Expanding the feasibility of fish and wildlife assessments with close-kin mark–recapture

Benjamin Marcy-Quay, Suresh A. Sethi, Nina O. Therkildsen, and Clifford E. Kraft

1Department of Natural Resources, Cornell University, Ithaca, New York 14853 USA
2U.S. Geological Survey, New York Cooperative Fish and Wildlife Research Unit, Cornell University, Ithaca, New York 14853 USA

Citation: Marcy-Quay, B., S. A. Sethi, N. O. Therkildsen, and C. E. Kraft. 2020. Expanding the feasibility of fish and wildlife assessments with close-kin mark–recapture. Ecosphere 11(10):e03259. 10.1002/ecs2.3259

Abstract. Close-kin mark–recapture (CKMR) is a powerful new method for the assessment of fish and wildlife population dynamics. Unlike traditional mark–recapture techniques, the use of kinship as an identifying mark is robust to many forms of capture heterogeneity including variation in gear efficiency and tagging-based effects such as loss and differential mortality. In addition, close-kin methods can be applied to a wider range of sampling designs than traditional methods (e.g., single-occasion surveys and lethal capture), can provide retrospective historical abundance estimates, and can produce survival estimates from as few as two sampling occasions. We evaluated the ability of CKMR to provide estimates of abundance and adult survival and then compared results to those from traditional mark–recapture. This analysis incorporated data from a three-year study of lake resident brook trout (Salvelinus fontinalis) where individuals were both physically (PIT) tagged and genotyped for 44 de novo developed microsatellites with high throughput sequencing. Traditional mark–recapture estimates were derived using Pollock’s Robust Design, relying upon three primary open sampling occasions and four secondary closed occasions. We found that close-kin methods produced contemporary estimates of adult abundance and survival that were similar to those produced by traditional mark–recapture in both magnitude and precision. Furthermore, CKMR provided abundance estimates for multiple years prior to sampling and, when restricted to data from a single year, still produced reliable abundance estimates for at least one and as many as three years. Retrospective abundance estimates corresponded with those from a separate historical two-sample mark–recapture dataset. This study provides support for the use of CKMR as a robust and sampling-efficient alternative to traditional mark–recapture methods of assessing population parameters.

Key words: close-kin mark–recapture; fish; genetic; population dynamics; robust design; amplicon sequencing; acidification recovery.

Received 20 December 2019; revised 21 April 2020; accepted 29 June 2020; final version received 12 August 2020.
Corresponding Editor: Wyatt F. Cross.
Copyright: © 2020 The Authors. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
† E-mail: bm455@cornell.edu

INTRODUCTION

The recovery of marked individuals has provided a foundation for estimating animal abundance and survival, providing tools to assess fish and wildlife populations that would be infeasible to census directly. Closed population estimation approaches, such as the two-sample Chapman estimator (Chapman 1951) and a wide variety of multiple-recapture models (e.g., Burnham and Overton, Schnable, etc.) (Sutherland (2006)), utilize short sampling intervals to estimate abundance with the assumption that no loss or gain of individuals has occurred. Open estimators like the
Cormack-Jolly-Seber model (Cormack 1964, Jolly 1965, Seber 1965) use longer intervals to estimate additional parameters such as survival. Pollock (1982) combined these approaches using a robust design that estimates abundance from a set of closed population surveys and survival from open periods between these surveys. Many subsequent modifications and applications continue to extend mark–recapture models (Lebreton et al. 1999, Williams et al. 2002, Cooch 2008, Amstrup et al. 2010), with some recent advances leading to integrated population models that combine mark–recapture analysis with other forms of demographic modeling and data (Zipkin and Saunders 2018).

Although widely used, mark–recapture models rely upon assumptions that inevitably bias population estimates when they cannot be met (Pollock 1980, Kendall 1999). At the heart of mark–recapture techniques is the idea of capture probability, that is, the probability of an individual being detected (captured) in a single unit of sampling effort. Indeed, one of the most challenging assumptions for any field survey to meet is to avoid unmodeled capture heterogeneity during survey efforts; that is, capture probability cannot vary within or between data collection periods in any way that is both systematic and not included within the underlying model (Link 2003). Yet, capture probability is almost always influenced by size-selectivity (Anderson 1995), gear efficiency (Seber 1970), or capture location (e.g., habitat type), violating this key assumption (Royle et al. 2013). In addition, model estimates ultimately rely on recapture events, requiring that previously handled individuals are catchable and distinguishable from previously uncaptured individuals, which introduces bias due to mortality from tag implantation, increased predation after handling, and tag loss (Pine et al. 2012). Similarly, behavioral responses to capture can also bias parameter estimates, as individuals may become more or less likely to be recaptured on subsequent occasions, referred to as trap-happiness or trap shyness (Geis 1955).

Molecular techniques offer a promising approach to minimize many of the factors that confound mark–recapture models. Genetic markers can provide individual identification while avoiding the negative impacts common to physical tags (Taberlet et al. 1999). Some techniques, such as hair snares, have been used for two decades to collect genetic samples for individual identification without any direct handling of animals by researchers (Woods et al. 1999). However, until recently these genetic techniques have largely been applied as a direct replacement for physical tags. Close-knit mark–recapture (CKMR) builds on this framework by using genotype-derived estimates of kinship (i.e., not just individual identification) to extend mark–recapture methods to circumstances where sparse sampling or lethal capture makes direct recapture unlikely or impossible (Bravington et al. 2016a, b).

CKMR compares the number of kinship pairs observed in a sample or set of samples to the total number of potential pairs (based on observed age structure), making possible an estimate of the total number of potential parents (i.e., mature adults) present during one or more mating-years. By relying upon an individual’s genotype as the initial mark and capture of closely related individuals the recapture events, CKMR obviates many challenges inherent to traditional mark–recapture modeling. For example, capture heterogeneity due to individual size is accounted for in CKMR by the overall number of potential kinship pair comparisons; a capture method that selects for larger individuals will result in a sample with fewer potential offspring, thereby decreasing both the total number of comparisons examined and the expected number of kin-pairs. Another confounding influence on traditional mark–recapture is alleviated with CKMR because individuals in a set of samples are only considered once for the purposes of comparison; therefore, effects that occur after initial capture, such as trap shyness, do not influence the overall number of comparisons or the probability of identifying a kin-pair. As unique individuals are only considered once, it becomes possible with CKMR to obtain abundance estimates from sampling designs that would traditionally be difficult or impossible to model with traditional MR including single capture occasions and from any combination of “dead recovery” events (i.e., lethal sampling).

The ability of CKMR to obtain estimates under challenging capture conditions and robustness to many variables that impact traditional MR studies has led this approach to be applied to wide-ranging and difficult to capture marine species such as southern bluefin tuna (Thunnus maccoyii; Bravington et al. 2014a) and Antarctic blue whale
Ruzzante et al. (2019) recently provided the first in situ validation in a freshwater system by comparing CKMR to contemporary Chapman abundance estimates from double-pass electrofishing for a set of stream-dwelling brook trout (Salvelinus fontinalis) populations and concluded that estimates from the two techniques were statistically indistinguishable. Their approach implemented an alternative parameterization of the CKMR statistical model described by Bravington et al. (2016b), using it to estimate population abundance for a set of discrete populations. While that work represents an important validation of CKMR, in this study we explore a suite of additional features useful for fish and wildlife assessments that are available through CKMR.

Population estimation techniques that make efficient use of small sample sizes and flexible study designs can be usefully applied to a wide range of animal populations. Coupled with the rapid decrease in genotyping cost, CKMR represents a technique that may be both efficient and cost-effective for biologists and managers seeking to better understand population demographics. To further assess the potential of this technique, we conducted a side-by-side comparison of estimates obtained by parent–offspring pair-based CKMR and traditional mark–recapture (Robust Design and Chapman models) using data from three sampling efforts on brook trout in an isolated Adirondack lake. The availability of a multi-year dataset from a long-lived population allowed us to evaluate the accuracy of estimates of adult survival and abundance from a CKMR analysis. We show that CKMR can achieve single-sampling occasion abundance estimates, survival estimates with only two sampling occasions, and depending on the ages of sampled specimens, can reconstruct abundances for previous years not directly sampled. These results demonstrate that CKMR is a promising technique for a wide range of systems that include small, spatially discrete populations as well as large, wide-ranging ones.

**METHODS**

**Overview**

Our study involved three main efforts. First, we sampled and tagged brook trout in order to obtain the tissue samples needed for CKMR, as well as the recapture data used in the traditional mark–recapture side of our comparison study. Next, we used genotype data to identify parent–offspring pairs (POPs) within our dataset. This involved three substeps: marker discovery, genotyping, and kinship analysis. Marker discovery is necessary when there are no pre-existing marker sets with sufficient power for parentage analysis. Although both microsatellites and RAD-seq derived single nucleotide polymorphisms (SNPs) have previously been developed for our target species (Perry et al. 2005, Lamaze et al. 2012), the possibility that a bottleneck caused by acid precipitation might have reduced allelic diversity within the population led us to develop a new set of 44 microsatellite markers that were highly polymorphic in our target population to increase our parentage assignment power. Each individual was genotyped for the new marker set through efficient and cost-effective amplicon sequencing. The genotype data were then combined with ancillary information (e.g., age) to determine whether each individual’s parents are included in the dataset. Inclusion of age information allowed impossible parent–offspring combinations to be removed from consideration. Finally, the resulting information (recaptures and POPs) was used to estimate survival and abundance within the traditional mark–recapture and CKMR model frameworks.

**Study area and sampling**

Honnéda Lake is located in the southwestern Adirondack Mountain region of New York State and has a surface area of 312 ha and a maximum depth of 55.8 m (Map 1). This lake was highly susceptible to acid deposition due to its location and small, poorly buffered watershed (Schofield 1965), which led to the near-extirpation the native brook trout population until regulation of air pollution spurred recovery in the early 2000s (Kraft 2019).

From 2015 to 2017, we surveyed brook trout during the first two weeks of May each year using ten Oneida-style trap nets (car 1.2 × 1.2m, lead 30 m, wings 1.6m, mesh 12 mm) and four hoop nets (4 hoops, diameter 1.5 m, mesh 6 mm) deployed at standardized sites throughout the lake (Fig. 1). Nets were allowed to soak for 48–72 h prior to each capture event and were
reset immediately after tending until a total of four sequential capture occasions had been conducted during each May sampling period.

Captured fish were measured, weighed, and scanned for previously deployed tags. Fish 150 mm or larger that had not previously been captured were tagged with a 23-mm HDX PIT tag (OregonRFID) and given a partial-adipose fin clip to allow tag loss to be detected. From 2016 to 2017, we preserved these clips in 95% EtOH and froze them for genetic analysis. Although genetic samples were not taken for 2015, we were able to use tag data to identify individuals that were recaptured and genotyped during subsequent years. For example, a fish that was tagged in 2015, recaptured in 2016, and genotyped from a tissue sample taken during that second occasion would be considered a genotyped fish for both years. Following recovery, we released fish in an offshore location within the general area of the lake where they were originally captured. Sampling events for the historical (pre-study) traditional mark–recapture estimates were similar and consisted of single-occasion fall (October) surveys from 2012 to 2018 using the same gear and sites. Fish captured during these surveys were either tagged with Hallprint (2012–2014) or PIT (2015–2018) tags. We assigned a probable age to each fish based on observed length distributions and length–age information from a set of 831 previously aged fish (Appendix S3: Fig. S1).

Samples were aggregated differently for each of the three estimation models (Fig. 2). For the robust design traditional mark–recapture model, each spring sample occasion and year were treated separately. Samples were compared within each year to estimate abundance for that year, while comparison between years was used to estimate survival. In the CKMR model, samples from all occasions within a year were pooled together. Estimated age and capture year were then used to estimate each individual’s year of conception. This allowed us to identify potential POPs based on relative conception years. For example, in a population that matures at age one, an individual from the 2010 cohort and one from the 2012 cohort would constitute a viable potential POP regardless of the capture year for either individual. Finally, the historical Chapman estimates were based on the pre-study dataset of single-occasion fall captures.

Molecular analysis
Genomic DNA was extracted from all tissue samples using the Truett et al. (2000) HotSHOT
protocol, which combines a 30-minute hot (95°C) incubation with sodium hydroxide followed by treatment with a neutralizing agent (Tris-HCL). We amplified 74 newly identified microsatellites in two multiplex polymerase chain reactions (PCRs) per individual, built a dual-indexed library with all the amplicons for each individual, and sequenced the libraries on an Illumina HiSeq 2500 instrument. The resulting sequence data were processed with a custom bioinformatic pipeline for automated quality filtering and genotyping. In total, 44 loci passed our filtering criteria and were used for kinship analysis. Detailed information about our microsatellite discovery, library preparation, and genotyping procedure is provided in Appendix S1.

**Genetic kinship analysis**

We used the software CERVUS 3.0 to analyze kinship and assign parent–offspring pairs because (1) the program allows for the assignment of parent pairs (i.e., parent–offspring trios),
(2) it accommodates parents of unknown sex, and (3) it allows for genotyping error (Marshall et al. 1998, Kalinowski et al. 2007). As CERVUS can accommodate some missing data, all well-performing loci typed for more than half \((n = 215)\) individuals were included in the parentage analysis \((n = 44; \text{Appendix S1: Table S1 and Table S2})\). We used likelihood ratio scores to assign likely trios and determined 99% confidence critical values using a simulation of 10,000 offspring, 2200 parents (the maximum allowed by the software given our dataset), and an error rate of 0.05 (chosen based on observed error rates, see Appendix S1). We chose to use a 99% confidence threshold as false positives are far more likely to negatively affect CKMR estimates than false negatives. Simulations suggested broad separation between likelihood ratio scores for true parent–offspring combinations and unrelated individuals (Appendix S2: Fig. S1), indicating that our genotype data have high power to correctly infer POPs. Offspring not assigned to a trio were re-evaluated for a single-parent analysis using the same simulation settings and acceptance standards. We validated our CERVUS parentage assignments by analyzing the same genotype data using COLONY (version 2.0.6.5) which uses a full-pedigree likelihood rather than pairwise comparisons (Jones and Wang 2010). For this cross-validation, we used similar settings including an error rate of 0.05, high-precision, and a long run length. As COLONY considers the entire pedigree simultaneously, it was not necessary to consider trios and pairs separately. Assignments from COLONY were nearly identical with the exception of four juveniles where COLONY identified the same most likely candidate parent as CERVUS, but without a high enough probability to assign it as a true pair. This uncertainty was likely a result of not knowing the sex of candidate parents, which can cause COLONY to miss smaller families (Wang 2018). We therefore considered the cross-validation successful and used the CERVUS assignments for all further analyses.

**Population estimation modeling**

Close-kin mark–recapture.—Close-kin mark–recapture is based on the idea that, given two individuals with sufficiently different ages to potentially constitute a POP the probability of the pair actually being a POP is the number of potential parent roles \((e.g.,\, one\, if \, the\, potential\, parent’s\, sex\, is\, known,\, two\, if\, not)\) divided by the total number of potential parents that were alive when the younger individual was spawned. Thus for non-lethal sampling with unknown sexes, the probability of potential parent \(i\) and offspring \(j\) with birth years \(y_i\) and \(y_j\) being a kin-pair \(K\) (i.e., a parent–offspring pair in this case) can be expressed as:

\[
\Pr[K_{ij} = \text{POP}] = \frac{2\mathbb{1}[y_i + \alpha \leq y_j]}{N_{\text{adult}y_j}}
\]

where the indicator function \(I\) determines whether \(i\) was old enough at \(y_j\), given the age-at-maturity \(\alpha\). This indicator is multiplied by two for unknown sexes because each potential offspring has two parents, that is, two chances for a given adult to be their parent. Because identity is based on the potential offspring’s genotype, the true parents can be considered marked regardless of whether they are actually observed. As the probability is conditioned on the fact that both compared individuals are alive at the time of capture (and therefore alive during the potential offspring’s conception year), adult mortality does not affect the estimate. The exception to this is when the potential parent was sampled prior to the potential offspring’s conception year, in which case the numerator must be discounted by the adult survival rate \(\varphi\) (i.e., by the chance that the potential parent died post-sampling and pre-sampling the potential offspring’s conception). This takes the form:

\[
\Pr[K_{ij} = \text{POP}] = \frac{2\mathbb{1}[y_i + \alpha \leq y_j]}{N_{\text{adult}y_j}} \times \begin{cases} 1 & t_i > y_j \\ \varphi^{y_j-t_i} & t_i < y_j \end{cases}
\]

Studies that include samples from more than one year can therefore be used to estimate adult survival as long as least one observed POP includes a parent that was captured before its offspring’s birth. Extended to all potential POPs, this probability function follows a binomial distribution (Hillary et al. 2018). Close-kin mark–recapture can also accommodate the inclusion of covariates that may influence an adult’s reproductive potential. For example, size is often closely related to fecundity and larger individuals may therefore be more likely to be potential
parents (Ruzzante et al. 2019). This approach is particularly necessary when the gear used to capture parents may be highly size selective, as CKMR assumptions are violated if a variable influences both the probability of sampling and parent and that parent’s expected reproductive output (Bravington et al. 2016b, Skaug 2017). However, selectivity of entrapment gear does not change with length once a fish is susceptible to capture (Kraft and Johnson 1992, Millar and Fryer 1999) and all adult fish were susceptible to capture by the gear used in this study. We therefore did not include any reproductive covariates in our model.

In our study, we used length-based estimated age and year of capture to determine the year class (i.e., spawning year) of each individual. Because potential POPs are determined by year class, fish can be compared between any number of sampling events within and between years. Thus for the analysis, we pooled fish from all surveys and years. Our offspring dataset was restricted to fish age-3 and younger at the time of capture, as we were less confident in our ability to assign length-based ages beyond age-4 (i.e., we could not confidently discriminate between an age-4 individual and an age-5 individual; Appendix S3: Fig. S1). This restriction reduced both the number of potential parent–offspring combinations and the expected number of identified POPs for earlier year classes, reducing the precision of estimates for those years but leaving them unbiased. This choice also allowed us to use the parent-centric formulation described by Bravington et al. (2016b) that is simpler but relies on predetermined ages rather than the offspring-centric approach described by Skaug (2017) and used by Ruzzante et al. (2019). Based on gonad development of previously captured and aged fish from fall samples, we set \( \alpha = 1 \) because 93% of the age-1 fish caught during the spawning season were reproductively mature. We compared each individual in this restricted dataset (potential offspring) to each individual in the full dataset (potential parent) to obtain our total number of combinations for the model.

Robust design mark–recapture.—We used Pollock’s closed robust design which uses datasets consisting of multiple closed capture occasions separated by larger open intervals to estimate both yearly abundance and survival from traditional tagging data (Pollock 1982, Kendall et al. 1995). Beyond abundance and survival, the model has been extended to accommodate generic capture heterogeneity among individuals through a temporary emigration process, that is, the probability that an individual leaves and re-enters the capturable population during the study period (Kendall et al. 1997), and to include various divisions of capture probability parameters to accommodate systematic capture heterogeneity such as trap shyness. For our model, we assumed that no effect of capture on capture probabilities and estimated a separate capture probability for each sampling occasion.

Chapman estimator.—To estimate historical annual abundance estimates for comparison to CKMR estimates for years preceding those for which robust design mark–recapture estimates could be generated (2012–2014), we used a separate dataset of yearly single-occasion fall trap net surveys, as described previously, in conjunction with the standard Chapman modification of the Lincoln-Petersen estimator (Lincoln 1930, Chapman 1951). The Chapman estimator carries an assumption of population closure, which may not be satisfied in our present study due to mortality and the potential for recruitment between years. However, under the assumption that mortality is equal across the tagged and untagged population and that immigration or emigration does not occur, the estimator can provide valid abundance estimates under specific population closure violations, providing a census at the time of the first sampling occasion for any pair of sampling occasions (Seber 1982). Long-term continued capture and tagging efforts on the Honnedaga Lake brook trout population indicate fish are robust to capture effects, with specimens being caught multiple times over sampling years. Thus, we believe there is little to no additional mortality associated with fish sampling and that the assumption of equal mortality among tagged and untagged fish is plausible. Similarly, as Honnedaga Lake is isolated, migration of fish into the lake is not feasible. Recruitment of fish from the Honnedaga Lake population between any two specific years would be a form of immigration; however, to eliminate the potential for recruitment between any pair of yearly sampling occasions that could bias estimates, we limited our secondary sample to fish that were at least
50 mm longer than the minimum tagging size for the primary sample (150 mm). Thus, consistent with our robust design estimates, abundance estimates were developed specifically for the portion of the Honnedaga Lake brook trout population >150 mm in size.

**Model implementation**

We used Bayesian implementation for all three models. Close-kin mark–recapture and Chapman models were run using rjags (Plummer et al. 2018), a Program R-based interface for the JAGS library (Plummer 2003), with the jagsUI wrapper (Kellner 2018) to allow for parallel processing which greatly increased analysis speed. Three MCMC chains were used and each run for 10,000 iterations. Our Bayesian robust design model used the code provided by Rankin et al. (2016) adapted for our data and, as with our CKMR model, was estimated in Program R using JAGS (Plummer 2003) through the rjags interface (Plummer et al. 2018) with dclone used for parallel processing (Solymos 2019). Three MCMC chains were run for 100,000 iterations each.

**RESULTS**

Sampling during the spring 2015–2017 field seasons resulted in capture and tagging of 716 unique fish (Table 1). In total, 44 of our newly developed microsatellites passed our data filtering criteria and were genotyped in the 384 individuals caught during the 2016–2017 spring sampling occasions. The mean number of successfully genotyped loci per individual was 36.2 loci (82.3%). The number of alleles per locus ranged from 2 to 11 (average of 7.07 alleles) with heterozygosity between 0.29 and 0.95 (average of 0.72). Mean observed genotyping error rates were 0.02 for known parent–offspring trios and 0.04 for replicate samples. More details of genotyping procedures and performance are available in Appendix S1.

Of the 384 genotyped individuals, 257 were young enough (age ≤ 3) for consideration as potential offspring. This yielded 98,431 potential combinations of which 67,840 had an age difference large enough for consideration as a POP. Parentage assignment with CERVUS identified 72 parent–offspring pairs from among the 384 specimens collected during 2016–2017 spring sampling, including two parent–offspring trios (offspring, mother, and father) that met the critical value for 99% confidence.

The presence of at least two POPs for each potential offspring cohort year allowed CKMR estimates of fall adult abundance for 2011–2015. All parameters for the CKMR model converged successfully with Gelman-Rubin $R$ statistics <1.1 (Table 2; Gelman and Rubin 1992). Estimates for

| Occasion        | Individuals captured† | Recaptures (within year) | Recaptures (total) | Individuals genotyped |
|-----------------|-----------------------|--------------------------|--------------------|-----------------------|
| 6 May 2015      | 90                    | N/A                      | N/A                | 16‡                   |
| 8 May 2015      | 86                    | 7                        | 7                  | 19‡                   |
| 11 May 2015     | 192                   | 8                        | 15                 | 44‡                   |
| 13 May 2015     | 168                   | 12                       | 21                 | 28‡                   |
| 4 May 2016      | 78                    | N/A                      | 3                  | 68                    |
| 6 May 2016      | 35                    | 2                        | 3                  | 33                    |
| 9 May 2016      | 84                    | 1                        | 11                 | 69                    |
| 11 May 2016     | 27                    | 4                        | 5                  | 25                    |
| 28 April 2017   | 32                    | N/A                      | 4                  | 28                    |
| 1 May 2017      | 38                    | 0                        | 1                  | 36                    |
| 3 May 2017      | 29                    | 1                        | 2                  | 25                    |
| 5 May 2017      | 20                    | 1                        | 1                  | 20                    |

**Note:** Also included is the number of individuals caught during each occasion that were later genotyped.

† Number of individuals captured reflects only those that were large enough to tag (≥150 mm).

‡ No tissue samples were taken during 2015. Number reflects tagged individuals that were recaptured and genotyped during subsequent years.
abundance during the study period displayed an upward trend from 2011 to 2014 with tight 95% confidence intervals. Chapman estimates for fall surveys over the 11-yr 2012–2018 period (Table 3) paralleled those for CKMR, displaying a similar upward trend (Fig. 3). The robust design traditional mark–recapture model also successfully converged for all parameters and provided similar estimates of abundance for 2015–2016 (Table 4). Although convergence was successful for the 2017 abundance estimate, the 95% confidence interval was extremely wide (2889–14,258) due to the low number of closed recaptures during the 2017 sampling occasions (Fig. 3).

An additional nine POPs were identified where the parent was sampled one year prior to their offspring’s spawning, allowing survival of mature fish (age-2 or greater) to be estimated using CKMR from these two-year data. The resulting estimate of annual adult survival was 0.80 (95% CI: 0.44–0.99). Traditional robust design estimates of adult survival were similar with a mean of 0.75 (95% CI: 0.49–0.99).

In addition to the analysis using the full dataset, CKMR estimates were successfully obtained for each of the single-year sampling datasets in isolation. Samples from 2015 were able to provide fall abundance estimates for 2011 and 2012, while samples from 2016 provided estimates for falls 2011, 2013, and 2014. Samples from the spring 2017 surveys could only provide an abundance estimate for fall 2015. Single-year abundance estimates were comparable to those obtained from the full dataset in both mean and 95% CI. However, survival estimates require at least one POP where the parent was sampled prior to the offspring’s conception; therefore, estimates of survival cannot be obtained from any single-year CKMR.

**DISCUSSION**

Our estimates of abundance for fall 2014–2015 and annual survival from the CKMR model were comparable to those for spring 2015–2016 obtained from the robust design traditional mark–recapture model despite relying upon almost entirely different sets of data. While correspondence between the models does not necessarily mean that either is free from bias, this broad agreement suggests that differences in bias between the two models are minimal. For example, 95% confidence intervals from the robust design model overlapped the CKMR estimates for both of the one-year periods for which estimates were available from the two models. In both cases, models suggested that ~2500 adult brook trout inhabit the study lake. The two models also provided median posterior estimates of

---

**Table 2. Parameter estimates for the Close-Kin Mark–Recapture model including adult abundance for all estimable years ("N-[year]") and annual survival with parenthetical values representing 95% credible intervals.**

| Parameter | No. comparisons | No. POPs | Estimate                  | R-hat |
|-----------|-----------------|----------|---------------------------|-------|
| N-2011    | 1066            | 2        | 602 (136–1733)            | 1.00  |
| N-2012    | 6370            | 16       | 1016 (479–1978)           | 1.00  |
| N-2013    | 14,490          | 19       | 1788 (1013–2875)          | 1.00  |
| N-2014    | 26,606          | 25       | 2126 (1410–3039)          | 1.03  |
| N-2015    | 17,850          | 10       | 2093 (1321–3119)          | 1.00  |
| Survival  | 5200            | 7        | 0.811 (0.46–0.99)         | 1.00  |

Notes: Also provided are the number of potential parent–offspring comparisons and number of identified parent–offspring pairs (POPs) for each parameter. R-hat refers to the Gelman-Rubin diagnostic which indicates poor MCMC convergence when values are above 1.1 (Gelman and Rubin 1992).

**Table 3. Summary of fall sampling returns for 2012–2018 including the number of individuals caught during each occasion as well as the number of recaptures from the previous year.**

| Year | Individuals captured† | Individuals recaptured |
|------|------------------------|------------------------|
| 2012 | 312                    | 25                     |
| 2013 | 217                    | 32                     |
| 2014 | 314                    | 36                     |
| 2015 | 526                    | 71                     |
| 2016 | 384                    | 79                     |
| 2017 | 330                    | 35                     |
| 2018 | 246                    | 24                     |

† Number of individuals captured reflects only those that were large enough to tag (≥150 mm).
adult survival with very similar (albeit low) precision. The estimated mean annual survival rates were approximately 0.80, consistent with a population inhabiting a lake without other piscine predators, few avian predators and ample thermal refuge. We consider it notable that the CKMR and robust design estimates had similar precision, though the CKMR model used a much lower sample size (Table 1). Tissue samples were not collected during spring 2015, which if available could have provided information about parents of younger age classes caught in 2016 and 2017.

Historical estimates of abundance from CKMR compared favorably with those obtained from between-year Chapman estimators using individuals collected during fall sampling occasions (Fig. 3). In both cases, an upward trend in abundance from 2012 to 2015 was evident and—while Chapman estimates were higher than those given by the CKMR model—the 95% confidence intervals of each overlapped. It is important to note that since fall surveys captured many more fish than did spring surveys (Tables 1 and 3), the Chapman estimates are based on a much larger number of fish than are the CKMR estimates.

**Table 4. Parameter estimates for the Robust Design model with number of potential parent–offspring comparisons and number of identified parent–offspring pairs for each parameter.**

| Parameter | Estimate         | R-hat |
|-----------|------------------|-------|
| N-2015    | 1729 (1042–3268) | 1.01  |
| N-2016    | 2979 (2289–3907) | 1.00  |
| N-2017    | 6141 (2889–14742)| 1.00  |
| Survival  | 0.74 (0.49–0.97) | 1.00  |

*Note: Parenthetical values represent 95% credible intervals.*

Fig. 3. Estimates of abundance and annual survival from the full-sample close-kin mark–recapture (circles) and Pollock’s closed robust design (squares) models for spring survey data as well as estimates of abundance from the fall survey-based between-year Chapman estimators (triangles).
Additionally, between-year Chapman estimates required several key assumptions: no migration, no difference in survival between tagged and untagged individuals, no negative effect of capture (i.e., trap shyness), and a 50 mm cutoff for the second capture occasion fish to eliminate incorporating new recruits in the data evaluated. Although decades of collective experience surveying this population lead us to consider these assumptions valid, we recognize that some degree of systemic violation may persist and that these assumptions are difficult to evaluate.

The ability of CKMR models to provide abundance estimates for years prior to sampling and thereby detect trends is a key advantage over traditional models, especially when applied to long-lived species. This contrasts with traditional mark–recapture models relying upon closed capture occasions that only provide estimates of abundance during the closed period. By contrast, CKMR models are able to estimate abundance for all breeding years when at least one POP was detected. Although this restricts CKMR estimates to the most recent breeding season, it also provides the ability to retrospectively assess changes in abundance. We were able to obtain historical abundance estimates from both our multi-year CKMR analysis and for analyses restricted to data from each of our sampling years in isolation. While the multi-year analysis provided the most comprehensive set of historical estimates (five years), our single-year analyses were able to estimate abundance for at least one and as many as three previous years (Appendix S2: Fig. S2).

This includes a relatively precise estimate of abundance for 2015 derived solely from 2017 captures, despite that fact that recapture data from 2017 were too sparse to estimate abundance using our robust design model. In this study of a relatively short-lived species, we restricted our use of offspring data to age-3 and younger fish due to low confidence in length–age assignment beyond age-4. This restricted our earliest estimate of abundance to 2012 (e.g., from age-3 fish caught in 2015). However, use of accurate aging techniques coupled with a longer-lived species could provide much longer-term estimates of historical abundance, making it possible to study the impact of historical drivers of population abundance based solely upon contemporary sampling. Potential also exists for the development of hybrid models that integrate data from both kinship and true recapture, boosting the precision of contemporary estimates while still providing information about historical abundances.

Close-kin mark–recapture requires genotyping, which introduces additional time, labor, and financial costs relative to mark–recapture methods that rely upon physical tags. We estimate that our genotyping using the microsatellite sequencing technique costs approximately US$8 per sample (~$6.50 excluding microsatellite discovery). However, genotyping techniques are widely recognized to be advancing rapidly (Meek and Larson 2019) and techniques such as GT-seq can currently be performed for as low as $4 per sample (Campbell et al. 2015). This approaches the cost of physical tags per fish evaluated, especially PIT tags that currently retail for $1–2 per unit. Nevertheless, studies using CKMR in the near-future will likely be more expensive than those employing traditional mark–recapture on a per-captured-individual basis. However, we expect that CKMR may still be cheaper than traditional mark–recapture studies in many instances due to its flexibility. Samples for genetic analysis can be preserved almost indefinitely with proper storage and then analyzed when interest or resources allow. CKMR sampling can also piggyback on other work (or use previously collected samples) and, without a requirement for true recapture, study designs employing as little as one capture occasion become feasible.

Close-kin mark–recapture's efficient use of sampling effort and robustness to many problems associated with capture heterogeneity offer the potential to estimate abundance and survival from populations and datasets for which this was previously infeasible. While the individuals included in our CKMR model were collected using consistent gear, effort, and timing, this was only necessary for accurate estimation using the traditional mark–recapture model. Because CKMR does not rely upon (and in fact ignores) true recapture, effects such as size-selectivity are only important inasmuch as they affect the overall age structure of the dataset. This provides a great deal of flexibility to accommodate opportunistic sampling with poorly standardized capture effort or even deliberate sampling with different gear types to target coverage over a
broader age structure or to adapt to changing personnel or survey conditions.

For example, we present three potentially desirable and pragmatic study assessment scenarios that are poorly suited for traditional mark–recapture but are feasible with CKMR methods, opening up opportunities to study a diverse range of animal populations:

1. Assessments employing a single, high-effort survey. This design is suited for evaluating populations where access is difficult or in situations where factors such as seasonality, personnel constraints, and expense limit the ability to repeatedly survey a population over time. In order for this type of study to be effective, survey gear must be capable of capturing a variety of ages of the target animal in order to yield a suitable number of POP comparisons. Unlike the other two examples listed below, this approach necessarily restricts the dataset to a single year and therefore can provide only abundance estimates, rather than both abundance and survival.

2. Assessments employing a large number of low-effort surveys. CKMR’s robustness to variation in effort and size-selectivity makes it ideal for situations that rely upon genetic samples collected from volunteers and therefore include extensive between-survey variation in effort. Examples include samples collected by recreational hunters and anglers for which effort and gear efficiency are unknown or poorly controlled.

3. Assessments using a combination of highly size-selective gears. As an example, genetic data collected from a coverboard or pitfall trap survey targeting adult amphibians could be combined with aquatic surveys of larval individuals. Traditional mark–recapture methods would be of little use in this scenario as one could not expect recaptures between the two gears; by contrast, CKMR methods using POPs would be highly effective. This approach may also be usefully combined with design #2, for example, samples from an adult-selective recreational fishery could be combined with those from a gear type optimized for juvenile individuals such as a larval fish survey.

Effective application of CKMR hinges upon the ability to accurately estimate relative ages and kinship status within a population. Techniques for the former are well established for many species, particularly a fast-growing, pulse-breeding species such as brook trout. However, most aging methods rely on whole organism measurements (e.g., length) or analysis of specific structures (e.g., otoliths, teeth), both of which require additional data or samples. Recently developed epigenetic aging techniques offer potential improvements that could allow determination of both age and kinship based solely upon tissue samples (Polanowski et al. 2014). This method has been successfully used for CKMR modeling of Antarctic blue whales (Bravington et al. 2014b), but additional work is required before these tools can be applied to other species. Simulation-based studies of how imprecise or biased age estimates effect the performance of CKMR could also be useful, especially for researchers seeking to apply the technique to new systems. While methods for genotyping and kinship analysis are well established, further advances would enable wider application of CKMR. Perhaps the greatest potential benefit from advances in genotyping would be further cost reduction, which could reduce the cost of evaluating each individual to the same or less than physical tagging. But even using current aging and genotyping techniques, this study provides empirical demonstration of the potential widespread utility of CKMR as a useful alternative to traditional mark–recapture methods and for extending abundance and survival estimation to populations for which traditional mark–recapture assessment is untenable.

ACKNOWLEDGMENTS

We thank K. Jirka, E. Randall, and D. Josephson for their assistance with field data collection, as well as private landowners for access to the study lake, S. Bogdanowicz provided extensive advice regarding our genotyping methodology, H. Borchardt-Wier facilitated our genetic laboratory work, and B. McFall assisted with DNA extraction. Any use of trade, firm or product names is for descriptive purposes only and does not imply endorsement by the US Government.
LITERATURE CITED

Amstrup, S. C., T. L. McDonald, and B. F. Manly. 2010. Handbook of capture-recapture analysis. Princeton University Press, Princeton, New Jersey, USA.

Anderson, C. S. 1995. Measuring and correcting for size selection in electrofishing mark–recapture experiments. Transactions of the American Fisheries Society 124:663–676.

Bravington, M. V., P. M. Grewe, and C. R. Davies. 2014a. Fishery-independent estimate of spawning biomass of Southern Bluefin Tuna through identification of close-kin using genetic markers. FRDC Report 2007/034. CSIRO, Canberra, Australian Capital Territory Australia.

Bravington, M. V., P. M. Grewe, and C. R. Davies. 2016a. Absolute abundance of southern bluefin tuna estimated by close-kin mark-recapture. Nature Communications 7:13162.

Bravington, M. V., S. N. Jarman, and H. J. Skaug. 2014b. Antarctic Blue Whale surveys: augmenting via genetics for close-kin and ordinal age. International Whaling Commission, Cambridge, UK.

Bravington, M. V., H. J. Skaug, and E. C. Anderson. 2016b. Close-Kin Mark-recapture. Statistical Science 31:259–274.

Campbell, N. R., S. A. Harmon, and S. R. Narum. 2015. Genotyping-in-Thousands by sequencing (GT-seq): a cost effective SNP genotyping method based on custom amplicon sequencing. Molecular Ecology Resources 15:855–867.

Chapman, D. G. 1951. Some properties of the hypergeometric distribution with applications to zoological sample censuses. University of California Publications Statistics 1:131–160.

Cooch, E. 2008. Program MARK, “A gentle introduction”. http://www.phidot.org/software/mark/docs/book/

Cormack, R. M. 1964. Estimates of survival from the sighting of marked animals. Biometrika 51:429–438.

Geis, A. D. 1955. Trap response of the cottontail rabbit and its effect on censusing. The Journal of Wildlife Management 19:466–472.

Gelman, A., and D. B. Rubin. 1992. Inference from iterative simulation using multiple sequences. Statistical Science 7:457–472.

Hillary, R. M., et al. 2018. Genetic relatedness reveals total population size of white sharks in eastern Australia and New Zealand. Scientific Reports 8:2661.

Jolly, G. M. 1965. Explicit estimates from capture-recapture data with both death and immigration-stochastic model. Biometrika 52:225–247.

Jones, O. R., and J. Wang. 2010. COLONY: a program for parentage and sibship inference from multilocus genotype data. Molecular Ecology Resources 10:551–555.

Kalinowski, S. T., M. L. Taper, and T. C. Marshall. 2007. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Molecular Ecology 16:1099–1106.

Kellner, K. 2018. jagsUI: a Wrapper Around “rjags” to Streamline “JAGS” Analyses. R Foundation for Statistical Computing, Vienna, Austria.

Kendall, W. L. 1999. Robustness of closed capture-recapture methods to violations of the closure assumption. Ecology 80:2517–2525.

Kendall, W. L., J. D. Nichols, and J. E. Hines. 1997. Estimating temporary emigration using capture-recapture data with Pollock’s robust design. Ecology 78:563–578.

Kendall, W. L., K. H. Pollock, and C. Brownie. 1995. A likelihood-based approach to capture-recapture estimation of demographic parameters under the robust design. Biometrics 51:293–308.

Kraft, C. E. 2019. Adirondack brook trout and acid rain: environmental legislation fosters successful restoration. Pages 155–180 in C. C. Krueger, W. W. Taylor, and S. J. Youn, editors. From catastrophe to recovery: stories of fishery management success. American Fisheries Society, Washington, D.C., USA.

Kraft, C. E., and B. L. Johnson. 1992. Fyke-Net and Gill-Net Size selectivities for Yellow Perch in Green Bay, Lake Michigan. North American Journal of Fisheries Management 12:230–236.

Lamaze, F. C., C. Sauvage, A. Marie, D. Garant, and L. Bernatchez. 2012. Dynamics of introgressive hybridization assessed by SNP population genomics of coding genes in stocked brook charr (Salvelinus fontinalis). Molecular Ecology 21:2877–2895.

Lebreton, J.-D., T. Almeras, and R. Pradel. 1999. Competing events, mixtures of information and multi-stratum recapture models. Bird Study 46(sup1): S39–S46.

Lincoln, F. C. 1930. Calculating waterfowl abundance on the basis of banding returns. US Department of Agriculture, 118, Washington, D.C., USA.

Link, W. A. 2003. Nonidentifiability of population size from capture-recapture data with heterogeneous detection probabilities. Biometrics 59:1123–1130.

Marshall, T. C., J. Slate, L. E. B. Kruuk, and J. M. Pemberton. 1998. Statistical confidence for likelihood-based paternity inference in natural populations. Molecular Ecology 7:639–655.
Meek, M. H., and W. A. Larson. 2019. The future is now: amplicon sequencing and sequence capture usher in the conservation genomics era. Molecular Ecology Resources 19:795–803.

Millar, R. B., and R. J. Fryer. 1999. Estimating the size-selection curves of towed gears, traps, nets and hooks. Reviews in Fish Biology and Fisheries 9:89–116.

Perry, G. M. L., T. L. King, J. St.-cyr, M. Valcourt, and L. Bernatchez. 2005. Isolation and cross-familial amplification of 41 microsatellites for the brook charr (*Salvelinus fontinalis*). Molecular Ecology Notes 5:346–351.

Pine, W. E., J. E. Hightower, L. G. Coggins, M. V. Laurretta, and K. H. Pollock. 2012. Design and analysis of tagging studies. American Fisheries Society, Bethesda, Maryland, USA.

Plummer, M. 2003. JAGS: a program for analysis of Bayesian graphical models using Gibbs sampling. http://genome.jouy.inra.fr/applibugs/applibugs.07_11_08.plummer.pdf

Plummer, M., A. Stukalov, and M. Denwood. 2018. rjags: Bayesian graphical models using MCMC. R Project for Statistical Computing, Vienna, Austria.

Polanowski, A. M., J. Robbins, D. Chandler, and S. N. Jarman. 2014. Epigenetic estimation of age in humpback whales. Molecular Ecology Resources 14:976–987.

Pollock, K. H. 1980. Capture-recapture models: a review of current methods, assumptions, and experimental design. North Carolina State University Department of Statistics, Raleigh, North Carolina, USA.

Pollock, K. H. 1982. A capture-recapture design robust to unequal probability of capture. Journal of Wildlife Management 46:752–757.

Rankin, R. W., K. E. Nicholson, S. J. Allen, M. Krützen, L. Bejder, and K. H. Pollock. 2016. A full-capture hierarchical Bayesian Model of Pollock’s closed robust design and application to dolphins. Frontiers in Marine Science 3:25.

Royle, J. A., R. B. Chandler, R. Sollmann, and B. Gardner. 2013. Spatial capture-recapture. Academic Press, Cambridge, Massachusetts, USA.

Ruzzante, D. E., G. R. McCracken, B. Ferland, J. MacMillan, D. Notte, C. Buhariwalla, J. Mills Flemming, and H. Skaug. 2019. Validation of close-kin mark-recapture (CKMR) methods for estimating population abundance. Methods in Ecology and Evolution 10:1445–1453.

Schofield, C. L. 1965. Water quality in relation to survival of Brook Trout, *Salvelinus fontinalis* (Mitchell). Transactions of the American Fisheries Society 94:227–235.

Seber, G. A. F. 1965. A note on the multiple-recapture census. Biometrika 52:249–259.

Seber, G. A. F. 1970. The effects of trap response on tag recapture estimates. Biometrics 26:13–22.

Seber, G. A. F. 1982. The estimation of animal abundance and related parameters. Blackwell Press, Caldwell, New Jersey, USA.

Skaug, H. J. 2017. The parent–offspring probability when sampling age-structured populations. Theoretical Population Biology 118:20–26.

Solymos, P. 2019. dclone: Data cloning and MCMC tools for maximum likelihood methods. R Project for Statistical Computing, Vienna, Austria.

Sutherland, W. J. 2006. Ecological census techniques: a handbook. Cambridge University Press, Cambridge, UK.

Taberlet, P., L. P. Waits, and G. Luikart. 1999. Noninvasive genetic sampling: look before you leap. Trends in Ecology & Evolution 14:323–327.

Truett, G. E., P. Heeger, R. L. Mynatt, A. A. Truett, J. A. Walker, and M. L. Warman. 2000. Preparation of PCR-quality mouse genomic DNA with hot sodium hydroxide and tris (HotSHOT). BioTechniques 29:52–54.

Wang, J. 2018. User’s guide for software COLONY Version 2.0.6.5. http://www.zsl.org/science/research-projects/software

Williams, B. K., J. D. Nichols, and M. J. Conroy. 2002. Analysis and management of animal populations. Academic Press, Cambridge, Massachusetts, USA.

Woods, J. G., D. Paetkau, D. Lewis, B. N. McLellan, M. Proctor, and C. Stroheck. 1999. Genetic tagging of free-ranging Black and Brown Bears. Wildlife Society Bulletin 1973–200627:616–627.

Zipkin, E. F., and S. P. Saunders. 2018. Synthesizing multiple data types for biological conservation using integrated population models. Biological Conservation 217:240–250.

**Supporting Information**

Additional Supporting Information may be found online at: http://onlinelibrary.wiley.com/doi/10.1002/ecs2.3259/full