

# Factors affecting the precision of age determination of sperm whales (*Physeter macrocephalus*)

KAREN EVANS\*, MARK A. HINDELL\*, KELLY ROBERTSON#, CHRISTINA LOCKYER+ AND DALE RICE\*\*

Contact e-mail: Karen.Evans@csiro.au

## ABSTRACT

Teeth from 92 sperm whales were prepared by etching for age determination. The total number of growth layer groups (GLGs) in the dentine of each tooth was determined from three to five reading sessions by a single reader. Four other readers, as part of a cross-reading experiment, read a subset of these teeth ( $n = 5$ ). This study investigated: (1) intra- and (2) inter-reader precision in GLG counts; (3) possible variation in growth structure deposition between different teeth within the same individual; (4) the use of photographs to identify and count GLGs and the effect of this technique on the precision of counts; and (5) mineralisation anomalies in tooth sections and the possible effects these may have on GLG count precision. Intra- and inter-reader precision was determined using coefficients of variation (CV) and indices of precision (D). Total numbers of GLGs estimated from individual teeth ranged from 0.75–64 ( $\bar{x} = 32.8$ ,  $n = 92$ ). Intra-reader mean CV was 10.6 and mean D was 4.8. Inter-reader mean CV ranged from 4.8–12.3 and mean D ranged from 2.8–7.1. Differences in final counts between readers appeared to be the result of differing interpretation of GLGs and this was the largest factor affecting the precision of GLG counts. While GLG counts between teeth in the same individual varied, it is possible that this variation was due to within reader variation rather than variation in the development of growth structures, but establishment of this cause is confounded by differential tooth wear. Use of photographs increased the definition of growth structures, decreasing the variation between GLG counts within reading sessions. The incidence of mineralisation anomalies and the closure of the pulp cavity increased with increasing GLG counts in individuals, but were not consistent between teeth from the same individual. These factors, while potentially affecting the accuracy of GLG counts in relation to age estimates, had little effect on the precision of GLG counts. The lack of an ability to validate age estimates in this species and the large inter-reader variation seen in this study suggests that age estimates based on GLG counts in this species are subjective and can only be regarded as relative. High-quality photographs of tooth sections should be used to verify GLG counts with other readers, resulting in 'consensus counts' generated by a number of readers, ensuring interpretation of the same structures and confidence in comparing GLG counts produced in different studies.

KEYWORDS: AGE DETERMINATION; SPERM WHALE; AUSTRALASIA; SOUTHERN HEMISPHERE; STRANDINGS

## INTRODUCTION

The determination of the age of animals is important in establishing the life history traits of individuals and populations. Integral to this is the development of an accurate age determination technique and the minimisation of any associated biases.

Growth layer groups (GLGs) in the teeth of sperm whales (*Physeter macrocephalus*) have been used to determine the age of individuals since the 1950s (Nishiwaki *et al.*, 1958; Ohsumi *et al.*, 1963; Gambell, 1977; Rice *et al.*, 1986). However, validation of the assumption that these GLGs are annual depositions, as is the case in most other marine mammals, has proven difficult. Validation techniques such as the use of 'known-age' individuals (Hohn *et al.*, 1989; Hohn, 1990) and tetracycline marking experiments (Myrick *et al.*, 1984; 1988; Brodie *et al.*, 1990) used in other species have not been feasible in sperm whales because of their size and the inability to keep captive individuals. Only limited mark-recapture studies investigating the accumulation rate of growth layers and studies calibrating seasonal changes in the thickness of the most recently formed dentine layer have been conducted on this species. These studies suggest that GLGs are deposited annually (Ohsumi *et al.*, 1963; IWC, 1967; 1971; Best, 1969) and as a result, studies involving the age determination of this species assume that each GLG represents one year's growth (Ohsumi, 1971; 1977; Lockyer, 1980; Rice *et al.*, 1986).

Another important concern associated with the aging of individual animals is that of the precision of counts of GLGs and, therefore age estimates (that is, the closeness of repeated GLG counts for the same individual). If final age estimates are the result of averaging the GLG counts from a number of reading sessions, the precision of GLG counts may have a major effect on the accuracy (the nearness of the final age estimate or GLG count to the actual age or number of GLGs) of the final estimate. As age increases, the pulp cavity in the tooth of a sperm whale fills in as a result of the deposition of further layers of dentine and eventually closes. Once the cavity is closed, the most recently deposited layers become compacted and are subsequently hard to discern. Mineralisation anomalies and dentinal resorption (Myrick, 1988; Lockyer, 1993) may also confuse the distinctiveness of GLGs, particularly in the recently deposited dentine of older animals, which may already be compromised by the closure of the pulp cavity. A number of publications have addressed variation in the accuracy of age determination from cetacean teeth associated with the preparation and reading techniques used (Anas, 1970; Hui, 1980; Hohn *et al.*, 1989; Hohn, 1990; Hohn and Fernandez, 1999). However, very few have addressed the problem of variation in precision (Donovan *et al.*, 1982; Mikhalev, 1982; Reilly *et al.*, 1983).

Variation in the number of GLGs in different teeth from the same individual may also be another source of bias in age determinations. Nishiwaki *et al.* (1958) found that teeth from

\* Antarctic Wildlife Research Unit, School of Zoology, University of Tasmania, GPO Box 252-05, Hobart, Tasmania, Australia, 7001. (Present address: Tropical and Pelagic Ecosystems, CSIRO Marine Research, GPO Box 1538, Hobart, Tasmania 7001, Australia).

# Southwest Fisheries Science Center, National Marine Fisheries Service, NOAA, PO Box 271, La Jolla, California, 92038, USA.

+ Danish Institute for Fisheries Research, Department of Fisheries, Charlottenlund Slot, Charlottenlund, DK2920, Denmark.

\*\* National Marine Mammal Laboratory, National Marine Fisheries Service, NOAA, 7600 Sand Point Way, Seattle, Washington, 98115, USA.

both the mandibular and maxillary jaw from the same individual in sperm whales contained similar numbers of growth layers. Conversely, the Workshop on Age Determination of Odontocete Cetaceans and Sirenians found that the number of GLGs varied between different teeth from one individual (Perrin and Myrick, 1980). However, the dataset was not large enough to test this statistically and it was recommended that the number of GLGs in complete series of teeth from both the mandibular and maxillary jaw of a number of individuals of varying ages be assessed. Bottlenose dolphins (*Tursiops truncatus*) were found to contain different numbers of GLGs in different teeth from the same individual (Hui, 1980), possibly because teeth in the anterior of the jaw ceased depositing dentine after 10 to 12 GLGs and posterior teeth ceased deposition of dentine at any time after 15 GLGs. However, both Myrick (1988) and Lockyer (1993) found that different teeth from the same individuals in spinner (*Stenella longirostris*), pantropical spotted (*Stenella attenuata*), common (*Delphinus delphis*) and bottlenose dolphins and long-finned pilot whales (*Globicephala melas*) showed similar growth patterns yielding similar age estimates (counts in *G. melas* differed by 0 to 4 GLGs).

Three mass strandings of sperm whales on the west and northwest coasts of Tasmania, Australia in 1998 provided material with which these problems could be investigated. This paper presents the results of investigations into variations in age estimates (1) within and (2) between readers; (3) between different reading methods; (4) between different teeth derived from the same individual; and (5) in relation to tooth morphology.

## MATERIALS AND METHODS

### Preparation of teeth

Near-complete or mid-sections of lower jaws with teeth were collected from 92 sperm whales involved in three mass strandings on the north and west coasts of Tasmania in 1998 (STR1: Ocean Beach, Strahan,  $n=56$ ; STR2: Greens Pt Beach, Marrawah,  $n=29$ ; STR3: Black River Beach, Stanley,  $n=7$ ). Other aspects of these strandings are reported in Evans *et al.* (2002). The least worn and straightest first or anterior-most mandibular tooth from each individual was sectioned along the bucco-lingual plane, and one half-section polished and then etched in 15% formic acid until clear, easily discernible dentinal layers or growth layer groups (GLGs) were produced. Teeth from calves were thin-sectioned, stained and mounted on microscope slides. Details of these methods are given in Evans and Robertson (2001).

### Age determination

The total number of GLGs in each of the 92 tooth sections was determined three times per session in three ( $n=3$ ), four ( $n=7$ ) or five ( $n=82$ ) sessions by a single reader (KE). The number of reading sessions was determined by the variability of GLG counts. For those teeth for which counts were not repeatable or at least two of the three counts were not close (within  $\pm 2$  GLGs), counts were repeated an additional one or two times. Time intervals between the sessions varied from seven to 92 days. Each reading was made without reference to previous readings or additional information on individuals (e.g. size, sex) and teeth were read in random order during each session.

Growth layer groups were interpreted as those identified in the report of the Workshop on Age Determination of Odontocete Cetaceans and Sirenians (Perrin and Myrick,

1980) as 'a repeating or semi-repeating pattern of adjacent groups of incremental growth layers within the dentine which is defined as a countable unit involving a change from a ridge to groove' in the case of etched teeth and 'intensely stained to lightly stained' in the case of thin-sectioned, stained teeth. For those specimens in which the neonatal line could be identified (many of the teeth had worn tips and therefore were missing the enamel, neonatal line and the first few GLGs), this was not included in the total number of GLGs.

The final age estimate for each individual was determined as either the most repeated GLG count (all session estimates pooled) or where there was no repeatability of counts ( $n=28$ ) the mean of all counts. It was assumed that skill in reading GLGs and estimating age increased with reader experience. To determine whether this had an effect on counts, a two-way ANOVA (with session number and tooth section as independent variables) was conducted on GLG counts to determine whether counts differed significantly between sessions.

### Assessment of intra-reader variation

For each tooth the standard deviation was calculated from all counts (the three counts from each of the three, four or five sessions all pooled). Following Chang (1982) and Reilly *et al.* (1983), the coefficient of variation (the standard deviation as a fraction of the mean expressed as a percentage:  $SD \times 100/\bar{x}$ ) and an index of precision (the percent error contributed by each observation to the average age class:  $D = CV/\sqrt{n}$ ) were calculated. The CV and D were plotted against GLG counts to determine if there was any effect of the number of GLGs (and therefore age) on the precision of counts.

### Assessment of inter-reader variation

A sub-sample containing five of the original 92 tooth sections and an associated photograph of each were supplied to four additional readers for comparative GLG determination. All readers had previous experience in counting GLGs from sperm whale teeth. No information on the animal or the stranding from which it was derived was supplied to the readers. Readers were supplied with a standard data form and were requested to estimate the number of GLGs directly from each tooth at least three times with a minimum of five to seven days between reading sessions and without reference to previous readings. Each reader was also requested to mark on the associated photograph of each tooth what they had interpreted and counted as GLGs, in an effort to establish areas in which variation, if it existed, occurred.

Individual reader CV and D were calculated for each tooth to quantify individual reader precision. Actual counts from individual teeth were compared between readers using a two-way ANOVA (with reader and tooth section as independent variables). Where this test revealed that there were significant differences in GLG counts between readers, the relevant photographs on which each reader had mapped their interpretation of GLGs were studied and any differences in the definition of GLGs noted.

### Assessment of counts from different teeth from the same animal

For seven whales, an additional 13 teeth were prepared for age estimation (providing a total of seven teeth from each side of the jaw and a total of 14 for each animal). These animals were selected randomly from a subset of the original that contained animals from which more than seven teeth on

each side of the jaw had been collected. The teeth selected were dependent on the number of teeth collected from the jaw. Where more than seven teeth from either jaw were collected, teeth were selected evenly along the length of the jaw. In all cases teeth from matching positions on both sides of the jaw were used. The number of GLGs in each tooth was estimated using the methods detailed above without reference to other teeth from each individual. GLG counts derived from teeth on the left and right sides of the jaw in an individual were compared for differences using a paired t-test.

To determine whether the numbers of GLGs in the 14 teeth of an individual were significantly different, an ANOVA with a Tukey HSD pairwise comparison was used. For each tooth the standard deviation, CV and D were calculated from all counts. To determine whether the number of GLGs did in fact vary between teeth in each animal, it was necessary to separate that variation associated with the reader from true differences in the number of GLGs present in each tooth. D values for each tooth from an animal were plotted with the mean D calculated from the assessment of within-reader variation (the mean overall D). Where D values for each tooth were lower than the mean overall D, any variation in GLG counts were regarded as true variation in the number of GLGs. Where D values for each tooth were the same or higher than the mean overall D, variation in GLG counts were regarded as a factor of reader variation. A one-way t-test was used to test for the presence of such differences.

**Assessment of direct tooth counts vs photo counts**

All teeth prepared (for both estimates of age of individuals in each stranding and for comparative counts of different teeth from the same individual) were digitally photographed ( $n = 171$ ). The number of GLGs in each tooth section was determined from these images. A paired t-test was used to compare GLG counts derived from photos against final GLG counts derived from direct readings. To determine if there was any difference in the precision of counts between this method and that from counts taken directly from teeth, a sub-sample of 50 randomly selected tooth images were read a further two times (for a total of three readings). Both CV and D were calculated for each method and then compared.

**Tooth morphology**

The number of pulp stones, the presence of mineralisation interferences (occlusions), and the state of the pulp cavity (whether it was open or closed) were determined for each tooth section. Anomalies were classified according to Lockyer (1993). The CV and D calculated during age determination were log-regressed against the state of the tooth cavity and against the presence of pulp stones and regressed against the number of pulp stones to determine whether tooth morphology factors effected CV and D.

**RESULTS**

**Assessment of intra-reader variation**

GLG counts from sperm whales in this study ranged from 0.75 to 64 GLGs (mean =  $32.8 \pm 13.2$ ,  $n = 92$ ). There were no significant differences among GLG counts estimated in the five sessions (ANOVA,  $F_{4,91} = 0.9$ ,  $P = 0.5$ ). For those estimates where there was no consensus of GLG counts between sessions, 89.3% contained estimates that differed by one GLG and 96.4% contained estimates that differed by two GLGs.

The mean CV was  $10.6 \pm 6.3$  and mean D was 4.8. There was no significant relationship between CV or D and the number of GLGs (Regression, CV:  $r^2 = 0.001$ ,  $F_{1,90} = 0.1$ ,  $P = 0.7$ ; D:  $r^2 = 0.004$ ,  $F_{1,90} = 0.4$ ,  $P = 0.6$ ; Fig. 1).

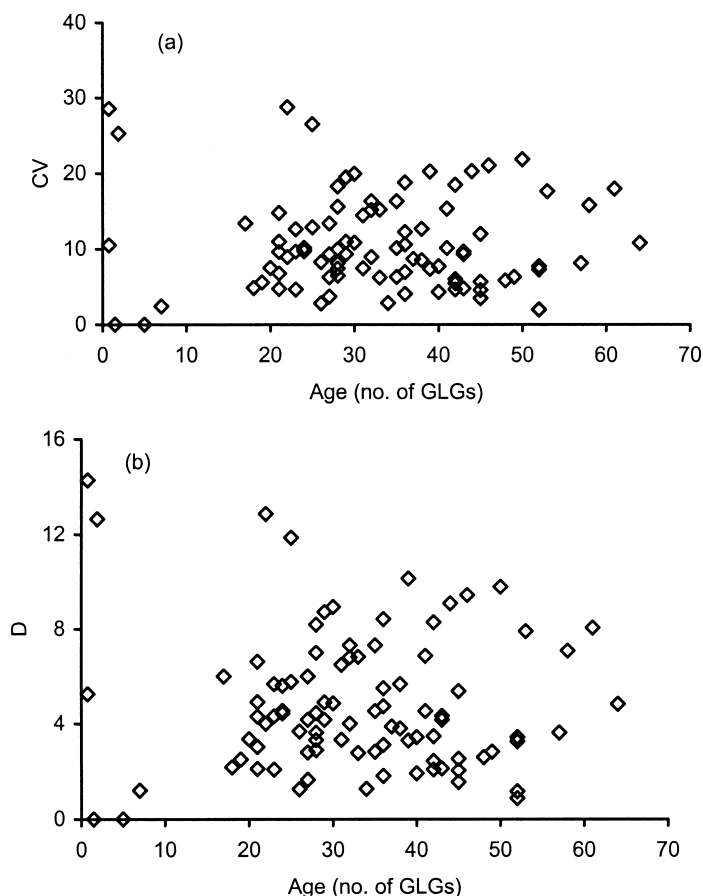


Fig. 1. CV and D calculated from estimated number of GLGs in teeth from sperm whales ( $n = 92$ ): (a) CV; (b) D.

**Assessment of inter-reader variation**

The difference in the number of GLGs estimated for each of the five teeth ranged considerably between readers, from one to 21 GLGs (means: 5.0-11.8 GLGs; Tables 1 and 2), increasing with teeth from older animals (Fig. 2). GLG counts were found to be significantly different between readers (ANOVA,  $F_{4,16} = 2.2$   $P = 0.02$ ). Mean CV ranged from 4.8-12.3 and mean D ranged from 2.8-7.1 across readers.

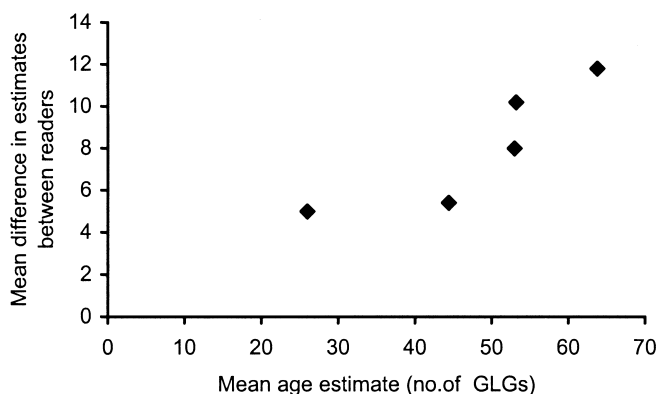


Fig. 2. Average difference in the estimates of the number of GLGs between five readers and the average age estimated for five sperm whale teeth.

Table 1

Estimated numbers of GLGs by five readers from three reading sessions of five sperm whale teeth.

Tooth	Number of estimated GLGs														
	Reader 1			Reader 2			Reader 3			Reader 4			Reader 5		
001	41	39	44	24	39	39	52	54	46	54	44	48	41	43	42
002	53	51	51	43	47	49	64	57	71	64	62	58	43	43	43
003	24	21	24	23	27	20	35	33	32	29	28	29	21	23	21
004	39	54	47	50	44	56	65	56	60	61	65	62	44	51	47
005	65	68	68	63	60	61	69	65	80	75	70	70	45	53	52

Table 2

Final estimates of GLG numbers for five sperm whale teeth by five readers and mean standard deviation, mean CV and mean D.

Tooth	Final estimate of the number of GLGs				
	Reader 1	Reader 2	Reader 3	Reader 4	Reader 5
001	41	39	51	49	42
002	52	46	64	61	43
003	24	23	33	29	21
004	45	50	60	63	47
005	65	61	71	70	52
Mean SD	3.2	4.6	5.0	2.7	2.0
Mean CV	7.6	12.3	8.4	4.9	4.8
Mean D	2.8	7.1	4.9	2.9	2.8

**Assessment of counts from different teeth derived from the same individual**

No significant differences were found in the number of GLGs, the CV or D between teeth in the left and right jaws of any of the individuals in the dataset. However, significant differences were found between GLG counts from teeth in different positions along the tooth row within an individual in six of the seven animals (Tables 3 and 4). When D values calculated from readings of each tooth in each individual were compared with the mean overall D, significant differences were found in only one individual (t-test,  $t_{13} = 4.9$ ,  $P < 0.001$ ; Fig. 3). GLG counts from different teeth in this animal ranged from 18-33. GLG counts varied increasingly with age (Fig. 4).

Table 3

Minimum, maximum and mean number of GLGs, mean SD, CV and D estimated from multiple teeth ( $n=14$ ) of seven sperm whales at STR2.

	Whale number						
	1	4	8	16	21	25	27
Min. no. GLGs	19	17	28	18	20	25	39
Max. no. GLGs	52	25	43	37	32	43	64
Mean no. GLGs	30.9	20.6	34.2	27.8	25.3	33.4	44.8
Mean SD	3.5	2.9	4.0	3.6	3.7	4.0	3.6
Mean CV	11.9	13.4	11.2	13.3	14.6	13.0	8.4
Mean D	5.4	6.0	5.1	6.0	6.5	6.1	3.7

Table 4

Results of ANOVAs conducted on GLG counts from teeth ( $n=14$ ) in the jaws of seven sperm whales at STR2.

	Whale number						
	1	4	8	16	21	25	27
df	13	13	13	13	13	13	13
F-ratio	19.6	1.7	3.0	9.6	2.1	10.8	7.6
P	<0.001	0.09	0.002	<0.001	0.03	<0.001	<0.001

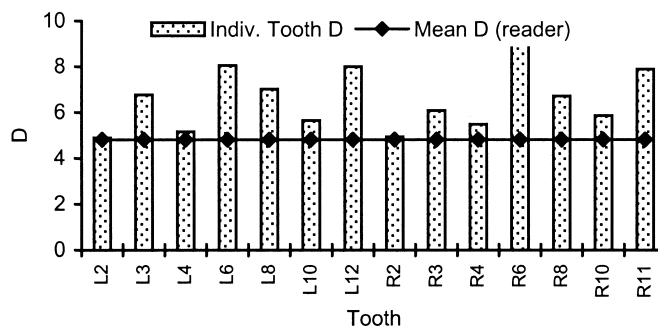


Fig. 3. D values for teeth ( $n=14$ ) in STR2(21) and the mean overall reader D.

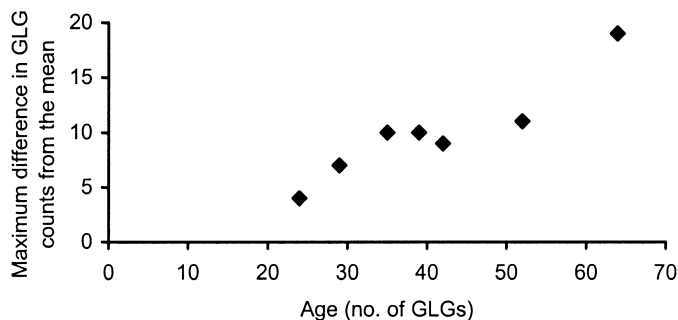


Fig. 4. Maximum difference in age estimates from the mean of age estimates derived from sets of 14 teeth prepared and examined in seven sperm whales.

**Assessment of direct tooth counts vs. photo counts**

GLG counts derived from photographs were only available for four of the five readers. Counts derived from photographs of individual teeth were significantly higher ( $\bar{x} = 36.1 \pm 10.7$  GLGs) than those derived from direct examination ( $\bar{x} = 32.9 \pm 9.5$ ) of teeth (t-test,  $t_{170} = 9.8$ ,  $P < 0.001$ ). The mean difference in GLG counts between these two methods was 3.2 GLGs.

The mean CV of the subset of photographs that were read several times was  $8.0 \pm 3.9$  and the mean D was  $4.6 \pm 2.3$ , while the mean CV derived from direct counts of these teeth was  $11.8 \pm 5.7$  and the mean D  $5.4 \pm 2.6$ .

**Tooth morphology**

The mean number of GLGs in teeth with pulp stones was  $36.2 \pm 11.5$  (range: 5-64 GLGs,  $n = 67$ ) and mean number of pulp stones present was  $6.4 \pm 9.7$  (range: 0-53,  $n = 92$ ). Both the presence and number of pulp stones in tooth sections were significantly related to the number of GLGs (Presence: log-regression,  $t_1 = 3.6$ ,  $P < 0.001$ ; Number: regression,  $r^2 = 0.04$ ,  $F_{1,90} = 4.4$ ,  $P = 0.04$ ; Fig. 5a). In those individuals where multiple teeth were examined, neither the presence nor the numbers of pulp stones were constant throughout different teeth (Fig. 6). The maximum range in pulp stone number between teeth in an individual was 0-32. There was no significant relationship between either CV or D and pulp stone presence or number (Figs 5b and 5c).

The incidence of a closed pulp cavity is related to increasing age. GLG counts from animals in which the tooth examined had an open pulp cavity were significantly lower ( $26.4 \pm 10.4$ ) than those of animals in which the tooth examined had a closed ( $45.5 \pm 8.0$ ) pulp cavity (t-test,  $t_{30} = -8.4$ ,  $P < 0.001$ ). In six of the seven individuals where multiple teeth were examined, the state of the pulp cavity was not consistent along the tooth row; instead each contained a mixture of teeth with open cavities and closed

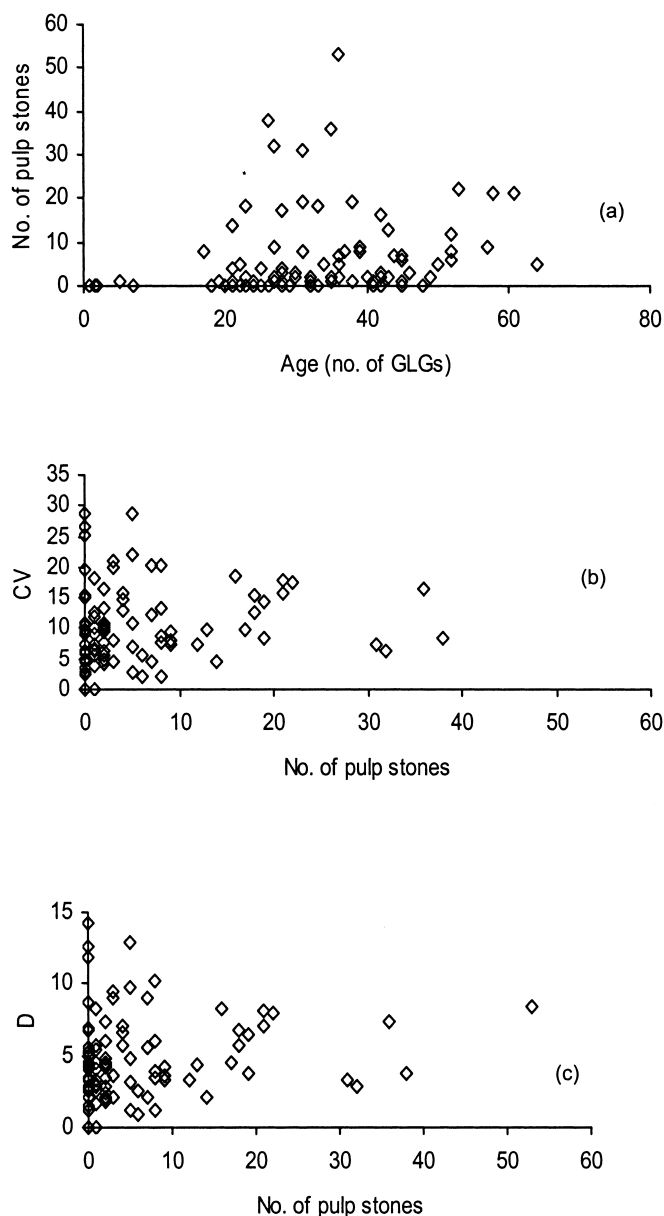


Fig. 5. Number of pulp stones present and the estimated age, CV and D calculated from sperm whale tooth sections ( $n = 92$ ): (a) age; (b) CV; (c) D.

cavities (e.g. Fig. 6). The number of teeth in the tooth row with closed pulp cavities increased with the number of GLGs. CV and D were not significantly related to the closure of the pulp cavity in tooth sections (log-regression, CV:  $t_1 = 0.004$ ,  $P = 0.1$ ; D:  $t_1 = 0.01$ ,  $P = 0.1$ ).

The mean number of GLGs in teeth containing occlusions was  $41.0 \pm 8.6$  (range: 30-61,  $n = 11$ ). Occlusions were not common throughout teeth from the same individual. Of two individuals where multiple teeth were examined and occlusions were present, one of 14 teeth contained an occlusion in one and four out of 14 teeth contained an occlusion in the other.

## DISCUSSION

### Assessment of intra- and inter-reader variation

Average CV and D calculated for intra-reader variation are similar to those presented in Reilly *et al.* (1983) for pantropical spotted dolphins and suggest that GLG counts from this dataset were relatively precise. Unlike in other

(a) L1



(b) L11



Fig. 6. Differences in mineralisation anomalies and state of pulp cavity between two teeth from STR2(25): (a) L1; (b) L11.

studies (Doubleday and Bowen, 1980; Reilly *et al.*, 1983; Bjørge *et al.*, 1995), this degree of precision did not decrease with increasing animal age. GLG counts did not appear to vary across reading sessions either, with no significant differences between session estimates. This suggests, at least in this study, that the precision of GLG counts was relatively constant throughout the age determination exercise.

However, GLG counts and average CV and D varied substantially between readers (by up to 21 GLGs) and this variation increased with increasing GLG number, although again, values for CV (4.8-12.3) and D (2.8-7.1) were similar to or lower than those calculated in other studies. Mean D values in Reilly *et al.* (1983) ranged between 2.8 and 6.6, while those in Chang (1982) ranged from 3.4-9.8. This variation has been found to increase with increasing specimen age in a number of other cetacean species (Reilly *et al.*, 1983; Bjørge *et al.*, 1995; Hohn and Fernandez, 1999) and is due to a decreasing ability to interpret growth structures in older animals. The deposition of growth layers becomes more highly compacted as the pulp cavity area fills in and its size decreases, making it harder to discern individual GLGs from one another.

None of the readings from direct examination of teeth coincided between readers, although in all at least two of the readings varied by less than three GLG. IWC (1969) reported that the average deviations from the mean of the age estimate ranged from +4.5 to -3.1 for 11 readers examining



the same teeth, although estimates for eight of the 11 readers were  $\pm 1$  GLG. Donovan *et al.* (1982) reported a significant difference (using Friedman's test) between the 'best' estimates of six readers when reading 50 etched teeth but no significant difference was found when four of the six readers (from the same 'school' of reading) were compared. When the age estimates of the remaining two readers were compared to the others, the average deviations from the mean were +1.42 and -1.76<sup>1</sup>. Mikhalev (1982) added the age estimates of two further readers experienced in reading sperm whale teeth to the results of Donovan *et al.* (1982) and observed further variation. However, he also noted that the average difference between the 'extreme' readers was only 3.2 GLGs; the maximum was 10 GLGs (in two teeth). These results all reveal a degree of subjectivity in the interpretation of growth layers in teeth. The implications of this subjectivity, particularly relevant for inter-study comparisons, depend on the use to which the age data are put and the way in which the 'best' estimate is arrived at. In this regard, the question of how to deal with worn teeth is important.

Examination of the associated photographs for the cross-reading experiment also highlighted this subjectivity, demonstrating that differences in GLG counts were due to differences in the interpretation of GLGs (i.e. what were regarded as accessory layers by one reader were regarded as GLGs by another; Fig. 7). This has substantial implications when comparing age estimates between studies. Attempts to standardise the definition and interpretation of GLGs in age determination studies were made during the International Whaling Commission's Workshop on Age Determination in Cetaceans and Sirenians. While the report of this workshop was published (Perrin and Myrick, 1980) and a number of

papers (Nishiwaki *et al.*, 1958; Ohsumi *et al.*, 1963; Best, 1969; Scheffer and Myrick, 1980) have provided photographs of sectioned teeth illustrating GLGs (as defined by the authors), no quantitative and objective method to assist researchers in the laboratory has yet been published. Definitions of GLGs depend, as a result, on the interpretation

(a) Reader 1. Estimated number of GLGs: 57.



Fig. 7a. Growth layer groups in a tooth from SPW2(25) as interpreted by three readers.

(b) Reader 2. Estimated number of GLGs: 50.



Fig. 7b.

(c) Reader 3. Estimated number of GLGs: 61.



Fig. 7c.

<sup>1</sup> For comparisons with present study, the mean CV for all readers was 7.55 and mean D was 1.26; for the four readers from the same school the values were 4.97 and 0.83, respectively (Donovan, pers. comm.).

of the individual or the laboratory at which age estimates are being determined and are therefore qualitative and subjective.

#### **Assessment of counts from different teeth from the same individual**

GLG counts from different teeth from individual whales differed significantly in six out of seven specimens. However, of these, the mean D generated from each set of 14 teeth was not significantly different from the mean D of the reader in all but one case, suggesting that variation in GLG counts between teeth was less likely to be the result of differences in the growth structures between teeth and may have been associated with intra-reader variation. The variation in GLG counts from teeth in the individual for which tooth D values differed significantly from the mean reader D could also be explained similarly. The mean D value generated from this individual was 6.5, a value higher than that of the overall mean reader D (4.81). However, attributing these differences to variation in the precision of GLG counts between teeth is confounded by the effects of differential tooth wear. Hui (1980) found that anterior teeth in bottlenose dolphins yielded lower numbers of GLGs than posterior teeth. Rather than these teeth containing varying growth structures, differential use and therefore differential wearing of these teeth may have resulted in varying GLG counts. If differential wear does have a significant impact on the determination of GLG counts in different teeth from the same individual, determining whether growth structure variation does occur between teeth becomes difficult.

Reducing intra-reader variation and thereby increasing the precision of GLG counts as well as devising some means by which tooth wear could be quantified would assist in establishing the source of this variation in GLG counts between teeth from the same individual. A larger formal internationally organised trial should be considered to quantify such factors and establish means by which these can be calibrated across studies.

#### **Assessment of direct tooth counts vs photo counts**

Differences of up to 21 GLGs between readers have serious implications for the validity of comparative studies, particularly in this long-lived species where there are no real means of verifying ages (e.g. via known-age animals or tetracycline experiments). However, the use of photographic techniques in the determination of GLG counts may serve to reduce this variation. Both the overall variation in estimates relative to the mean (CV) and the error contributed by each observation (D) decreased in GLG counts derived from the multiple readings of photographs in comparison to those derived from direct counts. The higher counts produced by all readers using photographs may be the result of two factors: (1) less confusion in interpreting between GLGs and accessory layers or (2) greater clarity of and contrast in growth structures causing accessory layers to appear as substantial a growth structure as GLGs. When counting GLGs, the reader must make a decision as to whether a growth structure is a GLG or an accessory layer. Readers may either be cautious, only interpreting the most clear structures as GLGs (and thereby perhaps underestimating the true number of GLGs) or may interpret most growth structures as GLGs (possibly including the clearest and most highly contrasted accessory layers as GLGs). This interpretation is highly subjective, but the fact that all readers counts increased while the individual reader CV values

decreased when using photographs suggests that the photographs resulted in the same effect on reader interpretation of growth structures and overall increased reader precision.

Hohn (1980) found that in comparing the use of polarised light, microradiography and scanning electron microscopy in age estimation techniques, scanning electron microscopy provided images in which GLGs were easiest to read. This was attributed to the higher contrast in topographic relief between the layers of each GLG. Bow and Purdy (1966) also found that the use of photographs of etched teeth increased contrast and maximised shadow detail between growth layer groups with the end result of decreasing errors in counts. While only the effect of the use of high quality photographs on GLG counts was studied here, other photographic methods such as the use of 3-D stereographic techniques should be considered in efforts to increase the clarity of and the contrast between individual GLGs and between GLGs and accessory layers, thereby increasing reader precision.

#### **Tooth morphology**

Mineralisation anomalies such as pulp stones and occlusion events have been documented in cetaceans on numerous occasions (Klevezal and Myrick, 1984; Myrick, 1988; Lockyer, 1993; 1995), but no assessment has been made on the effect of such anomalies on age estimation. Pulp stones are discrete events within the dentine of tooth sections, in most instances having little effect on the appearance of GLGs. Large pulp stones can bend GLGs, or may obscure that part of the GLG situated in the area of the pulp stone. Regardless of pulp stone size, GLGs can still be identified in the dentine of tooth sections. As a result, it would be expected that such events would have little effect on the precision (as found here) or the accuracy of GLG counts. Occlusions however, may obscure GLGs by disrupting lamina formation to the extent that they are no longer clearly defined. This may not affect the precision of GLG counts, since the same number of laminae actually defined within and outside the mineralisation interference area can be identified. However, such events have implications for the accuracy of GLG counts, especially in older animals in which both the incidence and the number of mineralisation anomalies are higher. Similarly, closure of the pulp cavity and the subsequent compacting and obscuring of GLGs is less likely to affect the precision of GLG counts (as found here), but is likely to affect the accuracy of GLG counts.

Even for the same individual, the presence and extent of mineralisation anomalies and the closure of the pulp cavity in differing teeth can be highly variable. Pulp stones form in the pulp and may not necessarily be incorporated into the dentine, or may spend varying amounts of time in the pulp before deposition in the dentine (Lockyer, 1993). As a result, varying numbers and positions of pulp stones in teeth from the same individual, as in this study are likely to occur (Fig. 6). If possible, rather than collecting a particular tooth from the jaw of an animal, several teeth should be collected and age estimates determined from the tooth with the least wear, the most highly defined growth layer groups, the minimum extent of mineralisation anomalies and if possible with an open pulp cavity (ensuring that GLGs have not become obscured with the closure of the cavity).

GLG counts and as a result, age estimates in this species are determined by an individual reader's interpretation of growth structures in tooth sections. Therefore, the largest factor affecting the precision of age estimates of individual animals is inter-reader variation in this interpretation. GLG

counts generated by a single reader can only be regarded as relative and comparable within a study, because any error introduced by the reader can be assumed to be relatively consistent across all estimates. However, large inter-reader variation compromises the ability to compare GLG counts and therefore age estimates between studies, especially when no indication of the precision of those age estimates is given. While there are currently no accurate means of determining the number of annual growth layers in this species, attempts should be made to increase the precision of age estimates, both within and between studies, and to devise more objective means by which GLG counts and therefore age estimates can be generated. The use of high quality photographs or other photographic techniques enabling clearer definition of GLGs should be investigated further as they may assist by increasing both intra- and inter-reader precision. Such photographic techniques could be used to verify GLG counts with other readers, ensuring interpretation of the same structures and facilitating 'consensus counts' generated by a number of readers, thereby increasing confidence in comparing age estimates between studies. Further studies investigating possible variability in growth structures between teeth from individuals and those enabling the separation of the effects of reader variability and the effects of differential wear in teeth on this variability should be initiated. Greater collaboration between investigators working on studies requiring age estimation of this species should be encouraged and is essential if standardisation of growth structure interpretation is to be achieved.

## ACKNOWLEDGEMENTS

We thank Margie Morrice, Deborah Thiele, John van den Hoff, Graham McKenzie and all the volunteers that helped with the collection of samples. The Parks and Wildlife Service, Tasmania, particularly Hans Wapstra and the School of Zoology, University of Tasmania are acknowledged for their help with logistics. The Australian Antarctic Division, particularly Harry Burton and Andrew Fleming and the Southwest Fisheries Science Center, California, USA, particularly Susan Chivers, kindly provided laboratory space and equipment for the preparation of samples. John Bannister and William Perrin provided valuable comments on earlier versions of the manuscript. This study was conducted under the Parks and Wildlife Service, Tasmania permit 97/16 and was funded by Environment Australia.

## REFERENCES

Anas, R.E. 1970. Accuracy in assigning ages to fur seals. *J. Wildl. Manage.* 34(4):844-52.

Best, P.B. 1969. The sperm whale (*Physeter catodon*) off the west coast of South Africa. 5 Age, growth and mortality. *Investl Rep. Div. Sea Fish. S. Afr.* 79:1-27.

Björge, A., Hohn, A.A., Kvam, T., Lockyer, C., Schweder, T. and Aarefjord, H. 1995. Summary Report from the Harbour Porpoise Age Determination Workshop, Oslo, 21-23 May 1990. *Rep. int. Whal. Commn* (special issue) 16:477-96.

Bow, J.M. and Purday, C. 1966. A method of preparing sperm whale teeth for age determination. *Nature, Lond.* 210(5034):437-8.

Brodie, P.F., Geraci, J.W. and St Aubin, D.J. 1990. Dynamics of tooth growth in beluga whales, *Delphinapterus leucas*, and effectiveness of tetracycline as a marker for age determination. pp. 141-8. *In:* T.G. Smith, D.J. St Aubin and J.R. Geraci (eds.) *Advances in Research on the Beluga Whale, Delphinapterus leucas*. Canadian Bulletin of

Fisheries and Aquatic Sciences 224, Department of Fisheries and Oceans, Ottawa.

Chang, W.Y.B. 1982. A statistical method for evaluating the reproducibility of age determination. *Can. J. Fish. Aquat. Sci.* 39:1208-0.

Donovan, G., Breiwick, J. and Best, P. 1982. A note on comparative readings of sperm whale teeth. *Rep. int. Whal. Commn* 32:251-2.

Doubleday, W.G. and Bowen, W.D. 1980. Inconsistencies in reading the age of harp seal (*Pagophilus groenlandicus*) teeth, their consequences and a means of reducing biases. NAFO SCR Doc. 80/XI/160. North Atlantic Fisheries Organisation. 24pp.

Evans, K. and Robertson, K. 2001. A note on the preparation of sperm whale (*Physeter macrocephalus*) teeth for age determination. *J. Cetacean Res. Manage.* 3(1):101-7.

Evans, K., Morrice, M., Hindell, M. and Thiele, D. 2002. Three mass strandings of sperm whales, *Physeter macrocephalus*, in southern Australian waters. *Mar. Mammal Sci.* 18(3).

Gambell, R. 1977. Dentinal layer formation in sperm whale teeth. pp. 583-90. *In:* M. Angel (ed.) *A Voyage of Discovery: George Deacon 70th Anniversary Volume*. Pergamon Press Ltd, Oxford.

Hohn, A.A. 1980. Analysis of growth layers in the teeth of *Tursiops truncatus* using light microscopy, microradiography, and SEM. *Rep. int. Whal. Commn* (special issue) 3:155-60.

Hohn, A.A. 1990. Reading between the lines: An analysis of age estimation in dolphins. pp. 575-86. *In:* S. Leatherwood and R.R. Reeves (eds.) *The Bottlenose Dolphin*. Academic Press, San Diego, CA. 653pp.

Hohn, A. and Fernandez, S. 1999. Biases in dolphin age structure due to age estimation technique. *Mar. Mammal Sci.* 15(4):1124-32.

Hohn, A.A., Scott, M.D., Wells, R.S., Sweeney, J.C. and Irvine, A.B. 1989. Growth layers in teeth from known-age, free-ranging bottlenose dolphins. *Mar. Mammal Sci.* 5(4):315-42.

Hui, C.A. 1980. Variability of dentin deposits in *Tursiops truncatus*. *Can. J. Fish. Aquat. Sci.* 37:712-6.

International Whaling Commission. 1967. Report of the Scientific Committee. Appendix E. Sperm Whale Sub-Committee Meeting Report. *Rep. int. Whal. Commn* 17:120-7.

International Whaling Commission. 1969. Report of the Scientific Committee, Appendix H. Report of the Meeting on Age Determination in Whales, Oslo, 1968. *Rep. int. Whal. Commn* 19:131-7.

International Whaling Commission. 1971. Report of the Scientific Committee, Annex D. Report of the Special Meeting on Sperm Whale Biology and Stock Assessments. *Rep. int. Whal. Commn* 21:40-50.

Klevezal, G.A. and Myrick, A.C., Jr. 1984. Marks in tooth dentine of female dolphins (genus *Stenella*) as indicators of parturition. *J. Mammal.* 65(1):103-10.

Lockyer, C. 1980. Age determination studies on *Physeter macrocephalus*. *Rep. int. Whal. Commn* (special issue) 3:216. [Abstract].

Lockyer, C. 1993. A report on patterns of deposition of dentine and cement in teeth of pilot whales, genus *Globicephala*. *Rep. int. Whal. Commn* (special issue) 14:137-61.

Lockyer, C. 1995. A review of factors involved in zonation in odontocete teeth, and an investigation of the likely impact of environmental factors and major life events on harbour porpoise tooth structure. *Rep. int. Whal. Commn* (special issue) 16:511-29.

Mikhalev, Y.A. 1982. Subjective and objective assessments of the lamination in sperm whale teeth. *Rep. int. Whal. Commn* 32:227-33.

Myrick, A.C. 1988. Is tissue resorption and replacement in permanent teeth of mammals caused by stress-induced hypocalcemia? pp. 379-89. *In:* Z. Davidovitch (ed.) *The Biological Mechanisms of Tooth Eruption and Root Resorption*. EBSCO Media, Birmingham, Alabama.

Myrick, A.C., Shallenbeger, E.W., Kang, I. and MacKay, D.B. 1984. Calibration of dental layers in seven captive Hawaiian spinner dolphins, *Stenella longirostris*, based on tetracycline labeling. *Fish. Bull.* 82(1):207-25.

Myrick, A.C., Yochem, P.K. and Cornell, L.H. 1988. Toward calibrating dentinal layers in captive killer whales by use of tetracycline labels. *Rit Fisk.* 11:285-96.

Nishiwaki, M., Hibiya, T. and Ohsumi, S.K. 1958. Age study of sperm whale based on reading of tooth laminations. *Sci. Rep. Whales Res. Inst., Tokyo* 13:135-53.

Ohsumi, S. 1971. Some investigations on the school structure of sperm whale. *Sci. Rep. Whales Res. Inst., Tokyo* 23:1-25.

Ohsumi, S. 1977. Age determination and growth of Cetacea. *Honyu-Kagaku* 34(4):54-63.



- Ohsumi, S., Kasuya, T. and Nishiwaki, M. 1963. The accumulation rate of dentinal growth layers in the maxillary tooth of the sperm whale. *Sci. Rep. Whales Res. Inst., Tokyo* 16:15-35.
- Perrin, W.F. and Myrick, A.C. (eds.). 1980. *Report of the International Whaling Commission (Special Issue 3). Age Determination of Toothed Whales and Sirenians*. International Whaling Commission, Cambridge, UK. 229pp.
- Reilly, S.B., Hohn, A.A. and Myrick, A.C. 1983. Precision of age determination of northern offshore spotted dolphins. NOAA Tech. Mem. NOAA-TM-NMFS-SWFSC-35. (Unpublished). iv+27pp. [Available from Southwest Fisheries Science Center, National Marine Fisheries Service, CA].
- Rice, D.W., Wolman, A.A., Mate, B.R. and Harvey, J.T. 1986. A mass stranding of sperm whales in Oregon: sex and age composition of the school. *Mar. Mammal Sci.* 2(1):64-9.
- Scheffer, V.B. and Myrick, A.C. 1980. A review of studies to 1970 of growth layers in the teeth of marine mammals. *Rep. int. Whal. Commn* (special issue) 3:51-63.