Conservation Biology

Contributed Paper

Reliability of Indicators of Decline in Abundance

ERIN J. PORSZT,*§ RANDALL M. PETERMAN,* NICHOLAS K. DULVY,† ANDREW B. COOPER,* AND JAMES R. IRVINE‡

*School of Resource and Environmental Management, Simon Fraser University, 8888 University Drive, Burnaby, British Columbia V5A 1S6 Canada

†Department of Biological Sciences, Earth to Ocean Research Group, Simon Fraser University, 8888 University Drive, Burnaby, British Columbia V5A 1S6 Canada

‡Fisheries and Oceans Canada, Pacific Biological Station, 3190 Hammond Bay Road, Nanaimo, British Columbia V9T 6N7 Canada

Abstract: *Although there are many indicators of endangerment (i.e., whether populations or species meet criteria that justify conservation action), their reliability has rarely been tested. Such indicators may fail to identify that a population or species meets criteria for conservation action (false negative) or may incorrectly show that such criteria have been met (false positive). To quantify the rate of both types of error for 20 commonly used indicators of declining abundance (threat indicators), we used receiver operating characteristic curves derived from historical (1938–2007) data for 18 sockeye salmon (*Oncorhynchus nerka*) populations in the Fraser River, British Columbia, Canada. We retrospectively determined each population's yearly status (reflected by change in abundance over time) on the basis of each indicator. We then compared that population's status in a given year with the status in subsequent years (determined by the magnitude of decline in abundance across those years). For each sockeye population, we calculated how often each indicator of past status matched subsequent status. No single threat indicator provided error-free estimates of status, but indicators that reflected the extent (i.e., magnitude) of past decline in abundance (through comparison of current abundance with some historical baseline abundance) tended to better reflect status in subsequent years than the rate of decline over the previous 3 generations (a widely used indicator). We recommend that when possible, the reliability of various threat indicators be evaluated with empirical analyses before such indicators are used to determine the need for conservation action. These indicators should include estimates from the entire data set to take into account a historical baseline.*

Keywords: COSEWIC, IUCN, receiver operating characteristic (ROC), sockeye salmon, threat indicators

Confiabilidad de Indicadores de Declinación de Abundancia

Resumen: *Aunque existen muchos indicadores de riesgo (i.e., sí las poblaciones o especies cumplen con criterios para justificar acciones de conservación), su confiabilidad ha sido probada pocas veces. Dichos indicadores pueden fallar al identificar que una población o especie cumple con criterios para acciones de conservación (negativo falso) o pueden mostrar incorrectamente que tales criterios se han cumplido (positivo falso). Para cuantificar la tasa de ambos tipos de error para 20 indicadores de declinación de abundancia (indicadores de amenaza) utilizados comúnmente, utilizamos curvas de características de operación de receptores derivadas de datos históricos (1937–2008) de 18 poblaciones de salmón (*Oncorhynchus nerka*) en el Río Fraser, Columbia Británica, Canadá. Retrospectivamente determinamos el estatus anual de cada población (reflejado en cambios en la abundancia en el tiempo) con base en cada indicador. Posteriormente comparamos el estatus de la población en un año determinado con el estatus de años subsecuentes (determinado por la magnitud de la declinación en abundancia en esos años). Para cada población de salmón, calculamos la frecuencia en que cada indicador de estatus pasado era igual al estatus subsecuente. Ningún indicador de amenaza proporcionó estimaciones de estatus libres de error, pero los indicadores que reflejaron*

§Current address: Fisheries and Oceans Canada, South Coast Area Office, 3225 Stephenson Point Road, Nanaimo, British Columbia V9T 6N7 Canada, email erin.porszt@dfo-mpo.gc.ca

Paper submitted June 1, 2011; revised manuscript accepted February 17, 2012.

Conservation Biology, Volume 26, No. 5, 894–904

DOI: 10.1111/j.1523-1739.2012.01882.x

la extensión (i.e., magnitud) de la declinación en abundancia pasada (mediante comparación de la abundancia actual con la abundancia histórica de referencia) tendieron a reflejar de mejor manera el estatus en años anteriores que la tasa de declinación en las 3 generaciones previas (un indicador ampliamente utilizado). Recomendamos que, cuando sea posible, se evalúe la confiabilidad de varios indicadores de amenaza con análisis empíricos antes de que esos indicadores sean utilizados para determinar la necesidad de acciones de conservación. Estos indicadores deben incluir estimaciones a partir del total de datos para considerar una referencia histórica.

Palabras Clave: característica de operación del receptor, COR, COSEWIC, indicadores de amenaza, IUCN, *Oncorhynchus nerka*

Introduction

To assign species or populations to categories of extinction risk, many agencies worldwide use the International Union for the Conservation of Nature's (IUCN) classification system (IUCN 2006). In the IUCN scheme, values of indicators, such as the rate of change in abundance, spatial distribution, and absolute abundance, are compared with a priori threshold values of that indicator (IUCN 2006). Indicators of declining abundance of a population, which we focus on here and call threat indicators, are especially widely used for evaluating, monitoring, and reporting on the status and trend of species or populations (Miller et al. 2006; Mace et al. 2008). Some organizations, for example, the Committee on the Status of Endangered Wildlife in Canada (COSEWIC), modify IUCN's criteria slightly and use the results of their classifications to recommend whether conservation action should be taken under the Canadian Species at Risk Act of 2005.

Despite the widespread use of indicators to inform decisions about designations of species' status, the reliability of these indicators is still uncertain. Mace et al. (2008, p. 1438) note that the transformation of data into indicators has ". . . unfortunately received much less critical external review than have the numerical thresholds in the criteria, although we believe they are often more significant." We conducted a case study to help address this need to test empirically the reliability of numerous indicators, including some suggested by IUCN (2006), Mace et al. (2002), and the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) (2010).

In recent years, there has been considerable debate over reliability of indicators of abundance in exploited populations of marine fishes (e.g., Hutchings 2001; Dulvy et al. 2005; Rice & Legacé 2007). A major concern is that abundance estimates are always imperfect, and the consequences of their being wrong are high (Peterman 1990). For instance, if a population is incorrectly classified as having a high chance of extirpation (a false positive), managers may reduce allowable catches, and short-term social and economic benefits may be reduced unnecessarily. Such false positives may also result in agencies taking unnecessary remedial actions with their limited funds. Conversely, if an exploited population erroneously is not categorized as high risk (a false negative), corrective action may not be taken and may result in increased probability of extinction, decreased probability of recovery, and reduced long-term social and economic benefits. Thus, methods for assessing threat must be reliable (i.e., have a high probability of correctly classifying a population's status) regardless of whether a population has a low or high probability of extinction (Table 1). Most research has focused on the power to detect declines and avoid false positives (e.g., Dulvy et al. 2006; Regan et al. 2009; Wilson et al. 2011), but a full analysis of the success and failure rates of a wide range of indicators of abundance has not been conducted.

We provide an example of how the reliability of such indicators can be evaluated with receiver operating characteristic (ROC) curves, which are widely used in ecology, conservation, and physical sciences (Pearce & Ferrier 2000; Burgman 2005; Baxter & Possingham 2011), as well as medicine (Hibberd & Cooper 2008). Receiver operating characteristic curves account for rates of occurrence of true and false positives and true and false negatives and produce an integrated measure of reliability.

We used sockeye salmon (*Oncorhynchus nerka*) as a case study because long time series of population-level data are available, and there is widespread concern about the status of socially and economically valuable salmon populations that occur from Korea to California (U.S.A.). We sought to determine which quantitative indicators of time trends in abundance of spawning sockeye salmon populations from the Fraser River, British Columbia, Canada are most reliable. Indicators other than trends in abundance exist for assessing the status of populations and species, and quantitative results from our case study will not necessarily apply to other taxa. Our aim was to devise a method for empirical evaluation of the reliability of indicators of species' status before those indicators are used to inform decisions.

Methods

The various indicators used to determine level of threat of extinction are in effect alternative measures or symptoms of extinction risk (Mace et al. 2008). We compared

Table 1. Terms used in this paper for 4 possible outcomes when the assessment of status of a population unit in a certain period before a given year (past status) is compared with the status in the period after that year (subsequent status).

	Past status[b]	
	not declining	*declining*
Subsequent status[a]	*fail to reject H_o*[c] *(no triggering event)*	*reject H_o (triggering event*[d]*)*
Not declining (i.e., H_0 true) subsequent trend not downward	true negative, correct conclusion [1-α][e]	false positive, incorrect conclusion [type I error, α][e]
Declining (i.e., H_0 false) subsequent trend is downward	false negative, incorrect conclusion [type II error, β][e]	true positive, correct conclusion [power = 1−β][e]

[a] *Status of a conservation unit derived from the unit's adult abundance in the years subsequent to the year when past status was evaluated. A declining conservation unit has a decreasing subsequent trend that is steeper than a given threshold, and a nondeclining unit has a subsequent trend that is constant, increasing, or declining at a rate less than the threshold.*
[b] *Estimated by applying a given threat criterion, each of which is derived by comparing the quantitative value of some indicator (such as rate of change in abundance) with some threshold (e.g., 50%) to generate a status (e.g., declining).*
[c] *Null hypothesis (H_0): adult abundance in the population (conservation unit) is not declining over time.*
[d] *An event that triggers conservation action occurs when the conditions of the threat criterion are met (e.g., a decline greater than the threshold extent of decline or rate of decline in abundance) and result in an estimated status of declining (assumption: triggering event is followed by appropriate management actions and without a triggering event, no management action is taken).*
[e] *Terms analogous to those used in statistical hypothesis testing are provided in square brackets to indicate for each quadrant of this table the probability of being in that quadrant (related to α or β) and the type of error (I or II) or conclusion that results.*

historical data on sockeye abundance before a given year with abundance after that year to determine how reliably various indicators of the past trend in abundance signal the subsequent trend. We assumed populations that were estimated as declining in the past and continued to decline have a greater chance of extirpation than populations that did not continue to decline (Mace et al. 2008).

Data

We used abundance data on spawning adults collected over as many as 70 years from 18 spatial conservation units (CUs) of sockeye salmon from the Fraser River, as designated by Fisheries and Oceans Canada. A CU is "... a group of wild salmon sufficiently isolated from other groups that, if extirpated, is very unlikely to recolonize naturally within an acceptable time frame, such as a human lifetime" (Fisheries and Oceans Canada 2005). We examined 4 run-timing groups of Fraser River sockeye salmon populations called Early Stuart, Early Summer, Summer, and Late by Fisheries and Oceans Canada. Each group migrates through fishing areas at a different time of year (Supporting Information). Methods used by Fisheries and Oceans Canada to estimate adult abundance within these populations include visual surveys, counts at weirs, and mark-recapture.

Sockeye salmon in the Fraser River are semelparous, and about 94% mature and die at age 4 years, which results in 4 distinct "cycle lines," one associated with every fourth year, and there is little gene exchange among the lines (Ricker 1997). In some of these sockeye salmon CUs, one line (dominant cycle line) has substantially higher abundance than other lines (subdominant and 2 off-cycle lines) in each 4-year period. Such variations and naturally varying survival rates with log-normal distributions (Peterman 1981) tend to mask underlying trends in abundance. Therefore, as described below, for many of the analyses of time trends, we used spawner abundance data from all cycle lines that were $\log_e$ transformed and smoothed with a moving average over 4-year generations.

Indicators of Status of Conservation Units

We evaluated 20 threat indicators (indicators of time trends in adult abundance). Each indicator was used to calculate either recent rate of decline or long-term extent (i.e., magnitude) of decline, and change in abundance was measured with regression or other methods (Table 2 & Supporting Information). For indicators of the recent rate of decline, we measured change in abundance over the past 3 generations (12 years), as is done for criterion A in COSEWIC (2010) and IUCN (2006). If an indicator reflects the extent or magnitude of decline over the long term, the baseline from which the decline is estimated is either abundance in the earliest years for which data exist (historical) or maximum abundance in the historical record (Table 2). Specifically, for indicators reflecting the long-term extent of decrease in abundance from some baseline level (as suggested by Mace et al. [2002]), we examined 5 types of baselines: first year of the data series, maximum abundance in the first 5 data points, geometric mean abundance of the first 4-year generation, maximum recorded abundance (Holt et al. 2009), and maximum geometric mean abundance of any 12-year (3 generations) period. For baselines derived from maximum abundance (either maximum of all years or maximum geometric mean abundance over 3 generations), we calculated the extent of decline in fewer cases than if the baseline had been calculated on the basis of early historical abundance. This smaller number of cases

Table 2. Summary characteristics of the 20 indicators (identified by indicator number)[a] used to classify status (abundance declining or not declining) of Fraser River sockeye salmon conservation units, where indicators differ in their combinations of periods of decline, baseline abundance, types of adult abundance estimates, and measures of decline.

			Definitions of historical baseline for calculating long-term extent of decline				
Type of adult abundance	*Measure of decline*	*Recent rate of decline (i.e., over past 3 generations)*	*earliest part of data series*	*maximum abundance in first 5 data points*	*geometric mean abundance of first 4-year generation*	*maximum abundance (all years)*	*maximum geometric mean abundance of any 3-generation (12-year) period*
$\log_e$ unsmoothed abundance	percent change in adult abundance estimated through linear modeling and robust regression	1[a]	3 (first year) 19 (first corresponding cycle year[b])	7		5	
	percent change between geometric means of 4-year generations				10 (moving window) 12 (nonoverlapping 4-year-generation blocks)		15 (moving window) 17 (nonoverlapping 4-year-generation blocks)
$\log_e$ smoothed abundance with a 4-year moving average	percent change in adult abundance estimated through linear modeling and robust regression	2	4 (first year) 20 (first corresponding cycle year[b])	8		6	
	percentage change between geometric means of 4-year generations				11 (moving window) 13 (nonoverlapping 4-year-generation blocks)		16 (moving window) 18 (nonoverlapping 4-year-generation blocks)
Raw	decline in raw abundance relative to historical baseline				9		14

[a] *Numbers in body of table are identifiers for threat indicators. See Supporting Information for descriptions of all 20 indicators. Where more than one indicator number is in a cell of the table, each is distinguished from the other in parentheses by how the time series data are used.*
[b] *Sockeye salmon in the Fraser River are semelparous, and about 94% mature and die at age 4 years, which results in 4 distinct "cycle lines"; one is associated with every fourth year. In some of these sockeye salmon conservation units, one line (dominant cycle line) has substantially higher abundance than other (subdominant and 2 off-cycle) lines in each 4-year period. Hence, indicators 19 and 20 use as the baseline abundance only data from the first corresponding cycle year up to the year of analysis (e.g., a dominant year is only compared with another dominant cycle year, and a subdominant year is only compared with another subdominant one, etc.).*

occurred because for some CUs, the maximum abundance occurred late in the time series and resulted in fewer than 10 subsequent data points, our minimum requirement for assessing abundance trends in subsequent years.

Adult abundance estimates were either raw values, values estimated by a linear time-trend model fit to $\log_e$ abundance (either smoothed or unsmoothed) over time (as used in COSEWIC 2003), or geometric mean abundance of 4-year generations (smoothed or unsmoothed) (Table 2). Regressions were performed over periods with at least 10 data points, adult abundances were $\log_e(x + 1.01)$ transformed, and only 5 missing data points out of several hundred were interpolated from surrounding data points. We used robust regression (Venables & Ripley 2002) to minimize the influence of outliers on estimates of the short-term rate of change or long-term extent of change. For indicators 19 and 20 (defined in Table 2 & Supporting Information), we estimated abundance changes only within the same category of year (i.e., dominant, subdominant, or off-cycle years) to account for some CUs' 4 distinct cycle lines.

We categorized the past status of CUs as declining if the estimated rate of decrease or magnitude of reduction in adult abundance up to the year of assessment was steeper or greater than a given threshold used to designate species' status. In separate iterations, we explored thresholds of decline in abundance ranging from 0% to 100%. That range encompassed all thresholds typically used to classify species as not at risk, of special concern, threatened, vulnerable, endangered, or critically endangered.

Subsequent Status of Conservation Units

The status of a CU in subsequent years was the CU's observed trend in adult abundance following the year for which we evaluated past status (hereafter, subsequent status). In preliminary analyses, we found that use of all remaining data in the time series, rather than just data on the subsequent 3 generations, best indicated whether abundance of a CU continued to decrease after the year for which past status was assessed (Supporting Information). We based that subsequent status (declining or not declining) on the spawner-to-spawner ratio (i.e., the ratio of estimated number of spawners at the end of the period subsequent to the year for which past status was assessed to the estimated number of spawners in that year of assessment of past status). We calculated this ratio as the change in best-fit estimates of spawner abundance from the robust regression of abundance on year. In a given year, we categorized subsequent status of a CU as declining if the spawner-to-spawner ratio was less than or equal to the ratio that corresponded to a given threshold of percent decline (e.g., a 30% decline would be equivalent to a spawner-to-spawner ratio of 0.7, or [(100-threshold)/100]). Such cases would indicate a percent decrease in abundance greater than or equal to the threshold over the subsequent period. Subsequent status (declining or not) was defined on the basis of whether the percent decline in abundance subsequent to the year for which past status was assessed was greater than 1 of 4 thresholds: 30%, 50%, 70%, or 90%.

Comparing Past and Subsequent Status

We evaluated the reliability of the 20 threat indicators by comparing each of the 20 estimates of past status (declining or not declining) in each year with the subsequent status of each CU. We first determined whether the indicator's value exceeded the threshold for classifying the CU as declining (a triggering event [Table 1]) or not (no triggering event [Table 1]). We then categorized each CU-year combination in the historical data as true positive, true negative, false positive (type I error), or false negative (type II error) (Table 1) on the basis of whether the past status matched (i.e., was a reliable reflection of) the subsequent status. For each threat indicator and threshold of decline that we used to define a triggering event, we calculated the probability of correctly identifying a declining trend (if the CU had a subsequent status of declining) (true positive rate, which is analogous to statistical power) as the number of cases of true positives across all years and CUs divided by the number of cases in which the CUs had a subsequent status of declining (i.e., true positives plus false negatives) (Vida 1993) (details in Supporting Information). We calculated the proportion of false positives (false positive rate) by dividing the number of false positives across all years and CUs by the number of cases in which the CUs had a subsequent status of not declining (Supporting Information).

Measuring Reliability with the Receiver Operating Characteristic

To compare reliability of different indicators, we used the ROC. An ROC analysis combines 4 probabilities (true and false positives and true and false negatives [Hibberd & Cooper 2008]) into a single measure of reliability (Supporting Information). Previous evaluations of threat indicators have been limited to only a few thresholds of estimated rates of decrease that result in designation of a particular status of endangerment (e.g., a decline >30%, 50%, or 70%). In contrast, ROC analysis summarizes the proportion of true and false positive estimates produced by a threat indicator across a wide range of thresholds, thus providing more insight into the overall reliability of that indicator (Supporting Information). We applied an ROC analysis here, the first time to our knowledge that it has been used to compare indicators of population decline.

An ROC analysis uses only the probabilities of occurrence of true positives and false positives, but it accounts for true and false negatives because each of the positive rates is the complement of its corresponding negative rate (Table 1). For a given threat indicator, the ROC curve shows the true positive and false positive rates across a range of thresholds of decline that result in an estimated classification of past decline. The area under the ROC curve (AUC) ranges from 0 to 1 and reflects the ability of a threat indicator to correctly distinguish between 2 states (Hibberd & Cooper 2008). Here, AUC is the probability that a given estimated past decrease in abundance is steeper (or greater) for populations that are classified as subsequently declining than for populations that are classified as subsequently not declining. A threat indicator that correctly identifies in past periods all situations in which abundance subsequently declined produces an AUC = 1, whereas an indicator that generates an ROC curve that falls close to the 1:1 line, where the false positive rate equals the true positive rate, has an AUC = 0.5 (Pearce & Ferrier 2000). Thus, the threat indicator with the largest AUC more reliably classifies the subsequent trend in abundance. We, therefore, ranked indicators by AUC. Two hypothetical sockeye salmon ROC curves

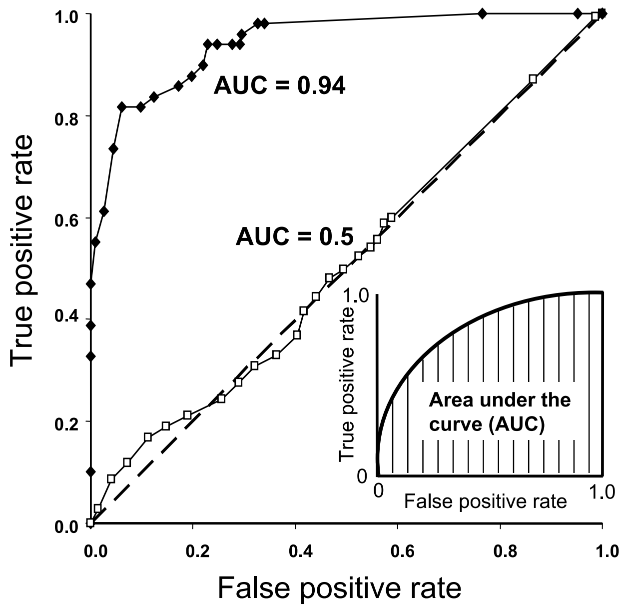

Figure 1. Two hypothetical receiver operating characteristic (ROC) curves for indicators of declining abundance used to designate conservation status of populations. Points are coordinates that result from a comparison of the value of the indicator with a given threshold (e.g., percent reduction in abundance beyond which a conservation unit is classified as declining). The first coordinate is the proportion of true positives (i.e., true positive rate, or proportion of years that a conservation unit was categorized as declining in the past and was also classified as declining in subsequent years) (Supporting Information). The second coordinate is the proportion of false positives (i.e., false positive rate, or proportion of years that a conservation unit was categorized as declining on the basis of past data, but in subsequent years data indicated the unit was not declining). The area under the ROC curve (AUC) is shown conceptually in the inset.

illustrate extremes in indicator reliability (Fig. 1). One curve has a large AUC (0.94), which signifies that the indicator is extremely reliable. The second curve has an AUC of 0.5 and roughly follows the line of equality; the latter means the indicator does not classify a population's subsequent status any better than a random classification.

Sensitivity Analyses

We performed 3 sensitivity analyses to evaluate the degree to which our findings were affected by, or an artifact of, changing harvest rates, quality of abundance estimates (defined below), and the definition of the population unit. First, we examined the influence of harvest rates of Fraser River sockeye salmon on our findings. Starting in 1995, annual harvest rates decreased from an average of 76% for 1952–1994 for the 4 run-timing groups to an average of 40% for 1995–2006 (Supporting Information). To examine whether this change affected our estimated reliability of threat indicators, we compared our initial results (for all years of data) with results of analyses limited to data from before 1995.

Second, we examined the effect of the quality of abundance estimates on the reliability of threat indicators. We compared the AUCs of threat indicators first with data from all 18 CUs and then with only the best data (i.e., taken from sites in which adult abundance was estimated from either counts at a weir or with mark-recapture methods). These latter methods tend to produce better estimates than visual surveys (Cousens et al. 1982). The 6 sites that had the best data were Pitt-ES, Chilko aggregate (Chilko-ES and Chilko-S), Fraser-S, Horsefly River (from Quesnel-S), Cultus-L, and Birkenhead River (from Lillooet-L) (T. Cone, personal communication & Supporting Information).

Third, we determined the effect of different definitions of sockeye salmon populations. Specifically, we compared the estimated status at the CU level with estimated status of 2 types of groups of CUs that covered larger areas: the 4 run-timing groups and a putative population that covered an even larger area, the 2008 IUCN subpopulation 68, which was composed of 10 CUs that spawned at 33 individual sites.

Results

The 20 indicators of past decline in adult abundance varied considerably in their reliability for classifying subsequent declines in the 18 Fraser River sockeye salmon CUs. In general, threat indicators that estimated the long-term extent of decline from a historical baseline were more reliable than indicators that measured either the rate of decline over the most recent 3 generations or the extent of decline from the maximum abundance, regardless of when that maximum occurred (Fig. 2a). The indicators of long-term extent of decline from a historical baseline had a 5% to 49% greater median AUC than indicators that were based on maximum abundance (Fig. 2a). This was also the result when we ordered indicators on the basis of number of times (out of the 4 thresholds that we used to define subsequent status) that an indicator was among the 5 indicators with highest AUC (Fig. 2a). One of the best indicators (13) measured the extent of decline from a historical baseline (i.e., percent decline between the geometric mean adult abundance in the first 4-year generation and the geometric mean abundance in a current status-assessment generation derived from $\log_e$-transformed abundances smoothed with a 4-year moving average [Table 2 & Fig. 3]).

The frequently used IUCN (2006) and COSEWIC (2010) decline criterion A (rate of decrease in

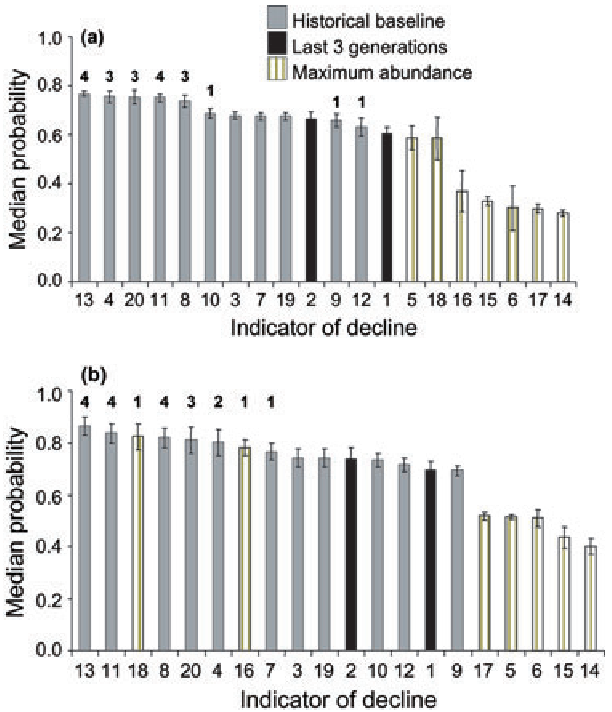

Figure 2. Threat indicators of Fraser River sockeye salmon arranged from left to right on the basis of their reliability (median probability area under the curve [AUC] [SE] that an indicator's estimated decrease in population abundance over past years is steeper for conservation units [CU] that subsequently declined than for CUs that subsequently did not decline) for data from (a) 1938 through 2007 and (b) before 1995. Indicators with higher AUCs rank higher. Numbers above bars are the number of cases (out of the 4 threshold cases) in which AUC values of a given threat indicator were among the highest 5 AUCs of the 20 indicators. The 4 cases of thresholds of decline (30%, 50%, 70%, and 90%) refer to what we used to determine whether a CU's subsequent time trend in abundance would be classified as declining. Bars with no numbers above them show that the indicator never ranked in the top 5. Numbers above bars for a given indicator should be compared between (a) and (b). Indicators are defined by number in Table 2 and in more detail in Supporting Information and fall into 3 categories: extent of decline from some specified historical baseline abundance, rate of decline over just the last 3 generations, and extent of decline from the maximum abundance.

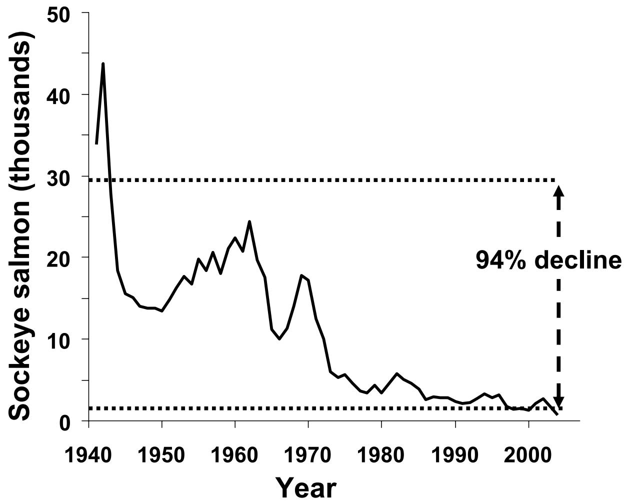

Figure 3. Extent of decline of Cultus Lake sockeye salmon on the basis of $\log_e$-transformed abundances that were smoothed with a 4-year moving average in nonoverlapping blocks (status assessed every 4 years), as determined by one of the best-performing threat indicators (indicator 13), which quantified the decrease between the geometric mean adult abundance in the first 4-year generation (historical baseline, upper dotted line) and the geometric mean adult abundance in the current 4-year generation (lower dotted line [Supporting Information]).

abundance over 3 generations calculated with smoothed $\log_e$-transformed data for indicator 2) performed only moderately well. Nine indicators were ranked higher and 10 were ranked lower than that indicator (Fig. 2a). Indicator 13 had a median AUC = 0.77 and indicator 2 had a median AUC = 0.66 across the 4 thresholds. Thus, indicator 13 had a higher probability than indicator 2 of identifying past declines for populations that subsequently declined than for populations that did not decline subsequently. The AUC of indicator 1, the other indicator of rate of decline over the last 3 generations (which was based on unsmoothed rather than smoothed data), was lower (median probability of 0.6). However, indicators that were a measure of extent of decline from the maximum abundance in the data series were the least reliable indicators. They had median probabilities (AUC) of 0.29–0.59 (i.e., 0.08–0.38 lower than indicator 2 and 0.18–0.49 lower than indicator 13). Neither of the indicators (1 or 2) of recent rate of decline ranked in the top 5 for any of the 4 thresholds used to classify subsequent status (Fig. 2a). Smoothing the time series of data on salmon abundance tended to increase the AUC of all indicators by 0.04–0.29 over indicators that were calculated from unsmoothed data.

Sensitivity Analyses

Our findings were largely insensitive to changes in harvest rate and abundance-estimation error; rankings of indicators on the basis of their reliability were generally quite similar to the initial analyses. This result held regardless of whether we ranked the indicators on the basis of median AUC or how frequently a given indicator had 1 of the 5 highest median AUCs. For effects of harvesting, regardless of whether we used all years of data (which included high and low harvest rates) or only pre-1995 data (only high harvest rates), the rank order of threat indicators was generally similar (compare Fig. 2a with 2b). This is not unexpected given that there was relatively low correlation among years in abundances of most CUs and those of the aggregate abundances of the run-timing groups and that management decisions have historically been made at the level of the run-timing groups, not at the level of their component CUs (Supporting Information). Thus, this low correlation reduced the potential confounding of interpretation of results by management interventions. Improved data quality led to an increased median AUC. Slight changes in the rank order of indicators occurred, but the general pattern remained the same regardless of data quality (Supporting Information).

The 2 types of spatial aggregation led to an underestimation of the status of Fraser sockeye CUs. For example, indicator 13 never showed IUCN subpopulation 68 as declining (COSEWIC's threatened or endangered rating), even though individual CUs were identified as declining on the basis of that indicator in 8 years of time-series data (Table 3). Similarly, the Late run-timing group was the only run-timing aggregate classified as declining on the basis of indicator 13, and that classification occurred in only 1 year. This result was low compared with analysis of individual CUs, which classified at least one CU as declining in a total of 23 years across 2 different run-timing groups (Table 3).

Discussion

The reliability of threat indicators is usually assumed and is rarely tested. In our case study of Fraser River sockeye salmon CUs, no indicator was a perfect measure of threat, but the most reliable threat indicators were those that were based on the overall extent of decline in abundance from the beginning of a CU's time series, even though the decline did not necessarily start then (long-term indicators 4, 8, 11, 13, and 20 in Table 2). These indicators consistently were more reliable than indicators that were based on either rate of decline over the most recent 3 generations or extent of decline from the maximum abundance in the time series. The reliability (measured by AUC) of recent rate of decline over 10 years or generations, whichever was longer (an indicator

Table 3. The number of years in which indicator 13 (one of the top-ranked threat indicators in our retrospective analyses) showed a decline in adult abundance of Fraser River sockeye salmon that was large enough to result in a classification as either threatened or endangered (on the basis of COSEWIC's [2010] method) for a population aggregate.

Population aggregate[a]	*Consevation units (CUs) in aggregate*[b]	*Number of years aggregate was classified as threatened endangered*[c]	*Number of years any individual CUs within aggregate were classified as threatened*[c]	*Number of years any individual CUs within aggregate were classified as endangered*[c]
International Union for Conservation of Nature subpopulation 68	11	0	1 (1 CU threatened)	7 (1 CU endangered)
Early Stuart run-timing group	Stuart EStu and Takla-Trembleur EStu	0	0	0
Early summer run-timing group	Chilliwack ES, Taseko ES, Nahatlatch ES, Fraser ES, Kamloops ES, Pitt ES, and Shuswap Complex ES	0	2 (1 CU threatened)	6 (1 CU endangered)
Summer run-timing group	Chilko aggregate, Takla-Trembleur S, Fraser S, Stuart S, Quesnel S, and Mckinley S	0	0	0
Late run-timing group	Lillooet L, Cultus L, and Seton L	1 (run-timing group threatened)	3 (1 CU threatened)	10 (1 CU endangered) 2 (2 CUs endangered)

[a]*A population aggregate for Fraser River sockeye salmon is a group of 2 or more conservation units, as shown in column 2. The International Union for the Conservation of Nature (IUCN) created an aggregate they called subpopulation 68. Agencies responsible for management of Fraser River sockeye salmon instead use 4 aggregates, with each one named according to the season in which adults return to the river. These run-timing groups are composed of between 2 and 7 conservation units, as shown.*
[b]*Abbreviations EStu, Early Stuart; ES, Early Summer; S, Summer; and L, Late. These are parts of the labels for the conservation units used by Fisheries and Oceans Canada.*
[c]*A status of threatened or endangered was assigned if the decrease in abundance over the period measured by threat indicator 13 was ≥30% or ≥50%, respectively. The number of years in which each population aggregate was classified as either threatened or endangered (column 3) should be compared with the number of years when individual conservation units within the population aggregates were classified as threatened (column 4) or endangered (column 5).*

often used by COSEWIC and IUCN), is 8–10% less reliable than long-term indicators of extent of decline. Indicators for which smoothed abundance data are used were more reliable than those with unsmoothed data probably because large fluctuations in abundance over time were reduced, thus making the underlying time trend clearer.

Simulations show that data quality can affect the reliability of threat indicators (Wilson et al. 2011), but we found that long-term indicators were consistently the most reliable, regardless of data quality. We suspect that the better ranking of long-term indicators of decline resulted from a combination of highly variable abundances over time and the fact that a long-term decrease in abundance is not necessarily discernible from a decline in abundance within the last 3 generations. Regardless of the cause, we recommend that when possible, indicators of declining abundance that are estimated from the entire data set should be included in assessments of status. Results of our case study suggest there is a need to evaluate reliability of alternative indicators of status before these indicators are used to determine status, and we encourage other researchers to explore ways to identify the best historical baseline in different situations. The estimate of stock biomass before the onset of fishing, which is commonly used as a historical baseline for marine fisheries, is analogous to abundance in the early part of our time series, which was the basis of our indicators of extent of decline from a historical baseline. However, unfished biomass is notoriously difficult to estimate reliably because in many fisheries, most data points are from periods well after fishing began (Walters & Martell 2004).

In previous studies in which frequencies of false positives or false negatives were estimated for various indicators, only one threshold of decrease in abundance was evaluated at a time (Dulvy et al. 2005; Rice & Legacé 2007), which is equivalent to estimating only one point on an ROC curve. In contrast, our ROC analysis examines the performance of decline indicators across a range of thresholds and thus provides more information about the overall reliability of any one indicator. For this reason, ROC approaches may help simplify evaluations of threat indicators for various plant and animal species, not just sockeye salmon.

By using the AUC measure to rank threat indicators, we implicitly assumed that both types of classification error (false positives and false negatives) were equally important, but in practice, equal weighting is not likely (Peterman 1990; Mapstone 1995). Instead, it is well known that some decision makers place more weight on reducing the chance of false negatives because that type of error may lead to corrective action not being taken when it should be. In contrast, other decision makers place more weight on reducing the chance of false positives because that type of error could lead to unnecessary corrective action and could reduce catches and short-term social and economic benefits.

More informed choices could be made about which threat indicators to use if the probability of each type of error were considered both separately and jointly through ROC evaluations of reliability of indicators. However, risk tolerance (which depends on the perceived cost of both types of errors and their probabilities of occurrence) is part of most management objectives, and is situation specific. It is not possible to have minimal chances of both false positives and false negatives because an unavoidable trade-off exists between these 2 error rates (Peterman 1990; Rice & Legacé 2007). For instance, an evaluation criterion that sets a threshold for classifying decline so stringently that false positives rarely occur will inevitably have a high chance of producing false negatives (failing to identify a decline). Decisions could also be improved if the benefits and costs of using various indicators of population decline were quantified.

The American Fisheries Society recommends that for marine fisheries the threshold at which a population is considered vulnerable should be a $\geq$70% reduction in abundance within a given period to reduce the probability of incorrectly classifying a species as vulnerable (Reynolds et al. 2005). Before any such changes are implemented, we recommend that our AUC methods be adapted to evaluate the reliability of different thresholds for classifying conservation status of fish populations.

Our results also highlight 3 key issues to consider when applying threat indicators to other species or populations: recent rate of decline in abundance versus longer-term extent of decline from some initial baseline, choice of baseline abundance, and spatial aggregation of populations. Our results may be specific to Pacific sockeye salmon, but our methods can be applied in other contexts to determine the reliability of indicators used to inform conservation decisions.

We also suggest researchers investigate the reliability of assessing multiple extinction-risk indicators together, instead of assessing indicators individually. For instance, if abundance were to drop substantially early in a data series and subsequently remained relatively constant, values of an extent-of-decline indicator might not elicit concern, whereas if absolute abundance were the indicator, the population might be considered to have a high probability of extinction.

We explored the effect of harvest rate and spatial delineation of populations on the reliability of threat indicators. For harvest rate, we evaluated threat indicators before a large and persistent reduction in harvest rate in 1995 and with data spanning this change (1952–2006). Ideally, when defining the status of a population, one should account for year-by-year effects of changing harvest rates to more clearly rule out the confounding effect of management on evaluations of threat indicators. Such a detailed analysis was not possible here because yearly harvest rates are not well estimated for sockeye CUs. Nonetheless, we found that the ranking of different threat indicators on the basis of their reliability was relatively unchanged when we took into account the 1995 reduction in harvest rates across all run-timing groups. This robustness of rankings likely emerged because harvest regulations for Fraser sockeye are usually set according to estimated abundances of run-timing groups, not their individual component CUs. Correlations in egg-to-adult survival among populations within run-timing groups are generally low (Peterman et al. 1998), as is the average correlation between inter-annual changes in abundances at the CU scale and the aggregate run-timing group.

Results of our sensitivity analyses of the effect of aggregation of CUs emphasize the need to designate spatial units carefully, otherwise incorrect conclusions could emerge. We found that larger aggregates of CUs (e.g., IUCN subpopulation 68 or management-defined run-timing groups) were less likely to be classified correctly than smaller units, presumably due to the masking effect of aggregating relatively independent asynchronous groups of fish. In a recent amendment to its original 2008 sockeye salmon assessment, the IUCN split various subpopulations in British Columbia including 68 (Rand 2011) (Supporting Information).

The 18 CUs for Fraser River Pacific sockeye salmon we examined encompass a wide range of geographic spawning sites, management actions, environmental conditions, productivities, quality of adult abundance estimates, and population trends. Thus, we were able to evaluate the overall effectiveness of each indicator across a wide range of conditions. We do not know whether our rank order of threat indicators, as determined by their reliability, can be applied to other species with different life histories, habitat variability, and population dynamics or to management systems with other risk weightings. Where possible, indicators should be evaluated either retrospectively to see how they would have performed relative to what subsequently happened or via simulation models that examine indicators under a wide range of situations (Punt 2000; Holt 2009; Regan et al. 2009).

Acknowledgments

We thank T. Cone and E. Grundmann from Fisheries and Oceans Canada (DFO) for providing the data and C. A. Holt for advice. We also thank H.R. Akçakaya, J.D. Reynolds, P. Mace, and 2 anonymous reviewers for comments on a draft manuscript. Funding was provided to E.J.P. from DFO, the Natural Sciences and Engineering Research Council of Canada, MITACS (Mathematics in Technology and Computer Science consortium), and Simon Fraser University.

Supporting Information

A summary of data sets for 18 CUs of sockeye salmon within the Fraser River watershed, British Columbia, Canada (Appendix S1), descriptions of the 20 threat indicators (Appendix S2), scenarios for determining status in subsequent years (Appendix S3), calculation of true and false positive rates (Appendix S4), analysis of the receiver operating characteristic (ROC) (Appendix S5), annual percent harvest rates on Fraser River sockeye salmon (Appendix S6), methods for sensitivity analyses on aggregations of CUs (Appendix S7), variations in abundances of CUs compared with run-timing groups (Appendix S8), and results for sensitivity analyses on the quality of data (Appendix S9) are available online. The authors are solely responsible for the content and functionality of these materials. Queries (other than absence of the material) should be directed to the corresponding author.

Literature Cited

Baxter, P. W. J., and H. P. Possingham. 2011. Optimizing search strategies for invasive pests: learn before you leap. Journal of Applied Ecology **48:**86–95.

Burgman, M. A. 2005. Risks and decisions for conservation and environmental management. Cambridge University Press, Cambridge, United Kingdom.

Committee on the Status of Endangered Wildlife in Canada (COSEWIC). 2003. COSEWIC assessment and status report on the sockeye salmon *Oncorhynchus nerka* Sakinaw population in Canada. COSEWIC, Ottawa, Ontario.

Committee on the Status of Endangered Wildlife in Canada (COSEWIC). 2010. COSEWIC's assessment process and criteria. COSEWIC, Ottawa, Ontario.

Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES). 2010. Criteria for amendment of Appendices I and II. CITES Secretariat, Geneva. Available from www.cites.org/eng/res/09/09-24R15.php (accessed April 2012).

Cousens, N. B. F., G. A. Thomas, C. G. Swann, and M. C. Healey. 1982. A review of salmon escapement estimation techniques. Canadian technical report of fisheries and aquatic sciences 1108. Fisheries and Oceans Canada, Nanaimo, British Columbia.

Dulvy, N. K., S. Jennings, N. B. Goodwin, A. Grant, and J. D. Reynolds. 2005. Comparison of threat and exploitation status in north-east Atlantic marine populations. Journal of Applied Ecology **42:**883–891.

Dulvy, N. K., S. Jennings, S. I. Rogers, and D. L. Maxwell. 2006. Threat and decline in fishes: an indicator of marine biodiversity. Canadian Journal of Fisheries and Aquatic Sciences **63:**1267–1275.

Fisheries and Oceans Canada (DFO). 2005. Canada's policy for conservation of wild Pacific salmon. DFO, Vancouver, British Columbia. Available from http://www.pac.dfo-mpo.gc.ca/publications/pdfs/wsp-eng.pdf (accessed January 2012).

Hibberd, P. L., and A. B. Cooper. 2008. Methodology: statistical analysis, test interpretation, basic principles of screening with application for clinical study. Pages 1517–1520 in R. E. Kleinman, O. Goulet, G. Mieli-Vergani, I. R. Sanderson, P. M. Sherman, and B. L. Shneider, editors. Walker's pediatric gastrointestinal disease: pathophysiology, diagnosis, management. 5th edition. B. C. Decker, Hamilton, Ontario.

Holt, C. A. 2009. Evaluation of benchmarks for conservation units in Canada's Wild Salmon Policy: technical documentation. Research document 2009/059. Canadian Science Advisory Secretariat, Ottawa, Ontario. Available from http://www.dfo-mpo.gc.ca/CSAS/Csas/Publications/ResDocs-DocRech/2009/2009_059_e.pdf (accessed January 2012).

Holt, C. A., A. Cass, B. Holtby, and B. Riddell. 2009. Indicators of status and benchmarks for conservation units in Canada's Wild

Salmon Policy. Research document 2009/058. Canadian Science Advisory Secretariat, Ottawa, Ontario. Available from http://www.dfo-mpo.gc.ca/CSAS/Csas/Publications/ResDocs-DocRech/2009/2009_058_e.pdf (accessed January 2012).

Hutchings, J. A. 2001. Conservation biology of marine fishes: perceptions and caveats regarding assignment of extinction risk. Canadian Journal of Fisheries and Aquatic Sciences **58:**108–121.

International Union for the Conservation of Nature (IUCN). 2006. Guidelines for using the IUCN Red List categories and criteria. IUCN, Gland, Switzerland.

Mace, G. M., N. J. Collar, K. J. Gaston, C. Hilton-Taylor, H. R. Akçakaya, N. Leader-Williams, E. J. Milner-Gulland, and S. N. Stuart. 2008. Quantification of extinction risk: International Union for the Conservation of Nature's (IUCN) system for classifying threatened species. Conservation Biology **22:**1424–1442.

Mace, P. M., et al. 2002. National Marine Fisheries Service (NMFS)/Interagency Working Group evaluation of CITES criteria and guidelines. Technical memorandum NMFS-F/SPO-58. National Oceanic and Atmospheric Administration, Rockville, Maryland.

Mapstone, B. D. 1995. Scalable decision rules for environmental impact studies: effect size, type I, and type II errors. Ecological Applications **5:**401–410.

Miller, R. M., et al. 2006. Extinction risk and conservation priorities. Science **313:**441.

Pearce, J., and S. Ferrier. 2000. Evaluating the predictive performance of habitat models developed using logistic regression. Ecological Modelling **133:**225–245.

Peterman, R. M. 1981. Form of random variation in salmon smolt-to-adult relations and its influence on production estimates. Canadian Journal of Fisheries and Aquatic Sciences **38:**1113–1119.

Peterman, R. M. 1990. Statistical power analysis can improve fisheries research and management. Canadian Journal of Fisheries and Aquatic Sciences **47:**2–15.

Peterman, R. M., B. J. Pyper, M. F. Lapointe, M. D. Adkison, and C. J. Walters. 1998. Patterns of covariation in survival rates of British Columbian and Alaskan sockeye salmon stocks. Canadian Journal of Fisheries and Aquatic Sciences **55:**2503–2517.

Punt, A. E. 2000. Extinction of marine renewable resources: a demographic analysis. Population Ecology **42:**19–27.

Rand, P. S. 2011. *Oncorhynchus nerka*. IUCN red list of threatened species. Version 2011.2. International Union for Conservation of Nature, Gland, Switzerland. Available from http://www.iucnredlist.org/apps/redlist/details/135301/0 (accessed January 2012).

Regan, T., B. Taylor, G. Thompson, J. Cochrane, R. Merrick, M. Nammack, S. Rumsey, K. Ralls, and M. Runge. 2009. Developing a structure for quantitative listing criteria for the U.S. Endangered Species Act using performance testing, phase 1 report. Technical memorandum NOAA-TM-NMFS-SWFSC-437. National Oceanic and Atmospheric Administration, Rockville, Maryland.

Reynolds, J. D., N. K. Dulvy, N. B. Goodwin, and J. A. Hutchings. 2005. Biology of extinction risk in marine fishes. Proceedings of the Royal Society of London, B **272:**2337–2344.

Rice, J. C., and E. Legacé. 2007. When control rules collide: a comparison of fisheries management reference points and International Union for Conservation of Nature (IUCN) criteria for assessing risk of extinction. ICES Journal of Marine Science **64:**718–722.

Ricker, W. E. 1997. Cycles of abundance among Fraser River sockeye salmon (*Oncorhynchus nerka*). Canadian Journal of Fisheries and Aquatic Sciences **54:**950–968.

Venables, W. N., and B. D. Ripley. 2002. Modern applied statistics with S. 4th edition. Springer Science, New York.

Vida, S. 1993. A computer program for non-parametric receiver operating characteristic analysis. Computer Methods and Programs in Biomedicine **40:**95–101.

Walters, C. J., and S. J. D. Martell. 2004. Fisheries ecology and management. Princeton University Press, Princeton, New Jersey.

Wilson, H. B., B. E. Kendall, and H. P. Possingham. 2011. Variability in population abundance and the classification of extinction risk. Conservation Biology **25:**747–757.