

**Appendix II.**

**Groundfish Assessment Review Meeting**

**August 15–19, 2005**

**Accuracy and Precision Exercises Associated with 2005 GARM Production Ageing**

GARM Draft Section

- 1.0 Introduction
- 2.0 Methods
- 3.0 Results and Discussion
- 4.0 GARM Discussion
- 5.0 References

Tables

Figures

## *Accuracy and precision exercises associated with 2005 GARM production ageing*

S. Sutherland, N. Munroe, N. Shepherd, V. Silva, S. Pregracke, and J. Burnett  
Fishery Biology Program  
Northeast Fisheries Science Center (NEFSC)  
Woods Hole, MA USA

### *1. 0. Introduction*

In production ageing programs, age reader accuracy can be thought of as how often the “right” age is obtained, and precision as how often the “same” age is obtained (Campana 2001). Both measures are important components of a quality control monitoring program. For the 2005 Groundfish Assessment Review Meeting (GARM), exercises were undertaken to estimate the accuracy and/or precision of production ageing by the Fishery Biology Program for cod *Gadus morhua*, haddock *Melanogrammus aeglefinus*, yellowtail flounder *Limanda ferruginea*, witch flounder *Glyptocephalus cynoglossus*, American plaice *Hippoglossoides platessoides*, winter flounder *Pseudopleuronectes americanus*, and Acadian redfish *Sebastes fasciatus*.

### *2. 0. Methods*

For all species, subsamples were selected to be re-aged to test age-reader accuracy or precision. Ageing accuracy is only presented for species that have reference collections already established, i.e. the Georges Bank stocks of cod and haddock. Precision data is presented for all species *versus* samples previously aged by the same reader. When re-aging fish, the age reader had knowledge of the same data as during production ageing, i.e. fish length, date captured, and area captured. Except in the case of cod, where two people aged the species, all exercises combined stock areas for each species.

All exercises were ‘one-shot’ deals, and no attempts were made to improve results by repeated readings. There was also no attempt to revise the original production ages in cases where differences occurred. Results are presented in terms of percentage agreement, total coefficient of variation (CV), age bias plots, and age agreement matrices (Campana *et al.* 1995, Campana 2001).

For Georges Bank cod, production ageing this year reverted to the previous age reader, who aged cod during the period 1984–2003. Following production ageing, age-reader accuracy was determined from a random subsample drawn from the NEFSC cod otolith reference collection. No precision estimates were attempted for this stock, due to time constraints.

For the Gulf of Maine cod stock, the current cod age reader completed precision exercises on two occasions. These were subsampled from U.S. commercial landings from the fourth quarter of 2003, and during all of 2004. A comparison between the two readers was also undertaken, with NEFSC 2004 autumn bottom trawl survey samples from the Gulf of Maine.

For haddock, age-reader precision was estimated on six occasions from second readings of random subsamples from each cruise (NEFSC 2004 autumn and 2005 spring bottom trawl surveys) and each quarter from U.S. commercial landings (2004). Following the completion of production ageing, age-reader accuracy was assessed by reading a random subsample from the NEFSC Georges Bank haddock otolith reference collection.

For yellowtail flounder, age-reader precision was estimated three times from second readings of random subsamples from 2004 Canadian landings, the Canadian 2005 bottom trawl survey, and a combination of recent U.S. samples (2004 autumn and 2005 spring bottom trawl surveys, plus 2004 commercial landings). These latter samples were also aged by a trainee who will soon assume yellowtail age-reader duties, following the recent retirement of the former age reader. The trainee worked with the former age reader during production ageing, and re-aged the same set of fish as the former reader.

For witch flounder, age-reader precision was estimated once from a combination of fish from both the NEFSC 2005 spring bottom trawl survey and Quarters 2 and 4 of 2004 U.S. commercial landings. Quarter 2 included fish from the large market category, while Quarter 4 was composed of both small and medium fish.

For American plaice, age-reader precision was estimated once from a combination of fish from both Quarter 1 of 2004 U.S. commercial landings and the NEFSC 2004 autumn bottom trawl survey.

For winter flounder, age-reader precision was estimated twice. One exercise used otoliths from NEFSC bottom trawl surveys, with equal numbers of Gulf of Maine fish from the 2004 autumn survey and Southern New England fish from the 2005 spring survey. The second exercise used scales from 2004 U.S. commercial landings, with Quarters 1 and 3 combined. In the commercial samples, Quarter 1 included Southern New England fish in both small and large market categories, and medium-sized Gulf of Maine fish. Quarter 3 was reversed in terms of stock areas and market categories.

For Acadian redfish, age-reader precision was estimated once from second readings of random subsamples from the NEFSC 2004 autumn bottom trawl survey.

### *3. 0. Results and Discussion*

The total sample sizes associated with the accuracy and precision exercises were as follows: 106 (Georges Bank cod), 217 (Gulf of Maine cod), 500 (haddock), 367 (yellowtail), 122 (witch flounder), 161 (American plaice), 225 (winter flounder), and 142 (redfish). Results for cod are presented in Figures 1–4, haddock in Figures 5–11, yellowtail flounder in Figures 12–15, witch flounder in Figure 16, American plaice in Figure 17, winter flounder in Figures 18 and 19, and redfish in Figure 20. All results are summarized in Table 1.

The accuracy estimate for Georges Bank cod was high (91% agreement) and the total CV (1.5%) was low. There was a slight tendency toward overageing by one year in the test readings (Fig. 1), but no ages differed by more than one year. Even so, accuracy was virtually the same as that obtained last year (91% agreement and 1.9% CV, Sutherland *et*

*al.* 2004, unpubl.), suggesting that the switch from the current to previous age reader was not problematic.

For the two Gulf of Maine cod exercises, precision levels of 86% and 94% agreement (total CVs of 2.7% and 0.8%, respectively) were attained (Figs. 2 and 3). While these values would seem to suggest an adequate level of consistency in age determinations, the age bias plots indicated that, for the first exercise (2003 Quarter 4 commercial samples; Fig. 2), the mean test age for Age 3 fish was significantly biased from the mean production age, necessitating remedial intervention. A comparison of ages performed by the two cod age readers resulted in 100% agreement (Fig. 4), indicating that the two age readers are consistent with each other in their age determinations.

For haddock, precision levels ranged between 91 and 98% agreement, with total CVs of 0.2–0.9%, between first and second readings (Figs. 5–10), indicating a high level of consistency in age determinations. No disagreement between readings was more than one year. This year's results showed an increase in precision from last year (median of 86% agreement and 2.0% CV, Sutherland *et al.* 2004, unpubl.). The relatively high accuracy estimate (94% agreement, 1.3% CV, Fig. 11) for samples from the Georges Bank reference collection, coupled with consistently high precision results, supports the conclusion that the haddock age reader, having just completed their second year of production ageing, has attained a reliable level of ageing capability.

For yellowtail flounder, precision levels were consistent between samples from the Canadian 2004 commercial landings and the Canadian 2005 spring survey (86 and 92% agreement and total CVs of 2.5 and 1.8%, respectively, Figs. 12–13). In the commercial samples, there was a slight tendency toward overageing in the second readings. The values obtained for U.S. samples, however, were less precise (71% agreement and 6.6% CV, Fig. 14), and revealed a bias towards underageing of older fish (age  $\geq 4$  years) in the second readings. Even so, no ages differed by more than one year. When the future age reader re-aged these same U. S. samples, similar precision levels were attained (73% agreement and 6.1% CV, Fig. 15), but no bias was apparent.

Observations of poor scale condition in yellowtail flounder from eastern Georges Bank, which began in 2002, have continued in these samples. The scales were characterized by actual holes and moderate to severe erosion of the anterior scale edges (illustrated in Sutherland *et al.* 2004, unpubl.). Causes for this condition remain unknown, but this may help to explain the reduced precision observed with yellowtail samples.

For witch flounder, the precision level was 80% agreement, with a total CV of 1.6%, between first and second readings (Fig. 16). This indicates a moderate level of consistency in age determinations for this long-lived species.

For American plaice, a precision level of 86% agreement (total CV of 1.7%) was attained between first and second readings (Fig. 17), indicating a moderate level of consistency in age determinations.

For the two winter flounder exercises, precision levels of 94% and 79% agreement (total CVs of 1.6% and 2.8%, respectively) were attained (Figs. 18 and 19). Much greater precision was obtained with otoliths from the survey samples than with the scales routinely collected from commercial landings. Neither exercise revealed a bias, although

there may have been an error in distinguishing ages 3 and 4 in the commercial production ages. This may be related to the lack of availability of sex data for commercial samples. Female winter flounder exhibit a strong check on their scales associated with the onset of maturation (about age 3), which cannot be distinguished from an annulus without data on fish sex.

For Acadian redfish, the precision level was 89% agreement, with a total CV of 1.0%, between first and second readings (Fig. 20), indicating a moderate level of consistency in age determinations for this long-lived species.

Acceptable levels of age determination accuracy and precision are highly influenced by species, age structure, and age reader experience. Even so, various ageing labs consider a total CV of under 5% to be acceptable for species of moderate longevity and ageing complexity (Campana 2001). Therefore, precision of recent age determinations appears to have been generally reliable for the GARM assessments. Completion of reference collections for additional species and continued training of new age readers are top priorities for the Fishery Biology Program in the coming year.

#### *4.0 GARM Discussion*

The GARM Panel suggested that tests of symmetry (Hoenig et al. 1995) may be a more appropriate method with which to evaluate age reader precision. For precision exercises presented above with age agreement less than 90%, Bowker's test of symmetry (Bowker 1948) was performed. Results are presented in Table 2. Only the exercise for 2003 Quarter 4 Gulf of Maine cod revealed a systematic difference between the two readings. Several exercises flagged as problematic from age bias plots or high CVs were not significantly asymmetrical. It appears that, for some data sets, the power of tests of symmetry may be low and sensitive to the degrees of freedom available in the analysis. However, the potential utility of the test as an additional diagnostic for age reader precision was accepted and will be routinely incorporated into the suite of precision evaluations conducted by the Fishery Biology Program.

#### *5. 0. References*

- Bowker, A.H. 1948. A test for symmetry in contingency tables. *Journal of the American Statistical Association* 43: 572-574.
- Campana, S.E., M.C. Annand, and J.I. McMillan. 1995. Graphical and statistical methods for determining the consistency of age determinations. *Transactions of the American Fisheries Society* 124: 131-138.
- Campana, S.E. 2001. Accuracy, precision, and quality control in age determination, including a review of the use and abuse of age validation methods. *Journal of Fish Biology* 59: 197-242.
- Hoenig, J.M., M.J. Morgan, and C.A. Brown. 1995. Analysing differences between two age determination methods by tests of symmetry. *Canadian Journal of Fisheries and Aquatic Science* 52: 364-368.

Sutherland, S., N. Shepherd, N. Munroe, V. Silva, and J. Burnett. 2004. Precision exercises associated with 2004 TAWG production ageing. Unpublished report.

Table 1. Results of all ageing exercises, with list of associated figures. Maximum age is the highest age found among the production ages within each exercise.

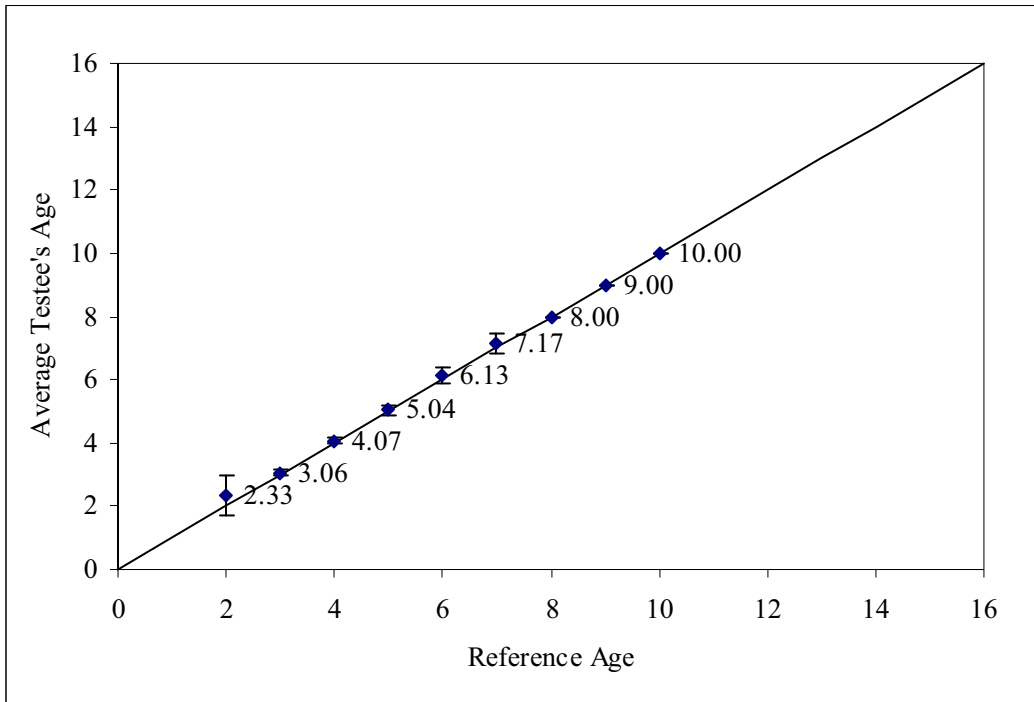
Fig.	Species & Exercise Type	N	Agreement (%)	CV (%)	Max. Age
1	Cod Accuracy (GB)	106	90.6	1.53	10
2	Cod Precision (GOM)	105	85.7	2.68	7
3	Cod Precision (GOM)	112	93.8	0.75	12
4	Cod Comparison (GOM)	60	100.0	0.00	10
5	Haddock Precision	102	95.1	0.71	10
6	Haddock Precision	76	96.1	0.89	9
7	Haddock Precision	56	92.9	0.69	11
8	Haddock Precision	55	90.9	0.94	11
9	Haddock Precision	44	97.7	0.25	11
10	Haddock Precision	60	95.0	0.44	12
11	Haddock Accuracy	107	93.5	1.26	12
12	Yellowtail fldr Precision	167	86.2	2.52	9
13	Yellowtail fldr Precision	100	92.0	1.79	7
14	Yellowtail fldr Precision	100	71.0	6.64	7
15	Yellowtail fldr Precision (trainee)	100	73.0	6.11	7
16	Witch fldr Precision	122	80.3	1.55	23
17	American Plaice Precision	161	85.7	1.69	13
18	Winter fldr Precision (Otoliths)	110	93.6	1.59	9
19	Winter fldr Precision (Scales)	115	79.1	2.81	9
20	Redfish Precision	142	89.4	0.99	19

Total	<b>2000</b>		
Average		<b>89.1</b>	<b>1.79</b>
Median		<b>91.5</b>	<b>1.40</b>

Table 2. Results of Bowker's test of symmetry for all precision exercises with age agreements of less than 90% (**bold** value indicates a systematic difference in the distribution of the two sets of ages).

Species	$\chi^2$	d.f.	P-value
Cod (GOM)	11.50	4	<b>0.02</b>
Yellowtail (2004 Canadian fishery)	16.47	11	0.12
Yellowtail (US survey & fishery)	10.45	6	0.11
Yellowtail (Trainee)	2.20	5	0.82
Witch flounder	19.33	18	0.37
American plaice	8.00	10	0.63
Winter fldr (Scales)	11.20	7	0.13
Redfish	10.33	11	0.50

Figure 1. Results of Georges Bank cod age-reader accuracy exercise against randomly selected samples from the NEFSC cod reference collection. Error bars indicate 95% confidence intervals.



N Aged	106	CV	1.53
N Agreed	96		
Disagreed	10	%Agreement	90.6%

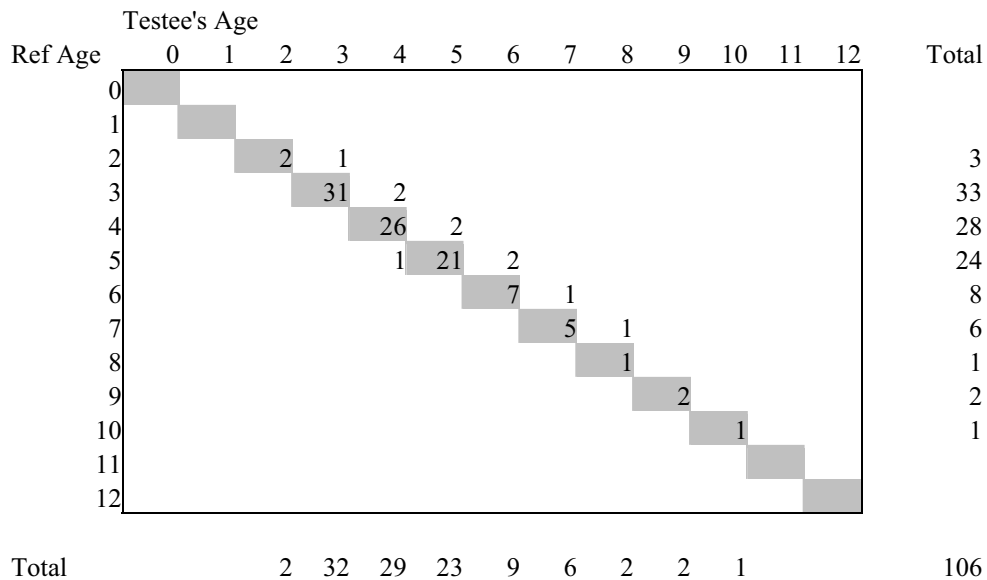
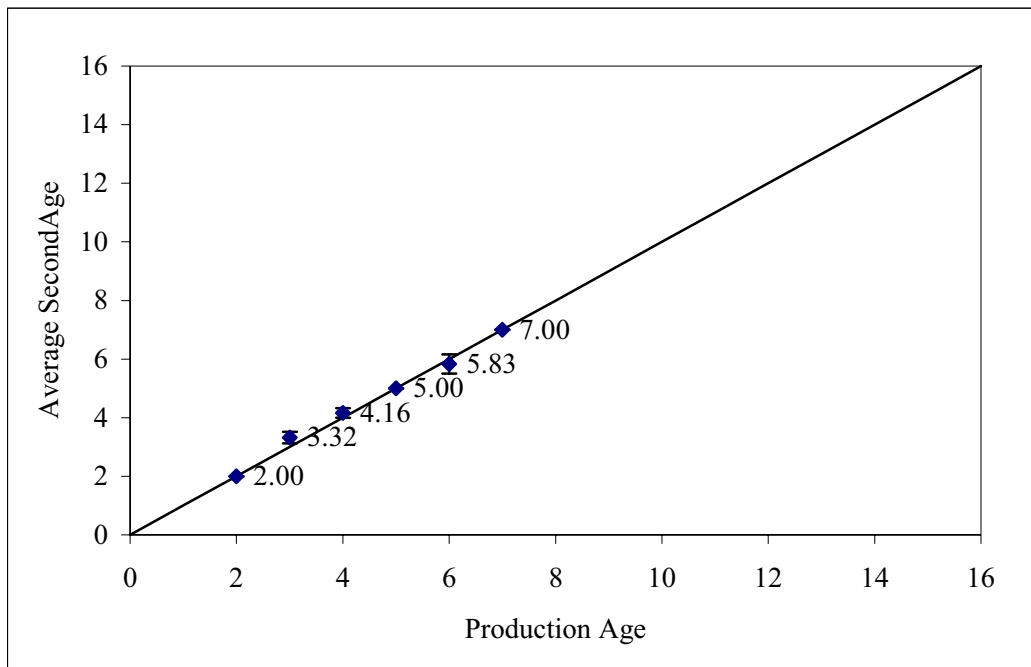




Figure 2. Results of Gulf of Maine cod age-reader precision exercise against randomly selected samples from Quarter 4 of 2003 U.S. commercial landings. Error bars indicate 95% confidence intervals.



N Aged	105	CV	2.68
N Agreed	90	%Agreement	85.7%
Disagreed	15		

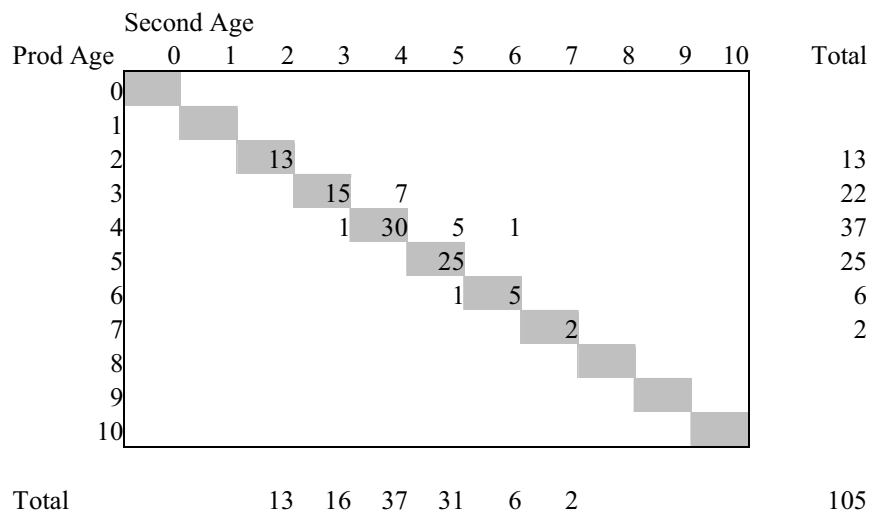
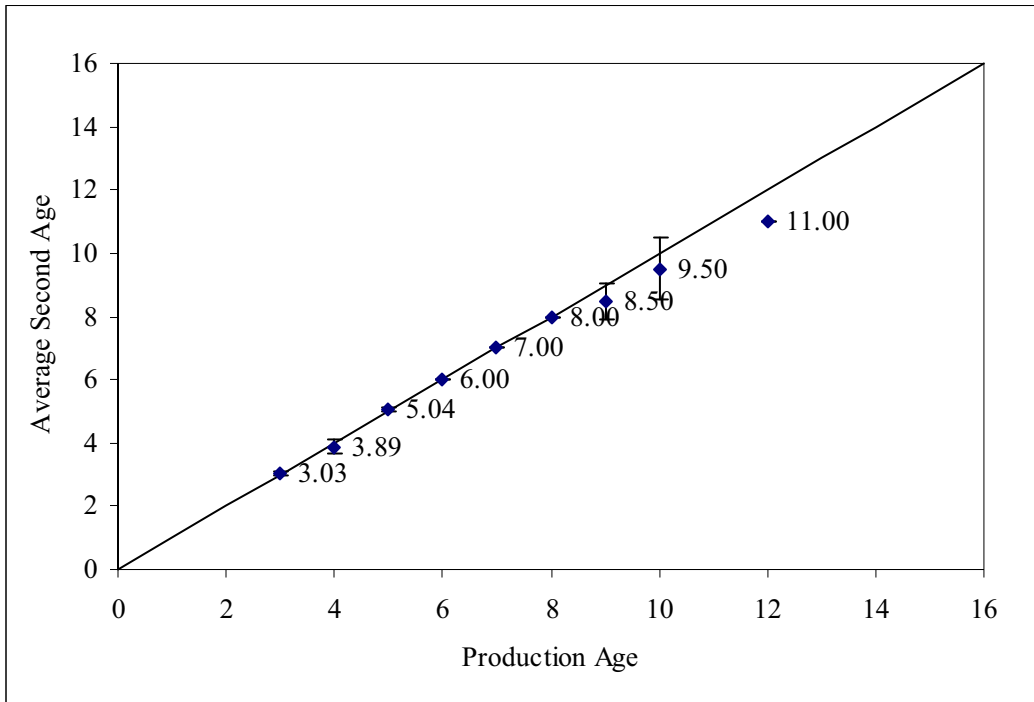


Figure 3. Results of Gulf of Maine cod age-reader precision exercise against randomly selected samples from Quarters 1–4 of 2004 U.S. commercial landings. Error bars indicate 95% confidence intervals.



N Aged	112	CV	0.75
N Agreed	105	%Agreement	93.8%
Disagreed	7		

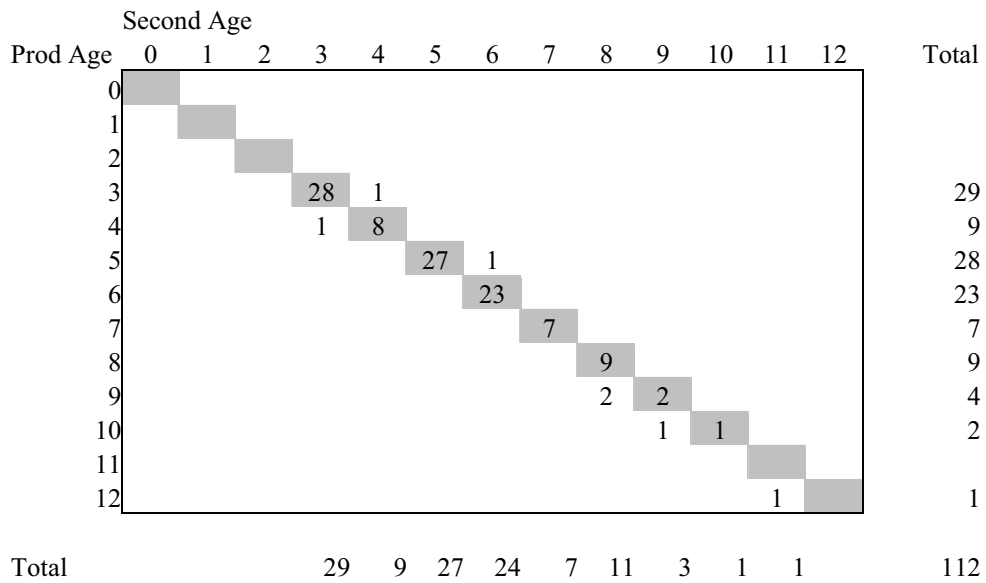
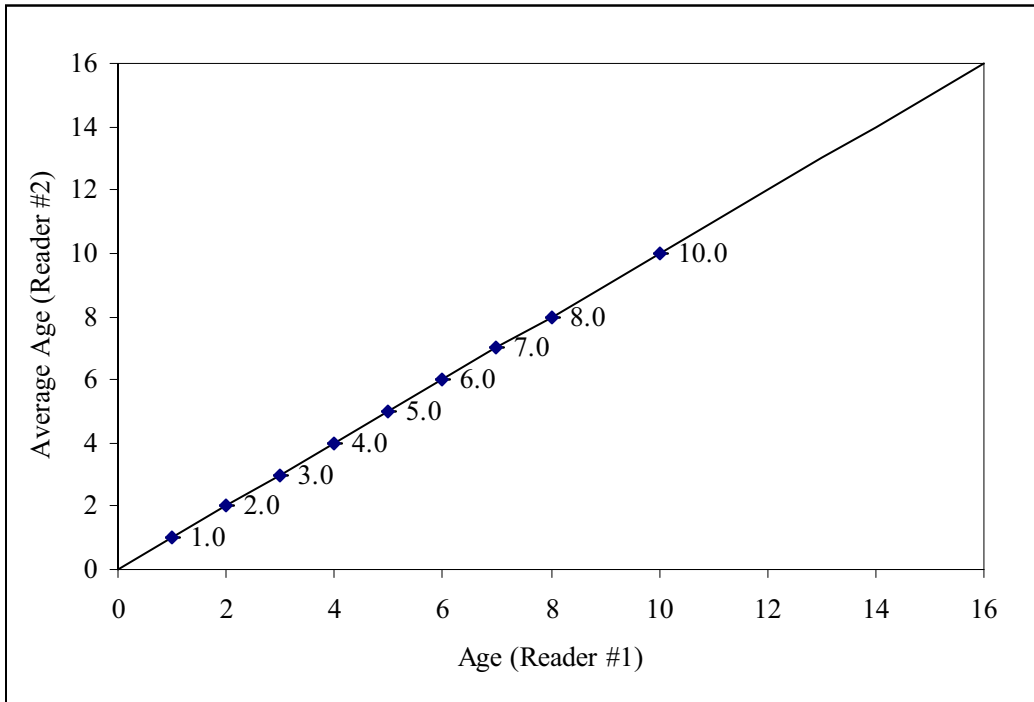


Figure 4. Results of cod age-reader comparison exercise using randomly selected Gulf of Maine samples from the NEFSC 2004 autumn bottom trawl survey. The current age reader is listed here as Reader #1. Error bars indicate 95% confidence intervals.



N Aged	60	CV	0.00
N Agreed	60		
Disagreed	0	%Agreement	100.0%

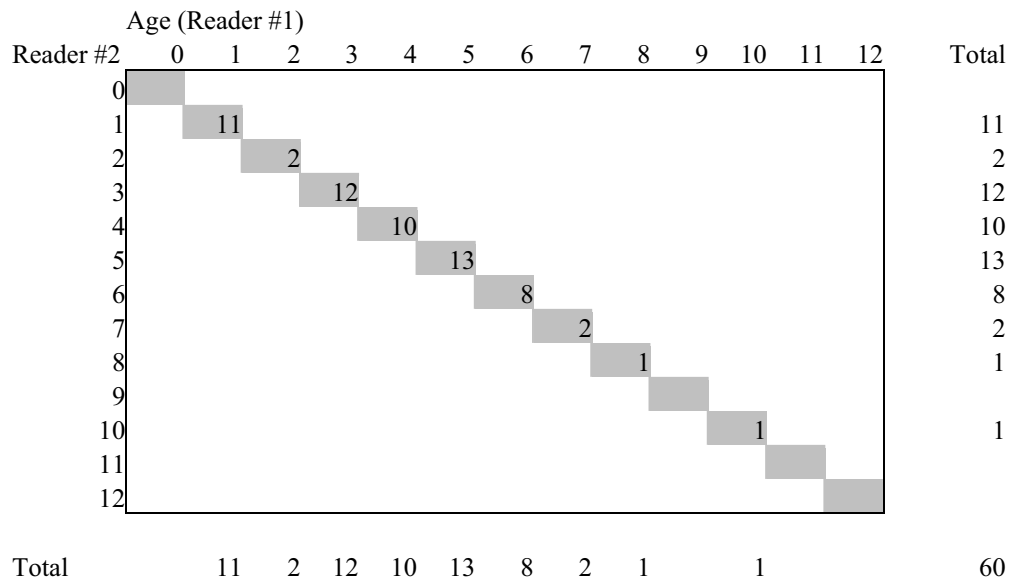
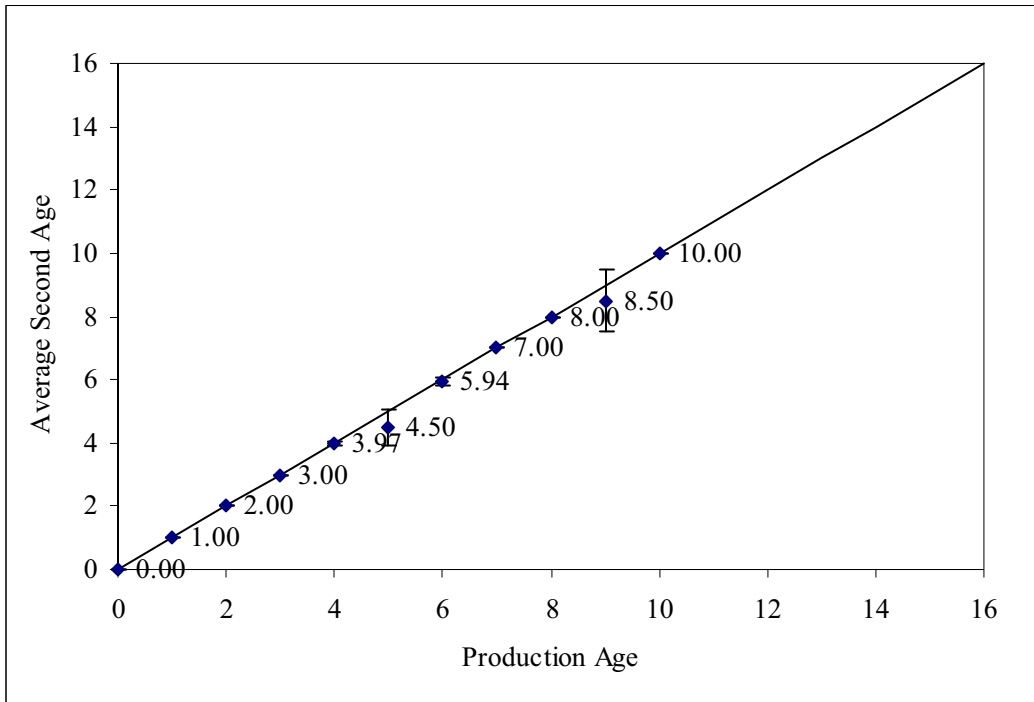


Figure 5. Results of haddock age-reader precision exercise against randomly selected samples from the NEFSC 2004 autumn bottom trawl survey. Error bars indicate 95% confidence intervals.



N Aged	102	CV	0.71
N Agreed	97	%Agreement	95.1%
Disagreed	5		

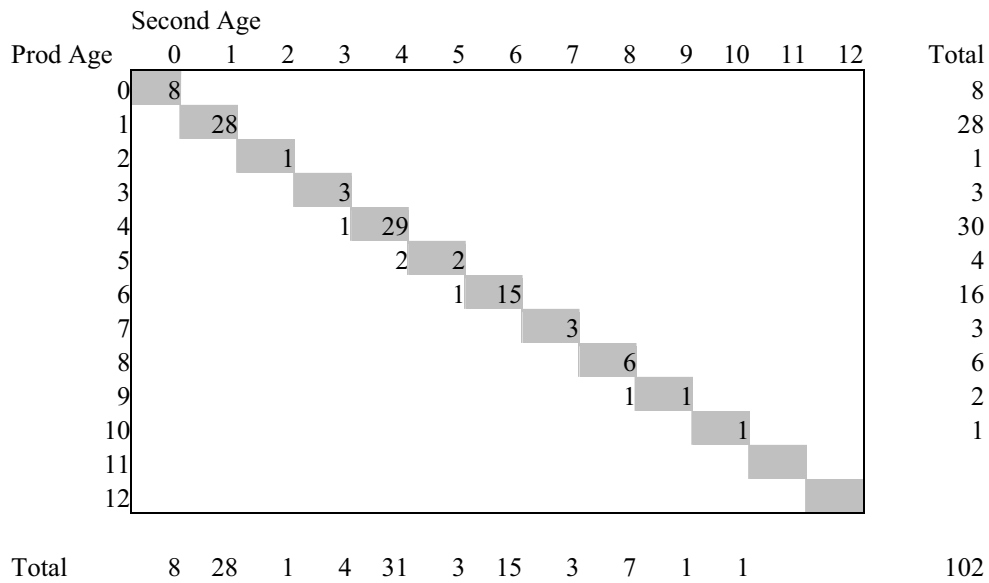
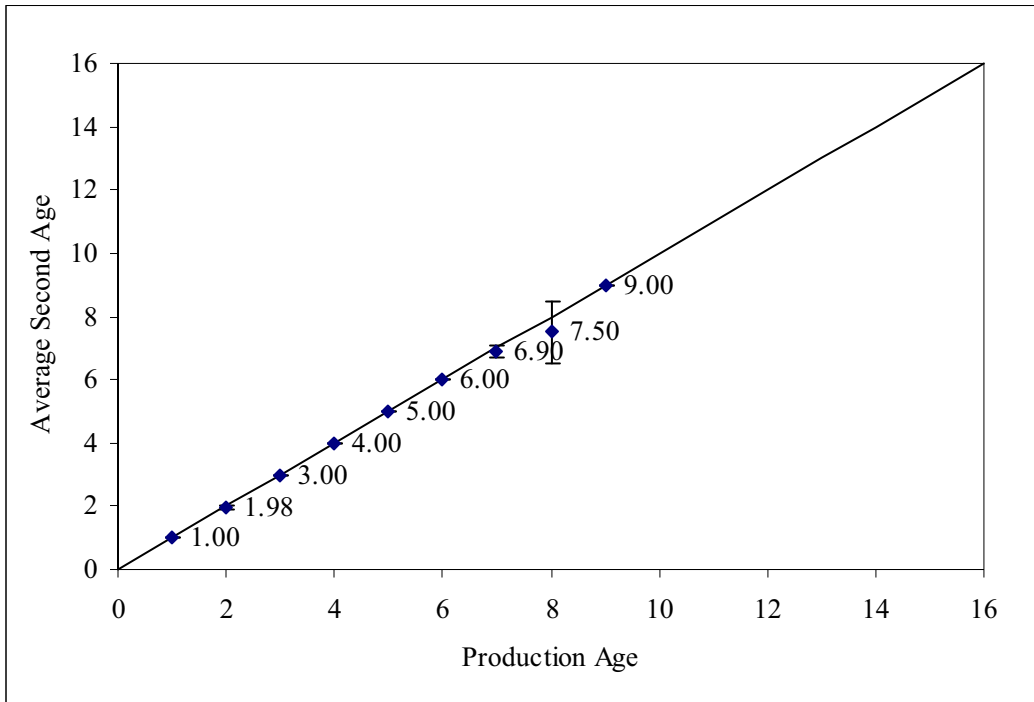


Figure 6. Results of haddock age-reader precision exercise against randomly selected samples from the NEFSC 2005 spring bottom trawl survey. Error bars indicate 95% confidence intervals.



N Aged	76	CV	0.89
N Agreed	73		
Disagreed	3	%Agreement	96.1%

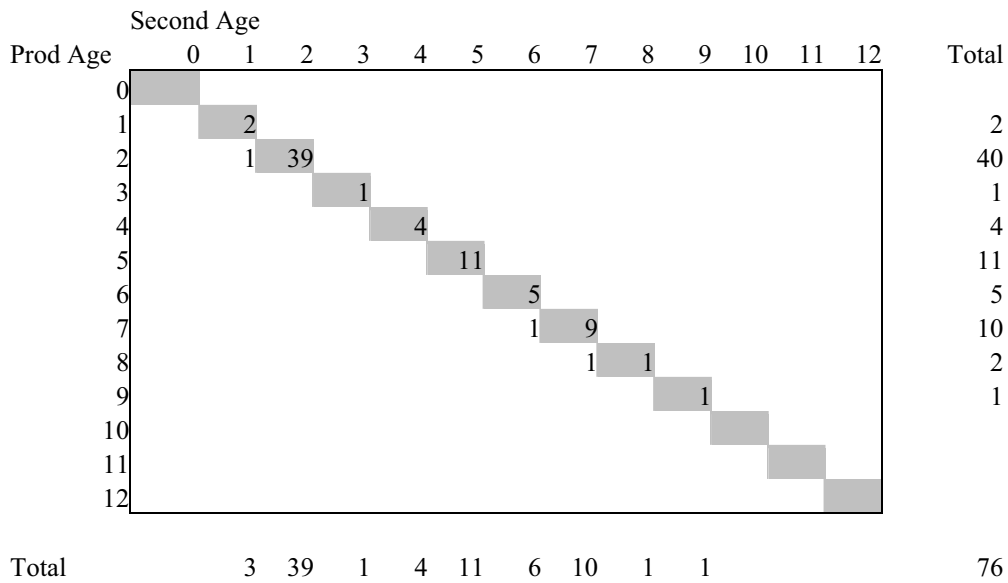
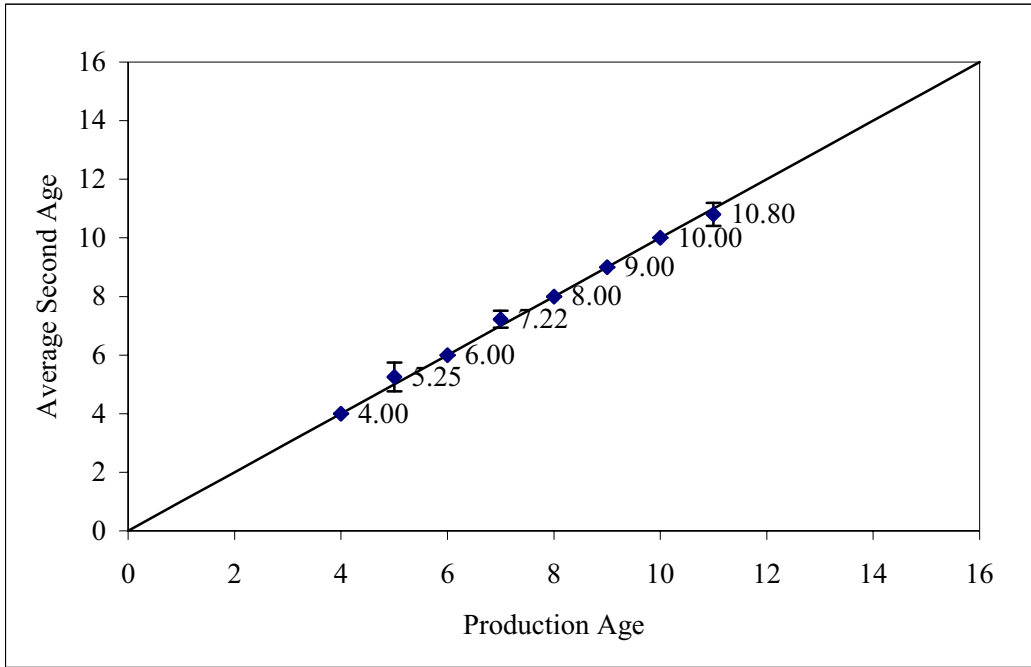


Figure 7. Results of haddock age-reader precision exercise against randomly selected samples from Quarter 1 of 2004 U.S. commercial landings. Error bars indicate 95% confidence intervals.



N Aged	56	CV	0.69
N Agreed	52	%Agreement	92.9%
Disagreed	4		

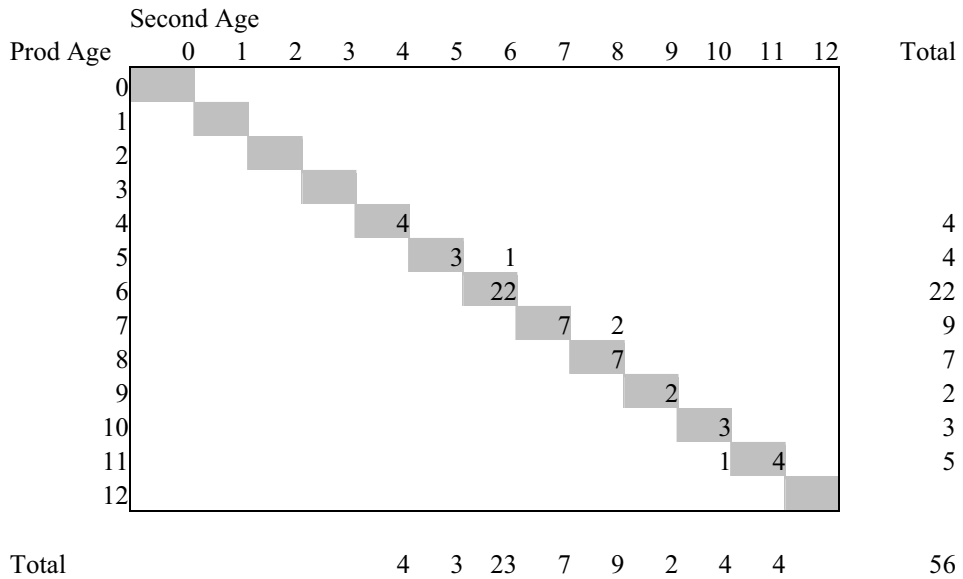
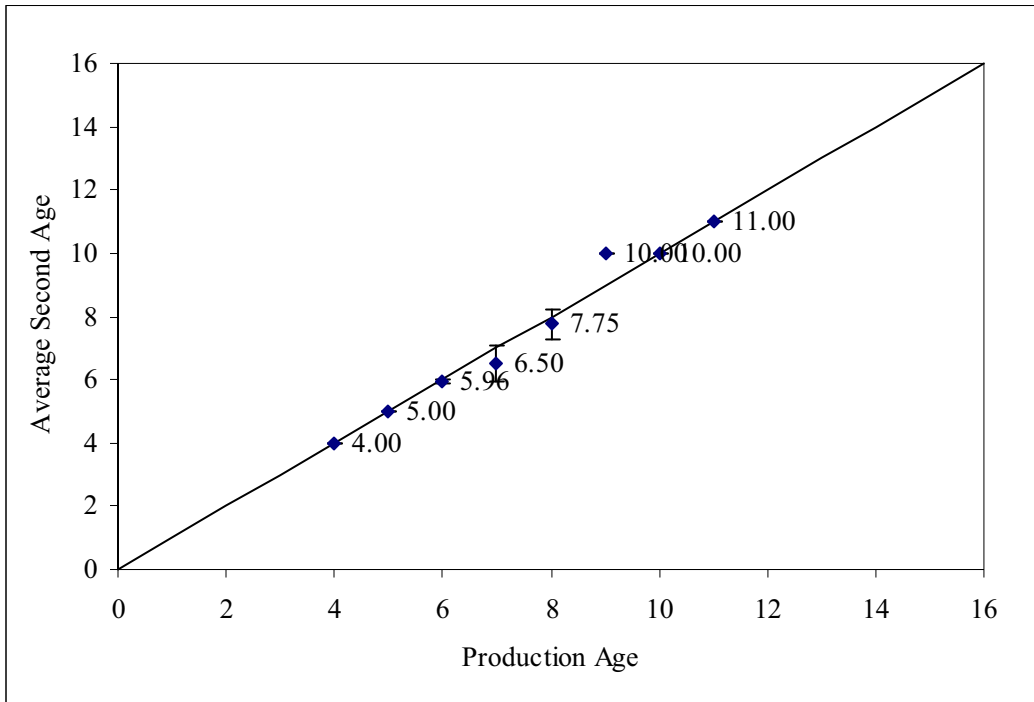


Figure 8. Results of haddock age-reader precision exercise against randomly selected samples from Quarter 2 of 2004 U.S. commercial landings. Error bars indicate 95% confidence intervals.



N Aged	55	CV	0.94
N Agreed	50		
Disagreed	5	%Agreement	90.9%

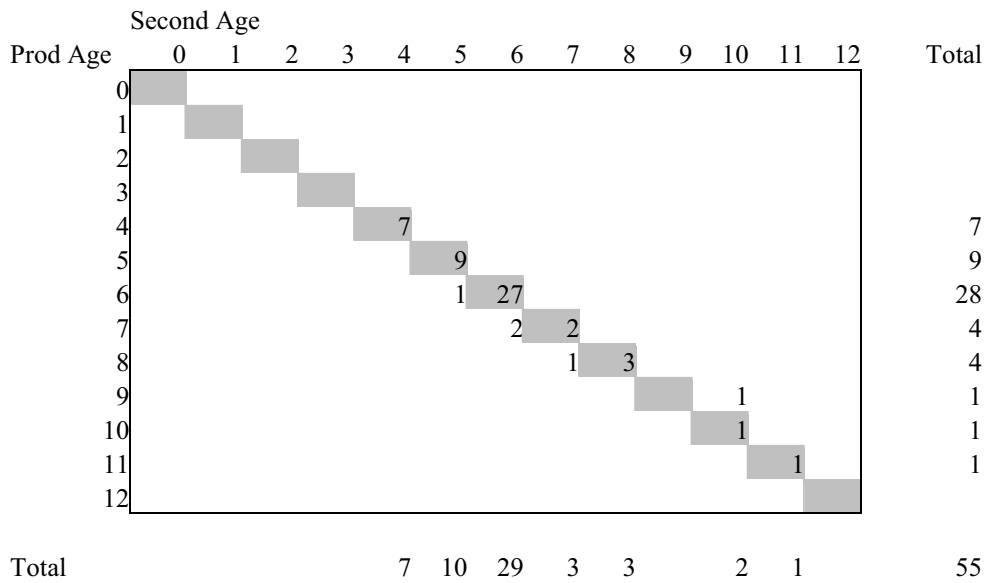
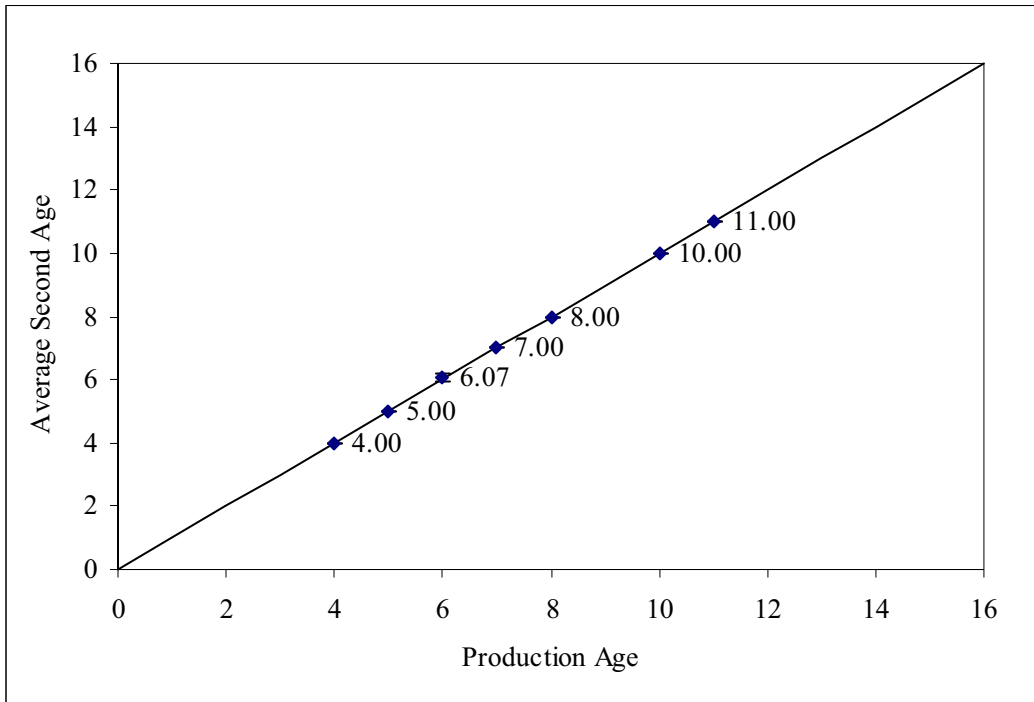


Figure 9. Results of haddock age-reader precision exercise against randomly selected samples from Quarter 3 of 2004 U.S. commercial landings. Error bars indicate 95% confidence intervals.



N Aged	44	CV	0.25
N Agreed	43		
Disagreed	1	%Agreement	97.7%

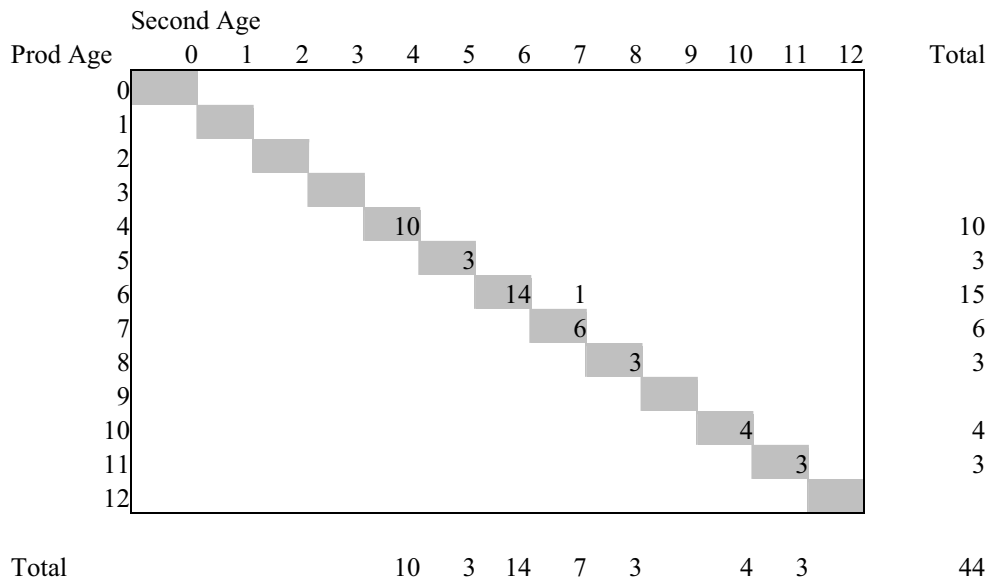
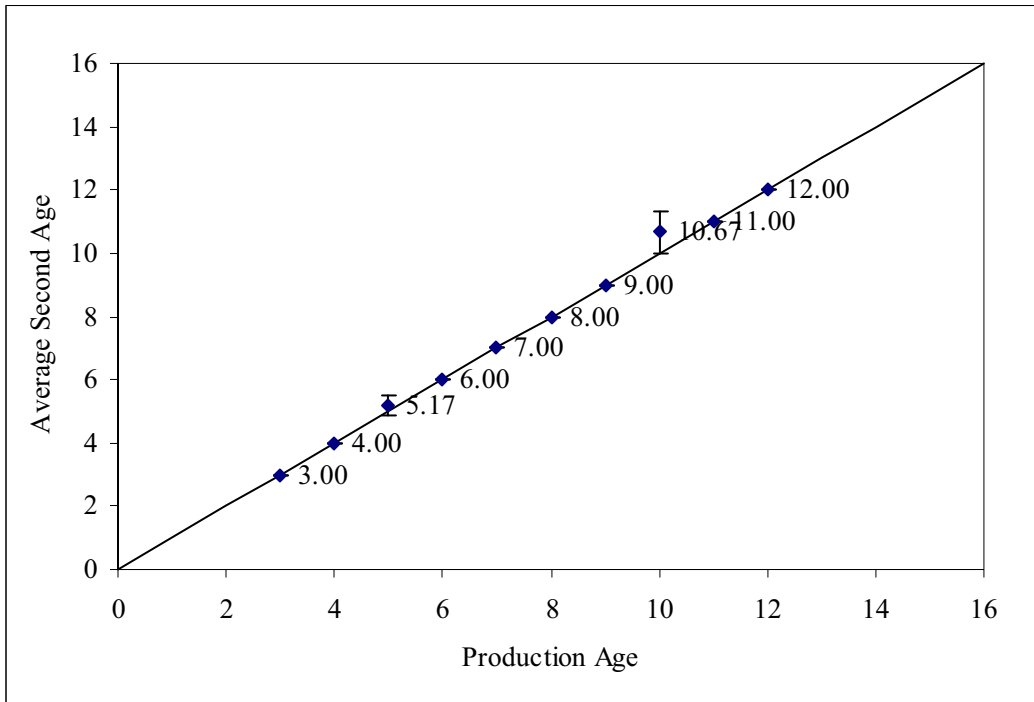




Figure 10. Results of haddock age-reader precision exercise against randomly selected samples from Quarter 4 of 2004 U.S. commercial landings. Error bars indicate 95% confidence intervals.



N Aged	60	CV	0.44
N Agreed	57		
Disagreed	3	%Agreement	95.0%

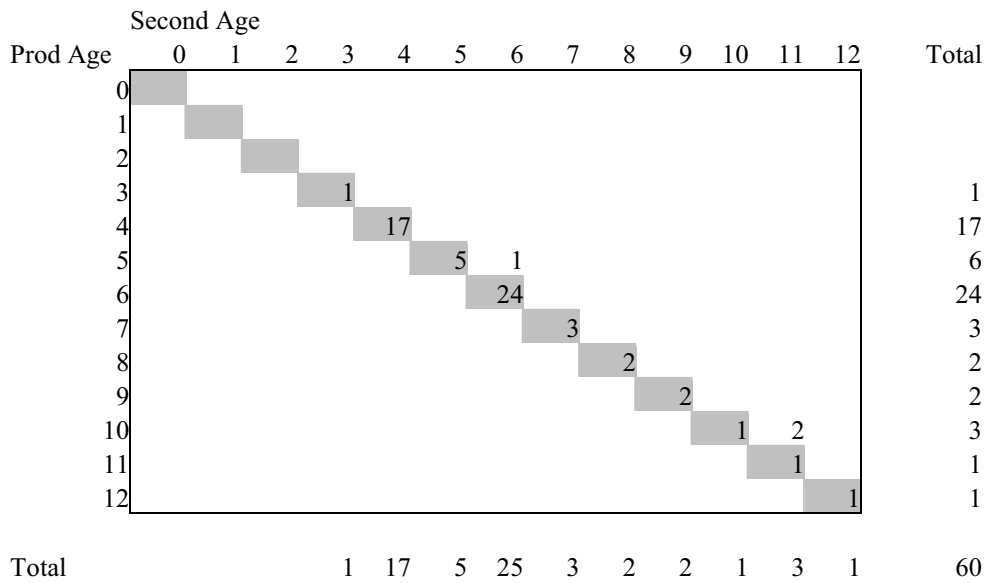
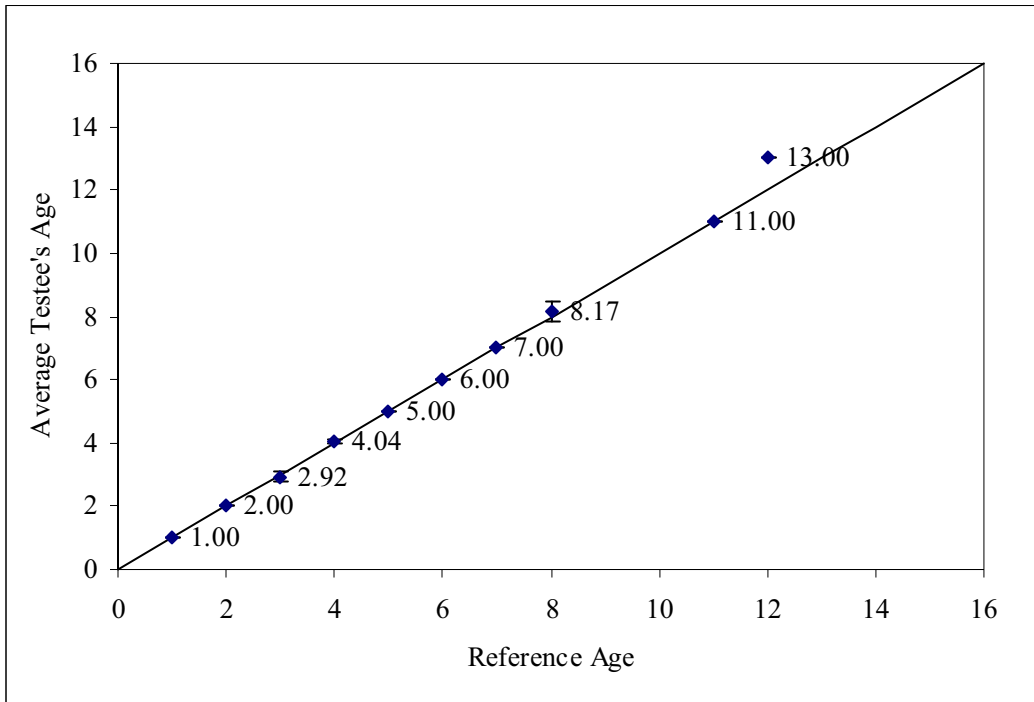


Figure 11. Results of haddock age-reader accuracy exercise against randomly selected samples from the NEFSC haddock reference collection. Error bars indicate 95% confidence intervals.



N Aged	107	CV	1.26
N Agreed	100		
Disagreed	7	%Agreement	93.5%

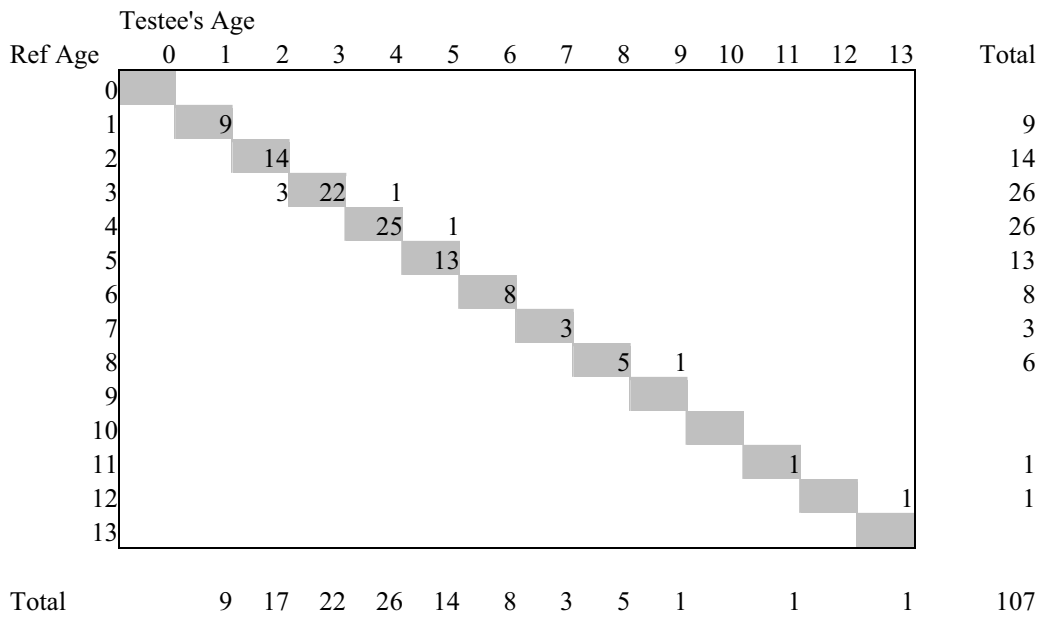
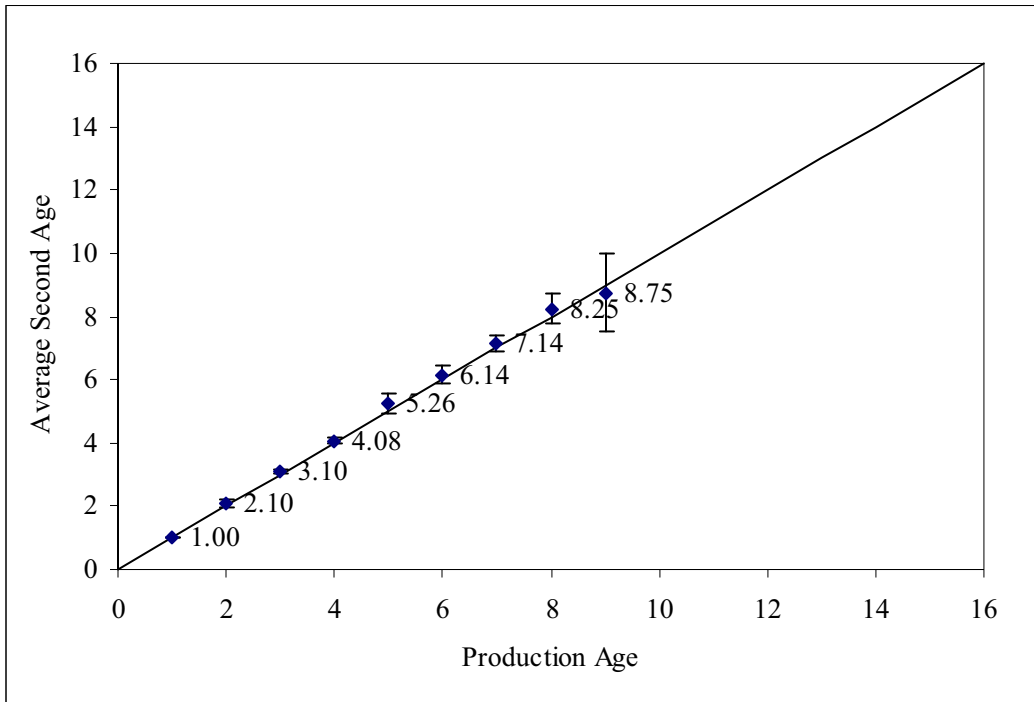


Figure 12. Results of yellowtail age-reader precision exercise against randomly selected samples from Canadian 2004 commercial landings. Error bars indicate 95% confidence intervals.



N Aged	167	CV	2.52
N Agreed	144		
Disagreed	23	%Agreement	86.2%

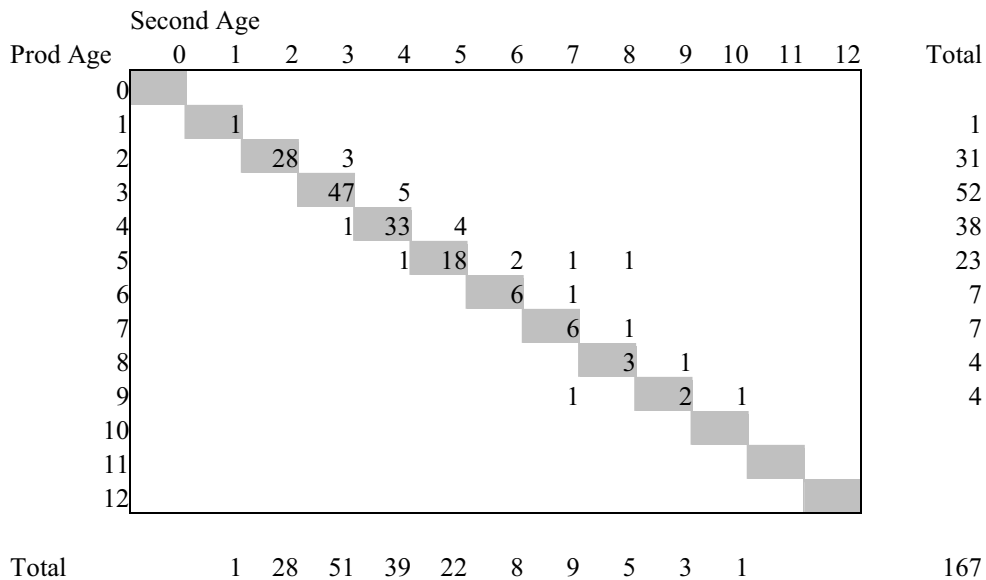
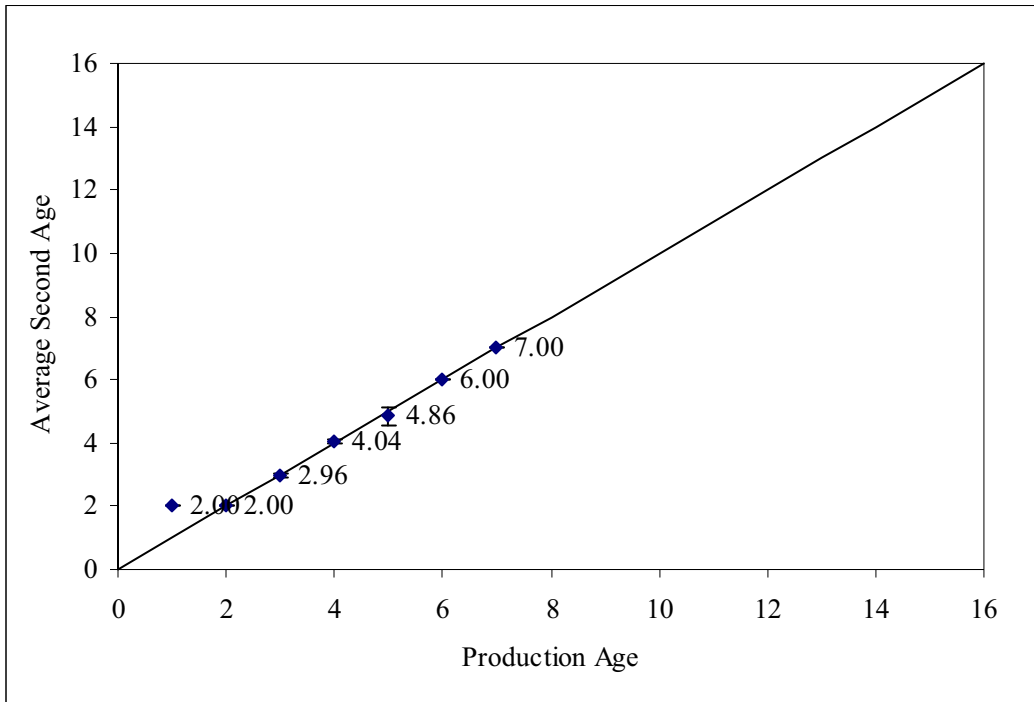


Figure 13. Results of yellowtail age-reader precision exercise against randomly selected samples from the Canadian 2005 bottom trawl survey. Error bars indicate 95% confidence intervals.



N Aged	100	CV	1.79
N Agreed	92		
Disagreed	8	%Agreement	92.0%

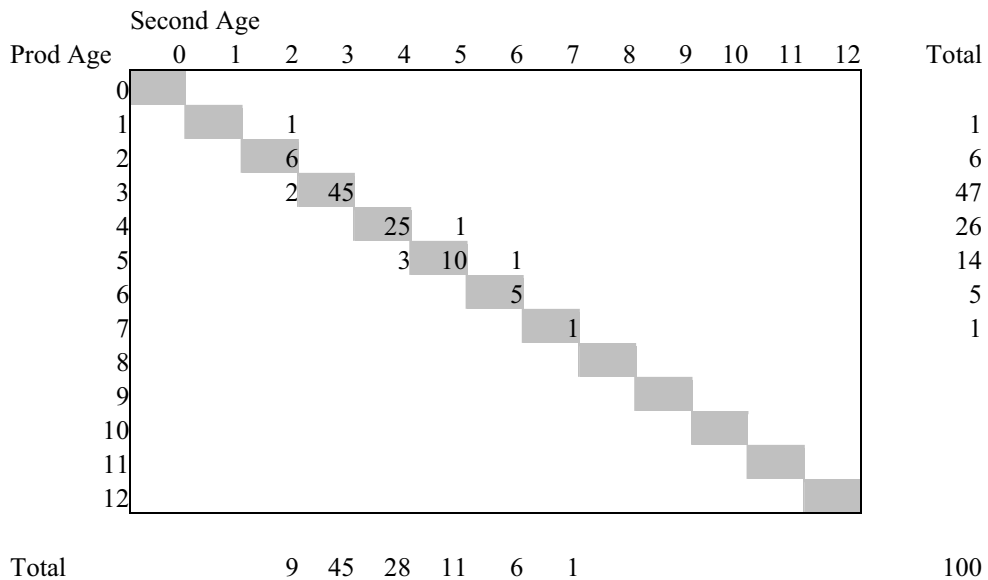
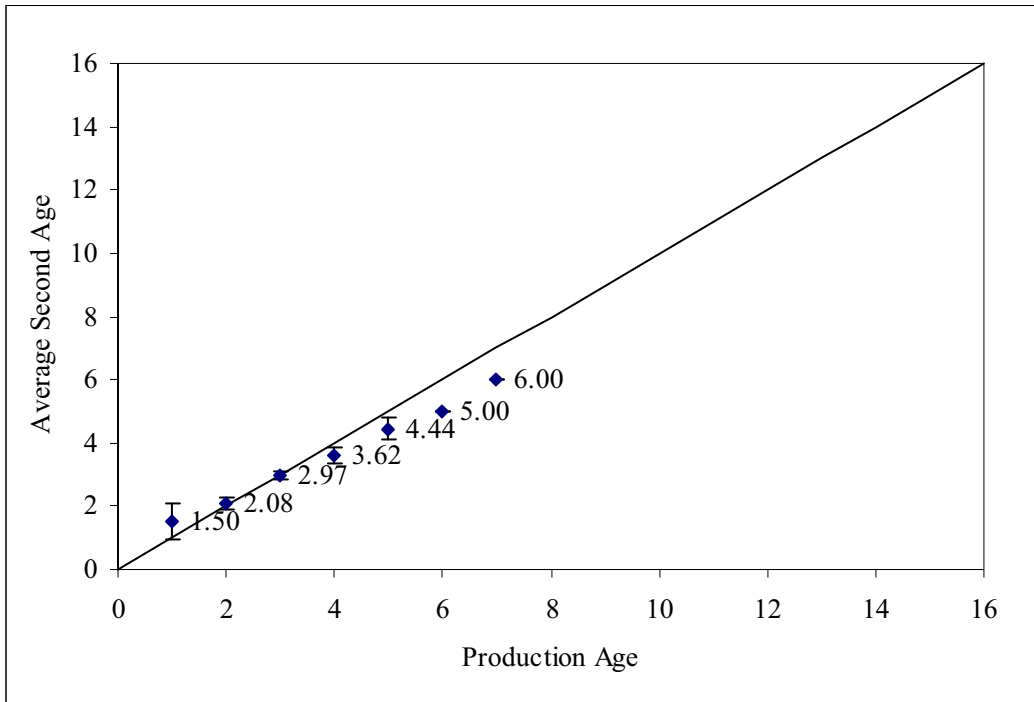


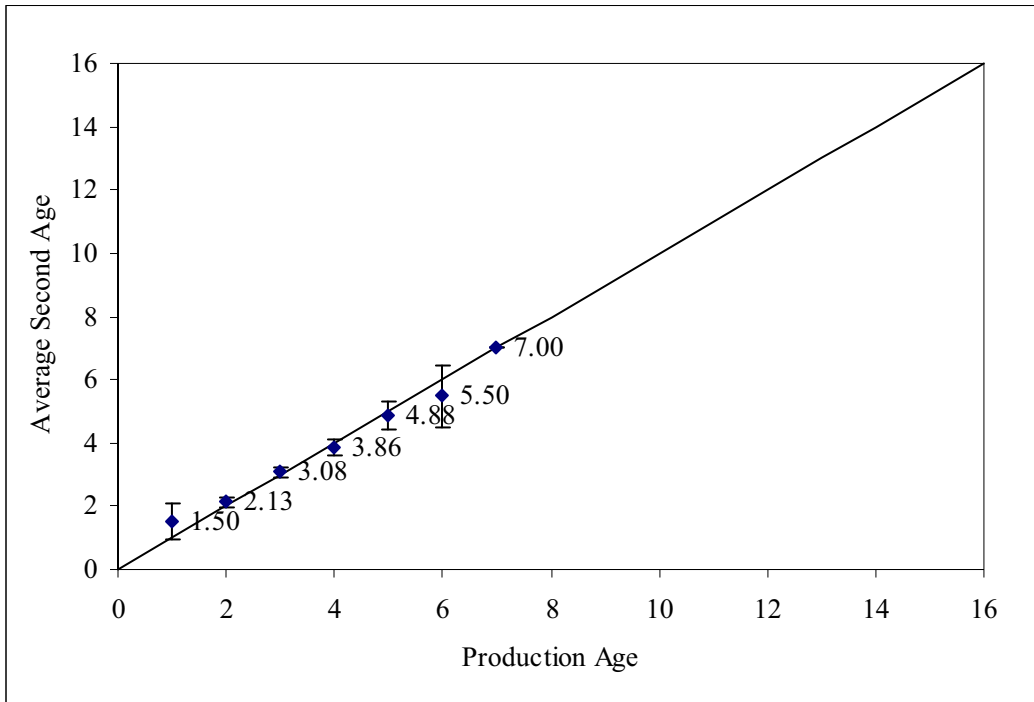
Figure 14. Results of yellowtail age-reader precision exercise against randomly selected samples from U.S. 2004 commercial landings, and NEFSC 2004 autumn and 2005 spring bottom trawl surveys. Error bars indicate 95% confidence intervals.



N Aged	100	CV	6.64
N Agreed	71		
Disagreed	29	%Agreement	71.0%

		Second Age															
		0	1	2	3	4	5	6	7	8	9	10	11	12	Total		
Prod Age	0	1													1		
	1		2	2											4		
	2		1	20	3										24		
	3			3	34	2									39		
	4				9	11	1								21		
	5					5	4								9		
	6						2								2		
	7							1							1		
	8								1								
	9									1							
	10										1						
	11											1					
12												1					
Total		3	25	46	18	7	1								100		

Figure 15. Results of trainee yellowtail age-reader precision exercise against randomly selected samples from U.S. 2004 commercial landings, and NEFSC 2004 autumn and 2005 spring bottom trawl surveys. Error bars indicate 95% confidence intervals.



N Aged	100	CV	6.11
N Agreed	73		
Disagreed	27	%Agreement	73.0%

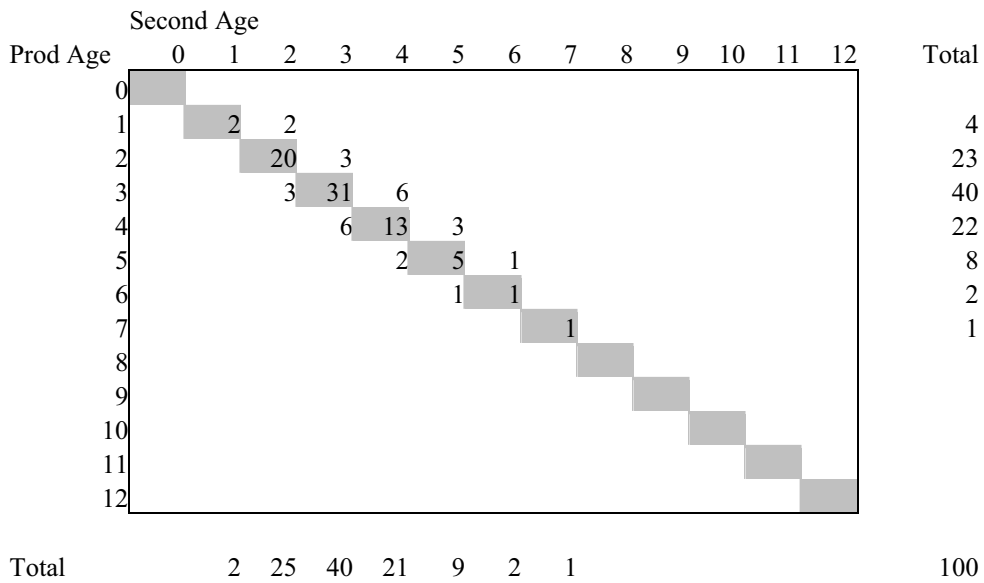
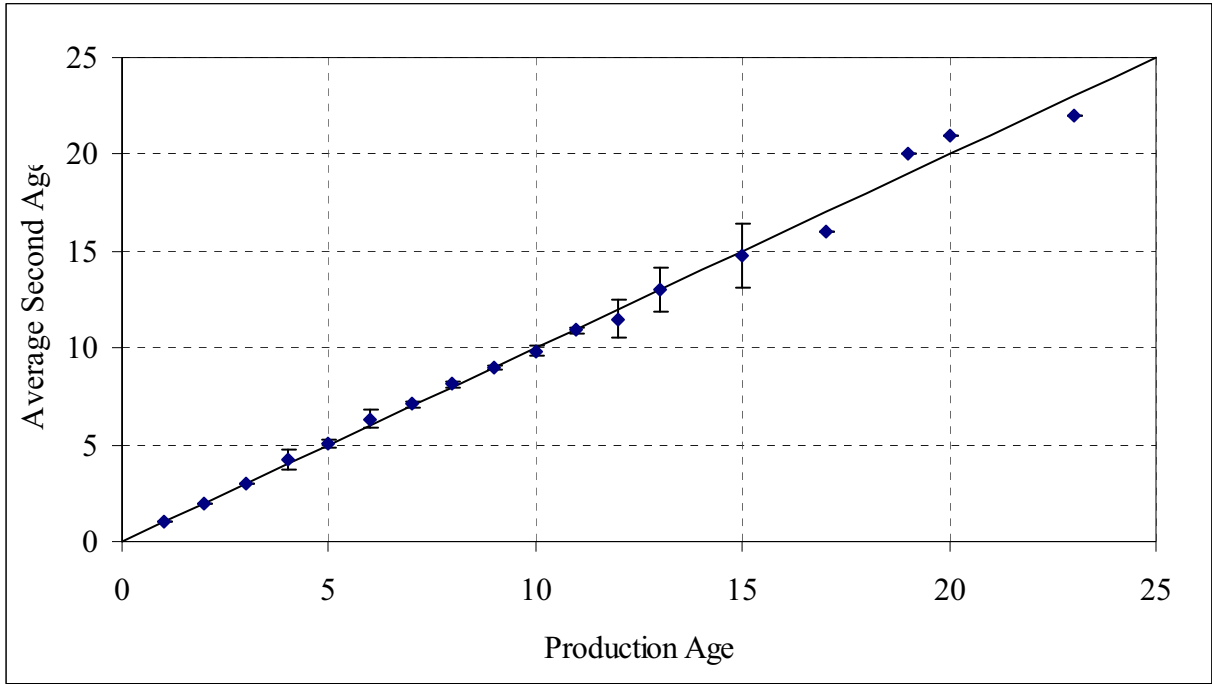


Figure 16. Results of witch flounder age-reader precision exercise against samples from Quarters 2 and 4 of 2004 U.S. commercial landings (N=60) and the NEFSC 2005 spring bottom trawl survey (N=62). Error bars indicate 95% confidence intervals.



N Aged	122
N Agreed	98
Disagreed	24

CV	1.55
%Agreement	80.3%

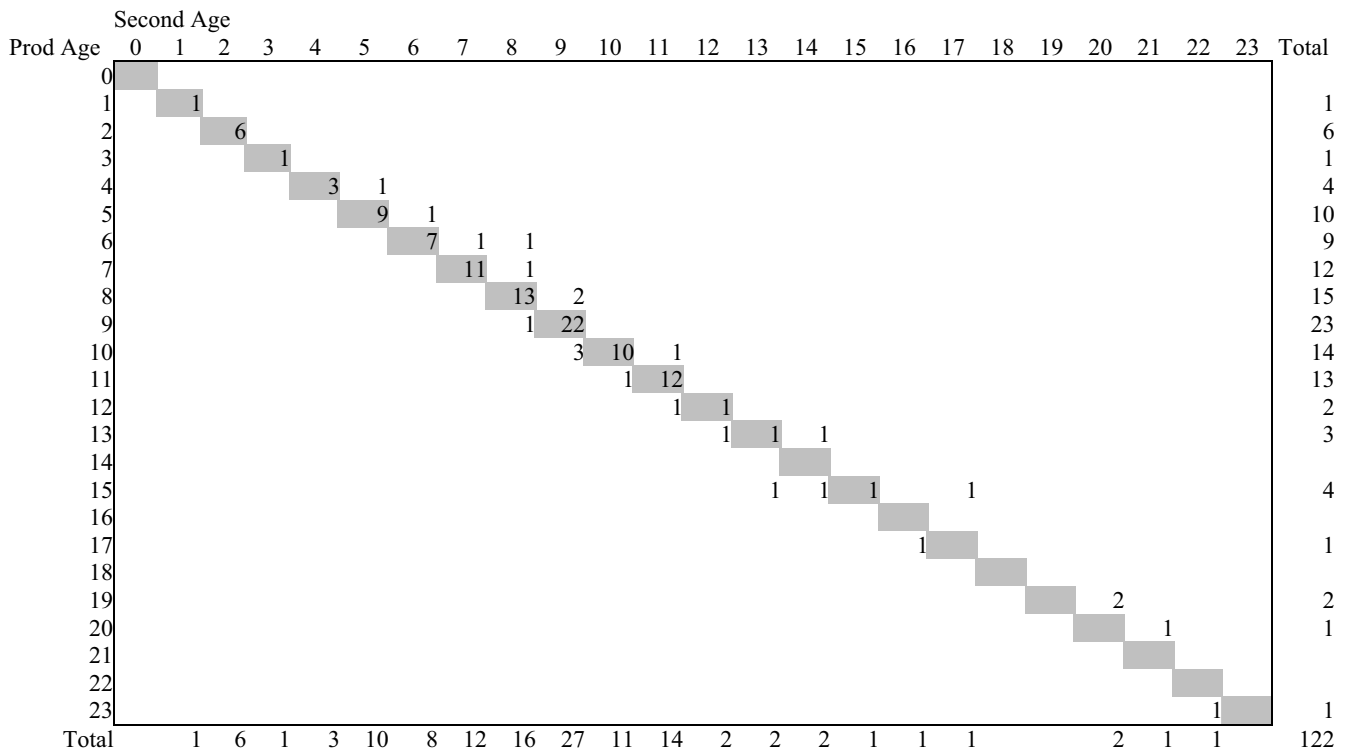
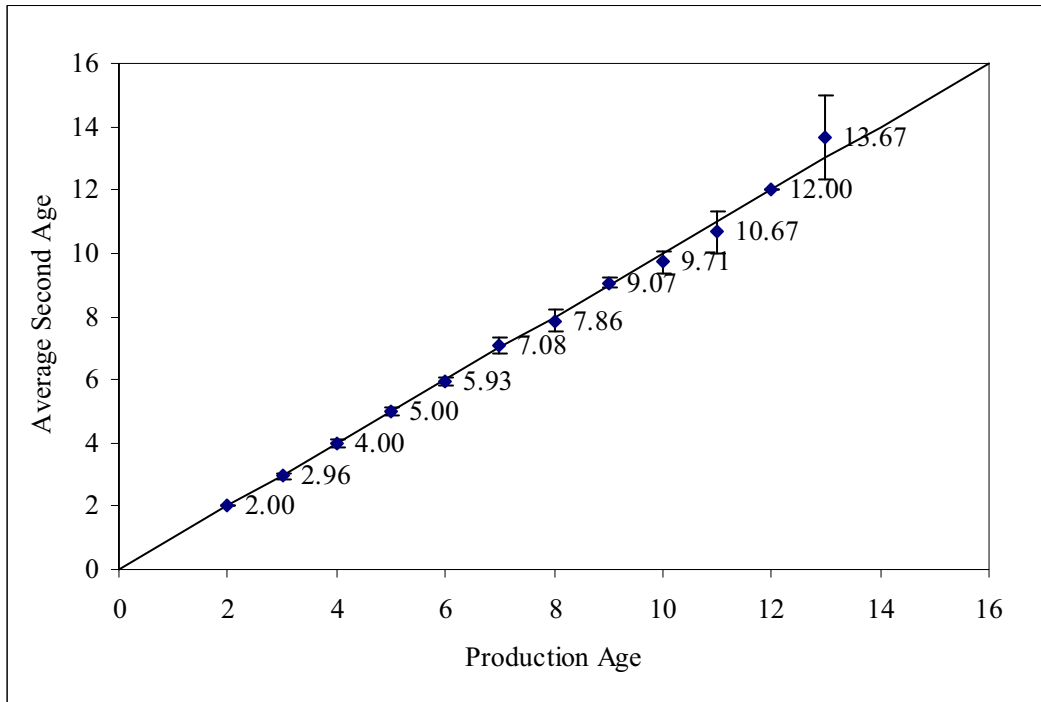


Figure 17. Results of American plaice age-reader precision exercise against samples from the Quarter 1 of 2004 U.S. commercial landings (N=82) and the NEFSC 2004 autumn bottom trawl survey (N=79). Error bars indicate 95% confidence intervals.



N Aged	161	CV	1.69
N Agreed	138	%Agreement	85.7%
Disagreed	23		

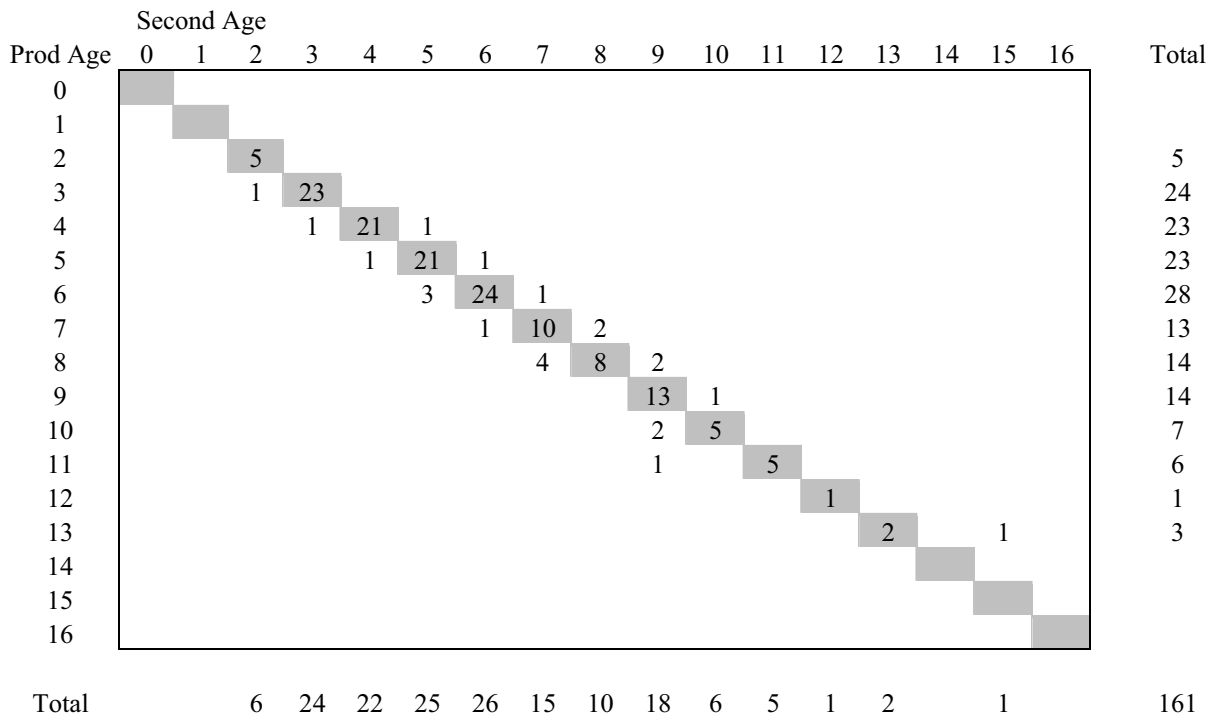
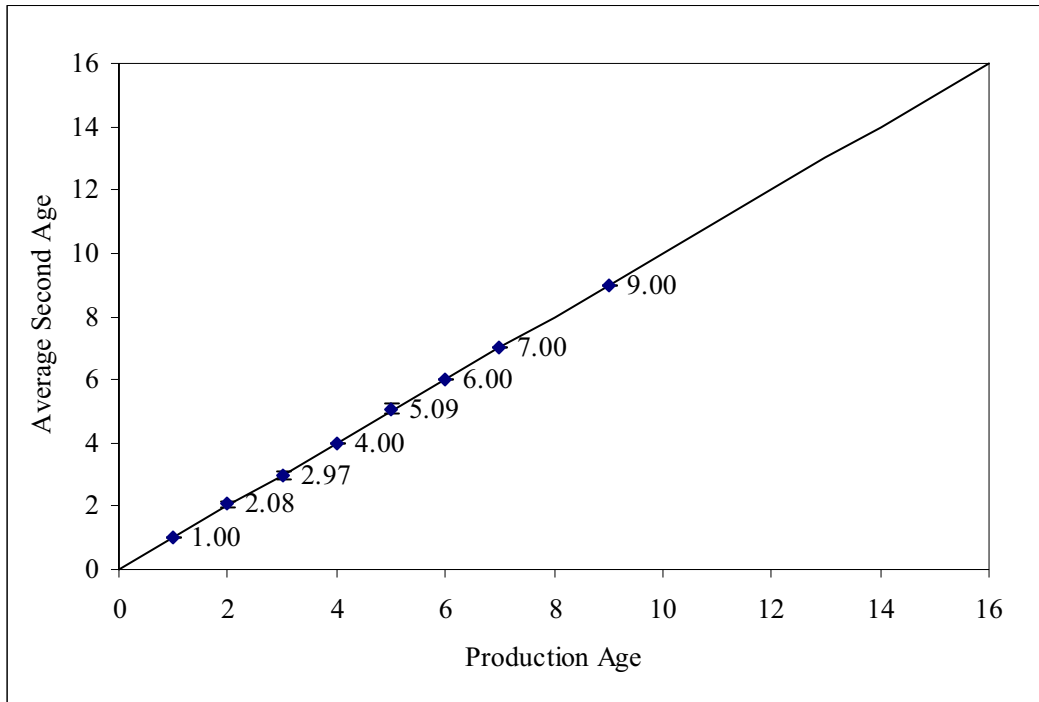




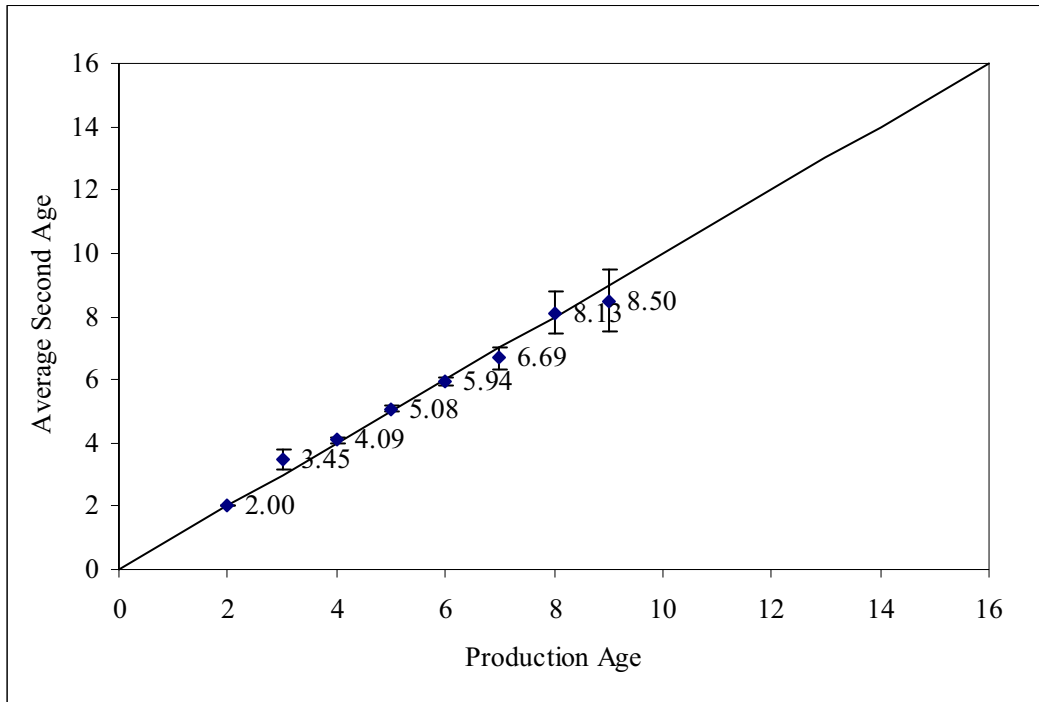
Figure 18. Results of winter flounder age-reader precision exercise against otolith samples from the NEFSC 2004 autumn (N=55) and 2005 spring (N=55) bottom trawl surveys. Error bars indicate 95% confidence intervals.



N Aged	110	CV	1.59
N Agreed	103	%Agreement	93.6%
Disagreed	7		

		Second Age											
		0	1	2	3	4	5	6	7	8	9	Total	
Prod Age	0	11											11
	1		34										37
	2			26									29
	3				16								16
	4					10							11
	5						2						2
	6							3					3
	7												
	8												
	9											1	1
Total		11	36	29	17	10	3	3			1	110	

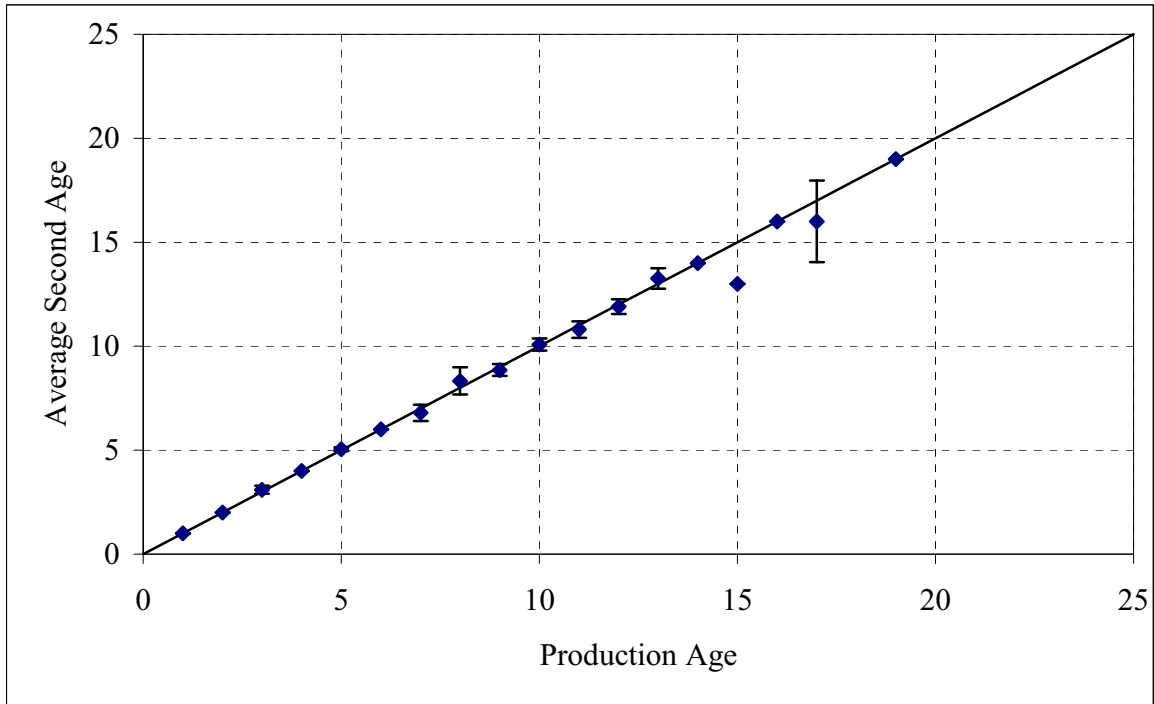
Figure 19. Results of winter flounder age-reader precision exercise against scale samples from Quarters 1 (N=55) and 3 (N=60) of 2004 U.S commercial landings. Error bars indicate 95% confidence intervals.



N Aged	115	CV	2.81
N Agreed	91	%Agreement	79.1%
Disagreed	24		

		Second Age											Total		
Prod Age		0	1	2	3	4	5	6	7	8	9	10			
Prod Age	0	11													11
	1		6												6
	2			1											1
	3				6										6
	4					5									5
	5						2								2
	6							23							23
	7								3						3
	8									28					28
	9										5				5
	10											7			7
Total															115

Figure 20. Results of redfish age-reader precision exercise against randomly selected samples from the NEFSC 2004 autumn bottom trawl survey. Error bars indicate 95% confidence intervals.



N Aged	142	CV	0.99
N Agreed	127	%Agreement	89.4%
Disagreed	15		

