Not to be cited without permission of the authors[1]

Ne pas citer sans autorisation des auteurs[1]

Canadian Atlantic Fisheries Scientific Advisory Committee

Comité scientifique consultatif des pêches canadiennes dans l'Atlantique

CAFSAC Research Document 91/59

CSCPCA Document de recherche 91/59

A Critical Review of Mark-Recapture Estimation of Northwest Atlantic Harp Seal Production 1977-1983

by

William G. Warren
Science Branch
Department of Fisheries and Oceans
P. O. Box 5667
St. John's, Newfoundland A1C 5X1

[1] This series documents the scientific basis for fisheries management advice in Atlantic Canada. As such, it addresses the issues of the day in the time frames required and the Research Documents it contains are not intended as definitive statements on the subjects addressed but rather as progress reports on ongoing investigations.

Research Documents are produced in the official language in which they are provided to the Secretariat by the author.

[1] Cette série documente les bases scientifiques des conseils de gestion des pêches sur la côte atlantique du Canada. Comme telle, elle couvre les problèmes actuels selon les échéanciers voulus et les Documents de recherche qu'elle contient ne doivent pas être considérés comme des énoncés finals sur les sujets traités mais plutôt comme des rapports d'étape sur les études en cours.

Les Documents de recherche sont publiés dans la langue officielle utilisée par les auteurs dans le manuscrit envoyé au secrétariat.

Abstract

A critical review was made of the mark-recapture estimates of Northwest Atlantic harp seal pup production in the period 1977-1983. It was concluded that, of the seven basic assumptions that must hold for the Lincoln-Petersen estimator to provide a reasonable estimate of population size, only those involving tag loss and recovery and the equiprobability of capture and random sampling might be of real concern. The adjustment made for tag loss appears reasonably sound, but that for the non-reporting of recovered tags is, if anything, somewhat optimistic. More seriously, there seems strong evidence that uniform mixing of marked and unmarked seals did not take place until after the bulk of the recaptures had been obtained; it is difficult to determine the effect that this has on the estimates although there is evidence that it can be substantial. Further, the estimates are predicated on the total kill statistics being error free, which seems rather doubtful. The wisdom of applying mark-recapture methods to such a widely dispersed and highly migratory species, where multiple recaptures (kills) are impossible, is questionable. Notwithstanding, one's best guess of harp seal pup production in the late 1970's and early 1980's would appear to be of the order of 500000 - 550000.

Résumé

On a procédé à un examen critique des estimations, obtenues par le marquage et la recapture, de la production de jeunes phoques du Groenland dans l'Atlantique nord-ouest dans la période 1977-1983. Il s'est avéré que des sept hypothèses fondamentales qui doivent se vérifier pour que la formule d'estimation de Lincoln-Peterson puisse fournir une appréciation raisonnable de la population, seules celles faisant appel à la perte et au recouvrement d'étiquettes, à l'équiprobabilité des captures et à l'échantillonnage au hasard semblent poser de véritables problèmes. La correction effectuée pour tenir compte des étiquettes perdues semble assez juste, mais celle qui porte sur la non déclaration d'étiquettes recouvrées paraît à tout le moins quelque peu optimiste. Plus encore, il semble très évident que le mélange uniforme des phoques marqués et de ceux qui ne l'étaient pas ne s'est produit qu'après le plus gros des recaptures. Il est difficile de déterminer les effets de ce phénomène sur les estimations, mais il apparaît qu'ils peuvent être considérables. De plus, les estimations sont fondées sur des statistiques de mortalité totale qui seraient libres d'erreur, ce qui semble plutôt improbable. On peut s'interroger sur le bien-fondé de l'application des méthodes de marquage-recapture à une espèce hautement migratoire et très dispersée, au sein de laquelle les recaptures multiples sont impossibles. Cela dit, on peut, au jugé, chiffrer la production de jeunes phoques du Groenland entre la fin des années 1970 et le début des années 1980 était de l'ordre de 500 000 à 550 000.

Introduction

Mark-recapture experiments for estimation of pup production of the harp seal (*Phoca groenlandica*) in the Northwest Atlantic were undertaken in 1978, 1979, 1980 and 1983 and reported by Bowen and Sergeant (1983, 1985). Bowen and Sergeant (1983) also report estimates based on the 1977 marking of 1230 seals in the Gulf. Because the question of whether the seal population is increasing has become an issue, the writer was requested to review the application of mark-recapture methodology as employed by the above authors. Further, Chapter 5 of Cooke, Trites and Larkin (1985) is a review of the mark-recapture estimates of Bowen and Sergeant, and raises several questions concerning the potential sources of bias in these estimates. The present report therefore draws heavily on these three documents which, for convenience will be denoted by BS1, BS2 and CTL.

The basic assumptions that must hold if the Lincoln-Petersen (or Chapman) formula is to provide a reasonable estimate of population size from a mark-recapture survey are listed by BS1 as:
1) The population is assumed closed, so that N [the population size] is constant.
2) All animals have the same probability of being caught in the first sample.
3) Marking does not affect catchability.
4) Animals are correctly classified as marked or unmarked on recovery.
5) Animals do not lose their marks in the time between the two samples.
6) All marks are reported accurately on recovery in the second sample.
7) The second sample is a simple random sample.

These assumptions are all, no doubt, violated to some extent. BS1 and BS2 give arguments, sometime backed quantitatively, that indicate the departures from (1), (3) and (4) would have no consequential effects on the estimates. Their arguments seem reasonable. With respect to (2) BS1 write "there was no significant heterogeneity in the sex ratios between years and area and so the data were pooled. Of 1759 double-tagged pups whose sex ratio was known, 914 (52.0%) were males and 845 (48.0%) were females. This ratio does not differ significantly from the population sex ratio of 1:1 ... reported by Sergeant (1966)". But this is not the point (although it does raise the question of what was the sex ratio of recaptured seals whose sex was known at time of tagging). The logistics of the situation demands that seals be tagged in groups (perhaps less so in 1983), i.e. all pups do *not* have the same probability of being caught in the first sample, and this will contribute to any non-uniform mixing problem (see below).

The incidence of non-reporting of recovered tags

It is certain that not all recovered tags are reported. BS1 present the results of interrogating cluster samples of sealers in 1979 and 1980 for the purpose of estimating the rate of non-reporting. BS2 report the results of a similar survey. Two surveys were conducted in 1979 yielding reporting rates of 0.726 (s.e. = 0.0692) and 0.624 (s.e. = 0.0979). The difference was judged not significant and a pooled estimate of 0.690 (s.e. = 0.0561) obtained. The 1980 survey yielded a reporting rate of 0.769 (s.e. = 0.0459). BS1 state that "There was a significant difference between the estimated reporting rate in 1979 and 1980 (t_s = 1.77, 1264 df, $P > 0.05$)" and go on to say "However, we required an estimate of the mean reporting rate for the period 1978-1981. Therefore we combined the 1979 and 1980 surveys, resulting in an average reporting rate of recovered tags of 0.751 ± 0.0345". The source of the 1264 d.f. is unclear and seems unreasonably large since, as I read it, a community formed a cluster and the number of communities surveyed was 45 in 1979 and 37 in 1980. A t-value of 1.77 does not exceed the 5% critical value but does exceed the 10% critical value; in other words the difference would not be judged as significant at the 5% level (but would be judged significant at the 10% level). On the other hand, the t statistic can be computed as

$$(0.769 - 0.690)/\sqrt{0.0561^2 + 0.0459^2} = 1.09$$

which fails to attain the 10% level. On the surface, the pooling of the 1979 and 1980 surveys seems justified. The survey of BS2 yielded a reporting rate of 0.693 (s.e. = 0.0351) which, by my calculation, does not differ from the pooled 1979-80 estimate at the 10% level.

On the other hand, BS1 present the reporting rates of subsets of communities for both 1979 and 1980 and concluded that "reporting rate was different between years, confirming the previous analysis [?], and was significantly different in different areas both within and between surveys". If sealing effort is variable over communities, the use of a single estimate of reporting rate may not be adequate (particularly if coupled with non-uniform mixing), although BS1 remark that "there is no significant correlation between the rate of reporting and the number of active sealers in a community, or the total number of tags recovered.

CTL note that the Bowen and Sergeant estimates "were based on the numbers of tags still in possession of sealers questioned" and that "no allowance appears to have been made for tags that were simply discarded. Hence the true rate of non-reporting may be higher than that estimated from such surveys".

The impression gained from all this is that the reporting rate is likely lower, and its precision less (i.e. standard error greater), than estimated.

Tag loss

There is no doubt that some tag loss occurs and, as with the non-reporting of recovered tags, BS1 and BS2 adjust the Lincoln-Petersen estimate for tag loss. CTL observe that Bowen and Sergeant's estimates of tag loss are 5.1% for the 1980 survey and 3.7% for the 1983 survey. Using an alternative method, CTL obtain estimates of 26.8% and 30.3%, respectively. They state that they were unable to identify the reason for this difference from the information available to them. Their estimates were based on the recapture of single- and double-tagged seals which had been tagged in the Gulf; they note that Front-tagging data were not available to them at the time.

The Bowen and Sergeant estimates of tag loss were generated from the proportion of double-tagged seals that retained only one tag on recapture. The probability of tag loss was estimated as $m_s/(m_s + 2m_d)$ where m_s is the number of recaptures of double-tagged seals that retain only one tag, and m_d the number with both tags retained. This follows immediately from, if the probability of tag loss is p, the probability of retaining both tags being $(1-p)^2$ and the probability of retaining exactly one of the two tags being $2p(1-p)$. One may then estimate p by setting $m_s/m_d = 2p(1-p)/(1-p)^2$.

On the other hand, CTL consider the recapture rates of each of single- and double-tagged seals. Of the 2711 effectively single-tagged seals in the Gulf in 1980 (number tagged less the pup kill of tagged seals) 275 had been recaptured by 1984, for a recapture rate of 275/2711 = 10.1% (standard error = 0.58%). For the 890 effectively double-tagged seals the recapture rate was 123/890 = 13.8% (s.e. = 1.16%). These recapture rates were judged, reasonably, to be significantly different and, under the assumption that the difference was due entirely to tag loss, the probability of tag loss was estimated as 1 - 10.1/13.8 = 26.8%, with a standard error of 7.4%. Likewise, for 1983, the recapture rates were estimated as 371/3230 = 11.5% (s.e. = 0.56%) and 74/448 = 16.5% (s.e. = 1.75%), and the probability of tag loss as 1 - 11.5/16.5 = 30.3% (s.e. 8.1%).

It is, however, instructive, to look more closely at the data that went into CTL's estimates. The recaptures by year and region of 1980 tagged animals are given in Table 1.

Table 1.

Tag	Number	1980		1981		1982		1983		1984	
Type	Tagged	G	F	G	F	G	F	G	F	G	F
S	2711	174	0	9	33	8	34	1	9	2	5
D	890	89	0	4	10	4	7	2	4	0	3

What is conspicuous about these data is that, in 1980 (the year of tagging) none of the recaptures occurs at the Front, but in subsequent years 81 out of the 101 (80.2%) single-tagged recaptures and 24 out of the 34 (70.6%) double-tagged recaptures occur at the Front. If the 1981-84 data are treated as a 2 by 4 contingency table one obtains, for single-tagged recaptures, $X^2 = 1.03$ (on 3. d.f.), and for double-tagged recaptures, $X^2 = 1.56$ (on 3 d.f.). If a 2 by 2 contingency table is formed from the 1981-84 totals for each of the single- and double-tagged seals, one obtains $X^2 =$ 1.36 (or 0.86 with Yates' correction for continuity) (on 1 d.f.). Thus, from 1981 mixing between the Front and Gulf seems uniform with about 78% of the recaptures (of Gulf-tagged animals) made at the Front (in contrast to 0 out of 263 in the year of tagging). One might argue, therefore, that the difference in recapture rates is a consequence of inadequate mixing in the first year, the year in which the bulk, 66%, of the recaptures occurs. The recapture rates of single- and double-tagged seals, for 1980 only, are 174/2711 = 6.42% (s.e. = 0.47%) and 89/890 = 10.00% (s.e = 1.01%), which, following CLT would imply a probability of tag loss of 35.8%. On the other hand, using only the recaptures from 1981-84 (after mixing has taken place) the recapture rates are 101/2711 = 3.73% (s.e. = 0.36%) and 34/890 = 3.82% (s.e = 0.64%). Clearly, these do not differ significantly, but even if they did, would imply a probability of tag loss of 1-3.73/3.82 = 2.4% with a confidence interval that would include zero. It can be argued that, after 1980, the effective populations of single- and double-tagged seals are not 2711 and 890; one should subtract the first-year recaptures. (These would, in all likelihood, still be overestimates of the population sizes). The revised values would be 101/2537 = 3.98% (s.e. = 0.40%) and 34/799 = 4.26% (s.e. = 0.75%), again not differing significantly but implying a point estimate of the probability of tag loss of 1-3.98/4.26 = 6.6%.

The recaptures by year and region of 1983 tagged animals are given in Table 2.

Table 2.

Tag	Number	1983		1984	
Type	Tagged	G	F	G	F
S	3231	291	33	5	42
D	448	63	3	0	8

These data show similar characteristics to those of the 1980 tagging. One difference is that there are some tagging-year recaptures in the Front, but only 36/354 = 10.2% compared with 50/55 = 90.9% in 1984. (Note that 3/66 does not differ significantly from 33/324; nor does 0/8 differ significantly from 5/47). For 1983 recaptures, the recapture rate is 334/3231 = 10.34% (s.e. = 0.54%) for single-tagged seals and 66/448 = 14.73% (s.e. = 1.67%) for double-tagged, implying a probability of tag loss of 1-10.34/14.73 = 29.8%. However, for the 1984 recaptures, the recapture rates are 47/3231 = 1.45% (s.e. = 0.21%) and 8/448 = 1.79% (s.e. = 0.63%) or, with the 1983 recaptures removed from the population, 47/2897 = 1.62% (s.e. = 0.23%) and 8/382 = 2.09% (s.e. = 0.73%). The difference is, clearly, not significant. Unfortunately, the inferences are limited by the post-1984 recapture data not being available at the time of writing.

The above analyses suggest that the relatively high rate of tag loss estimated by CTL arises from incompete mixing of single- and double-tagged seals in the year of tagging. By their criterion, there is little, if any, evidence of a difference in tag loss rate in subsequent years when, as might be expected and as suggested by the data, mixing would be more complete.

Front-tagged animals.

It is of interest to explore the Front-tagged data (not available to CTL) to see if the same feature is present. The recaptures by year and region are given in Tables 3, 4, and 5 (for 1979, 1980 and 1983 tagging, respectively).

Table 3.

Type	Site	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988
S	G	-	1	7	1	-	-	-	-	-	-
2017	F	86	48	10	9	6	3	1	1	-	-
	O	12	13	3	5	3	-	1	2	3	2
D	G	1	-	-	1	-	-	-	-	-	-
352	F	8	5	5	2	1	-	-	-	-	-
	O	6	6	3	1	-	3	-	1	1	2

where O refers to recapture sites other than the Gulf or Front, namely the Arctic and, more commonly, Greenland. The number effectively tagged is given with the tagging type.

Table 4.

Type	Site	1980	1891	1982	1983	1984	1985	1986	1987	1988	1989	1990
S	G	3	1	-	1	-	-	-	-	-	-	-
2545	F	644	30	21	17	6	3	1	3	2	-	-
	O	9	14	12	6	3	2	2	4	-	-	2
D	G	-	-	-	-	-	-	-	-	-	-	-
450	F	133	1	5	2	3	-	-	-	-	-	-
	O	2	-	1	-	-	-	-	1	1	1	-

Table 5.

Type	Site	1983	1984	1985	1986	1987	1988	1989	1990
S	G	2	-	-	-	-	-	-	-
7459	F	651	84	13	12	6	9	2	9
	O	15	27	12	26	13	10	4	4
D	G	1	-	-	-	-	-	-	-
762	F	69	6	1	-	3	-	-	-
	O	-	6	1	4	-	-	-	1

Most conspicuous here is that, unlike the Gulf taggings in which, after the year of tagging, there is evidence of considerable migration to the Front, very few Front-tagged animals are recaptured in the Gulf ($<$ 1% of the tagged recaptures) and almost half of these occur in 1981 from the 1979 tagging. On the other hand, for the 1979 tagging, 67 out of 263 (25.5%) of the recaptures occur in the Arctic or Greenland. For the 1980 and 1983 taggings the corresponding values are 60 out of 936 (6.4%) and 123 out of 991 (12.4%). (There is, no doubt, some migration from the Gulf to the Arctic and Greenland but recapture data of Gulf-tagged animals is not available for these sites at the time of writing).

First consider the 1980 tagging data (Table 4), and exclude the Gulf recaptures. Clearly those recaptured outside the Front in the year of tagging comprise a much smaller fraction relative to the subsequent years. The remaining years are, therefore, tested for homogeneity. The test is performed for single-tagged animals since the cell numbers of double-tagged recaptures are too small to permit a viable test. Also, the data for 1985-1990 have been combined. This yields the following contingency table

$$\begin{array}{ccccc|c} 30 & 21 & 17 & 6 & 9 & 83 \\ 14 & 12 & 6 & 3 & 10 & 45 \\ \hline 44 & 33 & 23 & 9 & 19 & 128 \end{array}$$

yielding $X^2 = 3.62$ (on 4 d.f.) so that the homogeneity assumption is clearly acceptable. The corresponding totals for the double-tagged seals are 11 and 4, and thus in essentially the same proportion (83:45); likewise for the 1980 recaptures (644:9 and 133:2).Again, it would appear that mixing may not be complete until the year after tagging, but seems reasonably stable from then on. In all cases, i.e. (i) all years combined, (ii) 1980 only, and (iii) all years combined excluding 1980, the recapture rates of single- and double-tagged seals do not differ significantly (5% level test) although the recapture rate of single-tagged seals is slightly greater than that for double-tagged for 1980 alone, but the reverse holds for the post-1980 period.

A similar situation exists for the 1983 tagging data (Table 5). The contingency table formed from the year subsequent to tagging (single-tagged only) is

84	13	12	6	9	2	9	135
27	12	26	13	10	4	4	96
111	25	38	19	19	6	13	231

yielding $X^2 = 34.07$ (6 d.f.). The homogeneity hypothesis, clearly, must be rejected. If the first column of the table is omitted, $X^2 = 7.90$ (on 5 d.f.) and the homogeneity hypothesis is, thus, acceptable under a 5% level test. The homogeneity of the pairs (651:15, 69:0), (51:69, 4:6) and even (84:27, 6:6) is also acceptable. One could interpret this as mixing not being complete until two years after tagging, the proportions for 1985 being intermediate between those for 1983 and 1986-90. Again, but with the exception of the post-1984 recaptures, the recapture rate of double-tagged seals is slightly greater than that of the double-tagged, in no case does the difference approach significance.

Finally, consider the 1979 tagging data (Table 3). Although less conspicuous, the proportion recaptured outside the Front in the year of tagging is again substantially less than in subsequent years. The contingency table, for single-tagged only, with the 1984-1988 data combined and 1979 excluded, is

48	10	9	6	5	78
13	3	5	3	8	32
61	13	14	9	13	110

yielding $X^2 = 9.03$ (on 5 d.f.) so that the homogeneity assumption is marginally acceptable (5% level test). The homogeneity assumption is, however, not acceptable between the single- and double-tagged animals, i.e. 82:12 v. 8:6 and 78:32 v. 13:18. It appears that, of the 1979 tagging, the proportion of double-tagged seals recaptured outside the Front is greater than the proportion of single-tagged. There is no obvious explanation for this and, perhaps, too much should not be read into it. As before, the recapture rate for double-tagged seals is greater than that for single-tagged but with the difference in no case being significant. The picture obtained is, however, qualitatively different in that, ostensibly, there is a greater difference, 3.0%, for the post-1979 recaptures than for the 1979 recaptures (0.6%).

One complication to the above is the decline in the Canadian seal hunt after 1983, so that in later years the Greenland recaptures may take on a somewhat inflated proportion. For the 1980 tagging data (Table 4) the number of recaptures post 1984, although slightly greater outside the Front than at the Front, is too small to have any consequential effect on the above analysis. The same applies to the post-1982 recaptures for the 1979 tagging (Table 3). The situation for the 1983 tagging (Table 4) is somewhat different since a substantial proportion of the data comprise post-1985 recaptures. The intermediate position of the 1985 recaptures (noted above) could be reasonably judged as a consequence of this change in hunting pattern.

General

Bowen and Sergeant's estimates, being based solely on recaptured initially double-tagged seals, is not affected by incomplete or non-uniform mixing. Accordingly, their values of the probability of tag loss would be the more realistic. BS1 gives two short term estimates of tagging loss, namely

6.31% (s.e. = 1.37%) after 1 month and 4.17% (s.e. = 0.98%) after 2-3 months. These do not differ significantly (5% level test); their pooled estimate is 5.10% (s.e. = 0.81%). These values are based on 1979 and 1980 tagging recaptures combined (for the Front in 1979 and the Gulf and Front in 1980); Table 6 of BS1 shows little evidence of heterogeneity between the results for these 3 classifications. BS2 give a short-term (1-3 months) estimate of tag loss, based on the 1983 tagging, of 1.02% (s.e. 0.50%). Clearly, this differs significantly from the 1979-1980 tagging estimate. Bowen and Sergeant offer no explanation for this discrepacy. One might conjecture that tagging was carried out more skillfully in 1983 than 1979-80, but without supporting evidence one is left in somewhat of a quandary as to the best estimate of tag loss and its precision.

A priori one might anticipate that the probability of tag loss would increase over time. BS1 give estimates of long-term tag loss as 4.35% (s.e. = 3.07%) (1 year) and 6.67% (s.e. = 4.71%) (2 years). BC2 give a 1-year estimate of 3.70% (s.e. = 3.70%). These values are consistent with Bowen and Sergeant's conclusion that "the loss of Roto-tags in harp seals occurs shortly after the tags are applied and subsequent losses are negligible". The relative lack of precision in these long-term estimates prohibits any definite conclusion. On the other hand, the relative homogeneity in the recapture rate of single- and double-tagged seals would tend to support BS1's contention, although they also note that "Future tag loss will most likely increase as tags may become worn or brittle and therefore more prone to failure".

The general conclusion from the above is that the CTL estimate of tagging loss is affected by the incomplete mixing of tagged and untagged seals, primarily in the year of tagging when, with the exception of the 1979 Front tagging, the great majority of the recaptures are obtained. Subsequent years show little, if any, difference in the recapture rates of single- and double-tagged seals. An exception is the 1979 tagging where a disproportionate number of double-tagged seals were recaptured outside the Gulf and Front but this again highlights a breakdown in the assumption of random mixing, the consequences of which are not clear.

We may conclude that, over the range of recaptures used for the estimation of pup production, tag loss is more of the order of 5% than the 25-30% or so suggested in Chapter 5 of CTL.

Non-uniform mixing

Non-uniform mixing may have a greater impact on the estimates of pup production than invalidating CTL's estimates of tag loss. BS1 recognize the potential for bias from violation of the assumption that the second sample is a simple random sample, but quote Seber (1963) in stating that the estimator is intuitively reasonable "when the sample proportion marked in the second sample faithfully reflects the population proportion", and go on to say that "this means that the Petersen estimate is valid even if assumption 7 is false, provided there is uniform mixing of marked an unmarked animals so that the proportion (m/N) of marked through the population is constant". They then carry out tests of homogeneity by doubly classifying tagged and not-tagged captures by time (two-week periods) and unit area.

(A) Short-term estimates

For the 1978 tagging they indicate that there were too few recaptures for adequate testing of uniform mixing; neverthess, after examining the data they concluded that "assumption 7 was not satisfied for short-term recoveries in 1978". For the 1979 tagging they note that few tagged beaters were captured in the Gulf, but they undertook a formal analysis, by time and unit area for the Front, and concluded that "a three-way G-test showed that significant heterogeneity exits in the proportion marked by area and date of recapture". Likewise, for the 1980 tagging, they concluded that "the assumption of uniform mixing was not upheld in either the Gulf or at the Front" and "It is clear that estimates of production based on these short-term recoveries could be seriously biased and unreliable". (Note that a test-statistic value was presented for the Front but not the Gulf).

BS2 undertook a similar analysis for 1983-tagged pups at the Front and again concluded that "a 3-way G-test indicated significant heterogeneity in the proportion marked in the population over time and area. Hence, to the extent that random sampling was not achieved the precision of our pup production estimate will be subject to bias".

It is of interest to consider the values of the test statistics. BS1 give $G = 1.4 \times 10^5$ (12 d.f.) (Table 9) and $G = 6.3 \times 10^3$ (24 d.f.) (Table 10). BS2 give $G = 3 \times 10^4$ (22 d.f.) and $G = 184$ (2 d.f.). Simply to indicate that these are significant at the 5% level, as done in Tables 9 and 10 (BS1), is something of an understatement. The 0.05% critical values are 15.2 (2 d.f.), 34.8 (12 d.f.), 50.5 (22 d.f.) and 53.5 (24 d.f.). In other words, it would be virtually impossible to obtain such test values under the null hypothesis. (There is some question as to what, exactly, was done here; after pooling as indicated, Tables 9 and 10 would have only 9 and 12 cells, respectively, yet the degrees of freedom for the G-test are given as 12 and 24).

(B) Long-term estimates

BS1 test the validity of assumption 7 in relation to long-term recoveries "by comparing the observed proportion of Gulf and Front tags recovered with that expected knowing the relative number tagged in each area". The found that "Pooled recoveries of the 1978 and 1979 cohorts were proportional to the effective number tagged in each area [their Table 12]. However, this was not the case for the recoveries of the 1980 cohort at age 1". They go on to conclude that "Hence, the pooled estimates of production in 1978 and 1979 are probably reliable and should be given more weight than the less reliable 1980 estimate". They add that "The 1980 estimate will likely improve over the next year or two as sample size increases". Although BS2 provide an updated estimate of 1980 production, they present no additional data on this point.

Bowen and Sergeant's conclusion for the 1978 cohort is based on total tag recoveries at the Gulf and Front of 178 and 176, respectively, versus expected values of 161 and 193 (based on numbers effectively tagged of 4170 and 4894, yielding a X^2 of 3.29 (1 d.f.), the 5% critical value being 3.84. (The 10% critical value is 2.71). The problem with this is that the recoveries for the Gulf and Front for the years 1979, 1980 and 1981 individually are (90, 124), (63, 42) and (25, 10). As a 2 by 3 contingency table, these exhibit considerable heterogeneity, thus

90	63	25	178
124	42	10	176
214	105	35	354

yielding a $X^2 = 16.1$ (on 2 d.f.) ($p < 0.0005$). Further, testing the recoveries for individual years against their expected values gives $X^2 = 1.20$, 8.63 and 9.33 for 1979, 1980 and 1981, respectively (each on 1 d.f.). Likewise, for the 1979 cohort, we have

76	14	90
42	24	66
118	38	156

The (90, 66) are consistent with their expected values (82, 74) based on effective numbers tagged of 2574 and 2365, $X^2 = 1.65$ (BS1 give 1.45 with Yates' correction). [I make the expected values (81, 75) for a X^2 of 2.08]. The test of homogeneity gives $X^2 = 7.85$ (with Yates' correction) (on 1 d.f.). Testing the recoveries for individual years gives $X^2 = 7.64$ and 3.80 for 1980 and 1981, respectively. The acceptance of the null hypothesis for the pooled recoveries of the 1978 and 1979 cohorts thus seems fortuitious. An argument that the pooled proportions are as expected is spurious since the proportions are not in agreement over the times the recoveries were made. CTL go further and show, from 2 by 2 tables, based on place of tagging (Gulf or Front) and place of recapture, that for most years there is a significant departure from the pattern expected under random mixing. Accordingly the random mixing assumption appears to be untenable for the long-term estimates as generated by Bowen and Sergeant.

In summary, the assumption of random sampling, or uniform mixing, is clearly violated, and at some times substantially so. Because of this, estimates based on short-term recaptures are suspect, whether or not they appear reasonable. Estimates based on long-term recaptures may be less affected but are also suspect.

Pup production estimates

Bowen and Sergeant's short-term estimates of pup production are

Year	Number(000)	Source
1978	2245	BS1
1979	698	BS1
1980	309	BS1
1983	534	BS2

The 1983 estimate has an estimated standard error of 33000. BS1 state that "The 1978 and 1979 estimates are clearly high and cannot be considered reasonable given the known history of the population ...". CTL give a set of somewhat different estimates but state also that "short term recoveries ... do not form a reliable basis for estimation of pup production".

The following long-term estimates of pup production have been given (in thousands)

Year[1]	BS1	BS2[2]	CTL unstratified	CTL stratified
1977	318		333 (31)	
1978	506	497 (34)	473 (25)	536
1979	489	478 (35)	463 (31)	532
1980	450	475 (47)	308 (19)	482
1983		136 (-)	131 (9)	141

[1] 1977 based on Gulf marking; [2] Also Roff and Bowen (1986).
Estimated standard errors are given in brackets. The standard errors in BS1 were incorrectly calculated are are not reproduced. BS2 estimates update the BS1 estimates with subsequently obtained recapture data. The difference between the unstratified CTL and BS2 estimates is unclear; it is conjectured that CTL did not have all the updates that went into BS2.

Bowen and Sergeant regard their 1983 estimate as unreasonably low. They note that in 1984 longliners accounted for 10% of the catch compared with 35% to 59% in previous years and conjecture that "estimates from long-term recoveries may only be reliable when a substantial proportion of the catch is taken by longliners operating in such a fashion that random sampling is approached more closely". They also conjecture that conditions were such as to permit sealers to selectively hunt for tagged animals.

On the other hand, Bowen and Sergeant regard the short-term estimate of the 1983 cohort as fairly reliable and list reasons why tags may have been more widely distributed thus inceasing the possibility of random mixing. Anecdotal evidence could likely be found to support or deny the credibility of the estimate in any one year. CTL comment as follows: "... it is not meaningful to state that the 1983 estimate is less variable than the others. Furthermore their analysis showed that the recoveries in 1983, as in the previous years, were significantly heterogenous between subareas within as well as between the Gulf and Front areas. There may be grounds for regarding the 1983 estimate as more credible than the others, but the fact remains that their method is inherently unreliable and may by chance have yielded an estimate that agreed with some prior expectation. However, if estimates are to be rejected or accepted according to whether they agree with some prior expectation, this begs the question as to what, if any, new information the tagging surveys provide". The writer concurs with this view; there are, however, ways by which expectations based on previous data may be formally combined with a current estimate to give, in general, a somewhat better estimate.

Stratified estimates

To counter to some extent the non-uniform mixing, CTL have computed estimates using a method given by Seber (1973) for populations stratified geographically with incomplete mixing between strata. (The method actually goes back to Darroch (1961) and, indeed, Schaefer (1951)). CTL use only two strata, the Gulf and the Front. Their long-term estimates are increased by 13%, 15%, 56% and 8% for the 1978, 1979, 1980 and 1983 cohorts respectively. The 56% increase for 1980 stems from their relatively low unstratified estimate, 308000 compared with Bowen and Sergeant's 475000. In all other years their unstratified estimates are reasonably close to those of Bowen and Sergeant. The cause of the 1980 discrepacy cannot be ascertained without going through the calculations of each in some detail.

Since non-random mixing occurs also within the Gulf and Front, at least over a period in which a large proportion of the recaptures are obtained, application of stratified method with only the Gulf and Front and strata, is not a panacea. The Gulf and Front are, in fact, divided into subareas (unit areas) so it becomes natural to ask whether stratified estimation on a unit area basis would be possible. While the unit area in which a seal is tagged is known, the location of the kill of both tagged and untagged seals is not so finely recorded. Since the Darroch-Seber method does not require the strata of the first sample to be coincident with those of the second, some modification may be possible along these lines. If it were possible to work from a finer stratifiaction, on the basis of CTL's results with two strata, one might conjecture a further increase in the estimates or pup production. On the other hand, because of the smaller numbers per stratum, this may be at the expense of a decrease in precision (increase in standard error).

With respect to the standard error of the stratified estimator, CTL remark that "Unfortunately, while giving an approximate formula for this variance of this estimator as a function of the unknown parameters, Seber does not give an estimator for the variance". In the writer's experience, based on a simulation study, substitution of the estimates of the unknown parameters does, however, provide a reasonable approximation.

Conclusion

To sum up, all the assumptions underlying the Lincoln-Petersen estimator are, no doubt, violated, however only those of tag loss, incomplete reporting, unequal probability of initial capture and non-uniform mixing (the last two being related) appear to be of potential importance. CTL's concern with respect to Bowen and Sergeant's estimate of tag loss seems to be without justification; also the probability of tag loss seems sufficiently small that any inaccuracy in its estimation would be of no practical consequence. There is some justifiable concern that reporting rates have been overestimated, along with their precision. The more serious problem, however, lies with the non-unifrom mixing aggravated by the fact that, for logistical reasons, the tagging of seals is clustered and all animals do not have the same chance of being marked. This appears to have had particularly severe consequences on estimates from short-term recaptures and persists, to some degree, for estimates based on the so-called long-term recaptures. Although the effect can be detected, it does not seem possible to determine its magnitude reliably. The magnitude depends of the conditions at the time and is far from constant. Because of this, estimates that appear reasonable must remain suspect and, until some means of resolving the problem can be devised, mark-recapture as a means of estimating seal pup production should be treated with extreme caution. Indeed, the wisdom of trying to apply mark-recapture methods to such a widely dispersed and highly migratory species, where multiple recaptures (kills) are impossible, could well be questioned, particularly when coupled with doubts that might exist about the accuracy of the total kill statistics, so far treated as error-free. (A study of the effects of uncertainty in estimated values of both the total number recaptured (i.e. number killed) and the number of these tagged is given as an appendix to Randall et al. 1991; the estimated values are, however, assumed to be unbiased).

There seems to be grounds for conjecturing that the non-uniform mixing would result in the population being somewhat underestimated. This could be compensated, to an unknown extent, by the probable overestimation of the tag-recovery rate. Further underestimation would occur if the total kill were greater than that reported. One should, therefore, be extremely reluctant to give a value for harp seal pup production in the late 1970's and early 1980's. If forced to do so, however, one's best guess would seem to be of the order of 500000 - 550000, with a standard error of at least 30000.

References

Bowen, W.D. and Sergeant, D.E. 1983. Mark-recaptures estimayes of harp seal pup (*Phoca groenlandica*) production in the Northwest Atlantic. Can J. Fish. Aquat. Sci. 40:728-742.

Bowen, W.D. and Sergeant, D.E. 1985. A mark-recapture estimate of 1983 harp seal pup production in the Northwest Atlantic. NAFO SCR Doc. 85/I/1.

Cooke, J.G. Trites, A.W. and Larkin, P.A. 1985. A review of the population dynamics of the Northwest Atlantic harp seal (*Phoca groenlandica*). Report submited to the Royal Commission on Seals and the Sealing Industry in Canada, June 30, 1985.

Darroch, J.N. 1961. Two-sample capture-recapture census when tagging and sampling are stratified. Biometrika 48:241-260.

Randall, R.G., Wright. J.A., Pickard, P.R. and Warren, W.G. 1991. Effect of run timing on the exploitation by anglers of the Atlantic salmon in the Miramachi River. Can. Tech. Rep. Fish. Aquat. Sci. No. 1790.

Roff, D. and Bowen, W.D. 1986 Further analysis of population trends in the Northwest Atlantic harp seal (*Phoca groenlandica*) from 1967 to 1985.
Can. J. Fish. Aquat. Sci. 43:553-564.

Scheafer, M.B. 1951. Estimation of size of animal populations by marking experiments. U.S. Fish and Wildlife Service Fisheries Bulletin 69:191-203.

Seber, G.A.F. 1973. The Estimation of Animal Abundance and Related Parameters. Griffin, London, 506 p.

Sergeant, D.E. 1966. On population dynamics of the western harp seal stocks. Int. Comm. Northwest Atl. Fish. Res. Doc. No. 4, Ser. No. 1749.