

Appendix. Evaluation of accuracy and precision of U.S. haddock aging during 2004–2005.

A.1. Introduction

For decades, US and Canadian scientists have routinely determined the ages of Georges Bank haddock (*Melanogrammus aeglefinus*) collected from the commercial fishery and research surveys and have used the same otolith aging approaches. These production aging data have provided important size-at-age information for conducting the analytical VPA assessment. In the spring of 2004, a new set of quality control tests were initiated to evaluate the amount of aging error and bias associated with production ages. In this context, tests of age reader accuracy can be thought of as how often the correct age is obtained while tests of precision reflect how often the same age is obtained (Campana 2001). Such measures are important components of a quality control monitoring program.

Acceptable levels of age determination accuracy and precision are highly influenced by species, age structure, and age reader experience. Although percent agreement is affected by these factors, the Northeast Fisheries Science Center (NEFSC) aging lab has long considered percent agreement levels above 85–90% to be acceptable for most species. In terms of the total coefficient of variation (CV), various aging labs around the world consider levels under 5% to be acceptable for species of moderate longevity and aging complexity (Campana 2001), such as haddock.

A.2. Methods

Accuracy of age determination was measured by re-aging a portion (50–100 fish) of the haddock reference collection. Accuracy tests were conducted after completion of production aging in August 2004 and June 2005. This allowed for detection and quantification of any age reader bias. In the future, accuracy tests will also be conducted prior to production aging.

The reference collection consists of sectioned otoliths from Georges Bank haddock captured on various NEFSC spring and autumn bottom trawl surveys. Fish included in the annual aging exchange between NEFSC and the Canadian Department of Fisheries and Oceans, for which the ages were agreed upon by both U.S. and Canadian age readers, were added to the NEFSC haddock reference collection. Therefore, the reference age for each fish was reached by consensus among at least two age readers.

Precision tests were conducted soon after completion of each segment (survey or commercial quarter) of samples. These exercises were separated by stock area if production ages for the two stocks were determined at different times.

For these tests, a random sub-sample of about 50 (range 30–100) fish was selected from the production sample. This subsample was then re-aged by the age reader. When re-aging fish, the same data were available as during production aging, i.e. fish length, date captured, and area captured. Precision measurements were calculated by comparing the test ages generated during the exercise against the ages obtained for the same fish during production aging. These measures can quantify the amount of random error associated with a given production age, and indicate if any change occurred in the way fish are aged between the time of production aging and the test exercise.

Age-reader precision was estimated for haddock for all sets of production ages generated since the start of 2004. This includes U.S. commercial landings for all quarters of both 2003, where the Georges Bank (GB) and Gulf of Maine stocks were aged separately, and 2004, where stock areas were mixed during aging. U.S. bottom trawl surveys for which precision exercises were completed are as follows: 2003 Fall (GB), 2004 Spring (GB), 2004 Fall (mixed stocks), and 2005 Spring (mixed stocks).

During age-testing exercises, no attempts were made to improve results with 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). In cases where the percent agreement dropped below 90%, a Bowker's test (Bowker 1948; Hoenig et al. 1995) was also used to measure symmetry.

A.3. Results and Discussion

A total of 157 haddock were used in the two accuracy tests, including fish with reference ages of up to 12 years. In these exercises, the age reader achieved total CVs of 2.3% and 1.3%, with 90% and 94% agreement respectively (Table A1, Figures A1–A2). This represents an acceptable level of accuracy for this species.

For the four precision exercises on survey samples, a total of 291 haddock were included, with production ages of up to 9–13 years in each exercise. For the commercial samples, 357 haddock were used in eight precision exercises. These exercises included fish with maximum production ages of 10–12 years. Results are presented in Figures A3–A14, and summarized in Table A2.

For all haddock precision exercises, a median total CV level of 0.8% and a median of 94% agreement were attained. These values ranged between 0.2% and 2.6% total CV, with 76% to 98% agreement (Table A2, Figures A3–A14). The agreement level fell below 90% in the four exercises on 2003 commercial samples. In all these cases, the Bowker's test found no significant deviation from symmetry ($P>0.25$). These results indicate a high level of consistency in age determinations for this species.

For the 49 fish (8% of all fish that were re-aged) in which the production and test ages differed, only three fish differed by 2 years, and none differed by 3 or more years. There may have been a modest bias toward subtracting one year while determining test ages, however, as there were 29 fish in which the age dropped by one year between the two readings, and only 17 fish for which the age was raised by one year.

The relatively high accuracy obtained in re-aging samples from the Georges Bank reference collection and the consistently high precision of age readings supports the conclusion that recent age determinations for Georges Bank haddock have been adequate for stock assessment purposes.

A.4. References.

- Bowker AH. 1948. A test for symmetry in contingency tables. *J. Am. Stat. Assoc.* 43:572–574.
- Campana SE, Annand MC, McMillan JI. 1995. Graphical and statistical methods for determining the consistency of age determinations. *Trans. Am. Fish. Soc.* 124:131-138.
- Campana SE. 2001. Accuracy, precision, and quality control in age determination, including a review of the use and abuse of age validation methods. *J. Fish. Biol.* 59:197-242.
- Hoenig JM, Morgan MJ, Brown CA. 1995. Analysing differences between two age determination methods by tests of symmetry. *Can. J. Fish. Aquat. Sci.* 52:364-368.

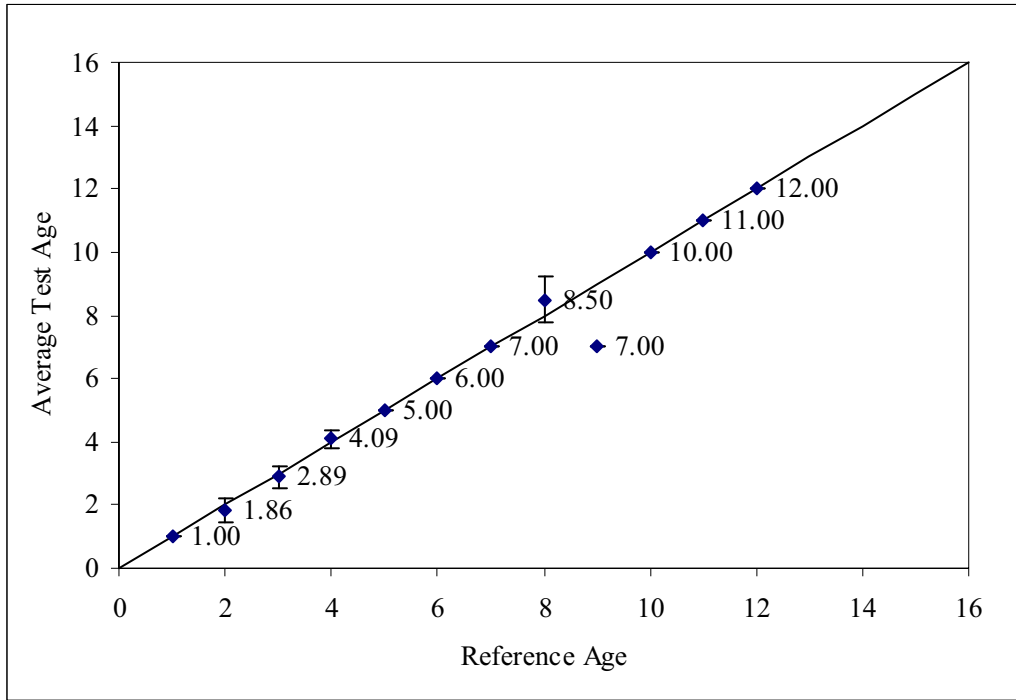
Table A1. Results of both haddock accuracy tests, listing associated figures.

Figure	Source	Date of Test	<i>N</i>	Agreement (%)	Total CV (%)
1	Reference Collection	Aug. 2004	50	90.0	2.34
2	Reference Collection	June 2005	107	93.5	1.26

Table A2. Results of all haddock precision exercises, with list of associated figures. The stock area column indicates whether each exercise was on Georges Bank (GB) samples only, or if the two stocks were mixed.

Figure	Source	Stock Area	Date of Test	<i>N</i>	Agreement (%)	Total CV (%)
<u>Survey Samples</u>						
3	Autumn 2003	GB	April 2004	48	95.8	0.75
4	Spring 2004	GB	May 2004	65	95.4	0.68
5	Autumn 2004	mixed	Jan. 2005	102	95.1	0.71
6	Spring 2005	mixed	May 2005	76	96.1	0.89
<u>Commercial Samples</u>						
7	1st Quarter 2003	GB	March 2004	34	82.4	2.20
8	2nd Quarter 2003	GB	April 2004	40	87.5	1.71
9	3rd Quarter 2003	GB	April 2004	33	75.8	2.58
10	4th Quarter 2003	GB	April 2004	35	88.6	1.71
11	1st Quarter 2004	mixed	Jan. 2005	56	92.9	0.69
12	2nd Quarter 2004	mixed	March 2005	55	90.9	0.94
13	3rd Quarter 2004	mixed	April 2005	44	97.7	0.25
14	4th Quarter 2004	mixed	April 2005	60	95.0	0.44

Figure A1. Results of August 2004 haddock age-reader accuracy exercise against randomly selected samples from the NEFSC haddock reference collection. Error bars indicate 95% confidence intervals.



N Aged	50	Total CV	2.34%
N Agreed	45	%Agreement	90.0%
Disagreed	5		

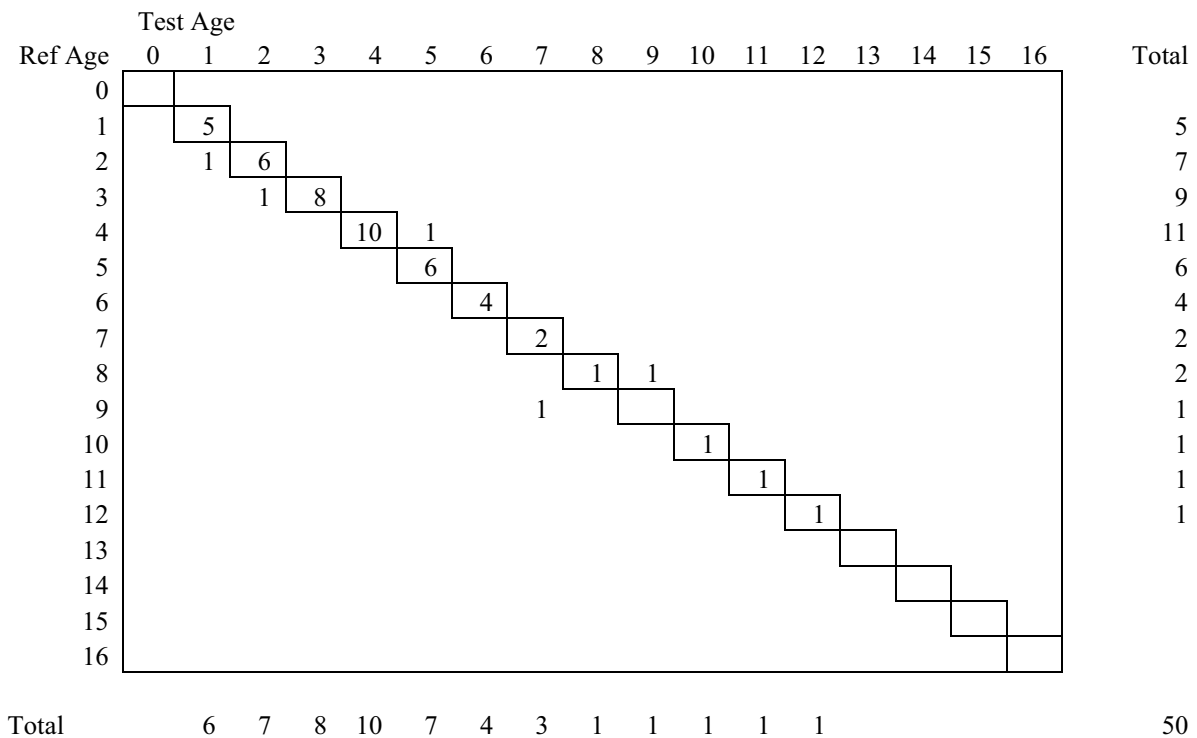
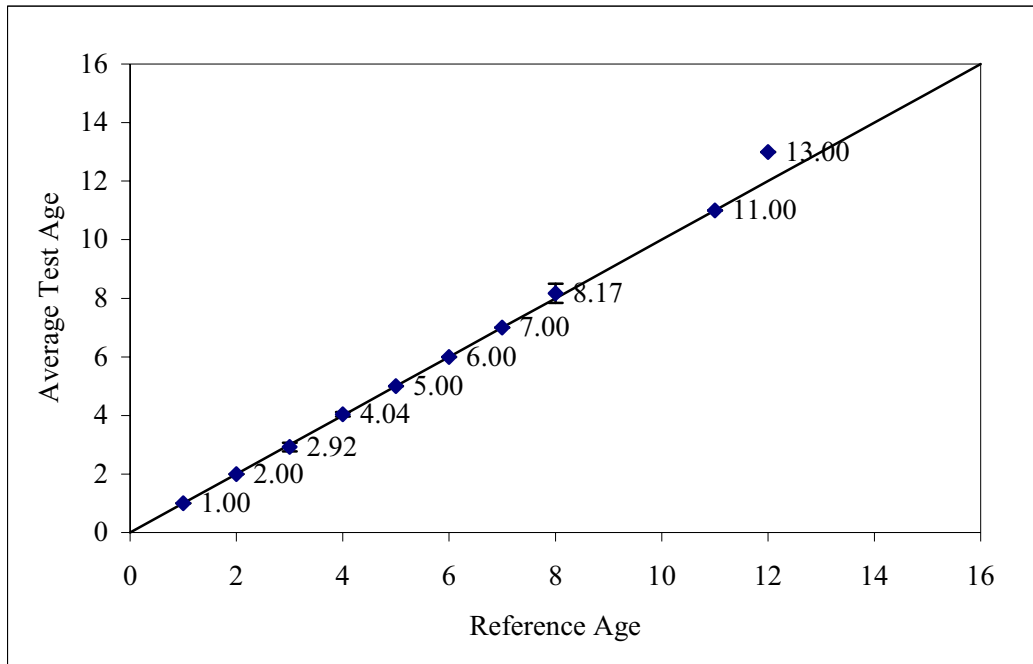


Figure A2. Results of June 2005 haddock age-reader accuracy exercise against randomly selected samples from the NEFSC haddock reference collection. Error bars indicate 95% confidence intervals.



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

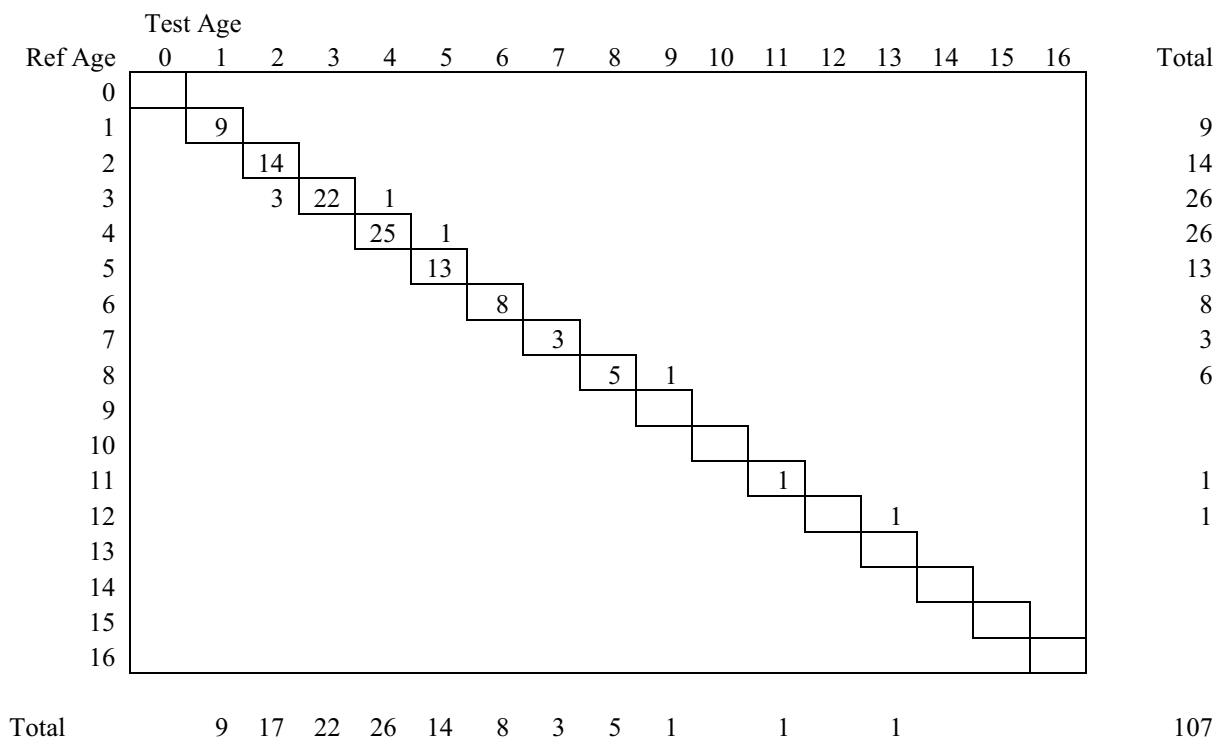
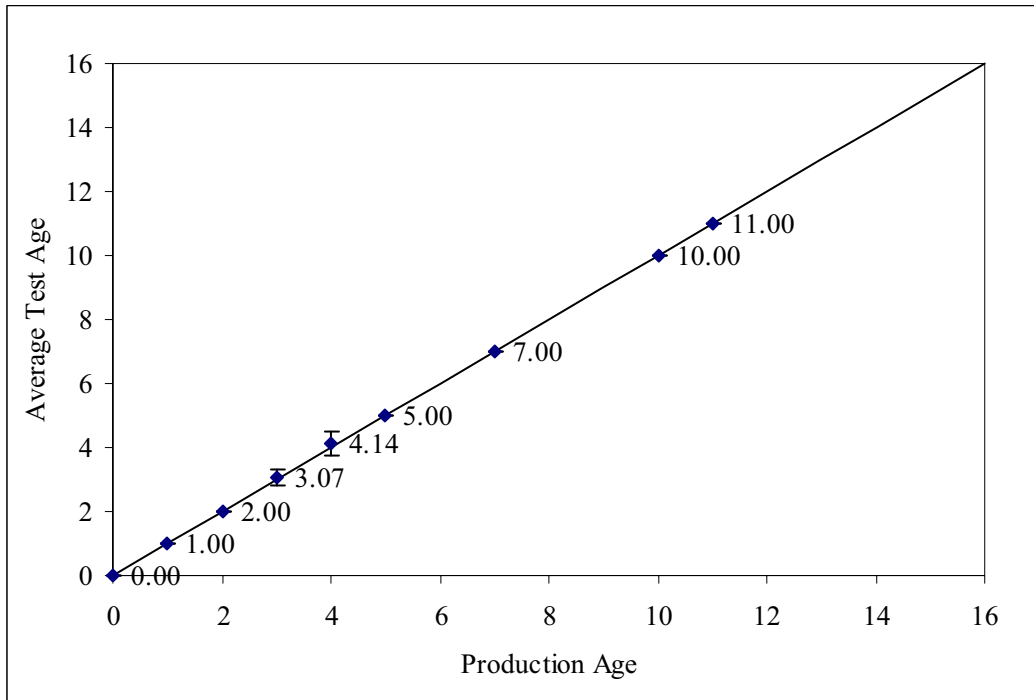


Figure A3. Results of haddock age-reader precision exercise against randomly selected samples from the NEFSC 2003 autumn bottom trawl survey, including only samples from Georges Bank. Error bars indicate 95% confidence intervals.



N Aged	48	Total CV	0.75%
N Agreed	46	%Agreement	95.8%
Disagreed	2		

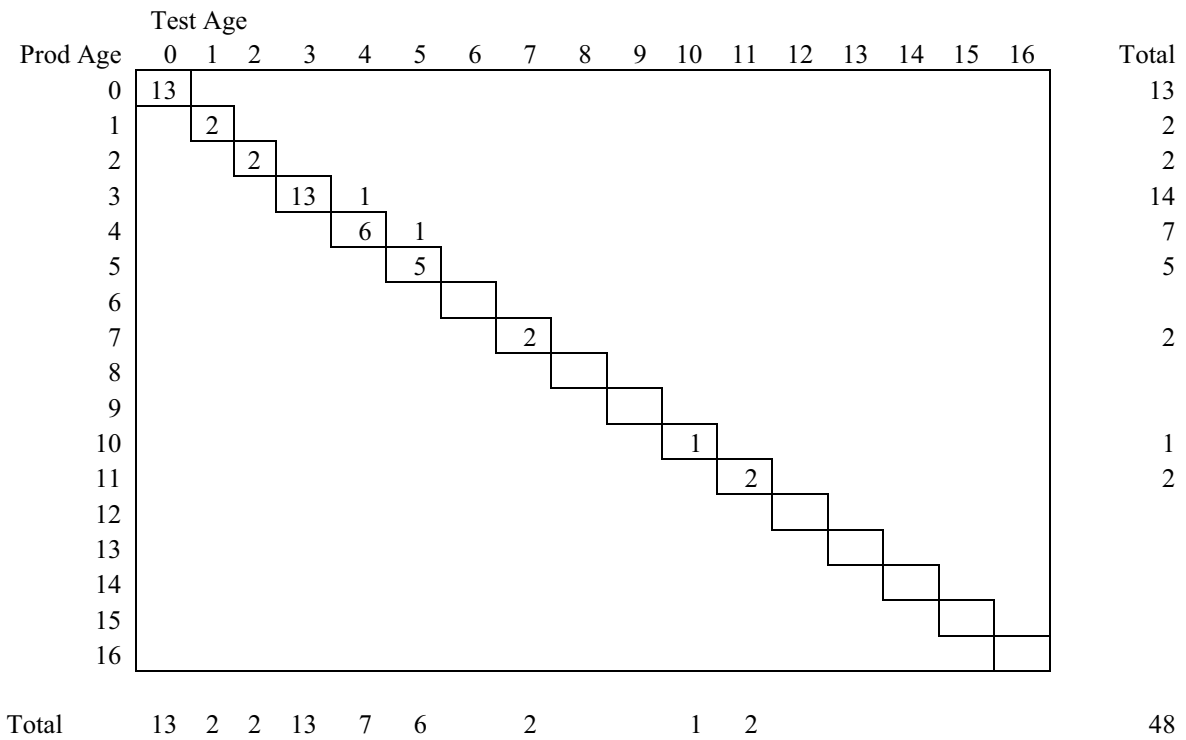
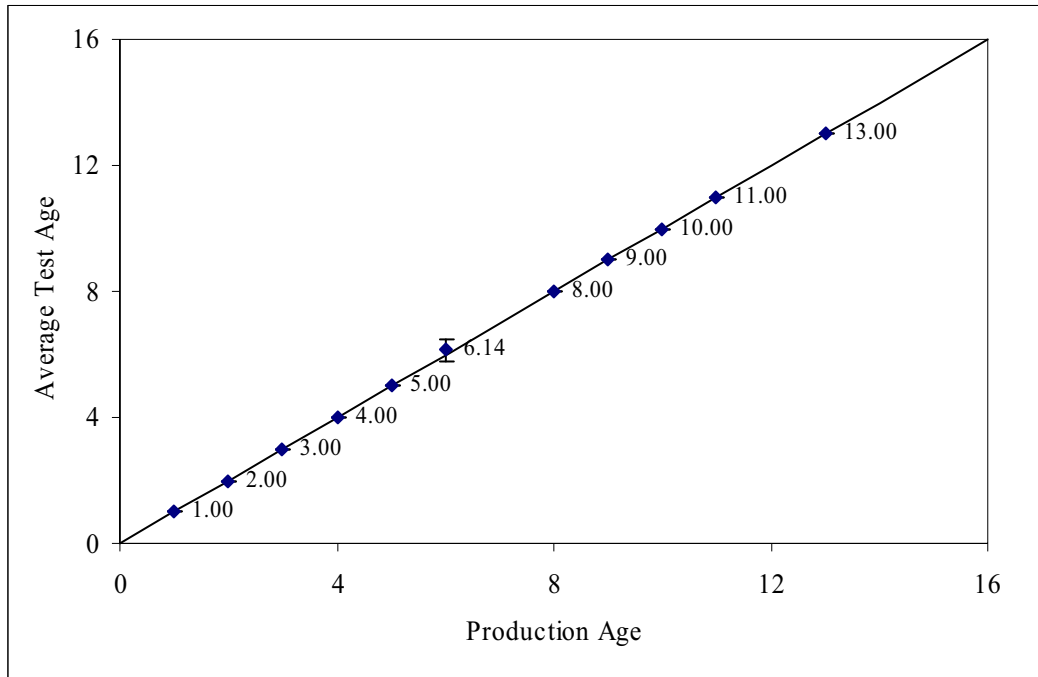


Figure A4. Results of haddock age-reader precision exercise against randomly selected samples from the NEFSC 2004 spring bottom trawl survey, primarily including samples from Georges Bank. Error bars indicate 95% confidence intervals.



N Aged	65	Total CV	0.68%
N Agreed	62	%Agreement	95.4%
Disagreed	3		

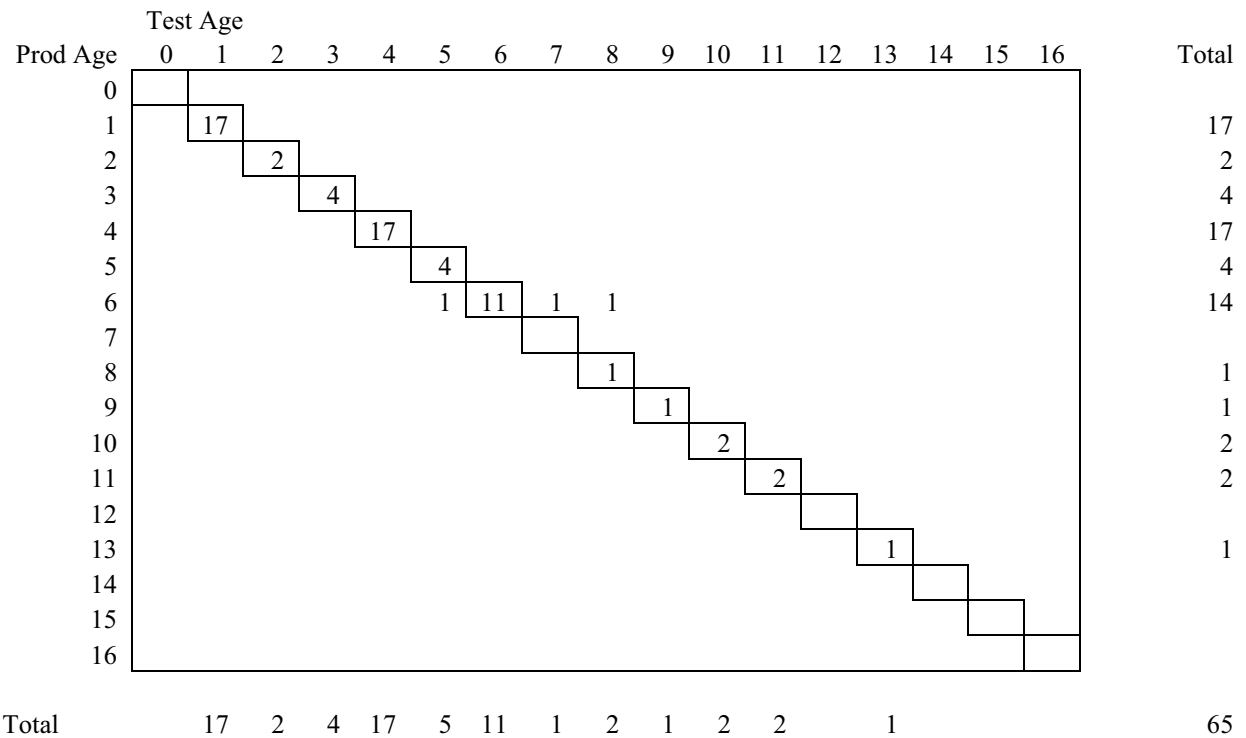
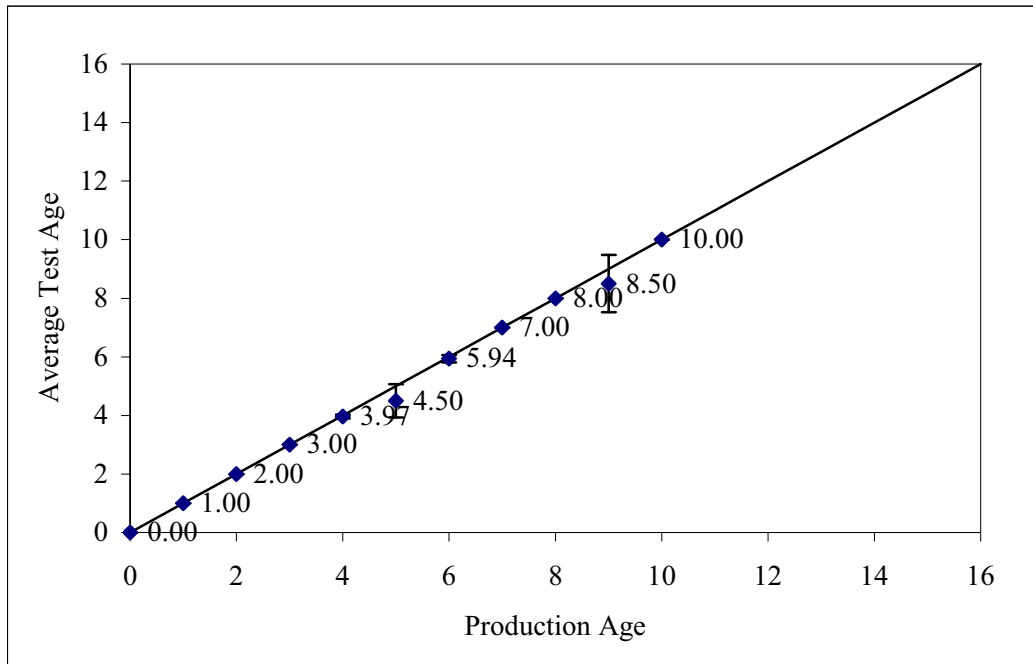


Figure A5. Results of haddock age-reader precision exercise against randomly selected samples from the NEFSC 2004 autumn bottom trawl survey, including samples from both stock areas. Error bars indicate 95% confidence intervals.



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

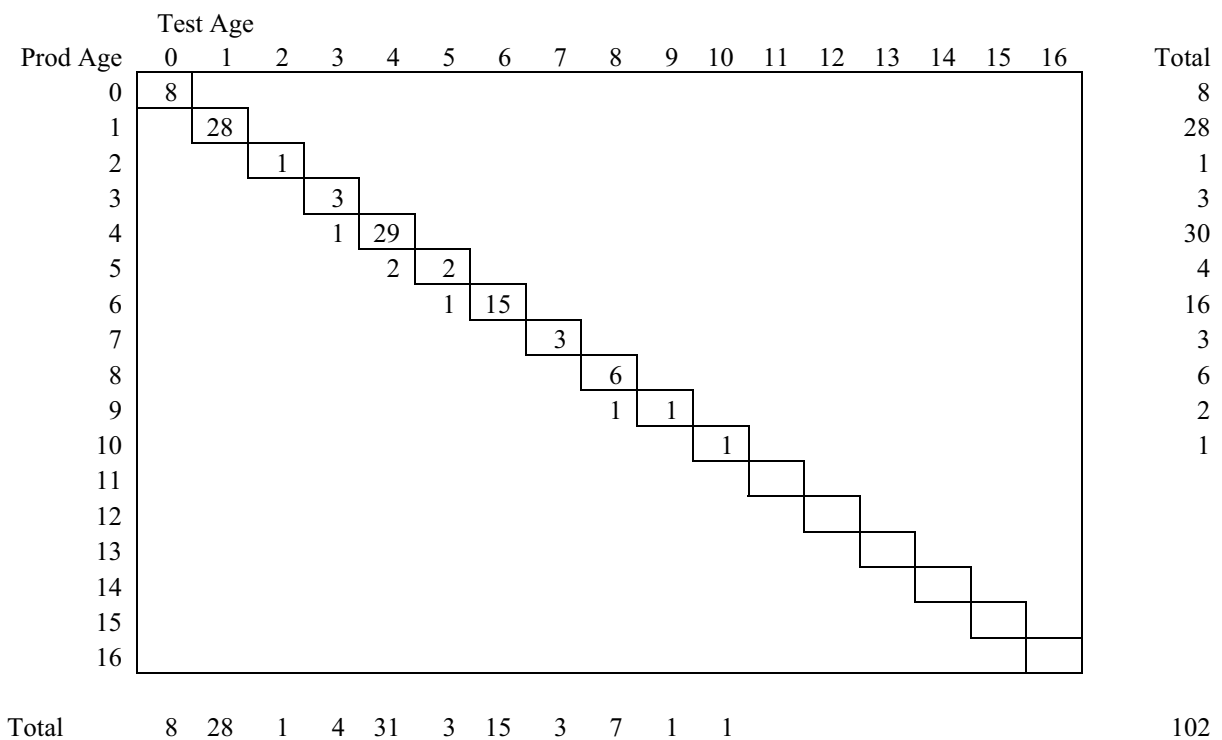
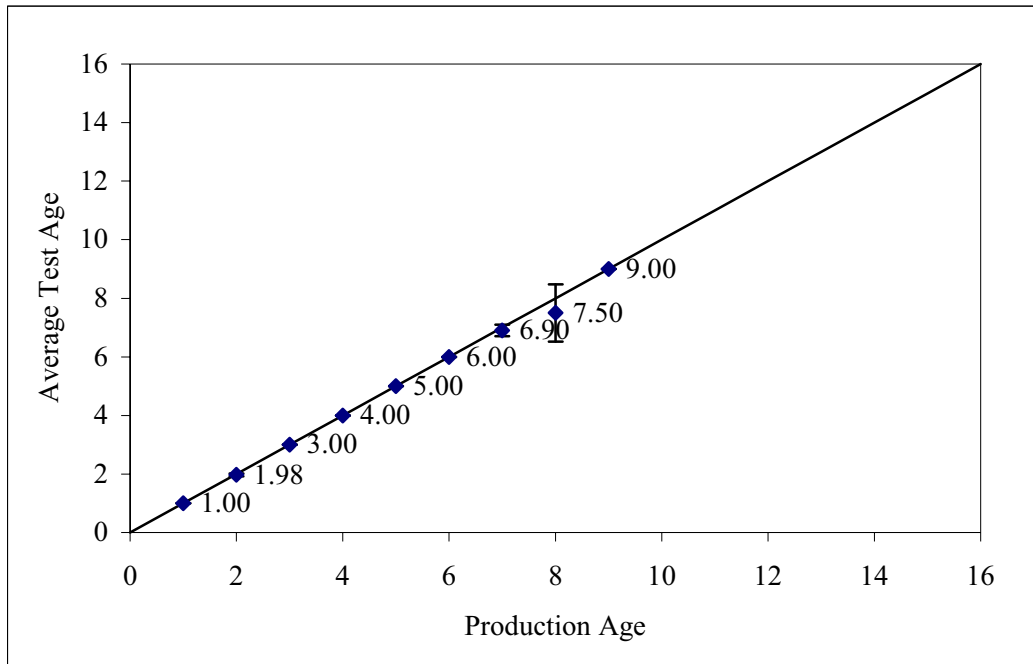


Figure A6. Results of haddock age-reader precision exercise against randomly selected samples from the NEFSC 2005 spring bottom trawl survey (both stock areas). Error bars indicate 95% confidence intervals.



N Aged	76
N Agreed	73
Disagreed	3

Total CV	0.89%
%Agreement	96.1%

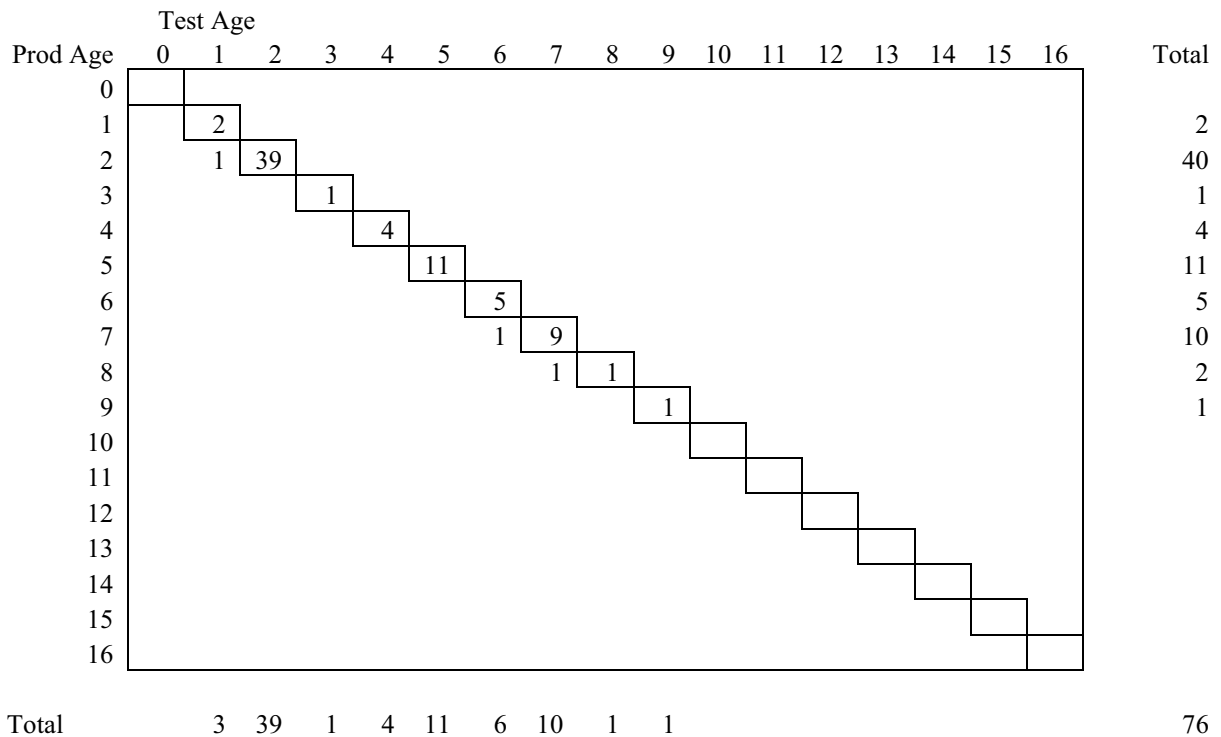
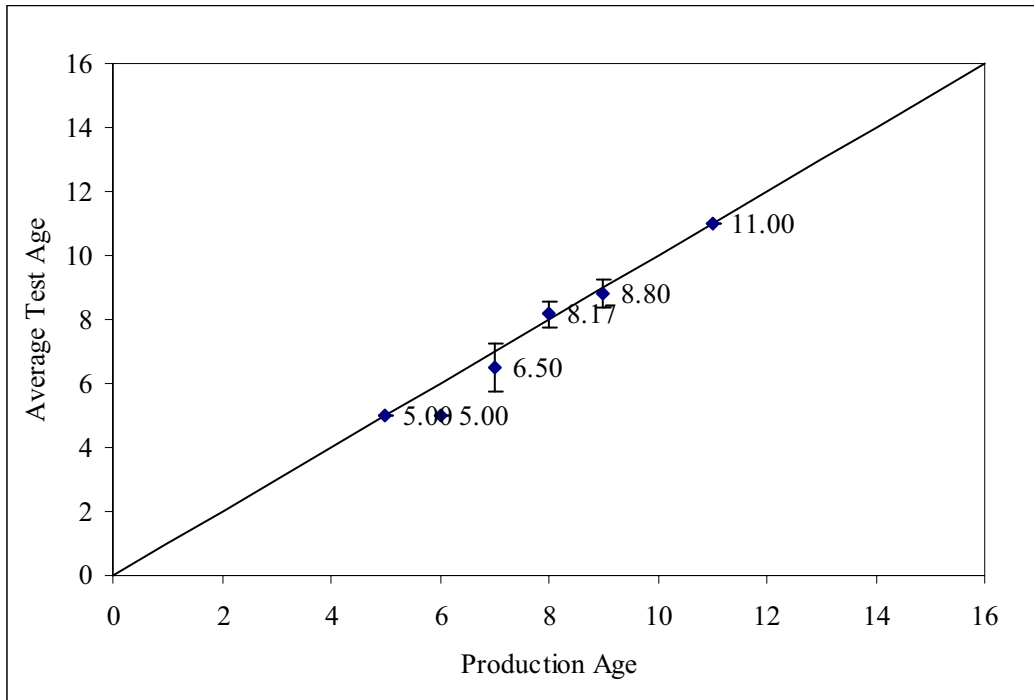


Figure A7. Results of haddock age-reader precision exercise against randomly selected samples from Quarter 1 of 2003 U.S. commercial landings for Georges Bank. Error bars indicate 95% confidence intervals.



N Aged	34
N Agreed	28
Disagreed	6

Total CV	2.20%
%Agreement	82.4%

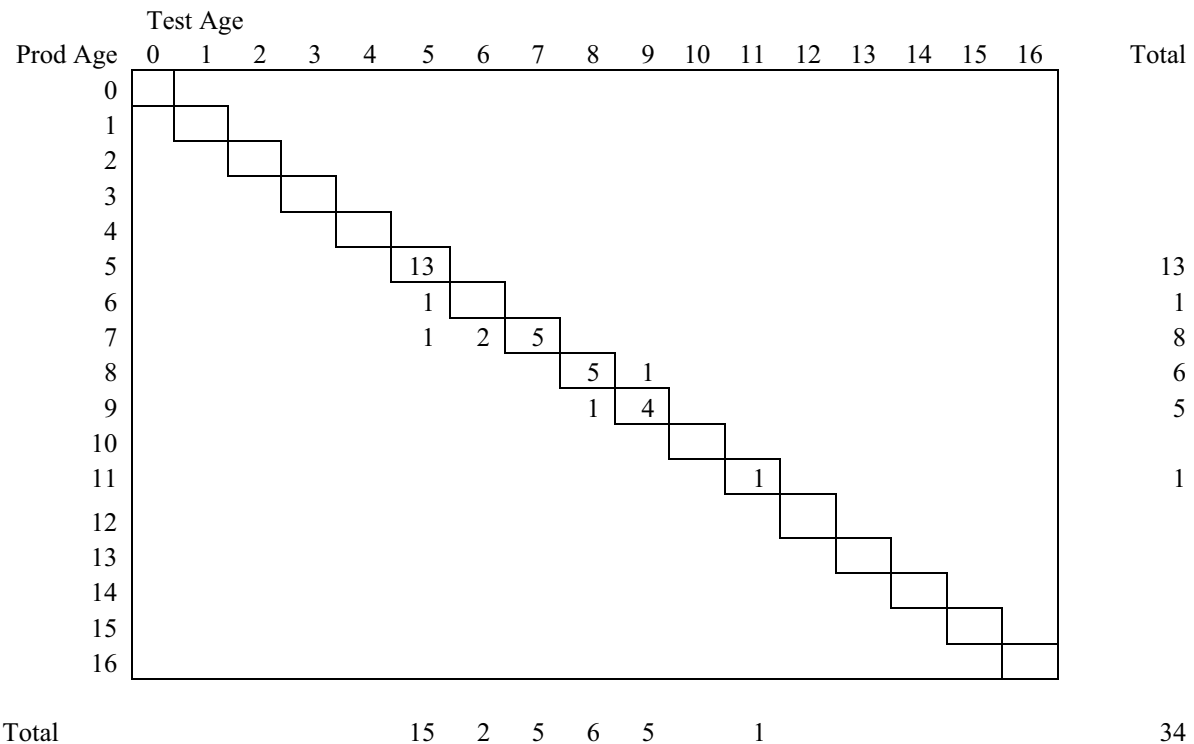
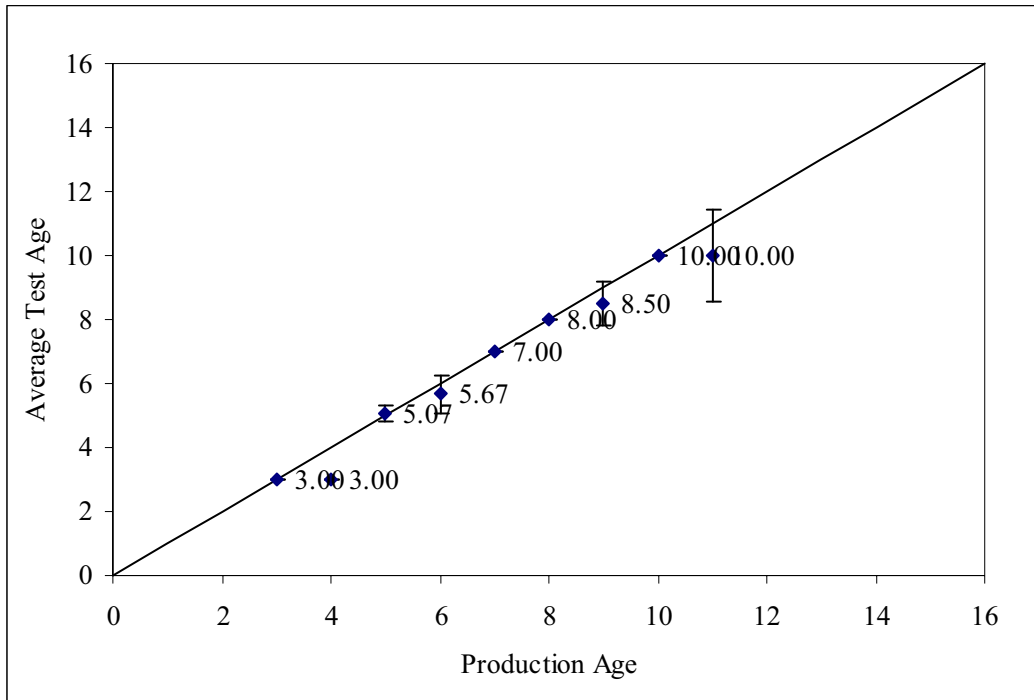


Figure A8. Results of haddock age-reader precision exercise against randomly selected samples from Quarter 2 of 2003 U.S. commercial landings for Georges Bank. Error bars indicate 95% confidence intervals.



N Aged	40
N Agreed	35
Disagreed	5

Total CV	1.71%
%Agreement	87.5%

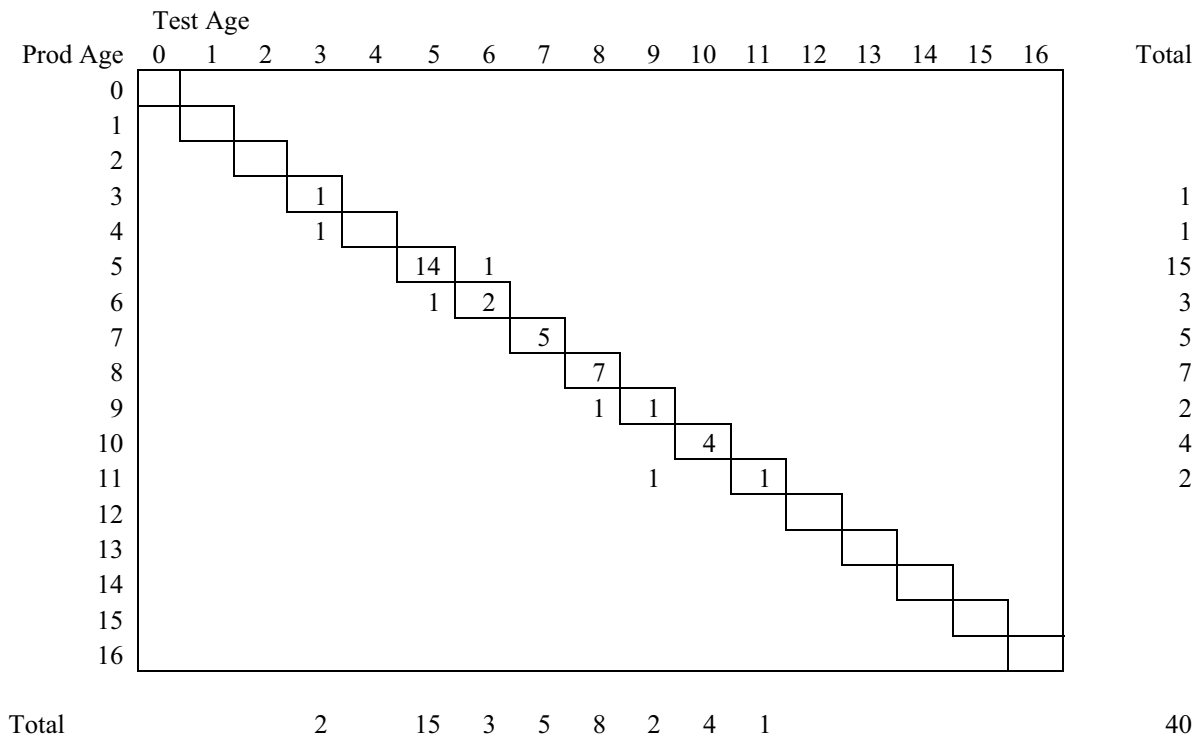
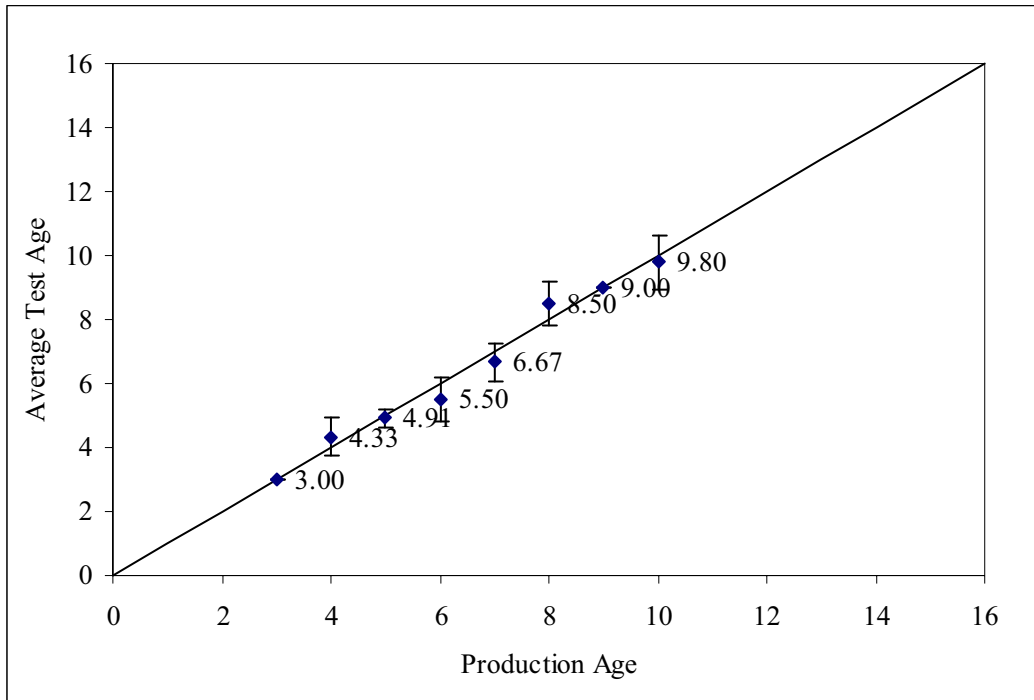


Figure A9. Results of haddock age-reader precision exercise against randomly selected samples from Quarter 3 of 2003 U.S. commercial landings for Georges Bank. Error bars indicate 95% confidence intervals.



N Aged	33
N Agreed	25
Disagreed	8

Total CV	2.58%
%Agreement	75.8%

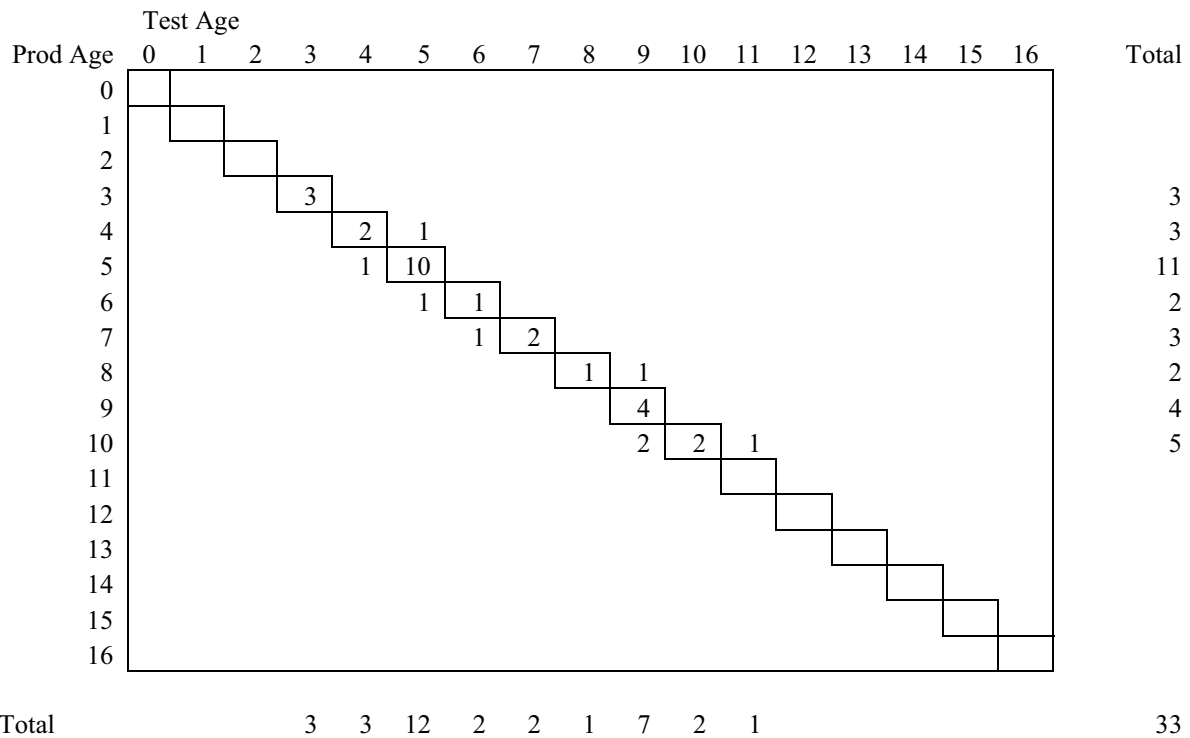
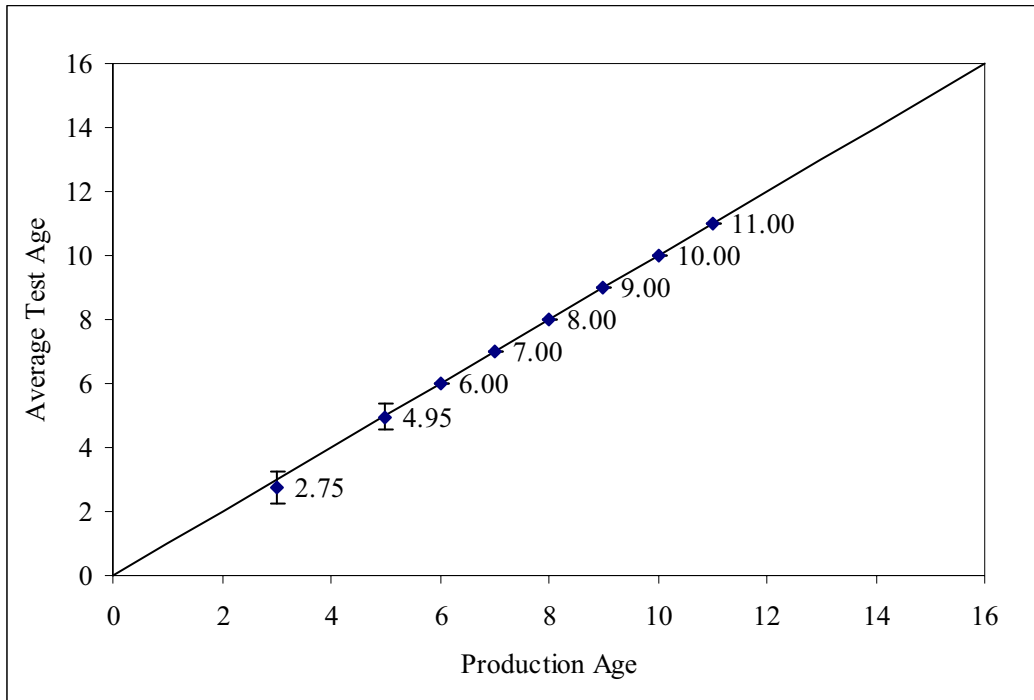


Figure A10. Results of haddock age-reader precision exercise against randomly selected samples from Quarter 4 of 2003 U.S. commercial landings for Georges Bank. Error bars indicate 95% confidence intervals.



N Aged	35
N Agreed	31
Disagreed	4

Total CV	1.71%
%Agreement	88.6%

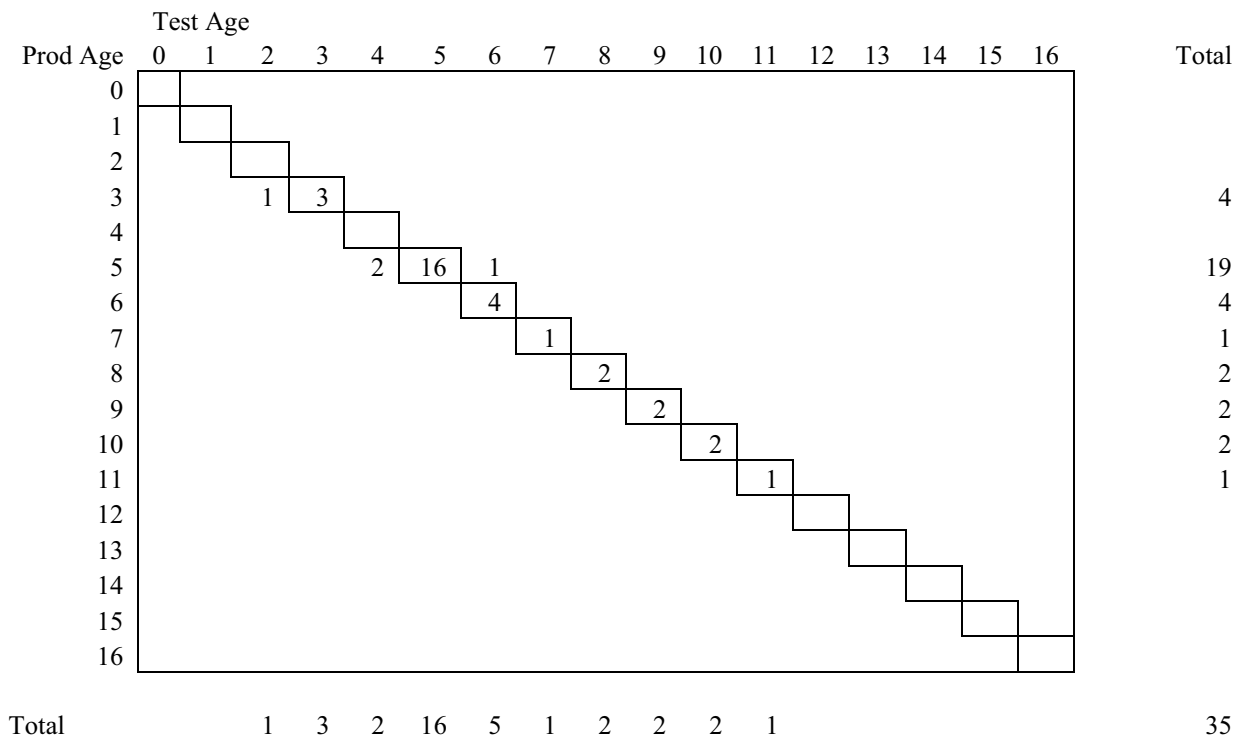
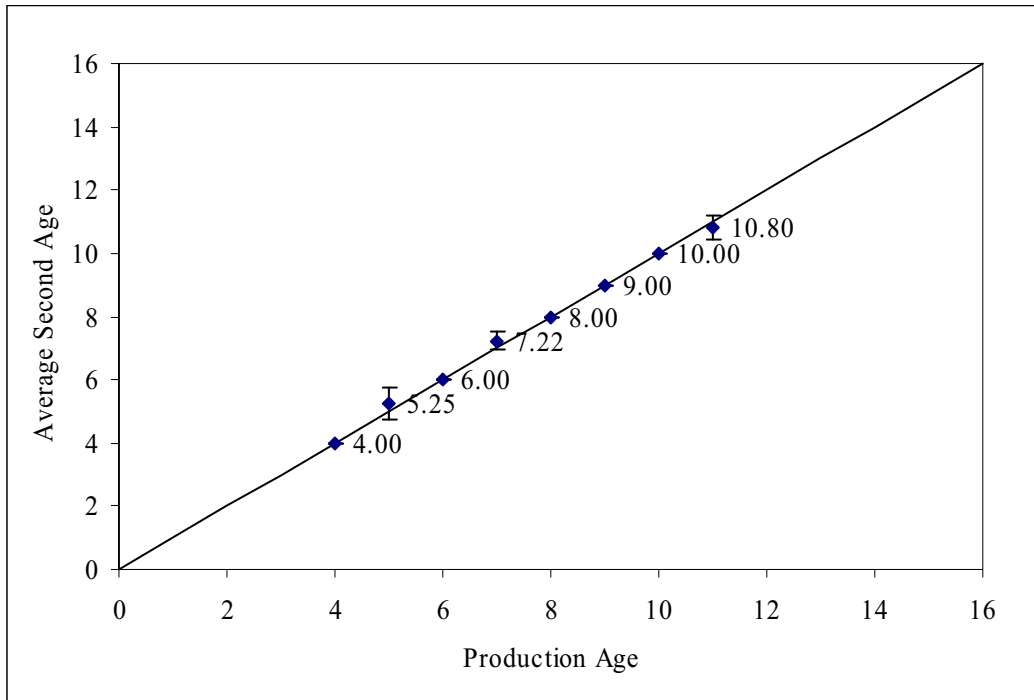


Figure A11. Results of haddock age-reader precision exercise against randomly selected samples from Quarter 1 of 2004 U.S. commercial landings (both stock areas). Error bars indicate 95% confidence intervals.



N Aged	56
N Agreed	52
Disagreed	4

Total CV	0.69%
%Agreement	92.9%

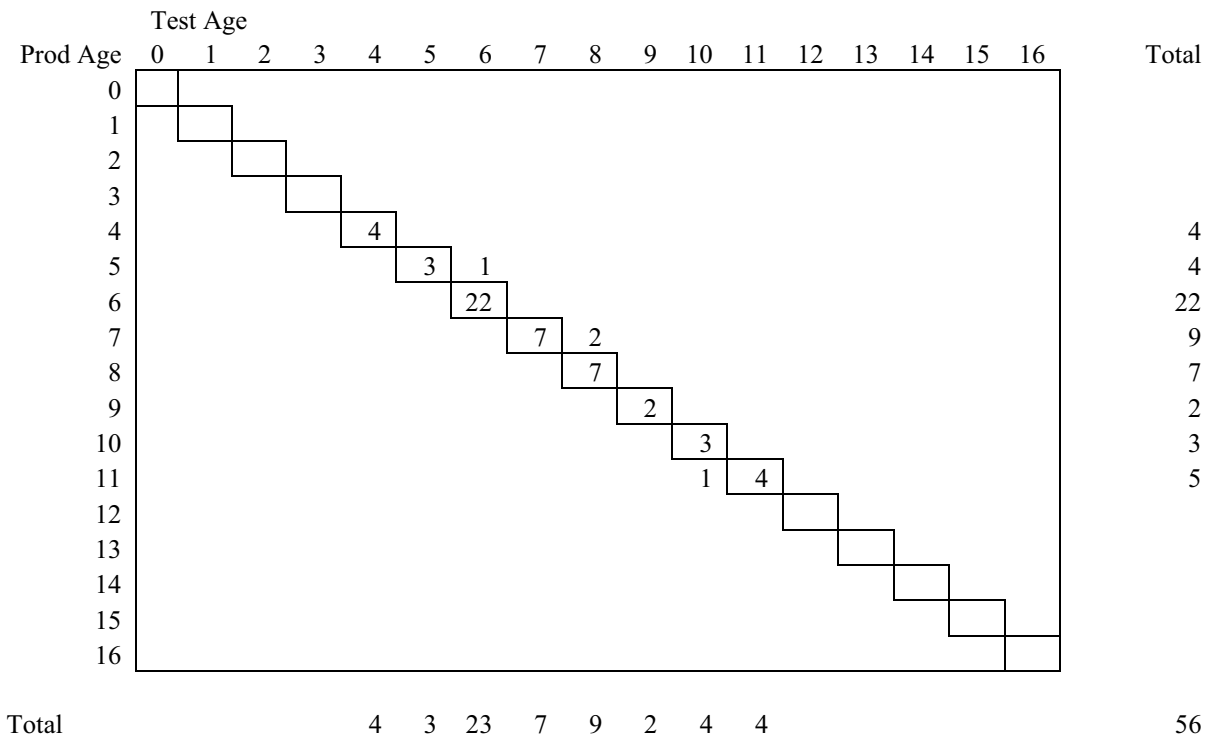
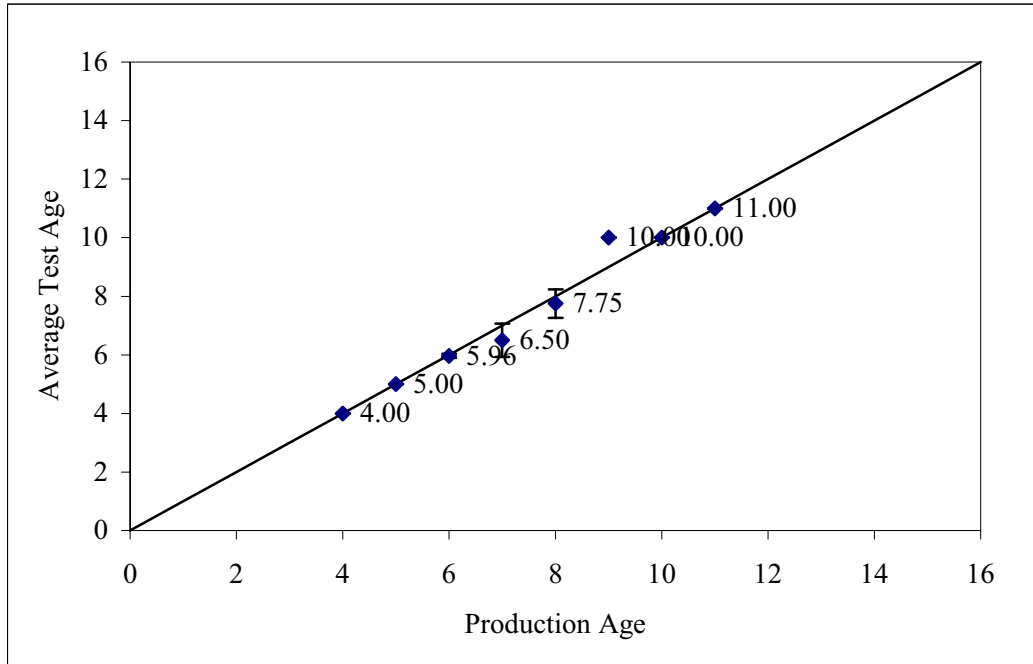


Figure A12. Results of haddock age-reader precision exercise against randomly selected samples from Quarter 2 of 2004 U.S. commercial landings (both stock areas). Error bars indicate 95% confidence intervals.



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

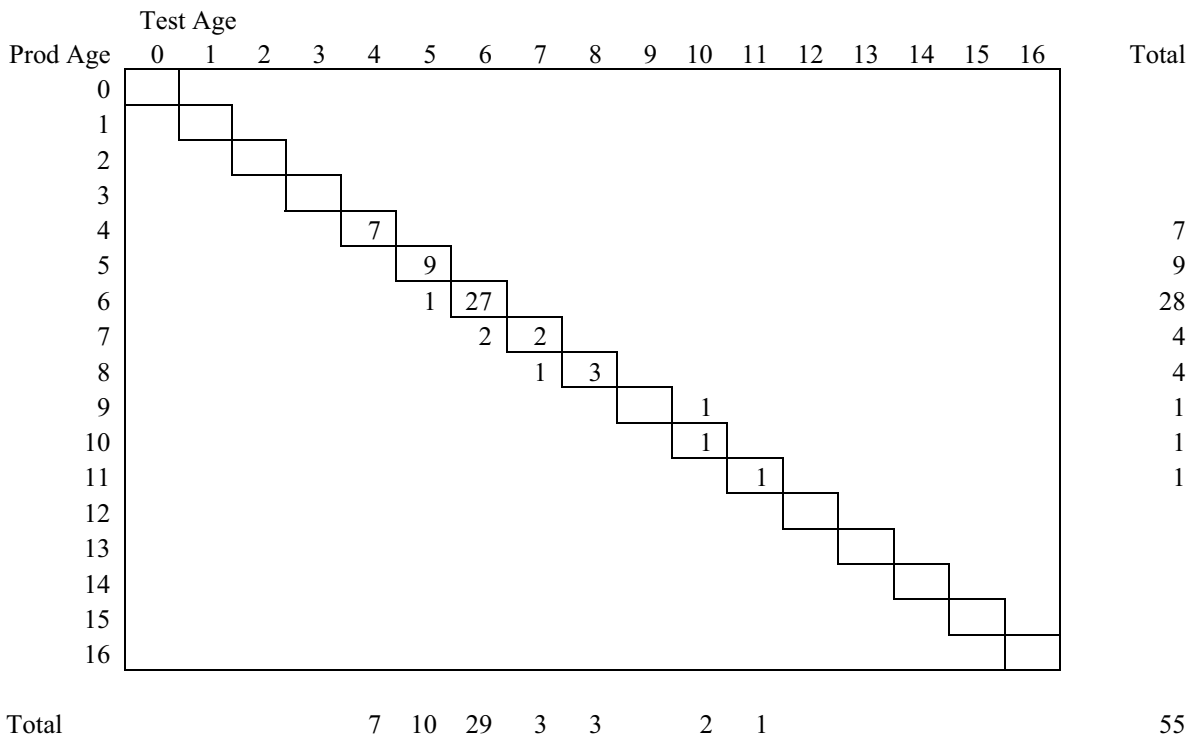
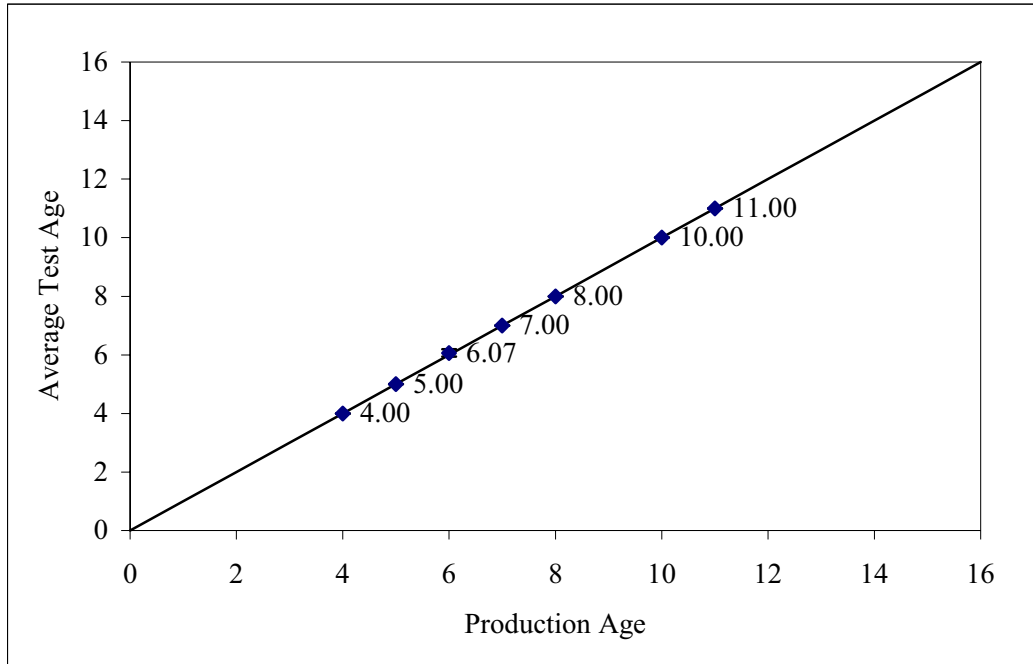


Figure A13. Results of haddock age-reader precision exercise against randomly selected samples from Quarter 3 of 2004 U.S. commercial landings (both stock areas). Error bars indicate 95% confidence intervals.



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

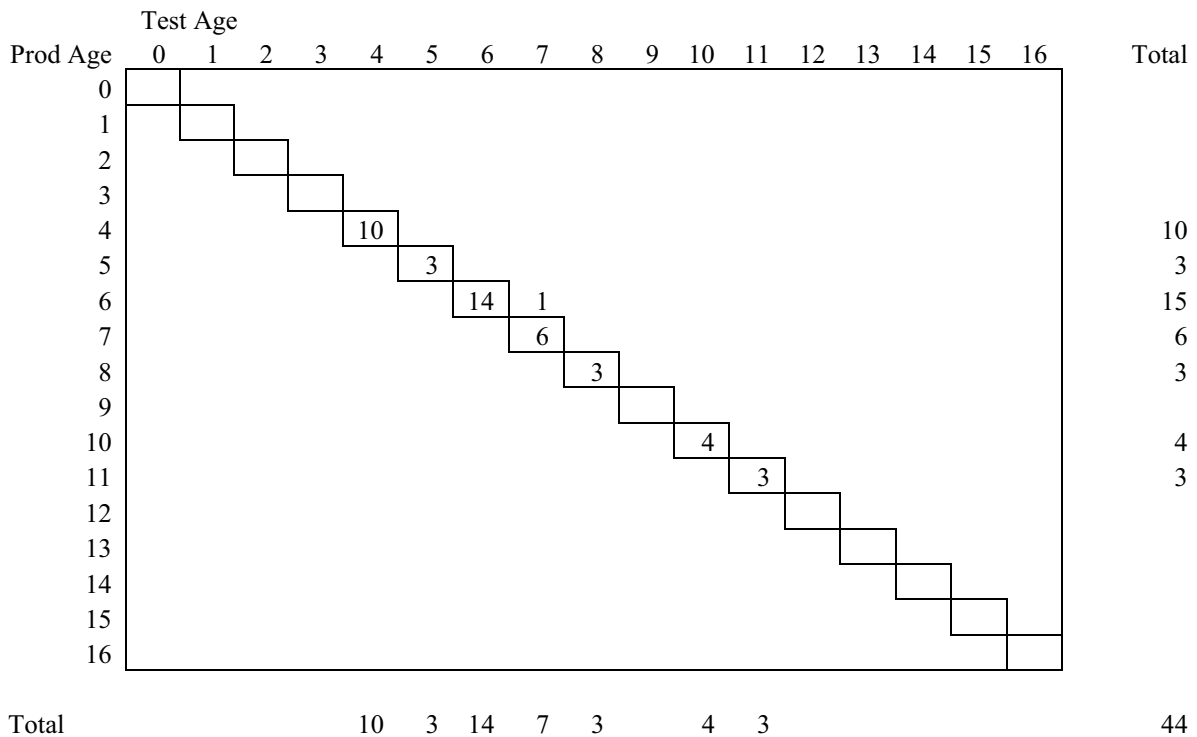
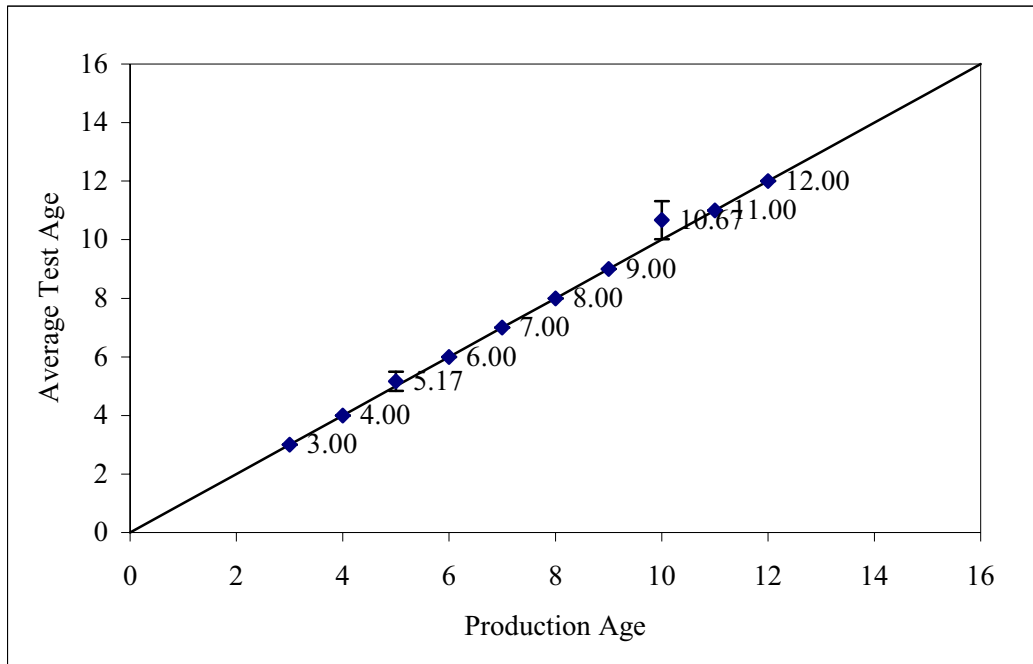


Figure A14. Results of haddock age-reader precision exercise against randomly selected samples from Quarter 4 of 2004 U.S. commercial landings (both stock areas). Error bars indicate 95% confidence intervals.



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

