North American Journal of Fisheries Management

ISSN: 0275-5947 (Print) 1548-8675 (Online) Journal homepage: http://www.tandfonline.com/loi/ujfm20

The Influence of Stock Assessment Frequency on the Achievement of Fishery Management Objectives

Yang Li, James R. Bence & Travis O. Brenden

North American Journal of Fisheries Management 36:793–812, 2016

ISSN: 0275-5947 print / 1548-8675 online
DOI: 10.1080/02755947.2016.1167145

ARTICLE

The Influence of Stock Assessment Frequency on the Achievement of Fishery Management Objectives

Yang Li,* James R. Bence, and Travis O. Brenden
Quantitative Fisheries Center, Department of Fisheries and Wildlife, Michigan State University, 480 Wilson Road, Room 13, East Lansing, Michigan 48824, USA

Abstract
Because of resource limitations with respect to both funding and staff expertise, there is growing interest among fishery management agencies in moving from annual to less-frequent assessments of fish stocks. We conducted simulations based on Lake Whitefish *Coregonus clupeaformis* populations in the Laurentian Great Lakes to evaluate (1) how statistical catch-at-age assessment frequency, the time lag between data collection and assessment, and approaches to setting target harvests in the years between assessments affected the achievement of management objectives; and (2) how the outcomes were influenced by the quality of assessment data, features of the populations, and characteristics of the fisheries exploiting the populations. We found that as assessments became less frequent, relative yields were reduced and the risk of stock depletion and interannual variation in yield increased. The effects of less-frequent assessments were ameliorated in populations with greater levels of productivity and when target mortality was lower. Conversely, the effects of assessment frequency were largely insensitive to changes in recruitment variation or the quality of assessment data. A 1-year lag between data collection and assessment when assessments were conducted annually primarily affected the risk of stock depletion and the interannual variation in yield. As recruitment variation increased, relative yield also became sensitive to the 1-year lag. Approaches to setting harvest targets in years between full assessments were less important than assessment frequency, and no single approach consistently outperformed other rules. Although populations with low productivity were the most sensitive to changes in assessment frequency and the lag between data collection and assessment, the management of those populations benefited to a greater extent from implementation of an appropriate target mortality rate than from more-frequent assessments or removal of the 1-year lag.

Modern stock assessment approaches, such as statistical catch-at-age (SCAA) models, are widely used to estimate the status of fish stocks and aid in determining appropriate levels of harvest to ensure that fishery management objectives can be met. As a general rule, timely (i.e., soon after observational data have been collected) and frequent assessments with high-quality data are considered important for effective and sustainable management (Mace et al. 2001). However, resources (e.g., funding and personnel effort) that can be devoted to stock assessment are limited; thus, assessments must provide high-quality information as cost efficiently as possible. With respect to stock assessment frequency, fishery scientists generally either choose the highest frequency that can be maintained or continue with the frequency that has been used historically. Our goal was to evaluate how the achievement of management objectives is influenced by the frequency of stock assessments, the lag between data collection and assessment, and the methods used for determining harvests in years between assessments.

A few published studies have evaluated how assessment frequency can affect the achievement of management objectives (Marchal and Horwood 1995; Marchal 1997; Kell et al. 1999), and this issue has been addressed in workshops and assessment model improvement plans (Mace et al. 2001; Woldt et al. 2006; ICES 2012). Previous research has found that multiannual assessments can lead to greater variability in catch and a greater risk that populations could decline below target threshold levels (Marchal and Horwood 1995; Marchal 1997). Consequently, multiannual assessments are not recommended for recovering

*Corresponding author: liyang11@msu.edu
Received November 18, 2015; accepted March 9, 2016

or rebuilding stocks (Kell et al. 1999). Mace et al. (2001) suggested that more-frequent assessments were better at avoiding overfishing because they allowed for the early detection of declining abundance levels. One benefit of less-frequent stock assessments is that effort and resources normally applied to assessment could instead be directed to tasks such as model improvement and model performance verification (Woldt et al. 2006). Indeed, it can be challenging to maintain the amount and quality of assessment work when assessments are performed annually (ICES 2012).

The International Council for the Exploration of the Sea (ICES; 2012) used a management strategy evaluation (MSE) approach to evaluate the risks engendered from conducting assessments multiannually (i.e., every 5 years). The ICES (2012) report focused on the short-term consequences of multiannual assessments in relation to particular starting population conditions. The results showed that with declining trends in population abundance at initialization, significant depletions in abundance could occur prior to the next assessment (ICES 2012); this result was avoided with annual assessments. Such findings partly reflected how the assessment results were used: the allowed harvest was fixed at a constant level that was estimated to produce the desired exploitation rate in the year after the assessment. Consequently, as weak year-classes entered the population, the exploitation rates increased.

We sought to build upon the ICES (2012) findings by evaluating how different assessment frequencies performed over the long term when populations were initially at an equilibrium state. We also examined population or fishery characteristics that might influence the effect of assessment frequency. According to the Northwest Atlantic Fisheries Organization (NAFO 2007), the consequences of assessment frequency for the management of fish stocks are influenced by two main factors: (1) the biological attributes of the assessed population, which determine the rate of change in stock size; and (2) the current or anticipated fishing pressure. Therefore, we evaluated how different assessment frequencies performed with respect to population productivity (i.e., stock–recruitment steepness), recruitment variation, target mortality level, and the quality of stock assessment input data.

We based our simulations on Lake Whitefish *Coregonus clupeaformis* populations in the 1836 Treaty-ceded waters of the Laurentian Great Lakes region. Although we used Lake Whitefish as the basis for our research, our intent was for the results to have broad relevance. Lake Whitefish stocks in treaty-ceded waters are assessed annually via SCAA models, with a 1-year lag between data collection and stock assessment (i.e., assessments conducted in a particular year incorporate data through the previous year and are used to set harvest levels for the next year; Ebener et al. 2005; MSC 2015). There are multiple assessed stocks in the 1836 Treaty-ceded waters; other stocks that occur in the region but are outside the ceded waters are assessed in a similar fashion. Our simulations were intended to represent generic Lake Whitefish stocks, and some factors (e.g., stock productivity and data quality) that are known to differ among stocks and that are suspected to potentially alter the influence of assessment timetables were set to different levels in our different scenarios. To the best of our knowledge, ours is the first study to evaluate the consequences of assessment frequency, assessment lag, and different methods for determining harvests in years between assessments for freshwater fish stocks based on an MSE approach. Although Marchal and Horwood (1995) and Marchal (1997) evaluated the performance of some multiannual management strategies, the estimated abundances in those studies were calculated by multiplying true population abundance by an observation error component rather than by estimating abundance through an assessment model. Both ICES (2012) and Methot (2015) proposed that the risk of applying multiannual assessments could be better evaluated with an MSE approach in which the dynamics of fish stocks are estimated through an assessment model, as such an approach more closely approximates the actual management process.

METHODS

Overall Simulation Framework

As in the ICES (2012) study, a simulation framework that is common to MSE studies was employed to evaluate the influence of stock assessment frequency on fishery performance (Kell et al. 2005; Punt 2008). The framework, which was similar to that of Li et al. (2015), included an operating model that represented the true stock and fishery dynamics. Simulated data from the operating model were then sampled (i.e., observed) and used as input to an SCAA model. A harvest control rule was applied to the SCAA model estimates to determine future harvest levels, which were in turn applied back to the operating model. With multiannual assessments, we evaluated three different approaches to setting the target harvest in years between assessments. Performance statistics based upon true stock conditions in the operating model were used to evaluate the fishery and conservation performance of each assessment frequency and method for setting target harvest in the years between assessments. Simulations were conducted in AD Model Builder (Fournier et al. 2012). All equations for the simulation framework are provided in Table 1. The values of life history parameters and stochastic modeling parameters are the same as used by Li et al. (2015) and are either included in the model description or with the equations in Table 1. Symbols are defined in Table 2. The parameter values were chosen to represent a generic Lake Whitefish stock in the Great Lakes. Past simulations showed that some life history characteristics known to vary among Lake Whitefish stocks did not have a substantial influence on the outcome of other management choices (i.e., the fishing rate or control rule; Deroba and Bence 2012); we concluded that those characteristics were unlikely to substantially alter the influence of assessment timetables, and therefore we fixed them at typical values.

TABLE 1. Equations used in the operating model and the assessment model. See Table 2 for symbol definitions.

Model name	Model equation	Equation number
Ricker stock-recruitment parameterization in the operating model		
Steepness	$h = \frac{R_{0.2S_0}}{R_{S_0}} = \frac{\alpha 0.2 S_0 e^{-\beta 0.2 S_0}}{\alpha S_0 e^{-\beta S_0}} = 0.2 S_0 e^{\beta 0.8 S_0}$, so $\beta = \frac{\log_e\left(h/0.2\right)}{0.8 S_0}$	1.1
Unfished abundance at age during the spawning season	$Nspawn_a = N_{a+1} = N_a e^{-(M)}$, where $N_{a=1} = R_{S_0}$	1.2
Unfished spawning stock	$S_0 = \sum_a Nspawn_a \cdot Fem \cdot m_a \cdot W_a$, where $Fem = 0.5$ (from Li et al. 2015)	1.3
Unfished stock per recruit	$\mathrm{SPR}_0 = \frac{S_0}{R_{S_0}} = \frac{S_0}{\alpha' S_0 e^{-\beta S_0}} = \frac{e^{\beta S_0}}{\alpha'}$, so $\frac{S_0}{R_{S_0}} = \frac{e^{\beta S_0}}{\alpha'}$	1.4
Solving for Ricker parameter α'	$\alpha' = \frac{e^{\beta S_0}}{S_0} = \frac{e^{\beta S_0}}{\sum_a Nspawn_a \cdot Fem \cdot m_a \cdot W_a}$ when $R_{S_0} = 1$	1.5
Biomass calculation in the operating model		
Ricker stock–recruit relationship	$R_y = \alpha \mathrm{SSB}_{y-1} e^{-\beta \mathrm{SSB}_{y-1}} e^{\varepsilon_{R,y}}$, where $\alpha = \alpha' e^{-\frac{\sigma_R^2}{2(1-\rho^2)}}$ $\varepsilon_{R,y} = \rho \times \varepsilon_{R,y-1} + \tau_{R,y}$ $\tau_{R,i,y} \sim \mathrm{N}(0, \sigma_R^2)$	1.6
Age-specific spawning stock biomass	$\mathrm{SSB}_y = \sum_a Fem.N_{y,a}.m_a.W_a$, where $Fem = 0.5$ (from Li et al. 2015)	1.7
Length at age	$L_a = L_\infty\{1 - \exp[-\kappa(a - t_0)]\}$, where $L_\infty = 60.9$ cm, $\kappa = 0.1689$ year^{-1}, and $t_0 = 0$ year (from Li et al. 2015)	1.8
Weight at age	$W_a = \gamma L_a^\psi$, where $\gamma = 8.06 \times 10^{-5}$ and $\psi = 2.45$ (from Li et al. 2015)	1.9
Maturity at age	$m_a = \frac{m_\infty}{1+\exp[-\vartheta(L_a - \delta)]}$, where $\vartheta = 0.315$ cm^{-1} and $\delta = 37.86$ cm (from Li et al. 2015)	1.10
Dynamics of postrecruitment abundance in the operating model		
Population-specific abundance at age	$N_{y+1,a+1} = N_{y,a}\exp\left(-M - F_{y,a}\right)$	1.11
Implemented total allowable catch with implementation error	$C_y = \mathrm{TAC}_y \exp\left(\upsilon_y - 0.5\sigma_c^2\right)$ $\upsilon_y \sim \mathrm{N}(0, \sigma_c^2)$	1.12
Baranov's catch equation to solve for F_y	$C_{j,y} = \sum_{a=1}^{n} \frac{s_a F_y}{M + s_a F_y}\left(1 - e^{-M - s_a F_y}\right) N_{y,a}$	1.13
Fishing mortality at age	$F_{y,a} = s_a F_y$	1.14
Selectivity at age	$s_a = \frac{a^\eta \exp(-\tau a)}{10^\eta \exp(-\tau 10)}$, where $\tau = 1.26$ year^{-1} and $\eta = 13.074$ cm (from Li et al. 2015)	1.15

TABLE 1. Continued.

Model name	Model equation	Equation number
	Deriving observations from the operating model	
Observed harvest	$\tilde{C}_y = C_y \exp\left(\tilde{\upsilon}_y - 0.5\tilde{\sigma}_c^2\right)$	1.16
	$\tilde{\upsilon}_y \sim \mathrm{N}(0, \tilde{\sigma}_c^2)$	
Observed fishing effort	$\tilde{E}_y = \frac{F_y}{q} \exp\left(\mu_y - 0.5\sigma_F^2\right)$	1.17
	$\mu_y \sim \mathrm{N}\left(0, \sigma_F^2\right)$	
	$\sigma_F^2 = 4\tilde{\sigma}_c^2$	
	Statistical catch-at-age assessment model	
Selectivity at age	$s_a = \frac{a^{\hat{\eta}} \exp(-\hat{\tau} a)}{10^{\hat{\eta}} \exp(-\hat{\tau} 10)}$	1.18
Fishing mortality at age	$F_{y,a} = F_y s_a$, where $F_y = \hat{q} \tilde{E}_y \widehat{\varepsilon F}_y$	1.19
Abundance at age	$N_{y+1,a+1} = N_{y,a} \exp\left(-\tilde{M} - F_{y,a}\right)$	1.20
	$N_{y,1} = \widehat{R\mu}\widehat{Rd}_y$	
	$N_{1,a} = \widehat{N\mu}\widehat{Nd}_a (a > 1)$	
Total catch	$\hat{C}_y = \sum_{a=3}^{12} \frac{F_{y,a}}{\tilde{M} + F_{y,a}} \left(1 - e^{-\tilde{M} - F_{y,a}}\right) N_{y,a}$	1.21
Age composition of the catch	$\hat{p}_{y,a} = \frac{\hat{C}_{y,a}}{\hat{C}_y} = \frac{\frac{F_{y,a}}{\tilde{M} + F_{y,a}} \left(1 - e^{-\tilde{M} - F_{y,a}}\right) N_{y,a}}{\hat{C}_y}$	1.22
Negative log likelihood for the total catch	$\ell_c = n_c \log_e(\hat{\sigma}_c) + \left(\frac{1}{2\hat{\sigma}_c^2}\right) \sum_y \left[\log_e\left(\frac{\tilde{C}_y}{\hat{C}_y}\right)\right]^2$	1.23
Negative log likelihood for annual fishing effort deviations	$\ell_{\varepsilon F} = n_{\varepsilon F} \log_e(\hat{\sigma}_F) + \left(\frac{1}{2\hat{\sigma}_F^2}\right) \sum_y \left[\log_e\left(\widehat{\varepsilon F}_y\right)\right]^2$, where $\hat{\sigma}_F^2 = 4\hat{\sigma}_c^2$	1.24
Negative log likelihood for the age composition of the catch	$\ell_p = -\sum_y n_{eff} \sum_a \tilde{p}_{y,a} \log_e\left(\hat{p}_{y,a}\right)$	1.25

TABLE 2. Definitions of the symbols used in Table 1 and Table A.1.

Symbol	Definition
	Index Variables
i	Population
j	Fishing ground
y	Year
a	Age
	Accents
$\sim$	Observed variable
$\frown$	Estimated variable
	Parameters
L	Length
L_∞	Asymptotic length parameter from the von Bertalanffy length-at-age relationship
κ	Body growth coefficient from the von Bertalanffy length-at-age relationship
t_0	Theoretical age at zero length from the von Bertalanffy length-at-age relationship
W	Weight
γ	Weight-at-age relationship parameter
ψ	Weight-at-age relationship parameter
m	Probability of maturity
m_∞	Maturity-at-age relationship parameter
ϑ	Maturity-at-age relationship parameter
δ	Maturity-at-age relationship parameter
s	Selectivity
η	Selectivity-at-age relationship parameter
τ	Selectivity-at-age relationship parameter
h	Steepness
R_{S_0}	Unfished recruitment
$R_{0.2S_0}$	Recruitment produced by 20% of unfished spawning biomass
α	Ricker stock–recruitment parameter
β	Ricker stock–recruitment parameter
S_0	Unfished spawning stock biomass
$Nspawn$	Abundance at the spawning season
M	Instantaneous natural mortality
F	Instantaneous fishing mortality
Fem	Female rate
SPR_0	Unfished stock per recruit
ρ	Autocorrelation coefficient in recruitment process error
σ_R	Standard deviation in recruitment process error
SSB	Spawning stock biomass
C	Actual catch
TAC	Total allowable catch
υ	Implementation error in TAC
σ_c^2	Variance of the implementation error in TAC
$\tilde{\upsilon}$	Random variable for observed harvest
$\tilde{\sigma}_c^2$	Variance in observed harvest
q	Catchability
μ	Random variable for observed fishing effort
σ_F^2	Variance in observed fishing effort
$R\mu$	Average recruitment over the years
Rd	Annual recruitment deviations
$N\mu$	Average number of age-2 and older fish in the first year
Nd	Deviation in the number of age-2 and older fish in the first year
εF	Assessment fishing mortality deviation
p	Age composition of the catch
ℓ_c	Likelihood component of the total catch
n_c	Length of the time vector in the time series of differences in catch estimation
$\hat{\sigma}_c$	Assessed standard deviation of the catch
$\ell_{\varepsilon F}$	A penalty for deviation from direct proportionality between fishing mortality and observed fishing effort
$n_{\varepsilon F}$	Length of the time vector in the time series of differences in effort estimation
$\hat{\sigma}_F$	Assessed standard deviation of fishing effort
ℓ_p	Likelihood component for the age composition of the fishery harvest
n_{eff}	Effective sample size

Background on Lake Whitefish in 1836 Treaty-Ceded Waters

The 1836 Treaty-ceded waters in the Great Lakes encompass portions of Lakes Huron, Michigan, and Superior (total surface area ≈ 5.8 million ha), and intensive fishery management in those waters has concentrated on Lake Whitefish and Lake Trout *Salvelinus namaycush*. The ceded waters include 18 separate Lake Whitefish management units (four were merged into a larger assessment unit). For most of the management/assessment units, SCAA models are used to annually estimate mortality rates and to project Lake Whitefish abundances at age (MSC 2015), as was described in detail by Truesdell and Bence (2016). The target harvest levels for both Lake Whitefish and Lake Trout are determined from a total mortality control rule. Use of the total mortality rate rather than the fishing mortality rate (F) as the control rule originally occurred because target mortality rates came from examination of the total mortality levels (estimated mainly from catch curves) that led to stock collapse (Healey 1975a, 1975b). Furthermore, for some stocks within treaty-ceded waters (most notably Lake Trout but also a few stocks of Lake Whitefish) there is evidence of time-varying natural mortality (M) due to predation by Sea Lampreys *Petromyzon marinus*. Sea Lamprey-related mortality in Lake Whitefish is not considered in this study, but when it is observed (wounds on surviving fish), its magnitude is estimated annually based on wounding data and is included as a separate mortality source in the assessments (Ebener et al. 2005). In the present study, the constant total mortality rate policy we applied was akin to a constant-F policy because M was assumed to be constant and known in both the operating model and the stock assessment model.

Operating model.—The basic structure of the operating model was based on that of Molton et al. (2012, 2013) and Li et al. (2015), but unlike these previous studies, the present operating model involved only a single fish stock. Recruitment of the simulated population was modeled through a Ricker stock–recruit function that was parameterized in terms of steepness and unfished spawning stock size (Table 1). The steepness parameter was used to represent population "productivity," with the unfished equilibrium spawning stock size fixed to the same value for all scenarios (Mangel et al. 2010). We chose the Ricker stock–recruit function because of empirical evidence for overcompensation in Lake Whitefish recruitment (Healey 1978; Henderson et al. 1983). Three steepness levels that were plausible for Lake Whitefish in the Great Lakes (Deroba and Bence 2012; Li et al. 2015) and that produced meaningful differences in equilibrium yield were chosen to represent low, medium, and high population productivity scenarios. Parameterization of the Ricker stock–recruit function was converted to standard form (Table 1, equations 1.1–1.5) based on equations given by Quinn and Deriso (1999). Unfished spawning biomass was based on recent assessment results for Lake Whitefish in the 1836 Treaty-ceded waters, although this only set the overall scale of the results and otherwise did not affect the study's outcomes.

Process errors about the Ricker stock–recruit relationship were assumed to be temporally autocorrelated (Table 1, equation 1.6). Stock–recruitment parameters (α', β) and process error parameters (ρ, σ_R) varied depending on the evaluated scenario (see the Experimental Design section). To generate random recruitment levels, the stock–recruitment parameter α' was scaled by

$$e^{-\frac{\sigma_R^2}{2(1-\rho^2)}}$$

so that the expectation of the stochastic form would equal the deterministic value and would not depend upon the assumed level of variation. The spawning stock biomass (SSB) was calculated as a function of abundance at age, maturity at age, and weight at age (Table 1, equation 1.7). Length at age was assumed to follow a von Bertalanffy growth model (Table 1, equation 1.8); weight at age and maturity at age were determined by power and logistic functions of length, respectively (equations 1.9 and 1.10), such that length at age, weight at age, and maturity at age were all fixed values.

Postrecruitment abundances at age (up to age 12, with the last age-class an aggregate group that included age-12 and older fish) were forward projected by using an exponential mortality model with a constant M of 0.25 and an F that was age and year specific (Table 1, equation 1.11). Total allowable catch (TAC) was determined via the management procedure described below. Actual harvest in each year was set equal to the TAC multiplied by a lognormal implementation error term with a coefficient of variation (CV) of 10% (Table 1, equation 1.12). The fully selected F that produced the actual harvest level given age-specific abundances was solved for by using a Newton–Raphson algorithm and Baranov's catch equation (Table 1, equation 1.13). Age-specific F-values were set equal to the solved F multiplied by age-specific selectivities (Table 1, equation 1.14). Selectivity was set to zero for age-2 and younger fish. Selectivities for other ages were calculated from a gamma function that produced a dome-shape selectivity pattern with peak selectivity at age 10 (Table 1, equation 1.15).

We used the same approach as Li et al. (2015) to determine initial abundances for each simulation. Specifically, initial abundances were set to the equilibrium values based on the target mortality rate and a deterministic version of our model. Also similar to the Li et al. (2015) study, during the initial 20-year period of each simulation, the harvest control rule based on the target mortality rate was applied to the actual abundances at age (i.e., the assessment modeling was skipped). The intent was to better ensure that the final 25 years of our 100-year simulations were from steady-state conditions. Furthermore, the initial 20-year period allowed the accumulation of data that were needed to perform a population assessment.

Assessment Model and Harvest Control Rule

We attempted to emulate key aspects of the management procedure used for Lake Whitefish in the 1836 Treaty-ceded waters, including data collection, stock assessment, and application of the harvest control rule. For data collection, the observed harvest, age composition, and effort data were collected annually and were available as assessment model inputs even though assessments were not necessarily conducted annually. Observed harvest differed from actual harvest as a result of observation error, which was modeled with a lognormal error term (Table 1, equation 1.16). The observed age composition of the harvest arose from a multinomial distribution with probabilities equal to the actual age composition. The observed fishing effort was a function of F, catchability, and lognormal observation error (Table 1, equation 1.17). The CVs of the error terms for observed harvest and effort and the effective sample size for the multinomial distribution used to produce the observed age composition were explored to represent the data quality factor in this research.

The SCAA assessment model involved fitting a population dynamics model to the observed data in order to estimate the parameters that were used to summarize stock status and determine target harvest (Table 1, equations 1.18–1.21). Only the most recent 20 years of data were used to fit the assessment models; recruitment, abundances at age in the first year, gamma function selectivity parameters, and catchability were estimated during model fitting. The M-values that were assumed in the SCAA assessment model were the same as those used for the true system. The parameters of the SCAA assessment model were estimated in AD Model Builder (Fournier et al. 2012). The objective function was the sum of three negative log-likelihood or log-prior components. Lognormal distributions were assumed for (1) the log-likelihood component associated with total fishery harvest and (2) the log-prior component associated with the relationship between F and effort (Table 1, equations 1.23 and 1.24; Fournier and Archibald 1982). The latter penalty is equivalent to predicting effort as proportional to estimated F and treating the deviations between observed and predicted fishing effort as lognormally distributed. A multinomial distribution was assumed for the log-likelihood component associated with the harvest age composition (Table 1, equation 1.25). By minimizing the negative log-likelihood (objective function), the assessment models were considered to have converged on a solution when the maximum gradient of the parameters was less than 0.001 and the Hessian matrix was positive definite.

We considered three assessment frequencies: annual, every 3 years (triennial), and every 5 years (quinquennial). We selected quinquennial as the lowest assessment frequency because other studies have suggested that 5 years generally would be sufficient for detecting and responding to one-time major changes (Shertzer and Prager 2007; De Leeuw et al. 2008; Brown et al. 2012). In the case of annual assessments, we considered scenarios with (L1) and without (L0) a 1-year lag between data collection and assessment—similar to the lag used in management of Lake Whitefish within the 1836 Treaty-ceded waters. For triennial and quinquennial assessments, a 1-year lag between data collection and assessment was always incorporated.

Regardless of assessment frequency, the target harvest levels were established annually. As a starting point, we first describe how target harvest levels were set for annual assessments, and we then explain how they were set for multiannual assessments. For the L1 scenario, there was a 1-year lag between data collection and incorporation in the stock assessment models, and the stock assessment estimated the abundances at age for the start of the year in which the assessment was done (the lag year). Abundances through the lag year were projected by using an exponential population model in which (1) the total mortality rate was assumed to be the mean of the last 3 years' values and (2) the recruits were assumed to be the mean from the most recent 10 years excluding the final year. During the year for which the target harvest was set, the projections assumed the same level of recruitment, but F was adjusted so that the target mortality rate was achieved. A similar process was used for the L0 scenario (i.e., annual assessment without a lag), but in that case, projection through a lag year was not required.

For the L1 and L0 scenarios, target harvests were determined in the same manner as that used for Lake Whitefish stocks in the 1836 Treaty-ceded waters. Target harvest was calculated from assessment-based estimates of abundance at age and the target F based on the total mortality target level (see the Experimental Design section) using Baranov's catch equation. The age-specific F-values were based on the estimated selectivity at age from the assessment multiplied by the target F.

For triennial and quinquennial assessments, we evaluated three approaches to setting target harvest levels in the years between assessments. In the first case, target harvests in the years between assessments were set equal to that determined from the last full assessment as described for L1 (a constant target [CT] harvest control rule). In the second case, target harvests in the years between assessments were based on multiyear projections after the 1-year data lag under the assumption that the stock experienced the target F in each of those years (a target-F [TF] harvest control rule). In the third case, target harvests in the years between full assessments were based on population projections that took into account each additional year of total harvest that was observed from the system. In each year, the fully selected F capable of producing the observed yield for that year was estimated by using the Newton–Raphson algorithm based on Baranov's catch equation and the assumed harvest weight at age and age-specific selectivities estimated by the last full assessment. The estimated fully selected F was then used to project abundance to the start of the next year, and the target harvest of the next year was then calculated in the same way as for L1 (adjusted by harvest [AH] control rule). Unlike the CT rule, the TF and AH control rules had the potential to produce different target harvests in the years between full assessments.

Experimental Design

In total, there were eight combinations of assessment frequency, lag interval, and approach for setting harvest levels in the years between assessments: L0, L1, CT3, CT5, TF3, TF5, AH3, and AH5 (Table 3); for example, "TF3" references a target-F harvest rule with triennial assessments. We refer to these eight combinations as "timetables." The timetables were evaluated for different combinations of data quality (two levels: low and high), population productivity (three levels: low, medium, and high), recruitment variation (two levels: average and high), and total finite mortality target (three levels: 0.45, 0.55, and 0.65; Table 4). The low and high values of data quality were chosen from within the ranges assumed in actual Lake Whitefish assessments, and we verified that in the simulations, these levels led to substantial differences in assessment error CV (e.g., CV = 18% [high quality] and 38% [low quality] for SSB in the last assessment year; and CV = 361% [high quality] and 402% [low quality] for the last estimated recruitment assuming an L0 timetable, medium productivity, average recruitment variation, and a total mortality target of 0.55). The total mortality target levels were selected because the current total mortality target for Lake Whitefish in the 1836 Treaty-ceded waters is 0.65, which previous research

TABLE 3. Description of timetables with respect to assessment frequency, length of the lag between data collection and assessment modeling, and approaches to setting harvest levels in years between assessments used in the simulations (NA = not applicable).

Timetable code	Assessment frequency (years)	Lag (years)	Harvest control rule
L0	1	0	NA
L1	1	1	NA
AH3	3	1	Adjusted by harvest information
CT3	3	1	Constant target harvest
TF3	3	1	Target fishing mortality (F)
AH5	5	1	Adjusted by harvest information
CT5	5	1	Constant target harvest
TF5	5	1	Target F

has identified as being too aggressive for low-productivity populations (Molton et al. 2012, 2013; Li et al. 2015). Because M was assumed to be constant at 0.25 and was known exactly in the stock assessment model, the three total mortality targets are the same as the F targets of 0.35, 0.55, and 0.80. Our preliminary results suggested that the influence of data quality on performance was minor; thus, for the sake of brevity, we chose to primarily focus on the results for the 18 combinations of factors with high-quality data. We also considered results for low data quality, medium productivity, average recruitment variation, and a total mortality target of 0.55 as well as results from four other combinations of low-data-quality scenarios (low or high population productivity in combination with a total mortality target of 0.45 or 0.65 and high recruitment variation). Overall, there were 23 combinations of data quality, productivity, recruitment variation, and total mortality target with which the timetables were evaluated (Table 5). To facilitate comparisons, we categorized the 23 scenarios into three groups: (1) baseline, (2) single-effect cases, and (3) combination-effect cases. The baseline category assumed a high level of data quality, medium productivity, average recruitment variation, and a total mortality target of 0.55. The baseline scenario was meant to evaluate the long-term performance of the different assessment frequencies and harvest level options (Table 5). We chose the total mortality target level of 0.55 for the baseline evaluations because this mortality rate has been identified as sustainable for many plausible levels of Lake Whitefish productivity (Molton et al. 2012, 2013; Li et al. 2015). The single-effect and combination-effect cases evaluated the robustness of the different assessment frequencies and harvest level options to differing levels of data quality, population productivity, total mortality target, and recruitment variation, both singly and in combination (Table 5).

Performance Statistics

For each evaluated scenario, 1,000 simulations were conducted. Models were run for 100 years, but only the results from the final 25 years were summarized to evaluate performance. Performance statistics were generated from the true system, which was used to describe the real status of the stock and the fishery. Performance statistics consisted of (1) the risk of low SSB (i.e., the proportion of simulation years [among the final 25 years] in which SSB was less than 20% of its unfished level, $P[\text{SSB} < B_{20\%}]$); (2) the mean annual relative yield (i.e., mean yield of the final 25 simulation years relative to the maximum sustainable yield [MSY] at equilibrium, which differed by productivity level); (3) the mean interannual variation (IAV) in yield over the final 25 years of the simulation; and (4) the mean realized F (the actual F experienced by the stocks) over the final 25 years. The median of the performance statistics calculated over all 1,000 simulations for an evaluated scenario was used to characterize the expected performance, and the interquartile range of the performance statistics was used to characterize variability.

Sensitivity Analysis

Sensitivity analyses were conducted to explore the robustness of study results to the assumed M, the functional form of the stock–recruitment relationship, and the assumed levels of recruitment variability and data quality in the operating model (Table 6). These were chosen to verify whether the results remained applicable for conditions that were not included in our base simulations but that might be relevant for other species or for some actual Lake Whitefish populations. The sensitivity analyses were performed under conditions similar to baseline (e.g., high data quality, medium productivity, average recruitment variation, and a total mortality control rule of 0.55) except in cases where the sensitivity analyses called for a departure from baseline conditions (e.g., very high recruitment variability).

For the first sensitivity analysis, time-varying M was incorporated into the operating model even though it was assumed constant at 0.25 in the assessment model. In the operating model, M was modeled through a lag-1 autoregressive process with an overall mean of 0.25, an autocorrelation coefficient of 0.3, and an SD of 0.15 for white noise process error (Table 6, equation 6.1).

TABLE 4. Coefficient values for parameters that were used to generate different levels of data quality, productivity, and recruitment variation in simulating data and in the stock assessment model (CV = coefficient of variation).

Coefficient name	Coefficient level: Average	Low	Medium	High	Definition
Data quality					
n		50		200	Effective sample size
Harvest CV (%)		20		10	CV for observed harvest about actual harvest
Effort CV (%)		40		20	CV for effort about its expectation
Productivity					
Steepness		0.7	1.3	1.9	
α'		5.23×10^{-4}	1.13×10^{-3}	1.82×10^{-3}	Ricker stock–recruitment parameter
β		1.51×10^{-10}	2.26×10^{-10}	2.72×10^{-10}	Same as above
MSY (metric tons)		135	258	370	Maximum sustainable yield
F_{MSY}		0.40	1.00	1.75	Fishing mortality that produces MSY
Recruitment variation					
ρ	0.45			0.58	Autocorrelation coefficient
σ_R	0.78			0.89	Standard deviation in recruitment process error

For the second sensitivity analysis, recruitment of the simulated population was modeled via a Beverton–Holt stock–recruit function. As with the Ricker stock–recruitment function, the Beverton–Holt function was parameterized in terms of steepness and unfished SSB. We increased the unfished SSB to 1,039.4 metric tons and set the steepness equal to 0.82 so as to achieve the same MSY (258 metric tons) as the medium productivity level for the Ricker stock–recruit function. Details of the parameterization process are provided in Appendix Table A.1.

For the next two sensitivity scenarios, we explored alternative recruitment variability levels. The first level represented very high recruitment variation, wherein we kept the autocorrelation coefficient at 0.58 but increased the SD in recruitment process error to 1.5. For the second level, we removed the autocorrelation component of recruitment variation (i.e., recruitment variation was simply white noise) and increased the SD in recruitment process error to 0.8734 so that the same stationary variance of the average recruitment variation level in Table 4 could be maintained.

For the last sensitivity scenario, we considered a situation of extremely poor data quality for the assessment model. An examination of retrospective errors for actual Lake Whitefish assessments was suggested by one of this paper's reviewers (e.g., converged estimates versus an estimate for the same year when no subsequent data were used); the results indicated that real assessment error was often consistent with the level of assessment error occurring in our simulations when based on high- and low-quality data (e.g., across stocks, over 60% of error values were less than 50%). However, there were cases

TABLE 5. Combinations of data quality, recruitment variation, productivity, and total mortality targets used in simulations to evaluate how assessment frequency influences the achievement of management objectives. See Table 4 for the coefficient values corresponding to the data quality, productivity, and recruitment variation levels.

Scenario	Scenario code	Data quality level	Productivity level	Recruitment variation level	Total mortality target
Baseline	B	High	Medium	Average	0.55
Single effects	D	Low	Medium	Average	0.55
	R	High	Medium	High	0.55
	P	High	Low and high	Average	0.55
	A	High	Medium	Average	0.45, 0.65
Combination effects	RP	High	Low and high	High	0.55
	RA	High	Medium	High	0.45, 0.65
	PA	High	Low and high	Average	0.45, 0.65
	RPA	High	Low and high	High	0.45, 0.65
	DRPA	Low	Low and high	High	0.45, 0.65

TABLE 6. Scenarios for sensitivity analyses (CV = coefficient of variation).

Sensitivity scenario	Model description	Parameter values	Equation number
Time-varying natural mortality (M_TimeV)	$M_{(y)} = \bar{M}e^{\varepsilon_{M(y)} - \frac{\sigma_M{}^2}{2(1-\rho_M{}^2)}}$, where $\varepsilon_{M(y)} = \rho_M \times \varepsilon_{M(y-1)} + \tau_{M(y)}$ $\tau_{M(y)} \sim \mathrm{N}(0,\sigma_M^2)$	Mean natural mortality rate $\bar{M} = 0.25$ (year^{-1}); SD in M perturbations (σ_M) = 0.15; autocorrelation in M perturbations (ρ_M) = 0.3	6.1
Beverton–Holt stock recruitment function (BH)	$R_{(y)} = \frac{S}{\alpha+\beta S}e^{\varepsilon_{R(y)} - \frac{\sigma_R^2}{2\left(1-\rho_R^2\right)}}$, where $\varepsilon_{R(y)} = \rho_R \times \varepsilon_{R(y-1)}$ $+ \tau_{R(y)} \tau_{R(y)} \sim N(0, \sigma_R^2)$	BH parameter (α) = 502.7; BH parameter (β) = 4.18×10^{-7}; ρ_R and σ_R = same as the average levels in Table 4	6.2
Very high recruitment variability (Rec_H)	See Table 4	$\rho = 0.58$; $\sigma_R = 1.5$	
Recruitment variability without autocorrelation (Rec_0)	See Table 4	$\rho = 0$; $\sigma_R = 0.8734$	
Very poor assessment data (Data_VLow)	See Table 4	$n = 25$; harvest CV = 40%; effort CV = 80%	

in which retrospective errors seemed too large to be consistent with the simulated assessments (J. R. Bence, unpublished analysis). The sensitivity analysis with extremely low data quality used (1) an effective sample size of 25 for generating observed age composition from a multinomial distribution and (2) CVs of 40% and 80% for the observed harvest and effort error terms, respectively. This translated into an assessment error CV of 88% for SSB in the final year of the simulated assessments and an assessment error CV of 796% for the final estimated recruitment in the simulated assessments; these error levels were consistent with the more extreme errors seen in the retrospective analysis.

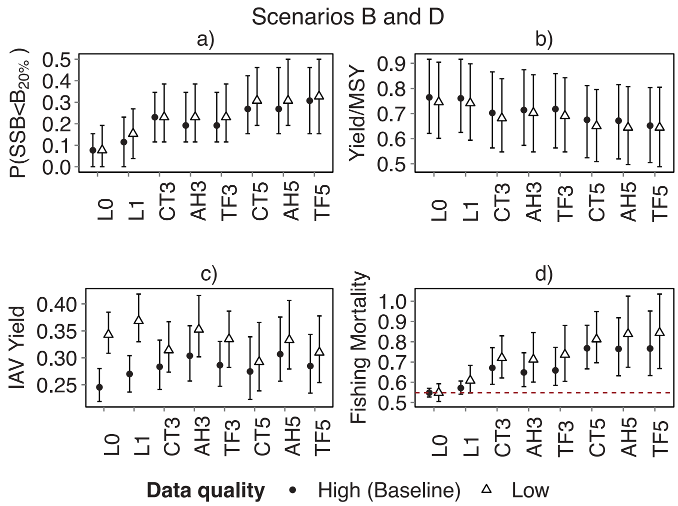

FIGURE 1. Simulation results (median ± interquartile range) for different data quality levels and eight different timetables (defined in Table 3; scenarios B and D, Table 5) that were used to evaluate how assessment frequency influences the achievement of management objectives: **(a)** the proportion of the final 25 simulation years in which spawning stock biomass (SSB) was less than 20% of unfished biomass, $P(\text{SSB} < B_{20\%})$; **(b)** mean annual relative yield (mean yield relative to the MSY at equilibrium); **(c)** mean interannual variation (IAV) in yield; and **(d)** mean realized fishing mortality rate F (dashed line = target F) averaged over the final 25 years of the simulations. Simulation conditions consisted of a medium productivity level, average recruitment variation, and a total mortality control rule of 0.55.

RESULTS

Simulation results for all combinations of data quality, population productivity, recruitment variation, total mortality target, and assessment timetable are given in Supplementary Table S.1 available in the online version of this article.

Baseline

For our baseline evaluations (i.e., high data quality, medium productivity level, average recruitment variation, and a total mortality target = 0.55), the incorporation of a 1-year lag between data collection and assessment increased the $P(\text{SSB} < B_{20\%})$ and mean IAV in yield by approximately 50% and 7%, respectively, compared to the use of annual assessments without a lag (Figure 1). Neither relative yield nor realized F was sensitive to the incorporation of a 1-year lag, as there were only minor differences in these performance statistics for the L0 and L1 timetables (Figure 1).

Switching from annual to multiannual assessments led to increases in $P(\text{SSB} < B_{20\%})$ and realized F and to decreases in relative yield (Figure 1). These effects were progressive, with $P(\text{SSB} < B_{20\%})$ being approximately 78% and 172% higher and relative yield being 7% and 15% lower for triennial and quinquennial assessments, respectively, than for annual assessments with a 1-year lag when averaged over the different harvest rules. The realized F was approximately 4% over the target mortality for annual assessments with a 1-year lag, and realized F increased to about 20% and 51% greater than the target for triennial and quinquennial assessments, respectively, when averaged over the harvest rules. The variability in the simulation results for realized F also increased progressively as assessments became less frequent.

The differences among the various harvest rules for a particular assessment frequency were generally smaller than the differences observed across the assessment frequencies. Nevertheless, there were still some noticeable differences among harvest rules for the baseline scenario. For triennial and quinquennial assessments, the CT rule produced the smallest IAV in yield (Figure 1). However, for triennial assessments, the CT rule also resulted in the highest $P(\text{SSB} < B_{20\%})$, the highest realized F, and the smallest relative yield. For quinquennial assessments, the TF rule resulted in the greatest $P(\text{SSB} < B_{20\%})$ and the highest realized F. The AH rule resulted in the greatest IAV in yield for both triennial and quinquennial assessments; however, the AH rule generally was among the better-performing rules for the other performance metrics.

Single Effects

Data quality.—Lower-quality assessment data generally increased the $P(\text{SSB} < B_{20\%})$, the IAV in yield, and the realized F and decreased the relative yield in comparison with the baseline scenario (Figure 1). For annual assessments without a lag, the expected $P(\text{SSB} < B_{20\%})$ and realized F were similar across the two data quality levels, although the variability in results (as measured by the simulation interquartile ranges [IQRs]) for these metrics increased with lower-quality data. The IAV in yield was the performance measure that appeared to be the most sensitive to data quality based on within-timetable differences. For each timetable, the IAV in yield was greater with lower-quality data, and the within-timetable difference decreased as the assessment frequency decreased.

With respect to the performance of different assessment frequencies with low-quality assessment data, results were similar to those found with high-quality data, with the exception of IAV in yield. Less-frequent assessments resulted in a

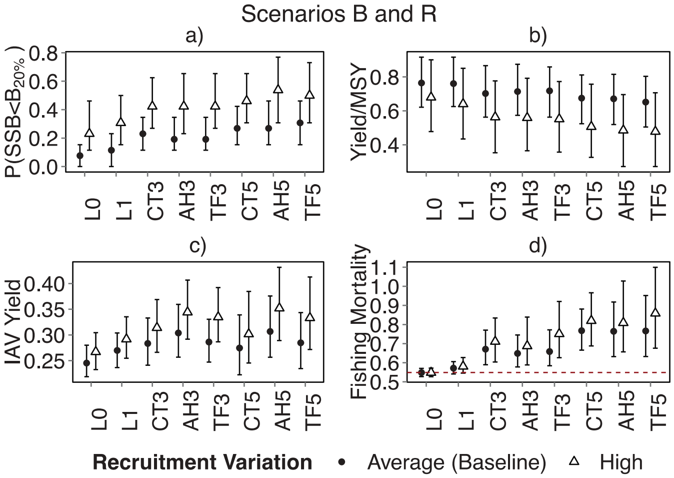

FIGURE 2. Simulation results (median ± interquartile range) for average and high recruitment variation and eight different timetables (defined in Table 3; scenarios B and R, Table 5) that were used to evaluate how assessment frequency influences the achievement of management objectives: **(a)** mean $P(\mathrm{SSB} < B_{20\%})$; **(b)** mean annual relative yield (mean yield relative to the MSY at equilibrium); **(c)** mean IAV in yield; and **(d)** mean realized F (dashed line = target F) averaged over the final 25 years of the simulations (acronyms and symbols are defined in Figure 1). Simulation conditions consisted of high data quality, a medium productivity level, and a total mortality control rule of 0.55.

greater $P(\mathrm{SSB} < B_{20\%})$, a greater realized F, and a smaller relative yield, with effects being progressive as assessments became more infrequent, just as with high-quality data. However, for IAV in yield, annual assessments with or without a 1-year lag performed similarly to—or in some cases worse than—the other timetables (Figure 1). With respect to the performance of different harvest rules for multiannual assessments, the results depended on the performance metric. For IAV in yield, the AH rule continued to result in the greatest IAV, whereas the CT rule continued to generate the smallest IAV in yield for both triennial and quinquennial assessments; however, the differences between the rules became larger for low-quality assessment data. For the other performance metrics, differences in the expected performance of the different rules became negligible under conditions of low-quality data (Figure 1).

Recruitment variation.—Greater recruitment variation increased the $P(\mathrm{SSB} < B_{20\%})$, the IAV in yield, and the realized F and decreased the relative yield (Figure 2). Greater recruitment variation also resulted in greater variability in simulation results for each of the performance metrics, as measured by the IQRs of the simulation results (Figure 2). Qualitatively, the patterns in results discussed for the baseline scenario were unaffected by greater recruitment variation: namely, less-frequent assessments resulted in a greater $P(\mathrm{SSB} < B_{20\%})$, greater IAV in yield, a greater realized F, and a smaller relative yield (Figure 2). With respect to performance of individual timetables, one consequence of high recruitment variation was that the incorporation of a 1-year lag with annual assessments resulted in a 6% decrease in expected relative yield in comparison with the use of annual assessments without a lag (Figure 2). In terms of harvest rules, for triennial assessments the harvest rules performed similarly with respect to $P(\mathrm{SSB} < B_{20\%})$ and realized yield; the AH control rule still produced the largest IAV in yield but also resulted in the lowest realized F-values. For quinquennial assessments, the AH rule resulted in the highest $P(\mathrm{SSB} < B_{20\%})$ and greatest IAV in yield, whereas the TF rule generated the greatest realized F (Figure 2).

Productivity of the population.—A low productivity level increased the $P(\mathrm{SSB} < B_{20\%})$ and decreased the relative yield in comparison with the performance from a medium productivity level (Figure 3). Low productivity also led to greater variability in results for these performance metrics (Figure 3). The effect of low productivity on IAV in yield and realized F depended on the timetable. For most timetables, low productivity decreased both the IAV in yield and the realized F (Figure 3). The main exceptions to this were (1) the AH rule for triennial assessments, where the IAV in yield was greater than that found under the medium productivity level; and (2) the TF harvest rule for quinquennial assessments, wherein the realized F was greater than that observed under medium productivity (Figure 3). Conversely, a high productivity level decreased the $P(\mathrm{SSB} < B_{20\%})$ and relative yield and increased the IAV in yield and realized F in comparison with medium productivity (Figure 3). Given that

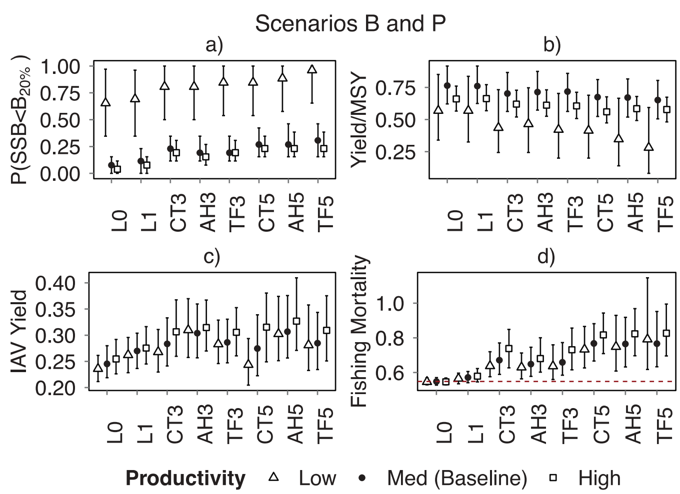

FIGURE 3. Simulation results (median ± interquartile range) for low, medium, and high productivity levels and eight different timetables (defined in Table 3; scenarios B and P, Table 5) that were used to evaluate how assessment frequency influences the achievement of management objectives: **(a)** mean $P(\mathrm{SSB} < B_{20\%})$; **(b)** mean annual relative yield (mean yield relative to the MSY at equilibrium); **(c)** mean IAV in yield; and **(d)** mean realized F (dashed line = target F) averaged over the final 25 years of the simulations (acronyms and symbols are defined in Figure 1). Simulation conditions consisted of high data quality, average recruitment variation, and a total mortality control rule of 0.55.

relative yield is the ratio between mean yield and MSY, the decrease in relative yield simply meant that a lower fraction of MSY was observed for the high productivity levels—not that the actual yield was lower.

The consequences of assessment frequency were not affected by differing levels of productivity. Less-frequent assessments still resulted in a higher $P(\mathrm{SSB} < B_{20\%})$, greater IAV in yield, greater realized F, and a smaller relative yield (Figure 3). In terms of harvest rules, the results largely varied depending on the productivity level, assessment frequency, and performance metric. The one consistent result was that the AH rule continued to produce the greatest IAV in yield (Figure 3). For low productivity levels, the CT rule generated the smallest $P(\mathrm{SSB} < B_{20\%})$ and IAV in yield for both triennial and quinquennial assessments; the AH control rule resulted in the largest relative yield for triennial assessments, while the CT rule produced the largest relative yield for quinquennial assessments. Much like the findings obtained from high recruitment variation, high productivity levels reduced the amount of difference observed among the harvest rules for particular assessment frequencies, such that the expected difference among the rules became negligible.

Mortality targets.—As the target mortality level increased, the $P(\mathrm{SSB} < B_{20\%})$, IAV in yield, and realized F increased in comparison with the values observed from a target total mortality level of 0.55 (Figure 4). The influence of target mortality rate on relative yield depended on the timetable. Relative to the 0.55 target mortality level, a total mortality target of 0.65 produced an increase in relative yield for annual assessments but a slight decrease in relative yield for multiannual assessments. Greater target harvest levels also increased the variability in results for each of the performance metrics (Figure 4).

The qualitative pattern in results with regard to the consequences of decreased assessment frequency was largely unaffected by the target mortality level. One consequence of a higher total mortality target level was a larger difference in relative yield between multiannual and annual assessments (Figure 4). Likewise, in terms of harvest rules, there was not a single harvest rule that consistently outperformed the other rules across all metrics. Rather, the results depended on the performance metric, the total mortality target level, and whether assessments were triennial or quinquennial.

Combination Effects

Productivity and recruitment variation.—The combination of low productivity and high recruitment variation resulted in greater differences among the different timetables for some of the performance metrics (Figure 5). For instance, median relative yield was 0.25 for the L0 timetable versus 0.03 for the TF5 timetable. Conversely, under baseline conditions, median relative yield ranged from 0.65 to 0.77 across the timetables. In addition, the effect of recruitment variation on relative yield and realized F was much greater under low-productivity conditions. For example, median relative yield for the L0 timetable decreased by 56% from average to high recruitment variation under low productivity, whereas only a 9% decrease was observed for the same change in recruitment

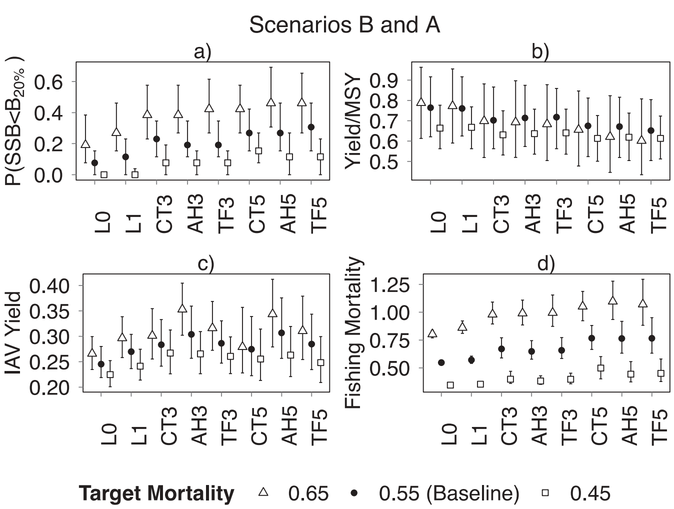

FIGURE 4. Simulation results (median ± interquartile range) for three different total mortality control rules (total mortality targets = 0.45, 0.55, and 0.65) and eight different timetables (defined in Table 3; scenarios B and A, Table 5) that were used to evaluate how assessment frequency influences the achievement of management objectives: **(a)** mean $P(\mathrm{SSB} < B_{20\%})$; **(b)** mean annual relative yield (mean yield relative to the MSY at equilibrium); **(c)** mean IAV in yield; and **(d)** mean realized F (dashed line = target F) averaged over the 25 years of the simulations (acronyms and symbols are defined in Figure 1). Simulation conditions consisted of high data quality, a medium productivity level, and average recruitment variation.

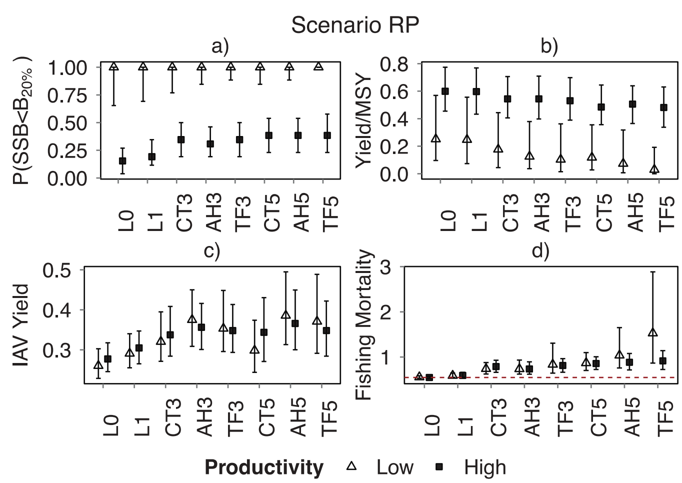

FIGURE 5. Simulation results (median ± interquartile range) for low and high productivity levels and eight different timetables (defined in Table 3; scenario RP, Table 5) that were used to evaluate how assessment frequency influences the achievement of management objectives: **(a)** mean $P(\mathrm{SSB} < B_{20\%})$; **(b)** mean annual relative yield (mean yield relative to the MSY at equilibrium); **(c)** mean IAV in yield; and **(d)** mean realized F (dashed line = target F) averaged over the 25 years of the simulations (acronyms and symbols are defined in Figure 1). Simulation conditions consisted of high data quality, high recruitment variation, and a total mortality control rule of 0.55.

variation under high productivity. Regardless of the timetable, there was a high risk of stock depletion for the low productivity level when recruitment variation was high (Figure 5). Apart from this, the consequences of switching from annual assessments to multiannual assessments remained unchanged. With respect to harvest rules for determining target harvest in between assessment years, there again was no rule that consistently outperformed the other rules (Figure 5). One consequence of high productivity and high recruitment variation was that the IAV in yield exhibited somewhat muted differences among rules, although the AH rule continued to generate the greatest variability in yield (Figure 5). The CT rule resulted in the greatest relative yield for triennial and quinquennial assessments (Figure 5) but also produced the greatest levels of realized F and, like the other rules, was associated with a high risk of stock depletion.

Mortality target and recruitment variation.—The influences of the target mortality level and assessment frequency when recruitment variation was high were consistent with results obtained for an average level of recruitment variation. The only exception was observed for relative yield under the high total mortality target level of 0.65 (Figure 6). In that case, the L1 timetable decreased relative yield by 16% compared with the L0 timetable. Additionally, for the multiannual assessment scenarios, the CT harvest rule provided the highest relative yield, followed by the AH and TF rules. The scale of relative yield also changed with high recruitment variation. For the L0 timetable and the total mortality target of 0.65, the expected relative yield decreased from 0.79 under average recruitment variation to 0.62 under high recruitment variation (Figures 4, 6). Conversely, when the total mortality target level was low (0.45), the expected relative yield only decreased from 0.66 to 0.60 in response to the same change in recruitment variation (i.e., compare Figure 5 with Figure 6).

Productivity and total mortality target.—The different combinations of productivity level (high and low) and total mortality target level (high and low) yielded results similar to those obtained from the other examined scenarios, with few exceptions. First, when productivity was low and the total mortality target level was high, all timetables resulted in a P (SSB < $B_{20\%}$) of around 100% (Figure 7). Additionally, whereas with the other examined scenarios the L1 timetable was expected to produce a greater relative yield than multiannual assessments regardless of harvest rule, both the CT3 and CT5 timetables produced median relative yields that were comparable to those generated by the L1 timetable under a combination of low productivity and high total mortality.

Recruitment variation, productivity, and total mortality target.—Under conditions of high recruitment variation, the risk of stock depletion increased considerably for some combinations of productivity level and total mortality target level relative to the results that were observed under average recruitment variation. With high recruitment variation, the P (SSB < $B_{20\%}$) increased for (1) the low productivity × low total mortality target combination and (2) the high productivity × high total mortality target combination (i.e., compare Figure 7 with

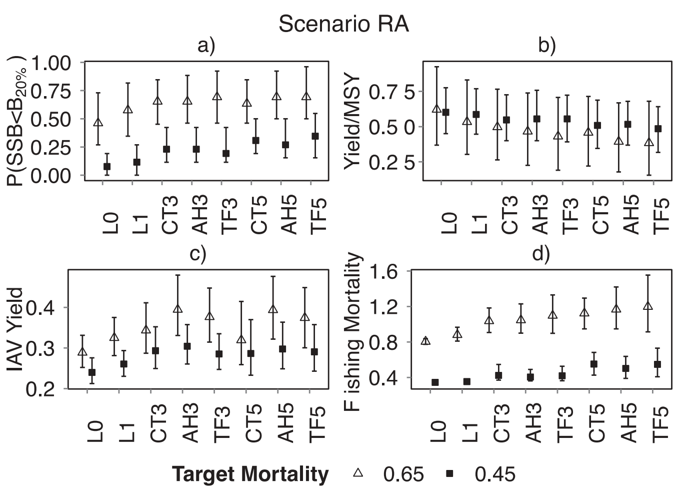

FIGURE 6. Simulation results (median ± interquartile range) for two different total mortality control rules (total mortality targets = 0.45 and 0.65) and eight different timetables (defined in Table 3; scenario RA, Table 5) that were used to evaluate how assessment frequency influences the achievement of management objectives: **(a)** mean P(SSB < $B_{20\%}$); **(b)** mean annual relative yield (mean yield relative to the MSY at equilibrium); **(c)** mean IAV in yield; and **(d)** mean realized F (dashed line = target F) averaged over the 25 years of the simulations (acronyms and symbols are defined in Figure 1). Simulation conditions consisted of high data quality, a medium productivity level, and high recruitment variation.

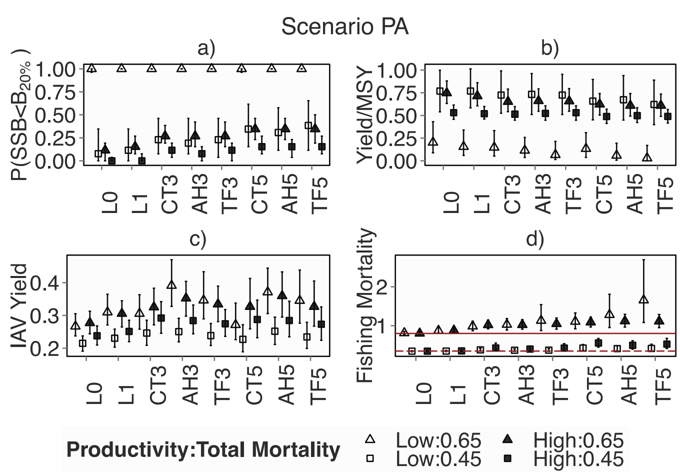

FIGURE 7. Simulation results (median ± interquartile range) for low and high productivity levels in combination with two total mortality control rules (total mortality targets = 0.45 and 0.65) and eight different timetables (defined in Table 3; scenario PA, Table 5) that were used to evaluate how assessment frequency influences the achievement of management objectives: **(a)** mean P (SSB < $B_{20\%}$); **(b)** mean annual relative yield (mean yield relative to the MSY at equilibrium); **(c)** mean IAV in yield; and **(d)** mean realized F (dashed line = target F) averaged over the 25 years of the simulations (acronyms and symbols are defined in Figure 1). Simulation conditions consisted of high data quality and average recruitment variation.

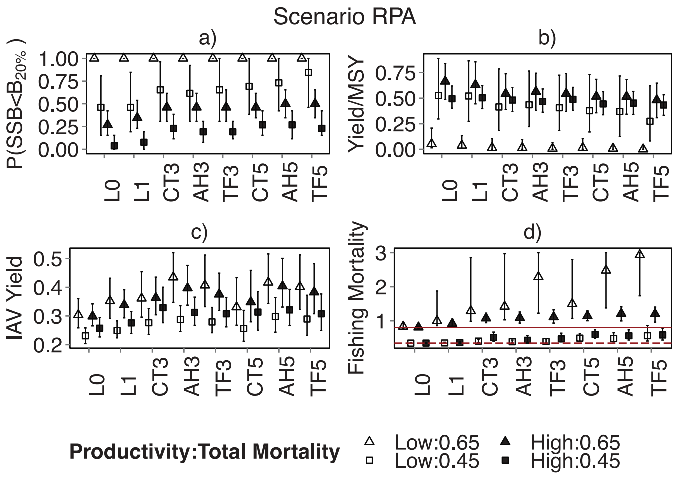

FIGURE 8. Simulation results (median ± interquartile range) for low and high productivity levels in combination with two total mortality control rules (total mortality targets = 0.45 and 0.65) for eight different timetables (defined in Table 3; scenario RPA, Table 5) that were used to evaluate how assessment frequency influences the achievement of management objectives: **(a)** mean $P(\text{SSB} < B_{20\%})$; **(b)** mean annual relative yield (mean yield relative to the MSY at equilibrium); **(c)** mean IAV in yield; and **(d)** mean realized F (dashed line = target F) averaged over the 25 years of the simulations (acronyms and symbols are defined in Figure 1). Simulation conditions consisted of high data quality and high recruitment variation.

Figure 8). Relative yield was also somewhat sensitive to recruitment variation; this was most noticeable for (1) the low productivity × high total mortality target combination and (2) the low productivity × low total mortality target combination. For the combination of low productivity and a high total mortality target, relative yield decreased substantially for high recruitment variation in comparison with that observed under average recruitment variation conditions, especially for multiannual assessment scenarios. The median relative yield was lower than 0.015 for all multiannual timetables, and none of the harvest rules outperformed the others. For the low productivity × low total mortality target combination, the only noticeable difference between the high and average levels of recruitment variation was for the TF rule, which resulted in the highest realized F for both triennial and quinquennial assessments and the highest risk of stock depletion and lowest relative yield for quinquennial assessments. Otherwise, the qualitative patterns in results were similar to the baseline scenarios.

Data quality, recruitment variation, productivity, and total mortality target.—The quality of assessment data had no influence on the expected performance of timetables and harvest rules when differing productivity levels and total mortality targets were considered under conditions of high recruitment variation (results not shown).

Sensitivity Analysis

Results with respect to assessment frequency and approaches to setting harvest levels in the years between assessments were generally insensitive to changes in assumptions about M, the functional form of the stock–recruit relationship, recruitment variation, and data quality, as evidenced by consistency in the resulting patterns relative to baseline conditions. When assuming (1) temporally autocorrelated M, (2) a Beverton–Holt stock–recruit relationship in the operating model, or (3) no temporal autocorrelation in recruitment, different approaches to setting harvest levels in the years between assessments produced very similar results, and the use of less-frequent assessments led to increases in the $P(\text{SSB} < B_{20\%})$, IAV in yield, and realized F and to decreases in relative yield (Figure 9). Under a scenario of very high recruitment variability, performance metrics exhibited large changes relative to the values observed under baseline conditions (Figure 9). All timetables resulted in a median $P(\text{SSB} < B_{20\%})$ near 100% and a relative yield lower than 10%. When assessment data quality was poor, the results of $P(\text{SSB} < B_{20\%})$ and relative yield were similar to those from the baseline scenario for all evaluated timetables (Figure 9). The IAV in yield and realized F were the most sensitive to very poor data quality. Compared with the baseline scenario in which less-frequent assessments resulted in slightly higher IAV in yield, very poor quality assessment data produced greater IAV in yield for more-frequent assessments. Very poor quality assessment data also resulted in greater IQRs for realized F than the baseline scenario.

DISCUSSION

Our goals for this study were to evaluate how changes in the frequency of stock assessment affected the achievement of management objectives and how this was influenced by factors such as population productivity, the target harvest level, recruitment variation, and the quality of the assessment data. Additionally, for multiannual assessments, we sought to determine how different approaches to setting target harvests for years between assessments performed relative to each other. We found that across most of the conditions considered herein, decreasing the assessment frequency resulted in a greater risk of stock depletion, lower relative yield, and greater realized F. Multiannual assessments also produced greater IAV in yield, but unlike the other performance metrics, IAV in yield did not necessarily get progressively worse when changing from triennial to quinquennial assessments. Although our simulations were based on Lake Whitefish in the Great Lakes, our sensitivity analyses suggested that the results can be generalized beyond this specific application.

If there were no costs to conducting assessments of fish stocks, then our results would generally support the use of annual assessments without a lag between data collection and assessment. However, it would be unrealistic to not consider our results in light of assessment costs. The trade-offs between the costs of conducting more-frequent assessments and the benefits of obtaining higher yields and avoiding the risk of

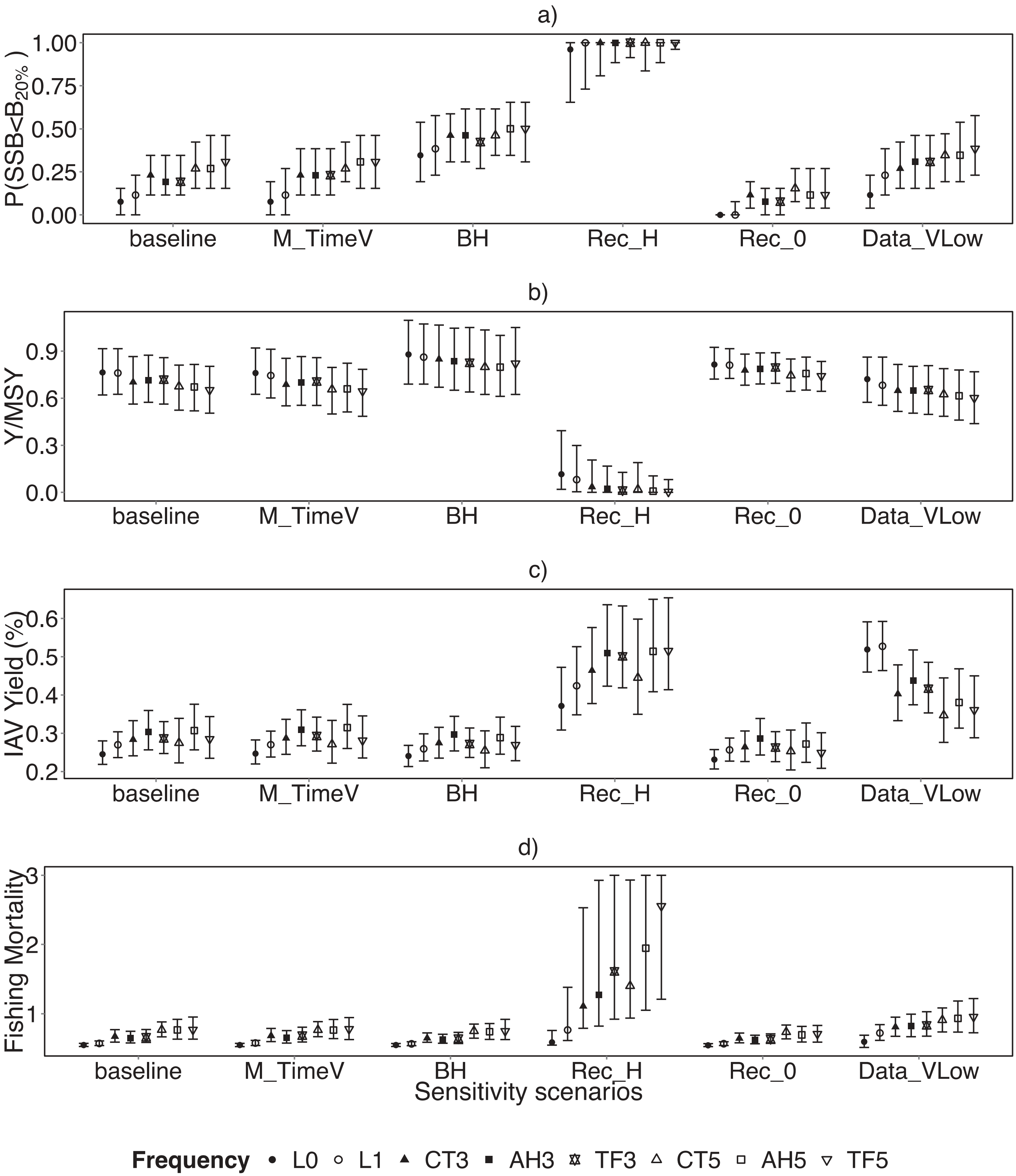

FIGURE 9. Results of sensitivity analyses: **(a)** the proportion of the final 25 simulation years in which spawning stock biomass (SSB) was less than 20% of unfished biomass, $P(\mathrm{SSB} < B_{20\%})$; **(b)** mean annual relative yield (mean yield [Y] relative to the maximum sustainable yield [MSY] at equilibrium); **(c)** mean interannual variation (IAV) in yield; and **(d)** mean realized fishing mortality rate averaged over the 25 years of the simulations (timetables [frequency] are defined in Table 3; sensitivity scenarios [x-axis] are defined in Table 6).

depletion must be considered on a system- and stock-specific basis. Based on our results, the obvious candidates for less-frequent assessments are those where target *F*-values are low, productivity is high, or both. In such situations, the risks of substantial depletion are generally low, and the costs in terms of foregone yield are lower than would be the case under conditions of lower productivity or higher target *F*. Our conclusions correspond to some extent with those from the ICES (2012) report, which concluded that stocks under modest exploitation should be considered as potential candidates for a lower assessment frequency (i.e., less than annual). An additional condition specified by ICES (2012) was that stocks should be stable, which is similar to our recommendation that low-productivity stocks are less-viable candidates for multiannual assessments.

Most of the published studies that have considered stock assessment frequency have focused on the consequences of management delays after major shifts in population dynamics and/or ecosystem characteristics (Shertzer and Prager 2007; De Leeuw et al. 2008; Brown et al. 2012). Shertzer and Prager (2007) and De Leeuw et al (2008) found that delays in management response could result in long-term losses and longer recovery times. However, those studies both focused on stocks that were substantially depleted, and the delays that resulted in the greatest loss were on the order of 10 years or more. Brown et al. (2012) noted that if environmental change caused population declines, delays greater than 5 years increased the chance of population collapse. Unlike these previous studies, our study was designed to explore the long-term consequences of assessment frequency under steady-state conditions rather than evaluating the responses to a one-time major change. Nevertheless, these other studies found that quinquennial assessments—the longest assessment frequency we considered—were sufficient for detecting and responding to one-time changes.

The 1-year lag between data collection and incorporation in the stock assessment model matches the process that has been used for Lake Whitefish stocks in 1836 Treaty-ceded waters of the Great Lakes since 2000. To our knowledge, this study is the first evaluation of the long-term influence of such a lag. Management agencies often expend considerable effort to avoid lags of this nature so that harvest recommendations can be made based on the most recent collected data. For example, there are no lags in the management system for Walleyes *Sander vitreus* in Lake Erie and Lake Trout in 1836 Treaty-ceded waters (Thomas and Haas 2005; MSC 2015). Our results suggest that for fish species with life histories similar to that of the Lake Whitefish, removal of the 1-year lag may slightly reduce the risk of stock depletion and moderately decrease the IAV in yield, but it would not be expected to exert a major impact on typical yield levels. The higher IAV in yield when a lag was used reflects somewhat poorer-quality assessments and thus a more variable realized *F*, which outweighs any added constancy resulting from an assumption that the most recent recruitment was equal to a historical average. Thus, with regard to stable stocks that are not in danger of being overfished and for which constancy in yield is not of overriding importance, avoidance of the 1-year lag may not be worth the associated costs, such as conducting assessments with data that have not been properly vetted.

The consequences of a 1-year lag became more pronounced as recruitment variation and mortality targets increased. We attributed this result to the process used in forward-projecting the population conditions through the lag (i.e., mean recruitment and *F* estimated as part of the stock assessment for the last 10 years [recruitment] or 3 years [*F*]). Mean recruitment generally exceeded the typical values due to the rightward skew of recruitment distributions, and the average *F* tended to be lower than typical values—also due to the skewed distributions of values. This mechanism additionally contributed to the performance of the multiannual assessments. An understanding of what causes some of the costs of lags and infrequent assessments suggests that these problems could be ameliorated by changes in the projection methods used when calculating harvest limits. In particular, future analyses could consider the potential benefits of using different projections of recruitment and *F* (e.g., using median rather than mean levels of recruitment or *F* so that values are less sensitive to outliers).

Although data quality did not influence our main conclusions about when to implement less-frequent assessments, clearly there are limits to the applicability of this result. The availability of higher-quality data than what was considered here or the use of different types of data (e.g., annual surveys providing absolute measures of population abundance) might lead to greater benefits for annual assessments. As evidenced by our sensitivity analysis, one consequence of very low data quality is that more-frequent assessments can generate greater IAV in yield; thus, there may be benefits to adopting a more conservative and constant strategy (i.e., less-frequent assessments) in such situations. Given the range of data quality for important fish stocks in the Great Lakes and elsewhere, our results appear robust to data quality. However, we must emphasize that we considered data quality in terms of precision and rather than evaluating a case in which data are of low quality due to bias. Likewise, we did not consider model misspecification except in the sensitivity analysis where *M* varied but was not accounted for in the assessment. We did not focus on bias or model misspecification because these effects are potentially of infinite variety (Maunder and Piner 2015). Some cases of bias and model misspecification might simply change the scale of population assessment; this would not change the relative performance of annual or multiannual assessments provided that the target *F* is adjusted to account for the bias. Some types of bias or model misspecification might, however, lead to such high estimation error that there is simply no point to frequent assessments, and one could just as well manage for average conditions (Chen et al. 2003; ICES 2012).

One consequence of multiannual assessments was that the realized F (i.e., the true F that the stock experiences) was greater than that observed when assessments were conducted annually. When we checked the trajectory of realized F for individual simulations under the multiannual stock assessment scenario, we found that the trend of realized F-values during non-assessment years depended on the realized F in the starting year within each assessment period. If the realized F was greater than the target F at the start, then the realized F continued to increase and result in ever-increasing discrepancies from the target F. Conversely, if the realized F started from a value lower than the target F, then the realized F tended to become even lower in later years. The consequences of these two situations, however, are asymmetrical. Severe overfishing could occur between assessments, leading to cascading increases in realized F as the stock becomes depleted. In contrast, fishing that was more conservative than intended led to more modest increases in stock size and decreases in true F.

Our results indicated that the manner in which harvest targets are set between full assessment years overall was less important than the assessment frequency. The approach was most important in influencing the IAV in yield. The two forward-projection approaches to setting harvest targets between full assessment years resulted in greater IAV in yield than using the same target from the last full assessment. Although the difference in other management performance indicators was modest, the use of the same target among non-assessment years had a lower risk of stock depletion and higher yield than the other two approaches when productivity was not high and when the mortality target was not low, especially for quinquennial assessments. The failure of the two forward-projection rules in those cases was largely due to the assessment error. The assessment error would be amplified in the forward-projection years, given that the target harvest in those years was always estimated based on the biased stock estimate. Regardless of the IAV in yield, calculating harvest targets during the years between assessments based on updating TACs to match observed harvest from previous lag years (AH control rule) generally was among the better-performing rules, especially when the target mortality rate was low. This may be attributable to the annual adjustment and early detecting mechanism for the AH rule, in which overfishing could be detected earlier than the other rules.

The basic premise of our study was that (1) annual harvest targets would be set on the basis of a fitted stock assessment model and (2) when the model was not fitted every year, target harvests between assessment years either would remain unchanged or would be based on model projections since the last stock assessment. Alternatively, one could view the periodic assessments as providing a calibration between some empiric quantity (such as CPUE) and an appropriate harvest level (Cox and Kronlund 2008; Holland 2010). One version of this simply might be to treat periodic assessments as a way of updating catchability estimates and thereby translate CPUE to abundance. Alternatively, one could revise our view of assessments to be primarily about gaining an understanding of how population dynamics work, with the intent of evaluating alternative strategies (McDonald et al. 1997; Punt 2008). In our study, the only use of data we considered for the period between assessments was in making projections that were consistent with observed yield. The higher variability we saw in this case likely reflects, in part, the fact that this procedure—unlike fitting a dynamic model—does not smooth out any of the measurement error. Empiric procedures also can experience higher variation for the same reason, but survey or fishery CPUEs might be more informative and potentially could extend the period between assessments if they are used as indicators either to adjust model-based target harvest or as a signal to do a new assessment earlier than planned. The ICES (2012) report also discussed the value of using key indicators as a guide to management between times of full assessments, and we believe that there is still much to be learned about the optimal approach to applying assessment efforts. We only considered harvest strategies involving a constant mortality rate, given the Lake Whitefish managers' strong preference for this type of strategy; however, we suspect that a state-dependent mortality rate, such as one that declines when biomass is low (Wilberg et al. 2008), could be more sensitive to changes in assessment frequency. We recommend additional studies to evaluate the influence of assessment frequency for fisheries that use such management strategies. Finally, we urge caution in attempting to extrapolate results beyond the assessment frequencies considered in this research, as it is conceivable that the robustness of the assessment models could quickly deteriorate with longer periods between assessments.

ACKNOWLEDGMENTS

Funding for this project was provided by the Quantitative Fisheries Center at Michigan State University and the Michigan Department of Natural Resources through the Partnership for Ecosystem Research and Management program and by grants from the Great Lakes Fishery Commission's Fishery Research Program and the Great Lakes Fishery Trust (Project 2012-1250). We acknowledge the support of the High-Performance Computing Center and the Institute for Cyber-Enabled Research at Michigan State University. We thank the Modeling Subcommittee of the Technical Fisheries Committee, Lansing, Michigan, for providing input on our research. This is Publication Number 2016-15 of the Quantitative Fisheries Center at Michigan State University.

REFERENCES

Brown, C. J., E. A. Fulton, H. P. Possingham, and A. J. Richardson. 2012. How long can fisheries management delay action in response to ecosystem and climate change? Ecological Applications 22:298–310.

Chen, Y., L. Q. Chen, and K. I. Stergiou. 2003. Impacts of data quantity on fisheries stock assessment. Aquatic Sciences 65:92–98.

Cox, S. P., and A. R. Kronlund. 2008. Practical stakeholder-driven harvest policies for groundfish fisheries in British Columbia. Canadian Journal of Fisheries and Aquatic Sciences 94:224–237.

De Leeuw, J. J., W. Dekker, and A. D. Buijse. 2008. Aiming at a moving target, a slow hand fails! 75 years of fisheries management in Lake IJsselmeer (The Netherlands). Journal of Sea Research 60:21–31.

Deroba, J. J., and J. R. Bence. 2012. Evaluating harvest control rules for Lake Whitefish in the Great Lakes: accounting for variable life-history traits. Fisheries Research 121–122:88–103.

Ebener, M. P., J. R. Bence, K. R. Newman, and P. J. Schneeberger. 2005. Application of statistical catch-at-age models to assess Lake Whitefish stocks in the 1836 Treaty ceded waters of the upper Great Lakes. Great Lakes Fishery Commission Technical Report 66:271–309.

Fournier, D. A., and C. P. Archibald. 1982. A general theory for analyzing catch at age data. Canadian Journal of Fisheries and Aquatic Sciences 39:1195–1207.

Fournier, D. A., H. J. Skaug, J. Ancheta, J. Ianelli, A. Magnusson, M. N. Maunder, A. Nielsen, and J. Sibert. 2012. AD Model Builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. Optimization Methods and Software 27:233–249.

Healey, M. C. 1975a. The dynamics of exploited Lake Trout populations. Journal of Wildlife Management 42:307–328.

Healey, M. C. 1975b. Dynamics of exploited Whitefish populations and their management with special reference to the Northwest Territories. Journal of the Fisheries Research Board of Canada 32:427–448.

Healey, M. C. 1978. Fecundity changes in exploited populations of Lake Whitefish (*Coregonus clupeaformis*) and Lake Trout (*Salvelinus namaycush*). Canadian Journal of Fisheries and Aquatic Sciences 35:945–950.

Henderson, B. A., J. J. Collins, and J. A. Reckahn. 1983. Dynamics of an exploited population of Lake Whitefish (*Coregonus clupeaformis*) in Lake Huron. Canadian Journal of Fisheries and Aquatic Sciences 40:1556–1567.

Holland, D. S. 2010. Management strategy evaluation and management procedures: tools for rebuilding and sustaining fisheries. Organization for Economic Cooperation and Development, Food, Agriculture and Fisheries, Working Paper 25, Paris.

ICES (International Council for the Exploration of the Sea). 2012. Report of the workshop on frequency of assessments (WKFREQ). ICES, C.M. 2012/ACOM:34, Copenhagen.

Kell, L. T., C. M. O'Brien, M. T. Smith, T. K. Stokes, and B. D. Rackham. 1999. An evaluation of management procedures for implementing a precautionary approach in the ICES context for North Sea Plaice (*Pleuronectes platessa* L.). ICES Journal of Marine Science 56:834–845.

Kell, L. T., G. M. Pilling, G. P. Kirkwood, M. Pastoors, B. Mesnil, K. Korsbrekke, and P. Abaunza. 2005. An evaluation of the implicit management procedure used for some ICES roundfish stocks. ICES Journal of Marine Science 62:750–759.

Li, Y., J. R. Bence, and T. O. Brenden. 2015. An evaluation of alternative assessment approaches for intermixing fish populations: a case study with Great Lakes Lake Whitefish. ICES Journal of Marine Science 72:70–81.

Mace P. M., N. W. Bartoo, A. B. Hollowed, P. Kleiber, R. D. Methot, S. A. Murawski, J. E. Powers, and G. P. Scott. 2001. Marine fisheries stock assessment improvement plan. NOAA Technical Memorandum NMFS-F/SPO-56.

Mangel, M., J. Brodziak, and G. DiNardo. 2010. Reproductive ecology and scientific inference of steepness: a fundamental metric of population dynamics and strategic fisheries management. Fish and Fisheries 11:89–104.

Marchal, P. 1997. Managing growth over-fishing with multi-annual compromise strategies. Canadian Journal of Fisheries and Aquatic Sciences 54:2255–2276.

Marchal, P., and J. W. Horwood. 1995. Multi-annual TACs and minimum biological levels. ICES Journal of Marine Science 52:797–807.

Maunder, M. N., and K. R. Piner. 2015. Contemporary fisheries stock assessment: many issues still remain. ICES Journal of Marine Science 72:7–18.

McDonald, A. D., A. D. M. Smith, A. E. Punt, G. N. Tuck, and A. J. Davidson. 1997. Empirical evaluation of expected returns from research on stock structure for determination of total allowable catch. Natural Resource Modeling 10:3–29.

Methot, R. D. Jr. 2015. Prioritizing fish stock assessments. NOAA Technical Memorandum NMFS-F/SPO-152.

Molton, K. J., T. O. Brenden, and J. R. Bence. 2012. Control rule performance for intermixing Lake Whitefish populations on the 1836 Treaty waters of the Great Lakes: a simulation-based evaluation. Journal of Great Lakes Research 38:686–698.

Molton, K., T. O. Brenden, and J. R. Bence. 2013. Harvest levels that conserve spawning biomass can provide larger and more stable and sustainable yields in intermixed fisheries. Fisheries Research 147:264–283.

MSC (Modeling Subcommittee of the Technical Fisheries Committee). 2015. Technical Fisheries Committee administrative report 2015: status of Lake Trout and Lake Whitefish populations in the 1836 Treaty-ceded waters of Lakes Superior, Huron, and Michigan, with recommended yield and effort levels for 2015. Michigan Department of Natural Resources, Technical Fisheries Committee, Lansing.

NAFO (Northwest Atlantic Fisheries Organization). 2007. Scientific Council reports 2006. NAFO, Dartmouth, Nova Scotia.

Punt, A. E. 2008. Refocusing stock assessment in support of policy evaluation. Pages 139–152 *in* K. Tsukamoto, T. Kawamura, T. Takeuchi, T. D. Beard, and M. J. Kaiser, editors. Fisheries for global welfare and environment, 5th World Fisheries Congress. Terrapub, Tokyo.

Quinn, T. J. II, and R. B. Deriso. 1999. Quantitative fish dynamics. Oxford University Press, New York.

Shertzer, K. W., and M. H. Prager. 2007. Delay in fishery management: diminished yield, longer rebuilding, and increased probability of stock collapse. ICES Journal of Marine Science 64:149–159.

Thomas, M. V., and R. C. Haas. 2005. Status of Yellow Perch and Walleye in Michigan waters of Lake Erie, 1999–2003. Michigan Department of Natural Resources, Fisheries Research Report 2082, Ann Arbor.

Truesdell, S. B., and J. R. Bence. 2016. A review of stock assessment methods for Lake Trout and Lake Whitefish in 1836 Treaty waters of Lake Huron, Lake Michigan, and Lake Superior. Michigan State University, Quantitative Fisheries Center Report T2016-01, East Lansing.

Wilberg, M. J., B. J. Irwin, M. L. Jones, and J. R. Bence. 2008. Effects of source–sink dynamics on harvest policy performance for Yellow Perch in southern Lake Michigan. Fisheries Research 94:282–289.

Woldt, A. P., J. R. Bence, and M. P. Ebener. 2006. Executive summary. Pages 4–9 *in* A. P. Woldt, S. P. Sitar, J. R. Bence, and M. P. Ebener, editors. Technical Fisheries Committee administrative report 2005: status of Lake Trout and Lake Whitefish populations in the 1836 Treaty-ceded waters of Lakes Superior, Huron, and Michigan, with recommended yield and effort levels for 2005. Michigan Department of Natural Resources, Lansing.

Appendix: Beverton–Holt Stock–Recruit Function

TABLE A.1. Parameterization of the Beverton–Holt (BH) stock–recruit function (symbols are defined in Table 2).

Model name	Model equation	Equation number
BH steepness	$h = \frac{R_{0.2S_0}}{R_{S_0}} = \frac{\frac{0.2S_0}{\alpha+\beta 0.2S_0}}{\frac{S_0}{\alpha+\beta S_0}} = \frac{\alpha+\beta S_0}{5\alpha+\beta S_0}$, so $\alpha = \frac{(1-h)\beta S_0}{5h-1}$	A.1.1
Unfished abundance at age during the spawning season near end of the year	$Nspawn_a = N_{a+1} = N_a e^{-(M)}$ $N_{a=1} = R_{S_0}$	A.1.2
Unfished spawning stock	$S_0 = \sum_a Nspawn_a \cdot Fem \cdot m_a \cdot W_a$, where Fem = 0.5 (from Li et al. 2015)	A.1.3
Unfished stock per recruit	$\mathrm{SPR}_0 = \frac{S_0}{R_{S_0}} = \frac{S_0}{\left(\frac{S_0}{\alpha+\beta S_0}\right)} = \alpha + \beta S_0$, so $\alpha = \mathrm{SPR}_0 - \beta S_0$ On the other hand, $\mathrm{SPR}_0 = \frac{S_0}{R_{S_0}} = \sum_a Nspawn_a \cdot Fem \cdot m_a \cdot W_a$ when R_{S_0}= 1	A.1.4
Solving for the BH parameter β based on equations A.1.1 and A.1.4	$\alpha = \frac{(1-h)\beta S_0}{5h-1} = \mathrm{SPR}_0 - \beta S_0$, so $\beta = \frac{(5h-1)\mathrm{SPR}_0}{4hS_0}$ $= \frac{(5h-1)}{4hS_0} \sum_a Nspawn_a \cdot Fem \cdot m_a \cdot W_a$ when R_{S_0}= 1	A.1.5
Solving for the BH parameter α based on equations A.1.4 and A.1.5	$\alpha = \mathrm{SPR}_0 - \beta\ S_0 = \mathrm{SPR}_0 - \frac{(5h-1)\mathrm{SPR}_0}{4hS_0} S_0$ $= \frac{(1-h)\mathrm{SPR}_0}{4h} = \frac{(1-h)}{4h} \sum_a Nspawn_a \cdot Fem \cdot m_a \cdot W_a$ when R_{S_0}= 1	A.1.6