogate endpoint, and DCS as the true endpoint; for interventions that act via tissue bubbles and result in a corresponding effect on both VGE and DCS, VGE may meet the operational definition for a surrogate for DCS; for interventions that do not act on tissue bubbles, illustrated with the large X, VGE do not meet the operational definition for a surrogate for DCS



respectively.[3] In the equations, *DCS* is the true endpoint which is binary (0,1); VGE_i is the surrogate endpoint which can be one of $i = 1...m$ ordinal VGE grades, and X_j is the j^{th} intervention.[3] This operational definition of a surrogate endpoint for the particular case of VGE as a surrogate endpoint for DCS is illustrated in Figure 1 which shows the intervention must have a corresponding effect on both VGE and DCS, and VGE must be prognostic for DCS.

Mechanistically, relevant interventions act via effects on tissue gas kinetics and bubble dynamics; VGE arise from tissue bubbles and centrally detected VGE numbers are presumed proportional to tissue bubble numbers; both VGE and tissue bubbles can cause manifestations of DCS. Interventions for which VGE would not meet the operational definition for a surrogate for DCS are interventions acting on processes 'downstream' of centrally-detected VGE, for instance on bubble-tissue complexes at DCS sites.

Equation [1] provides a link between the null hypothesis that the intervention has no effect on the true endpoint (DCS) and the null hypothesis that the intervention has no effect on the surrogate endpoint (VGE). Proof of this relationship has been given for failure rates and binary true endpoints.[3,4] This proof is reprised here for the specific case of a binary true endpoint (DCS) and an ordinal surrogate endpoint (VGE). Since the $i = 1...m$ *VGE* grades partition the sample space for *DCS*, a link between *DCS* and *VGE*, conditional on the intervention (X_j), can be obtained from the Law of Total Probability.

$$P(DCS|X_j) = \sum_{i=1}^{m} P(DCS|X_j, VGE_i)\, P(VGE_i|X_j) \quad [3]$$

The null hypotheses that the intervention has no effect on VGE is:

$$P(VGE_i|X_j) = P(VGE_i). \quad [4]$$

Substitution of Equations (1) and (4) into Equation (3) gives:

$$P(DCS|X_j) = \sum_{i=1}^{k} P(DCS|VGE_i)\, P(VGE_i) = P(DCS) \quad [5]$$

which is the null hypothesis that the intervention has no effect on DCS.

Equation [2] ensures that rejection of the null hypothesis on VGE (Eq. [4]) implies a rejection of the null hypothesis on DCS (Eq. [5]). Equations [1] and [2] provide guidelines for validating potential surrogate endpoints.

VALIDATION DATA

The data required to validate a surrogate endpoint are large numbers of observations with both the surrogate and true endpoints for any specified intervention. The Navy Experimental Diving Unit (NEDU) has measured VGE using two-dimensional (2-D) echocardiography as a secondary outcome measure following experimental dives in which DCS was used as the primary endpoint. Among these data are two decompression trials of single, air, decompression bounce dives that will be used as validation data.[5,6]

VGE were measured and graded in the same manner in two decompression trials.[5,6] With subjects in the left decubital position, the heart chambers were imaged (apical long-axis four-chamber view) using a 2-D echocardiograph (Siemens Medical Solutions® Acuson Cypress Portable Colorflow Ultrasound System). VGE in the right heart chambers were graded according to the ordinal scale defined in Table 1.

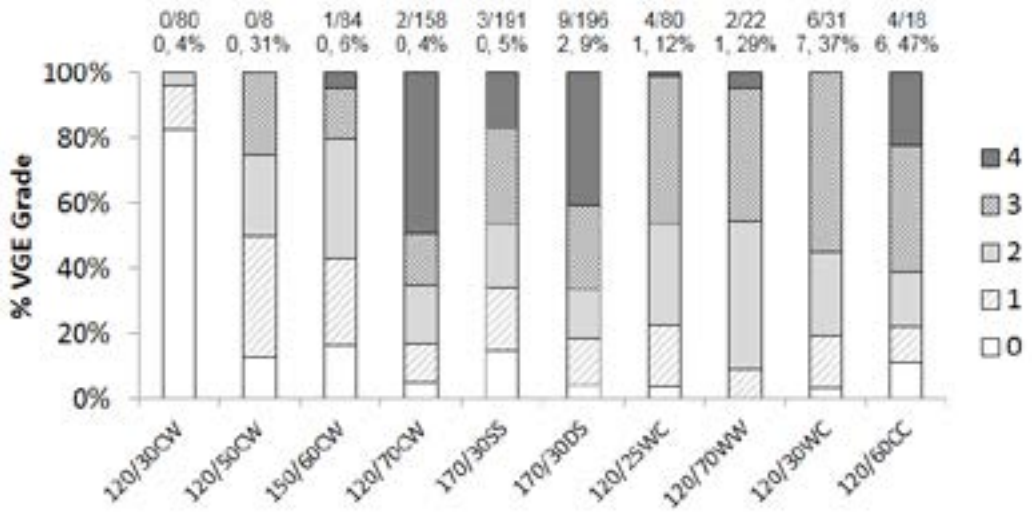
At each examination, VGE were measured five times: after the subject had been at rest for approximately one minute and then after forceful limb flexions around each elbow and knee. VGE were examined at about 30 minutes and two hours post dive in both trials. VGE were additionally examined at four hours post dive in the trial of diver thermal status (profiles with 'C' and 'W' labels in Figure 2). Only the maximum VGE grade observed at any time (rest or limb flexion, any examination) were used in this report, and will be referred to as 'VGE grade' without qualification. Maximum observed VGE grades have previously been shown to have the strongest relationship with cumulative incidence of DCS.[2] DCS was diagnosed by the duty Diving Medical Officer.

Table 1
NEDU 2-D Echocardiography VGE scale

Grade	Description
0	No bubble seen
1	Rare (< 1 per 5 heart beats), individual bubble seen
2	Several discrete bubbles visible in the same image
3	Multiple bubbles in most cardiac cycles, but not obscuring image
4	Bubbles in all cardiac cycles, bubbles dominate image and may blur chamber outlines

Figure 2

Summary of VGE grades and DCS for individual dive profiles (interventions); the stacked bars illustrate the percentage of man-dives in each dive profile that resulted in each VGE grade; the labels above the bars give the number of DCS cases / number of man dives and the 95% binomial confidence limits of the resulting estimate of *P(DCS)* as per cent; the labels along the horizontal axis identify the individual dive profiles in the original technical reports (see text for more details)



Individual VGE grades and descriptions of each DCS case are given in the original reports.[5,6]

Table 2 summarizes these data for single air decompression bounce dives.[5,6] These data are a unique resource for validating VGE as a surrogate endpoint for DCS because enough dives have been conducted on most dive profiles that the observed cumulative incidences of DCS provide credible point estimates of the probability of DCS, *P(DCS)*, for those dive profiles (Figure 2). Each distinct dive profile can be considered a distinct intervention (X) that modifies gas kinetics or bubble dynamics, as illustrated in Figure 1. All dive profiles were air decompression dives. The labels along the horizontal axis in Figure 2 identify the individual dive profiles in the original technical reports. For each dive profile the first part of the label gives the bottom depth in feet' sea water / bottom time in minutes. The two dives to 170/30 had 174 minutes of decompression stops in a shallow stops (SS) or deep stops (DS) distribution.[6]

The remaining dive profiles are all from a trial in which diver thermal status was manipulated independently for the bottom time and decompression.[5] For instance divers might be cold (C) on the bottom and warm (W) during decompression, indicated by 'CW'. All dives to 120 feet' sea water had 87 minutes of decompression stops. The dive to 150/60 had 110 minutes of decompression stops. These data exclude six man-dives for which VGE measurements were not available. Five of these were man-dives that resulted in onset of symptoms of DCS before VGE measurements were made. Two of these missing DCS cases are from 170/30DS, two from 120/70WW, and one from 120/30WC, resulting in cumulative incidences of DCS of 5.6%, 17%, and 22%, respectively. VGE data were lost for one man-dive which did not result in DCS from dive profile 170/30SS; this had little effect on the cumulative incidence of DCS.

Table 2

NEDU 2-D echocardiography VGE data for single bounce dives; DCS - decompression sickness; CL - confidence limits

Grade	# Dives	# DCS	% DCS	95% CL
0	134	0	0	(0, 2)
1	141	2	1	(0, 5)
2	178	4	2	(1, 6)
3	215	15	7	(4, 11)
4	200	10	5	(2, 9)
Total	868	31	4	(2, 5)

Validation

The two Prentice criteria, Equations [1] and [2], are assessed by first fitting the following nested logistic regressions models to these validation data.[4]

$$\ln\left(\frac{P(DCS)}{1-P(DCS)}\right) = \beta_0 + \beta_1 VGE_i + \beta_2 X_j + \beta_3 VGE_i \times X_j \quad [6]$$

$$\ln\left(\frac{P(DCS)}{1-P(DCS)}\right) = \beta_0 + \beta_1 VGE_i + \beta_2 X_j \quad [7]$$

$$\ln\left(\frac{P(DCS)}{1-P(DCS)}\right) = \beta_0 + \beta_1 VGE_i \quad [8]$$

If VGE contributes significantly to the fit of these models to the validation data, Eq. [2] is satisfied. If the intervention factor X or the interaction of X with VGE contribute significantly to the fit of models [6] or [7] to the validation data, Eq. [1] is not satisfied. Failure to find a significant contribution of X does not prove that Eq. [1], which is a

null hypothesis, is satisfied. Equation [1] implies that the surrogate fully captures the intervention effect on the true endpoint. Realistically, a surrogate endpoint will capture a proportion of the intervention effect on the true endpoint. This proportion can be assessed by fitting the model [9], which has the intervention as the only independent variable,

$$\ln\left(\frac{P(DCS)}{1-P(DCS)}\right) = \alpha_0 + \alpha_2 X_j \qquad [9]$$

and model [7], which includes the intervention and the surrogate, to the same validation data. The proportion of the intervention effect explained by including the surrogate endpoint in the model is assessed as the proportional decrease in the estimated coefficient for the intervention, $(\alpha_2 - ß_2)/\alpha_2$, where α_2 is the unadjusted coefficient for the intervention in model [9] and β_2 is the coefficient for the intervention factor adjusted for the effect of VGE in model [7].[4]

The coefficient vectors ($\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$) of the logistic models [6], [7], [8], and [9], as well as a null model in which *logit(DCS)* equals a constant (β_0), were estimated by fit to the data illustrated in Figure 2. The X_j were dummy coded and VGE grades were treated as interval data. Similar results, which are not presented, were obtained if VGE grades were grouped into zero, low (grades 1 and 2) and high (3 and 4) grades, or if the five VGE grades were linearized to values of 0, 0.1, 0.4, 2, and 10.[1] The contributions of the variables *VGE* and *X* to the fit were assessed by comparing the log-likelihood of nested models. The log-likelihood of the full model (LL_f) with p_f adjustable coefficients and the log-likelihood of the reduced model (LL_r) with p_r adjustable coefficients were compared by using the likelihood ratio test with goodness-of-fit statistic $G = -2(LL_r, -LL_f)$. If $P(\chi^2 > G) \leq 0.05$, $df = p_f - p_r$, the extra factors in the full model were considered to contribute significantly to the fit. Models were fitted to different subsets of the data.

Results

The logistic models were fitted to several subsets of the data, starting first with the eight dive profiles with non-zero cumulative incidence of DCS. All data from the two dive profiles with zero cumulative incidence of DCS were excluded to avoid the numerical problems that arise with fitting to data with covariate patterns (e.g., dive profiles) that have zero or 100% occurrence of a binary outcome (e.g., DCS). The likelihood ratio tests are shown in Table 3. The interaction of *X* with *VGE* did not contribute significantly to the fit of this model to this data subset or any subsequent data subsets investigated, as indicated by no significant difference between model [6] with the interaction term and model [7] without the term. The intervention factor *X* did contribute significantly to the model fit to this data subset as indicated by the significantly improved fit of model [7] with this factor over model [8] without this factor. This is evidence that Eq. [1] is not satisfied for this data subset. VGE grades contributed significantly to explaining the *P(DCS)* as indicated by the significant improvement of model [8] over the null model; therefore, Eq. [2] is satisfied for this data subset.

Table 3
Likelihood ratio tests of logistic models fit to data subsets; a,b; b,c; c,d – models compared

Model	LL		df	$P(\chi^2>G)$	
X_j: all dive profiles with DCS >0					
[6]	-105	a	16	0.1877	a,b
[7]	-110	b	9	< 0.0001	b,c
[8]	-126	c	2	0.0040	c,d
Null	-130	d	1		
X_j: 150/60CW, 120/70CW,170/30SS, 170/30DS, 120/25WC,120/70WW					
[6]	-82	a	12	0.0869	a,b
[7]	-86	b	7	0.0813	b,c
[8]	-91	c	2	0.0047	c,d
Null	-95	d	1		
X_j: LR: (150/60CW, 120/70CW,170/30SS); HR: (170/30DS, 120/25WC,120/70WW)					
[6]	-88	a	4	0.4540	a,b
[7]	-88	b	3	0.0055	b,c
[8]	-91	c	2	0.0007	c,d
Null	-95	d	1		
X_j: 170/30SS, 170/30DS					
[6]	-50	a	4	0.2618	a,b
[7]	-51	b	3	0.1947	b,c
[8]	-52	c	2	0.0572	c,d
Null	-53	d	1		

The fit of model [7] to the eight dive profiles with non-zero cumulative incidence of DCS produced significant Wald statistics, (not shown), for the X_j corresponding to 120/30WC and 120/60CC, the two dive profiles with the highest observed cumulative incidences of DCS. This finding indicates that the VGE grade alone does not explain the cumulative incidence of DCS on these dive profiles. The reason for this is apparent by examining the fitted *P(DCS)* from model [8] in which VGE are the only independent variable. These fitted values of the *P(DCS)* range from 1.1% for VGE grade 0 to 6.8% for VGE grade 4. Thus, the highest possible cumulative incidence of DCS estimated by model [8] is 6.8%, for a dive profile that results in grade 4 VGE after every dive. This latter value is a ceiling imposed by the data, and is obvious from inspection of Table 2 in which 7% is the highest cumulative incidence of DCS associated with any VGE grade. Dive profiles 120/30WC and 120/60CC have observed cumulative incidence of DCS much higher than can be predicted by any model based on VGE grade alone.

The logistic models were next fitted to a data subset that omitted dive profiles 120/30WC and 120/60CC, as well as the two dive profiles with zero cumulative incidence of DCS. In this case, the intervention factor (*X*) did not contribute

Table 4
Model coefficient estimates and proportion of intervention effect explained for two data subsets

Model	Variable	Coeff.	Estimate	S.E.	$(\alpha_2 - \beta_2)\ \alpha_2$
[7]	Intercept	β_0	-5.7703	0.8289	
	VGE	β_1	0.5457	0.2292	
	X: HR	β_2	1.2359	0.4912	
[9]	Intercept	α_0	-4.2650	0.4111	
	X: HR	α_2	1.3276	0.4891	0.0690
[7]	Intercept	β_0	-5.1030	0.9734	
	VGE	β_1	0.3940	0.2846	
	X: 170/30DS	β_2	0.8551	0.6941	
[9]	Intercept	α_0	-4.1431	0.5819	
	X: 170/30DS	α_2	1.1093	0.6746	0.2291

significantly to the model fit to this data subset (Table 3). Therefore, there is insufficient evidence to reject Eq. [1] for this data subset. However, because the *P*-value of the likelihood ratio test was only 0.0813, it was investigated if the lack of significance was due to the number of degrees of freedom associated with the six levels of the intervention factor. The six X_j were recoded into two levels, LR and HR, indicating all dive profiles with cumulative incidence of DCS less than 2% and greater than 2%, respectively. The cumulative incidences of DCS in the resulting LR and HR groups were 1.4% and 5.0%, respectively. The recoded intervention factor (*X*) did contribute significantly to the model fit to these data, indicating Eq. [1] is not satisfied for this recoded data subset.

Finally, the logistic models were fitted to a data subset comprising only dive profiles 170/30SS and 170/30DS. These dive profiles have the most precise estimated *P(DCS)* and the same ultrasound operator graded the VGE on all the dives. The intervention factor did not contribute significantly to the fit of model [7] to this data subset; however, model [8] just failed to reach significance compared to the null model.

The proportion of the intervention effect on the *P(DCS)* explained by VGE grade was assessed in the two data subsets in which the factor *X* has only two levels. Table 4 shows the estimates of the coefficients for models [7] and [9] and the proportion of the intervention effect explained by VGE. These proportions were quite small even for the data set for which there was insufficient evidence to reject the first Prentice criterion. The reference level of the intervention factor *X* was the group with the lower cumulative incidence of DCS, so the estimated coefficient is the effect of being in the group with higher cumulative incidence of DCS. In model [7], the estimated coefficients for *X*, adjusted for VGE, are positive for both data subsets, indicating a greater increase in *P(DCS)* than can be explained by the increase in VGE grade.

Discussion

The 'gold standard' data showing increasing cumulative incidence of DCS with increasing VGE grades following diving is the compilation of data arising from the development of the DCIEM decompression tables.[1,2] Those VGE data are Kisman-Masurel grades determined from the bubble noises in ultrasonic Doppler flow transducer signals. The present NEDU data are the first to show a similar association between cumulative incidence of DCS and VGE grades measured using 2-D echocardiography following diving. The present NEDU data are the only published data suitable for assessing the first Prentice criterion, and therefore validating VGE as a surrogate for DCS. In most of the subsets of the data examined the first Prentice criterion was rejected because differences in VGE grades only explained a small proportion of the differences in *P(DCS)* between dive profiles. This has implications for the interpretation of experimental findings arising from using VGE as a surrogate endpoint for DCS.

Inspection of Table 2 shows that detecting no VGE is strongly negatively predictive of DCS, but there is no VGE grade that has both good sensitivity and specificity for DCS, and it is well known that VGE are not a surrogate for DCS in the individual diver.[5,7] Nevertheless, the increasing cumulative incidence of DCS with increasing VGE grades, consistent with the second Prentice criterion, can allow comparison of decompression procedures in sufficiently large samples of divers.[8]

If a significant difference in the distribution of VGE grades is found between decompression procedures, this likely indicates a difference in the *P(DCS)* between the procedures. This is particularly true if there is a difference in the distribution of VGE grades among zero, low (grades 1 and 2) and high (3 and 4) grades, since there are substantive differences in cumulative incidences of DCS between these groups of VGE grades.[8] However, the difference in the *P(DCS)* may be greater than indicated by differences in VGE grades. Therefore, difference in VGE grades can be used to rank decompression procedures in order of relative *P(DCS)* but not to reliably quantify the difference in *P(DCS)*. The range of *P(DCS)* that can be estimated from VGE grades is the range of cumulative incidences of DCS associated with those grades, as illustrated here and described previously.[9] Therefore all *P(DCS)* estimated from the present VGE data must be compressed into the range 0–7% shown in Table 2.

For several reasons, failure to detected difference in VGE grades between decompression procedures is insufficient evidence to retain the null hypothesis of no difference in *P(DCS)*. First, VGE data cannot be used to distinguish differences between procedures which have *P(DCS)* outside the range of cumulative incidence of DCS associated with those grades. The maximum cumulative incidence of DCS associated with any 2-D echocardiographic VGE grade in the present data is 7%, and the maximum cumulative incidence

of DCS associated with any Doppler-detected VGE grade in the DCIEM air diving data set is about 10%.[1,2] Therefore it is not possible to distinguish between decompression procedures with actual *P(DCS)* above about 10%. This ceiling may be of little consequence for normal exposure diving; however *P(DCS)* at or above 7–10% is associated with higher risk, but operationally relevant, military, exceptional-exposure diving or DISSUB procedures.[10,11]

Second, it has been shown that for decompression procedures with *P(DCS)* in the range that is potentially distinguishable by VGE, 80% power to detect one-grade differences in VGE requires a paired comparison of about 50 subjects.[8] Smaller sample sizes may fail to detect a one-grade differences in VGE grades that can indicate a difference in *P(DCS)*. Finally, because VGE grades only capture a small proportion of the intervention effect, even between decompression procedures with *P(DCS)* less than 7–10%, an operationally relevant difference in *P(DCS)* may exist between procedures that does not manifest as a difference in VGE grades.

The implications of the present analysis are relevant to the other commonly used methods to detect and grade VGE because 1) there is good agreement between VGE scores measured using 2-D echocardiography and Doppler[12], and 2) the present analysis is based on the NEDU VGE scale that is broadly similar to other VGE scales. With respect to this latter point, the five NEDU VGE grades (0–4) were designed to be similar to the five grades in the Spencer and Kisman-Masurel Doppler scales.[1,5,11] The NEDU VGE grades differ slightly from those of the more widely used Eftedal-Brubakk scale for grading VGE in 2-D echocardiography images.[1,12] The principal difference is that the NEDU grade 4 covers what is described as grades 4b, 4c, and 5 in the recently proposed expanded version of the Eftedal-Brubakk scale.[13]

A strength of the Eftedal-Brubakk scale is that the grades are unambiguously defined, which facilitates inter-rater reliability. However, it is worth noting that a Medline search for studies that graded VGE in 2-D echocardiographic images after diving in humans found 13 papers reporting 12 dive trials identifying a total of 384 man-dives. All these trials use the original or expanded Eftedal-Brubakk scale. Thus, fewer man-dives have been evaluated using alternative 2-D echocardiography VGE scales than the number of man-dives reported here and evaluated using the NEDU VGE scale.

The present findings should be relatively broadly applicable to studies using VGE as an endpoint, but caution should be used when extrapolating findings from one set of experimental conditions to another, First, the present dives had higher *P(DCS)* than studies that use VGE to evaluate decompression procedures intended not to result in DCS. The present data are from experiments which used DCS as the primary endpoint and the tested decompression procedures were designed to have a measurable DCS incidence. The procedures were designed with predicted *P(DCS)* approaching the maximum for normal exposure military diving, and the actual observed incidences occasionally exceeded those predicted. However, the present data show a similar distribution of cumulative incidence of DCS among VGE grades as the larger DCIEM data set for air decompression dives, a data set which has an overall cumulative incidence of DCS half that of the present data.[1,2] This suggests that the present analysis is relevant to evaluation of decompression procedures with lower *P(DCS)* than are in the present data set.

Second, the present data are exclusively from wet working dives, but show a similar distribution of cumulative incidence of DCS among VGE grades as the DCIEM data set which includes both wet, working, and dry, resting dives.[1,2] This suggests the present findings are applicable to a range of diving conditions.

Finally, many of the decompression procedures in the present data involved manipulation of diver thermal status.[5] These manipulations presumably modified the *P*(*DCS*) by modifying tissue blood flow and, consequently, tissue gas kinetics and bubble dynamics. Therefore, manipulation of thermal status is a suitable intervention for using VGE as a surrogate endpoint, in accord with the model illustrated in Figure 1. The present data should be relevant to interventions that manipulate gas kinetics and bubble dynamics by other methods.

The present data have some limitations. First, these data were assembled from dive trials not designed for the present analysis, and any such retrospective analysis must acknowledge the possibility of confounding factors. Second, five dives which resulted in DCS (14% of the total DCS cases) were excluded because of the onset of DCS symptoms before VGE measurements could be made. This may have influenced the distribution of DCS with VGE grades in Table 2. Such data loss is inevitable in decompression trials.

Finally, although VGE measurements were made throughout the period during which the maximum VGE grade typically occurs following long bounce dives, the measurements were made relatively infrequently following each dive, and it is possible that maximum VGE grade was not always captured.[14] It is uncertain that this influenced the overall distribution of DCS with VGE grades and interventions, since, presumably, the likelihood of missing the maximum VGE grade was no different for any intervention or outcome in this large data set.

Conclusions

VGE grades are in imperfect surrogate endpoint for DCS and data using VGE must be interpreted cautiously. VGE cannot be used to diagnose DCS but can be used for comparisons of decompression procedures in samples of subjects. Whereas a significant difference in VGE grade probably indicates a difference in the *P(DCS)*, failure to find a significant

difference in VGE grades does not necessarily indicate no difference in *P(DCS)*.

References

1 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.

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

3 Prentice RL. Surrogate endpoints in clincal trials: definition and operation criteria. *Stat Med*. 1989;8:431-40.

4 Freedman LS, Graubard BI, Schatzkin A. Statistical validation of intermediate endpoints for chronic diseases. *Stat Med*. 1992;11:167-78.

5 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 from: htpp://archive.rubicon-foundation.org/5063 [cited 2015 December 29].

6 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 from: htpp://archive.rubicon-foundation.org/10269 [cited 2015 December 29].

7 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.

8 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 Hyperb Med*. 2014;44:14-9.

9 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.

10 Gerth WA, Doolette DJ. *VVal-18 and VVal-18M Thalmann algorithm air decompression tables and procedures*. Technical Report. Panama City (FL): Navy Experimental Diving Unit; 2007 May. Report No.: NEDU TR 07-09. Available from: htpp://archive.rubicon-foundation.org/8349 [cited 2015 December 29].

11 Latson GW, Flynn ET, Gerth WA, Thalmann ED. *Accelerated decompression using oxygen for submarine rescue - summary report and operational guidance*. Technical Report. Panama City (FL): Navy Experimental Diving Unit; 2000 Dec. Report No.: NEDU TR 11-00. Available from: htpp://archive.rubicon-foundation.org/3582 [cited 2015 December 29].

12 Brubakk AO, Eftedal O. Comparison of three different ultrasonic methods for quantification of intravascular gas bubbles. *Undersea Hyperb Med*. 2001;28:131-6.

13 Ljubkovic M, Dujic Z, Møllerløkken A, Bakovic D, Obad A, Breskovic T, et al. Venous and arterial bubbles at rest after no-decompression air dives. *Med Sci Sports Exerc*. 2011;43:990-5.

14 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.

Acknowledgements

This work was supported by the Office of Naval Research project grant No.: N0001414WX20596. Aspects of this work were supported by Naval Sea Systems Command Deep Submergence Biomedical Research Program. The views presented are those of the author and do not necessarily represent the views of the U.S. Department of the Navy. This work was presented at the 2015 International Meeting on Ultrasound for Diving Research, Karlskrona, Sweden.

Conflict of interest: nil

Submitted: 17 October 2015; revised 28 December 2015
Accepted: 01 January 2016

David J Doolette, Navy Experimental Diving Unit, Panama City, Florida, USA

Address for correspondence:
David J. Doolette
Research Physiologist
Navy Experimental Diving Unit
321 Bullfinch Road
Panama City, FL 32407, USA
Phone: *+001-(0)1850-230-3179*
E-mail: *<david.doolette.as@navy.mil>*