Nonparametric tests of double-tagging assumptions

George M. Leigh[a,*], William S. Hearn[b,c]

[a] Agri-Science Queensland, Department of Agriculture and Fisheries, 306 Carmody Road, St Lucia, Queensland 4067, Australia
[b] Sheridan College, 18/7 Aberdeen St, Perth WA 6000, Australia
[c] Formerly of CSIRO Marine and Atmospheric Research
[*]Corresponding author, email address george.leigh@daf.qld.gov.au (G. M. Leigh)

Abstract

Shedding rates of tags on fish are commonly estimated from double-tagging experiments, for which an assumption of independence between the two tags on a fish is required. For tags of qualitatively different types, a nonparametric test for this assumption was proposed by Myhre (1966: International Pacific Halibut Commission Report No. 41), making use of concurrent double- and single-tagging of fish. We extend Myhre's test by developing a nonparametric Bayesian test which is also applicable to the common situation where the two tags attached to a fish are identical and assumed to shed at the same rate; the validity of this assumption can be checked by an extra test that we supply in the case that each tag is identified uniquely. In addition to dependence between tags, the dependence test can also be triggered by departures from other experimental assumptions, such as marked variation in the expertise of taggers. We recommend the dependence test for monitoring tag-return data on an ongoing basis during an experiment. We apply our test to Atlantic cod tagging data listed by Barrowman and Myers (1996: Biometrics 52, 1410–1416). Frequentist tests based on Fisher's Exact Test are also presented.

Keywords: Bayesian methods; Monte Carlo simulation; Nonparametric statistics; Tag dependence; Tag shedding

1. Introduction

Wild fish are often tagged to allow estimates of mortality rates or population sizes (see review in Pine et al., 2003). However, individual fish that lose all their tags are seldom detected after recapture, which leads to biases in the mortality rate estimates. In response, shedding rates are mainly estimated from double-tagging experiments, for which an assumption of independence with respect to tag-

shedding between the tags on individual fish is needed. Models for estimating shedding rates from double-tagging experiments using conditional likelihoods, i.e., without the need to estimate mortality rates or the fraction of fish caught, have been developed by Gulland (1961), Chapman et al. (1965), Bayliff and Mobrand (1972) and Kirkwood (1981).

If tags on the same fish exhibit dependence with respect to tag shedding, linking or merging the above models with models that estimate mortality rates and population size may lead to seriously erroneous estimates of these important population parameters. Beverton and Holt (1957, ch. 14) discuss the possibility that shedding of a tag may depend on whether the other tag attached to the same fish is still attached. For example, an infection may begin at one tag site and spread to the other, so both tags shed at similar times; or an animal may try hard to remove a tag, and if successful will know how to quickly remove the other tag. Some events may remove all tags at once from a fish. Other authors have proposed that there could be dependence between tags with respect to deaths due to tagging; e.g., Myhre (1966), Barrowman and Myers (1996) and Hampton (1997) noted that attaching two tags might be more harmful to a fish than attaching only one tag. Hampton (1997) and Cadigan and Brattey (2006) considered dependence between tags with respect to the reporting rate, in that two tags on a recaptured fish might be more frequently noticed than one tag. Also the financial incentive to return two tags is higher than for one tag, as the reward for returning two tags is double that for returning one tag.

Studies that include auxiliary information have shown that the independence assumption often fails (see details below), and some have proposed methods to deal with dependence between tags on the same animal. Ideally, every double-tagging study should be designed to allow statistical assessment of the independence assumption, although it is also important that the proportion of double-tagged fish be kept high to maximize the number of reported recoveries and facilitate estimation of population parameters.

When dependence is detected, it is commonly dealt with by allowing different parameter values for the shedding risks before and after any tags have been shed. This has been employed by Barrowman and Myers (1996, Appendix) for Northwest Atlantic cod, Diefenbach and Alt (1998) for black bears, Rivalan et al. (2005) for leatherback turtles, McMahon and White (2009) for southern

elephant seals, and Laake et al. (2014) for black bears. Direct evidence of dependence mechanisms is provided by Siniff and Ralls (1991) who observed some Californian sea otters biting off their tags, presumably both in quick succession, while others ignored their tags, and Bradshaw et al. (2000) who found higher rates of tag shedding by New Zealand fur seal pups whose nursery sites were rocky, implying that some pups would quickly lose both their tags by snagging against rocks, while those in non-rocky nursery sites were likely to retain both their tags. Hearn, Leigh, and Beverton (1991) found dependence caused by variation in the skills of tagging operators, which they dealt with by fitting a separate set of parameter values to each operator (Chambers et al., 2015; see also Shaughnessy, 1994).

Other studies concentrate on reporting rates. Myhre (1966) assumes that dependence is due to a higher reporting rate for double-tagged fish and describes estimation of the ratio. Barrowman and Myers (1996, Section 5) describe a model for different types of tag in which one type has 100% visibility and the other on its own has some lower rate: the resulting parameter estimates could be included in mortality models.

The standard double-tagging field experiment without auxiliary information allows evaluation of the tag-independence assumption only in extreme cases where the proportion of single-tag returns decreases over some time interval and tags therefore appear to spontaneously reattach themselves to fish (Hearn et al., 1991). Myhre (1966), however, proposed an experimental design that included a concurrent single-tagging experiment and derived a nonparametric statistical test for dependence. A "nonparametric" test is one that does not require the tag shedding rate to be a parametric function of time at liberty: such a test allows assessment of the data before fitting any complex mathematical model.

Myhre's test can also detect non-compliance with experimental protocols, such as non-uniformity in tagging or recovery practices as described above, or the extra potential problem that some fishers may believe that it is sufficient for them to return only one tag and either retain the second tag as a souvenir or return the fish to water with the second tag still attached.

We extend Myhre's test by deriving a nonparametric test for tag dependence when the two tags on a double-tagged fish are completely interchangeable and assumed to have the same shedding rate. Our test can also be triggered by failure of the assumption that the two tags are governed by the same

shedding rate. If each tag is identified uniquely, the equal-shedding-rate assumption can be tested separately and we supply a suitable testing method.

We apply our test to data from Atlantic cod tagged during 1979 in the Northwest Atlantic and analyzed by Barrowman and Myers (1996). We find dependence between the spaghetti tags used in that experiment, but not between the Petersen disc tags.

2. Tags of different types: Existing theory

2.1. Single- and double-tagging

To allow comparability between single- and double-tagging, the ideal experimental design requires that the different categories of tagged fish be released on a "rotation basis" (Myhre, 1966). For example, the first fish may be tagged with an A tag, the next with an A and a B tag, and the following fish with a B tag, with this pattern repeated until the target numbers are reached. The A and B tags may be different tag types: a Petersen disk tag and a spaghetti tag were used in the experiment analyzed by Barrowman and Myers (1996). Alternatively, they could be of the same type but fall into different categories: for example, information may have been recorded about the attachment position for each tag on a fish. One position may be more difficult for the tagger to reach, or may have a higher shedding rate even when tags are correctly attached.

Suppose N^{AB} fish are double-tagged with an A and a B tag, N^{A} fish with a single A tag and N^{B} fish with a single B tag, and that r_{AB}^{AB}, r_{A}^{AB} and r_{B}^{AB} fish are recovered and reported from the double-tagged fish, and r_{A}^{A} and r_{B}^{B} fish are recovered and reported from the single-tagged fish: superscripts indicate tag types upon release and subscripts tag types upon recapture. The rotation basis on which tagging is conducted (see above) ensures that the time periods over which recoveries are made will be the same for the two tag types and for single- and double-tagged fish. The number of fish that have lost both their tags is usually not observable. We also use the subscript "+" to denote that a recovered fish may or may not retain its tag of the other type: e.g., the number of fish originally double-tagged that are recovered with an A tag is $r_{A+}^{AB} = r_{AB}^{AB} + r_{A}^{AB}$.

If the shedding of A tags is independent of the presence or absence of B tags, the expected proportion of A tags reported for all or part of the experiment is the same for both release subsets, AB and A (Myhre, 1966), i.e.,

$$E\left(r_{A+}^{AB}\right)/N^{AB} = E\left(r_{A}^{A}\right)/N^{A}. \quad (1)$$

Similarly, if the shedding of B tags is independent of the presence of A tags,

$$E\left(r_{B+}^{AB}\right)/N^{AB} = E\left(r_{B}^{B}\right)/N^{B}. \quad (2)$$

The equality of these ratios can readily be tested using the Pearson chi-square test (Myhre, 1966, pp. 12–13).

We note that this test can also be conducted by Fisher's Exact Test (Fisher, 1935a, 1935b) (see description in Dickinson Gibbons, 2004). We will present results from both the Bayesian tests that we describe below and from Fisher's Exact Test and related tests. We prefer the Bayesian tests because we believe them to be more logical and in some cases more powerful than the equivalent tests based on Fisher's Exact Test.

2.2. Bayesian version

The test of equations (1) and (2) has a Bayesian version (Good, 1957; Lindley, 1964) which, given a computer and statistical software, is easy to apply and very accurate for any sample size; other common tests such as the chi-square test are accurate only for large sample size. Taking (1) as an example, the number of recaptures $r = r_{A+}^{AB}$ or r_{A}^{A} follows a binomial distribution with parameters N and p, where $N = N^{AB}$ or N^{A} respectively, and $p = p_{A+}^{AB}$ or p_{A}^{A} is the probability that a fish tagged at the beginning of the experiment is later recovered with an A tag (with or without an accompanying B tag) and reported. The Bayesian test treats p as a random variable, specifying some prior distribution in the absence of any data and then combining this with the data to derive a posterior distribution. Monte Carlo simulation of the posterior distributions of p_{A+}^{AB} and p_{A}^{A} can be used to test the null hypothesis that p_{A+}^{AB} does not differ consistently from p_{A}^{A}.

For the prior distribution, Good (1957) proposed the beta distribution, which is the conjugate prior for the binomial distribution (Raiffa and Schlaifer, 1961; G. A. Barnard in Wetherill, 1961). The beta distribution, denoted B(α, β), has the probability density

$$f(p;\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1}$$

over the range $0 < p < 1$, where α and β are the parameters of the distribution. If the prior distribution is beta with parameter values α_0 and β_0, the posterior is also beta with parameter values $\alpha = \alpha_0 + r$ and $\beta = \beta_0 + N - r$. If there is no prior information about recovery rates, the parameters α_0 and β_0 can both be set to zero, yielding an improper prior which does not take finite positive values anywhere in the interval $0 < p < 1$ but still yields sensible posterior distributions. Then the posterior distributions are

$$p_{A+}^{AB} \sim \mathrm{B}\left(r_{A+}^{AB}, N^{AB} - r_{A+}^{AB}\right) \tag{3}$$

and

$$p_{A}^{A} \sim \mathrm{B}\left(r_{A}^{A}, N^{A} - r_{A}^{A}\right). \tag{4}$$

Monte Carlo simulations of p_{A+}^{AB} and p_{A}^{A} can easily be run in software such as *R* (R Core Team, 2016) which contains the function "rbeta" to simulate beta-distributed random variables. The statistical test becomes simply a count of how often $p_{A+}^{AB} < p_{A}^{A}$, and sufficiently high or low counts (e.g., > 97.5% or < 2.5%) result in rejection of the null hypothesis that p_{A+}^{AB} does not differ consistently from p_{A}^{A}.

3. Identical tags: New theory

3.1. Test for whether tag categories shed at the same rate

In section 3.2 below, we will introduce a test for tag-shedding dependence between tags attached to the same fish when the tags are assumed identical and their data can be pooled. By "identical" we mean that the two tags attached to a fish are completely interchangeable and only details such as their serial numbers differ. Even in this situation, however, tags may be distinguished by information such as the position on the fish in which each tag was attached. For such a case we propose a preliminary test of whether it is valid to combine these two tag categories.

Before pooling the data from two tag categories A and B, the following Bayesian test can be conducted for whether their shedding rates are in fact identical. If the test finds no difference in shedding rates, the analysis in section 3.2 can proceed with confidence that numbers of tags can be pooled to provide greater sample size and hence a more sensitive test for dependence between tags on the same fish.

The test described in this section requires detailed records of the tagging procedure; e.g., for a test of identical tags attached to different positions on the fish, the position to which each tag was attached. The numbers of tags returned from experiment components A, B and AB follow independent binomial or multinomial distributions. Excluding priors and multinomial coefficients provides the following overall likelihood:

$$\Lambda = \left(p_{AB}^{AB}\right)^{r_{AB}^{AB}} \left(p_{A}^{AB}\right)^{r_{A}^{AB}} \left(p_{B}^{AB}\right)^{r_{B}^{AB}} \left(1-p_{AB}^{AB}-p_{A}^{AB}-p_{B}^{AB}\right)^{N^{AB}-r_{A}^{AB}-r_{B}^{AB}-r_{AB}^{AB}}$$

$$\times \left(p_{A}^{A}\right)^{r_{A}^{A}} \left(1-p_{A}^{A}\right)^{N^{A}-r_{A}^{A}} \left(p_{B}^{B}\right)^{r_{B}^{B}} \left(1-p_{B}^{B}\right)^{N^{B}-r_{B}^{B}} \tag{5}$$

Considering the AB component alone, i.e., only the first four factors of the likelihood (5), and conditioning on the observed numbers of recoveries, p_{AB}^{AB}, p_{A}^{AB}, and p_{B}^{AB} follow a Dirichlet distribution which is a multivariate generalization of the beta distribution (see, e.g., Ratnaparkhi, 2004). The trivariate Dirichlet distribution for hypothetical variables p_1, p_2 and p_3 with parameters (α_1, α_2, α_3, α_4) has probability density

$$f(p_1,p_2,p_3;\alpha_1,\alpha_2,\alpha_3,\alpha_4) = \frac{\Gamma(\alpha_1+\alpha_2+\alpha_3+\alpha_4)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\Gamma(\alpha_3)\Gamma(\alpha_4)} p_1^{\alpha_1-1} p_2^{\alpha_2-1} p_3^{\alpha_3-1} (1-p_1-p_2-p_3)^{\alpha_4-1}.$$

We use the uninformative prior for p_{AB}^{AB}, p_{A}^{AB}, and p_{B}^{AB} as in section 2.2,

$$f_0(p_A^{AB}, p_B^{AB}, p_{AB}^{AB}) = 1/\left\{p_{AB}^{AB} p_A^{AB} p_B^{AB} \left(1-p_{AB}^{AB}-p_A^{AB}-p_B^{AB}\right)\right\}.$$

Then p_{AB}^{AB}, p_{A}^{AB}, and p_{B}^{AB} follow a trivariate Dirichlet distribution with parameters (r_{AB}^{AB}, r_{A}^{AB}, r_{B}^{AB}, $N^{AB}-r_{AB}^{AB}-r_{A}^{AB}-r_{B}^{AB}$). Using standard software, this distribution can be simulated as a sequence of beta distributions:

$$p_{AB}^{AB} \sim \mathrm{B}\left(r_{AB}^{AB},\ N^{AB}-r_{AB}^{AB}\right) \tag{6}$$

$$p_{A}^{AB} / \left(1-p_{AB}^{AB}\right) \sim \mathrm{B}\left(r_{A}^{AB},\ N^{AB}-r_{AB}^{AB}-r_{A}^{AB}\right) \tag{7}$$

$$p_B^{AB} \big/ \left(1 - p_{AB}^{AB} - p_A^{AB}\right) \sim \mathrm{B}\left(r_B^{AB},\ N^{AB} - r_{AB}^{AB} - r_A^{AB} - r_B^{AB}\right). \tag{8}$$

For the purpose of the overall test of equality of the two tag shedding rates, we assume that tags attached to the same fish are independent. Then $p_A^A = p_{AB}^{AB} + p_A^{AB} = p_{A+}^{AB}$ and $p_B^B = p_{B+}^{AB}$. The single-tagging components of the likelihood (5) can be included by the technique of rejection sampling (see, e.g., Hoaglin, 2004, p. 4). Together with the Dirichlet simulation (p_{AB}^{AB}, p_A^{AB}, p_B^{AB}) of the double-tagging parameters, we generate a random uniform variable u, where $0 < u < 1$. We accept the simulation if

$$u < \left(p_{A+}^{AB}\right)^{r_A^A} \left(1 - p_{A+}^{AB}\right)^{N^A - r_A^A} \left(p_{B+}^{AB}\right)^{r_B^B} \left(1 - p_{B+}^{AB}\right)^{N^B - r_B^B} \big/ k, \tag{9}$$

where k is the maximum value of the numerator. Otherwise we reject it and try again. The maximum value k occurs at $p_{A+}^{AB} = r_A^A \big/ N^A$ and $p_{B+}^{AB} = r_B^B \big/ N^B$.

The test for equal shedding rates is a count of how often $p_A^{AB} < p_B^{AB}$. Significantly high or low counts result in rejection of the null hypothesis that p_A^{AB} does not differ consistently from p_B^{AB}.

3.2. Dependence test when tags are identical

Let N^{AA} fish be double-tagged with two identical A tags, N^A fish be single-tagged with an A tag, and numbers r_{AA}^{AA}, r_A^{AA} and r_A^A fish be recovered and reported. Designators A and B can be assigned arbitrarily to the positions of two tags on a fish. Then the right-hand sides of equations (1) and (2) are identical, and averaging the left-hand sides leads to the independence condition

$$E\left(r_{AA}^{AA} + \tfrac{1}{2} r_A^{AA}\right) \big/ N^{AA} = E\left(r_A^A\right) \big/ N^A. \tag{10}$$

The factor of $\frac{1}{2}$ in the left-hand side of (10) precludes the Pearson chi-square test used by Myhre (see section 2.1), but the Bayesian test is straightforward to apply. Conditional on the observed numbers of recoveries and using the uninformative prior as above, p_{AA}^{AA} and p_A^{AA} follow a bivariate Dirichlet distribution with parameters (r_{AA}^{AA}, r_A^{AA}, $N^{AA} - r_{AA}^{AA} - r_A^{AA}$). Again the Dirichlet distribution can be simulated by a sequence of beta simulations:

$$p_{AA}^{AA} \sim \mathrm{B}\left(r_{AA}^{AA},\ N^{AA} - r_{AA}^{AA}\right) \tag{11}$$

$$p_A^{AA} \big/ \left(1 - p_{AA}^{AA}\right) \sim \mathrm{B}\left(r_A^{AA},\ N^{AA} - r_{AA}^{AA} - r_A^{AA}\right). \tag{12}$$

Equation (4) is unchanged. The statistical test is a count of how often $p_{AA}^{AA} + \frac{1}{2} p_{A}^{AA} < p^{A}$; significantly high or low counts result in rejection of the null hypothesis that $p_{AA}^{AA} + \frac{1}{2} p_{A}^{AA}$ does not differ consistently from p^{A}.

4. Application to cod tagging data

4.1. Background

Barrowman and Myers (1996) presented data from an experiment originally described by Lear (1984) on Northwest Atlantic cod that were tagged during 1979 (see Table 1). There were five different release types: two Petersen discs (AA), one Petersen disc and one spaghetti tag (AB), two spaghetti tags (BB), a single Petersen disc (A), and a single spaghetti tag (B). Totals of 999 double-tagged and 3018 single-tagged fish were released, of which 231 originally double-tagged and 564 originally single-tagged fish were recovered.

4.2. Test for independence of tags of different types

For the AB, A and B tagged fish, i.e., tags of qualitatively different types, a test for independence was described by Myhre (1966) and can be conducted by his chi-square test or by the Bayesian Monte Carlo test described in section 2.2 using equations (3) and (4). We used the Bayesian test and found counts of $8{,}982{,}419 / 10^7$ occurrences (89.8%) in which $p_{A+}^{AB} < p_{A}^{A}$ and $5{,}226{,}331 / 10^7$ (52.3%) in which $p_{B+}^{AB} < p_{B}^{B}$. Neither of these results was statistically significant even at a level of $P = 0.2$, as both ratios were between 0.1 and 0.9. Fisher's Exact Test provides significance levels of 0.120 and 0.519, which are similar to the Bayesian test levels of 0.102 and 0.477.

4.3. Test of shedding rates for different tag categories

Our test from section 3.1 for equality of shedding rates would be best applied to the AA and A tagged cod, or alternatively the BB and B tagged cod, if detailed information on the position of each tag on the fish were available. In the absence of this information, for the purpose of illustration we apply our test to the AB, A and B tagged fish, so that we are actually testing whether the Petersen discs and the spaghetti tags have equal shedding rates. We exclude the data on AA and BB tagged cod.

We simulated p_{AB}^{AB}, p_{A}^{AB}, and p_{B}^{AB} from equations (6)–(8). Then we included the likelihood for the single-tagging data by applying rejection sampling as in (9). The total number of simulations was increased by a factor of 20 to 2×10^8 so as to produce at least 10^7 simulations that were not rejected.

For the double-tagging component only, i.e., using all of the original simulations from (6)–(8), the frequency with which $p_{A}^{AB} < p_{B}^{AB}$ was $15{,}156{,}330/(2 \times 10^8) = 0.0758$ which is generally not regarded as statistically significant. An exact binomial test, conditioning on the value of the sum $r_{A}^{AB} + r_{B}^{AB}$, gives a significance level of 0.1148 for a one-sided test, which is somewhat different, possibly due to the discrete level of the exact test: more than half of the probability is in the largest term of the sum of binomial terms.

Including the single-tagging components in this test, i.e., applying the rejection sampling (9), and selecting the first 10^7 simulations that were not rejected, the frequency was $11{,}732/10^7 = 0.00117$ which is significant. An exact test combining the binomial distribution for the double-tagging component with an independent hypergeometric distribution for the single-tagging component, as in Fisher's exact test, yields a significance level of 0.00082. Therefore by both methods we find that the B tags (spaghetti tags) are shed at a higher rate than A tags (Petersen discs), and that including the single-tagging data makes a more powerful statistical test. These results agree with Barrowman and Myers (1996). It is possible that this finding actually relates to the reporting rate rather than the shedding rate, i.e., the Petersen discs may be more highly visible to fishers, as the experiment cannot separate these effects.

4.4. Test for dependence with identical tags

To test for dependence of tags for the case of AA and A tagged fish, and that of BB and B tagged fish, we used our test from section 3.2 with equations (10)–(12) and (4). For AA and A tagged fish the observed proportions $\left(r_{AA}^{AA} + \tfrac{1}{2} r_{A}^{AA}\right) \big/ N^{AA}$ and $r_{A}^{A} \big/ N^{A}$ from equation (10) were 0.1875 and 0.2179 respectively. The frequency of occurrence with which $p_{AA}^{AA} + \tfrac{1}{2} p_{A}^{AA} < p_{A}^{A}$ was $8{,}505{,}177/10^7 = 0.8505$, giving a significance level of 0.1495, not statistically significant. We defined an exact test in which the number of recoveries with one tag that were observed was binomial with parameters $(r_{A}^{AA}, 0.5)$;

the number of recoveries could then be compared to the single-tagging component by Fisher's Exact Test. This test gave a significance level of 0.2017. For BB tagged fish the observed proportions were $\left(r_{BB}^{BB} + \frac{1}{2} r_{B}^{BB}\right) / N^{BB} = 0.2238$ and $r_{B}^{B} / N^{B} = 0.1712$. The frequency with which $p_{BB}^{BB} + \frac{1}{2} p_{B}^{BB} < p_{B}^{B}$ was $54{,}351 / 10^{7} = 0.0054351$ which is significant; see Fig. . 1. We note that the numbers of both double-tagged and single-tagged releases were higher for the B tags, about double the numbers for A tags, which gives the test for B tags more power.

Our exact test gave a one-sided significance level of 0.01125, more than double the Bayesian level. The exact test for this problem appears to be less powerful than the Bayesian test, probably due to the preliminary binomial split that is needed to directly compare the observed recoveries from the double-tagged and single-tagged components.

The result for B tags (spaghetti tags) using the Bayesian test is significant at a level of about 0.01 (0.005 for a one-sided test). There are significantly more recoveries of fish that were originally double-tagged with two spaghetti tags than could be expected from the recoveries of single-tagged fish. This finding leads to a need to revise the Barrowman and Myers (1996) analyses.

A similar result for spaghetti tags on Northwest Atlantic cod was found from a later tagging experiment by Cadigan and Brattey (2003) . They found that the forward of the two tagging positions on the fish, which was the only one used in single-tagging, produced lower recovery rates. They hypothesized that this could be due to tags in the forward position catching on the fishing gear and being torn off as a fish was recaptured.

5. Discussion and recommendations

Properly designed and implemented concurrent double- and single-tagging components of an experiment enable nonparametric tests of whether the tags on a double-tagged fish are shed and returned independently of each other. We have extended the test of Myhre (1966), which is applicable when tags on a double-tagged fish are qualitatively different and not assumed to shed at the same rate, to a nonparametric Bayesian test which covers the case of identical tags that shed at the same rate. The Bayesian test is easy to use and provides precise significance levels for any sample sizes. We

have also included a nonparametric Bayesian test of whether potentially different tag categories, in which identical tags are attached to different parts of the fish, shed at the same rate.

We recommend the use of the nonparametric Bayesian test to monitor tag-return data on an ongoing basis during an experiment. With such monitoring it may be possible to resolve anomalies while the experiment is still in progress.

Despite the number of cases of dependence reported (see our Introduction above), our test found no evidence for dependence in Petersen disc tags in the Northwest Atlantic cod data of Barrowman and Myers (1996). This tag type may have some advantage over the spaghetti tag, e.g., by being easier to properly attach on each side of a fish.

Judging from the conclusion of Cadigan and Brattey (2003) that the position of a spaghetti tag makes a big difference, our example from Barrowman and Myers (1996) detects a departure from experimental assumptions rather than actual dependence between tags. This example also highlights the desirability of recording the sequence number of the tag attached to each position on each fish: with this information, our test for equality of the shedding rates may well have found a discrepancy and the experiment could have been re-analyzed to account for the unexpected difference between tag positions.

The R code and cod data used for both the Bayesian tests and tests based on Fisher's Exact Test are available from Harvard Dataverse (Leigh and Hearn, 2017). The R code was written for input files in CSV format (download option "Original File Format"). The Bayesian versions are set up to perform 10^7 simulations, and 20 times that number as input to the rejection sampling; these numbers may need to be reduced on some computers due to memory limitations.

Acknowledgements

Two anonymous referees made very helpful comments on an earlier draft of this paper, one of which was the suggestion to include Fisher's Exact Test, which we had not previously considered. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

Barrowman, N.J., Myers, R.A., 1996. Estimating tag-shedding rates for experiments with multiple tag types. Biometrics 52, 1410–1416.

Bayliff, W.H., Mobrand, L.M., 1972. Estimates of the rates of shedding of dart tags from yellowfin tuna. Inter-Am. Trop. Tuna Comm. Bull. 15, 439–462.

Beverton, R.J.H., Holt, S.J., 1957. On the Dynamics of Exploited Fish Populations. Chapman and Hall, London.

Bradshaw, C.J.A., Barker, R.J., Davis, L.S., 2000. Modeling tag loss in New Zealand fur seal pups. J. Agric. Biol. Environ. Stat. 5, 475–485.

Cadigan, N.G., Brattey, J., 2006. Reporting and shedding rate estimates from tag-recovery experiments on Atlantic cod (*Gadus morhua*) in coastal Newfoundland. Can. J. Fish. Aquat. Sci. 63, 1944–1958. doi:10.1139/f06-096

Cadigan, N.G., Brattey, J., 2003. Semiparametric estimation of tag loss and reporting rates for tag-recovery experiments using exact time-at-liberty data. Biometrics 59, 869–876. doi:10.1111/j.0006-341X.2003.00101.x

Chambers, M.S., O'Neill, B., Sidhu, L.A., 2015. Frailty models for the estimation of tag shedding rates with tagger effects. Fish. Res. 161, 303–311. doi:10.1016/j.fishres.2014.08.019

Chapman, D.G., Fink, B.D., Bennett, E.B., 1965. A method for estimating the rate of shedding of tags from yellowfin tuna. Inter-Am. Trop. Tuna Comm. Bull. 10, 333–352.

Dickinson Gibbons, J., 2004. Fisher's Exact Test, in: Encyclopedia of Statistical Sciences. John Wiley & Sons, Inc. doi:10.1002/0471667196.ess0791.pub2

Diefenbach, D.R., Alt, G.L., 1998. Modeling and evaluation of ear tag loss in black bears. J. Wildl. Manag. 62, 1292–1300. doi:10.2307/3801993

Fisher, R.A., 1935a. The Logic of Inductive Inference. J. R. Stat. Soc. 98, 39–82. doi:10.2307/2342435

Fisher, R.A., 1935b. The Design of Experiments. Oliver & Boyd, Edinburgh.

Good, I.J., 1957. Saddle-point methods for the multinomial distribution. Ann. Math. Stat. 28, 861–881.

Gulland, J.A., 1961. On the analysis of double tagging experiments. Presented at the North Atlantic Fish Marking Symposium, International Commission for the Northwest Atlantic Fisheries, Special Publication No. 4 (1963), Dartmouth, NS, pp. 228–229.

Hampton, J., 1997. Estimates of tag-reporting and tag-shedding rates in a large-scale tuna tagging experiment in the western tropical Pacific Ocean. Fish. Bull. 95, 68–79.

Hearn, W.S., Leigh, G.M., Beverton, R.J.H., 1991. An examination of a tag-shedding assumption, with application to southern bluefin tuna. ICES J. Mar. Sci. 48, 41–51.

Hoaglin, D.C., 2004. Random number generation. Encycl. Stat. Sci. doi:10.1002/0471667196.ess0872.pub2

Kirkwood, G.P., 1981. Generalized models for the estimation of rates of tag shedding by southern bluefin tuna (*Thunnus maccoyii*). J. Cons. Int. Pour Explor. Mer 39, 256–260.

Laake, J.L., Johnson, D.S., Diefenbach, D.R., Ternent, M.A., 2014. Hidden Markov model for dependent mark loss and survival estimation. J. Agric. Biol. Environ. Stat. 19, 522–538. doi:10.1007/s13253-014-0190-1

Lear, W.H., 1984. Discrimination of the stock complex of Atlantic cod (*Gadus morhua*) off southern Labrador and eastern Newfoundland, as inferred from tagging studies. J. Northwest Atl. Fish. Sci. 5, 143–159.

Leigh, G.M., Hearn, W.S., 2017. Nonparametric tests of double-tagging assumptions. Harv. Dataverse V1 UNF:6:M9LYpRrOB8MHkZflOkmayg==. doi:10.7910/DVN/XWMSHO

Lindley, D.V., 1964. The Bayesian analysis of contingency tables. Ann. Math. Stat. 35, 1622–1643.

McMahon, C.R., White, G.C., 2009. Tag loss probabilities are not independent: Assessing and quantifying the assumption of independent tag transition probabilities from direct observations. J. Exp. Mar. Biol. Ecol. 372, 36–42. doi:10.1016/j.jembe.2009.02.006

Myhre, R.J., 1966. Loss of tags from Pacific halibut as determined by double-tagging experiments. International Pacific Halibut Commission, Report No. 41, Seattle.

Pine, W.E., Pollock, K.H., Hightower, J.E., Kwak, T.J., Rice, J.A., 2003. A review of tagging methods for estimating fish population size and components of mortality. Fisheries 28, 10–23. doi:10.1577/1548-8446(2003)28[10:AROTMF]2.0.CO;2

R Core Team, 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna.

Raiffa, H., Schlaifer, R., 1961. Applied Statistical Decision Theory. Harvard University Press, Boston.

Ratnaparkhi, M., 2004. Liouville-Dirichlet Distributions (Multivariate Liouville Distributions). Encycl. Stat. Sci. doi:10.1002/0471667196.ess1480.pub3

Rivalan, P., Godfrey, M.H., Prévot-Julliard, A.-C., Girondot, M., 2005. Maximum likelihood estimates of tag loss in leatherback sea turtles. J. Wildl. Manag. 69, 540–548. doi:10.2193/0022-541X(2005)069[0540:MLEOTL]2.0.CO;2

Shaughnessy, P.D., 1994. Tag-shedding from South African fur seals *Arctocephalus pusillus pusillus*. South Afr. J. Mar. Sci. 14, 89–94. doi:10.2989/025776194784287148

Siniff, D.B., Ralls, K., 1991. Reproduction, survival and tag loss in California sea otters. Mar. Mammal Sci. 7, 211–299. doi:10.1111/j.1748-7692.1991.tb00099.x

Wetherill, G.B., 1961. Bayesian sequential analysis. Biometrika 48, 281–292. doi:10.2307/2332750

Table 1: Numbers of tagged cod released and recovered, summarized from Barrowman and Myers (1996) . Tag type A is a Petersen disc, while tag type B is a spaghetti tag.

Tags at release	Number released	Tags at recovery	Number recovered
AA	200	AA	32
		A	11
AB	399	AB	59
		A	16
		B	9
BB	400	BB	75
		B	29
A	1014	A	221
B	2004	B	343
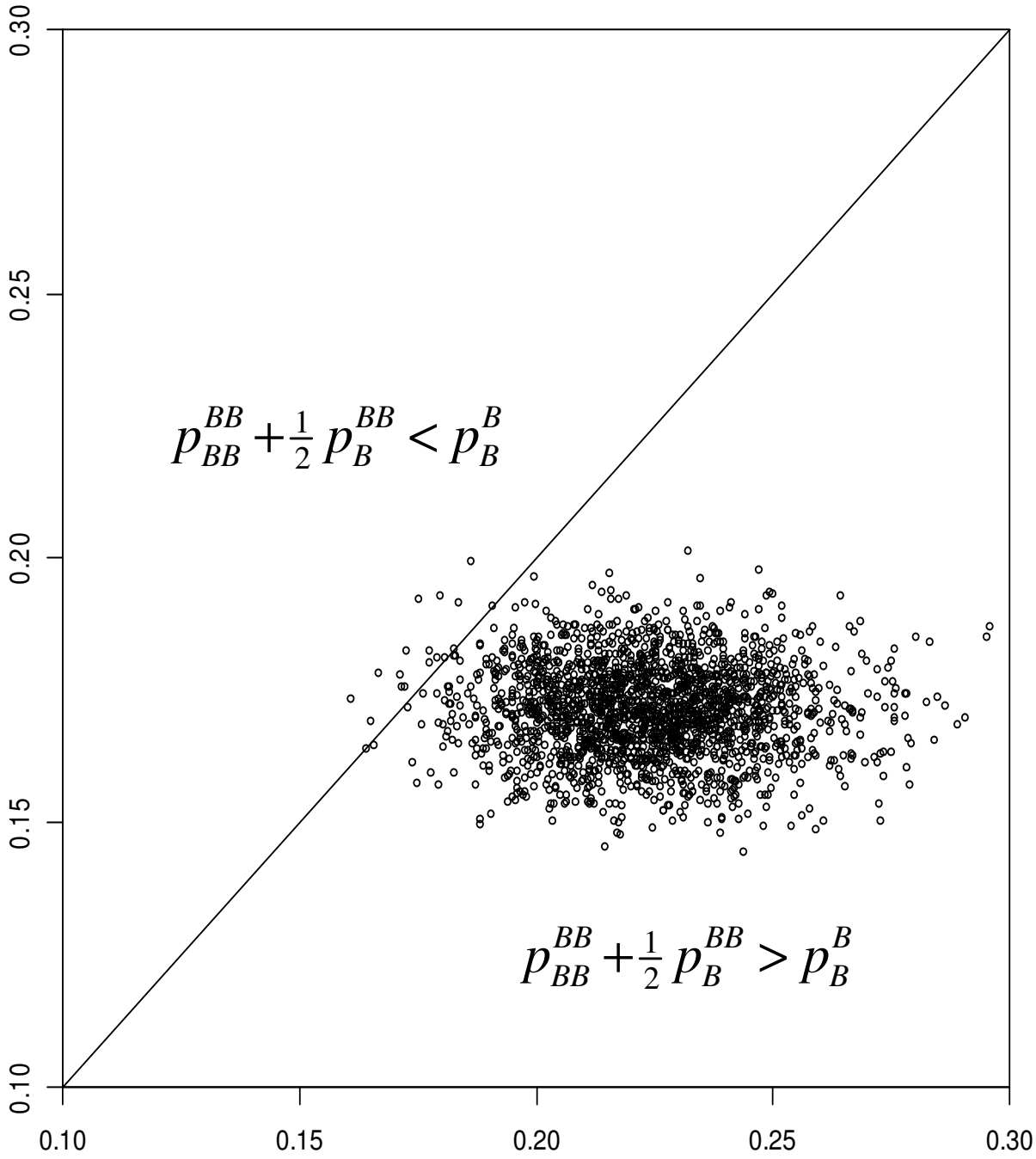
Fig. 1: Scatter plot of the first 2000 simulations of $p_{BB}^{BB}+\frac{1}{2}p_B^{BB}$ and p_B^B in the cod tagging data from the B tag versions of equations (10), (11), (12) and (4), showing that very few have $p_{BB}^{BB}+\frac{1}{2}p_B^{BB} < p_B^B$. The test for dependence between tags on the same fish is significant at level $P \approx 0.005$ (one-sided test) or $P \approx 0.01$ (two-sided test).