EAST Validation Sample Studies
Description Results provided by Prof. Ritter C1, C2 and Combined Data Set
(redacted to remove library names)
==
Cohort 1 - From report dated August 12, 2016

Description Results: Table 1 provides summary results on rates of missing for the 40 study libraries . Table 1 results, which are calculated based on equal weights for all participating libraries, show a mean estimated rate of missing of 3.0% accurate to within 1.0% with 99% likelihood. The table also shows that the distribution of estimated rates of missing has a 95th percentile of 7.4% (indicating that all but two libraries had missing rates of 7.4% or less), a 75th percentile of 4.2% (indicating that three quarters, or 30 libraries, had a missing rate of 4.2% or less), and a median of 2.8% (indicate that half of the libraries, or 20 of them, had estimated missing rates of 2.8% or less). In addition, all participating libraries had estimated missing rates under 10.0%, which means there is 99% likelihood that all estimates are accurate to within 1.0%.

Table 1: Distribution in Rates of Missing Monographs among 40 EAST Libraries

	mean	std[1]	min	5th pctl	25th pctl	50th pctl	75th pctl	95th pctl	95th pctl
Rate	3.0%	2.1%	.3%	.6%	1.3%	2.8%	4.2%	7.4%	9.7%

[1] The standard deviation (std) gives some idea of the spread on the values. Roughly 80% of all library rates should be with one std of the mean, and roughly 95% of all rates should be within two std of the mean.

Table 2 provides summary results on the distributions for the conditions of the monographs inspected at the 40 study libraries (note: Appendix 2 provides the individual condition rates for all 40 EAST libraries). Table 2, again based on equal weights, indicates that the mean poor condition rate among titles was 10.7%, the mean acceptable condition rate was 55.7%, and the mean rate for excellent condition was 33.6%. All three estimates were accurate to within .27% with 99% likelihood. Table 2 also provides the distributions for various percentile values, including that the 5th and 95th percentiles for poor condition titles were 1.3% and 29.3% respectively (indicating that all but four libraries had between 1.3% and 29.3% of their titles in poor condition), and 5th and 95th percentiles for excellent condition were 7.8% and 70.5% (indicating that all but four libraries had between 7.8% and 70.5% of their titles in excellent condition).

Table 2: Distribution in Conditions of Monographs among 40 EAST Libraries

	mean	std	min	5th pctl	25th pctl	50th pctl	75th pctl	95th pctl	max
Poor	10.7%	9.8%	1.1%	1.3%	3.0%	8.2%	15.2%	29.3%	44.2%

Acceptable	55.7%	17.3%	34.8%	34.8%	45.5%	54.1%	65.9%	88.1%	91.1%
Excellent	36.6%	17.7%	6.8%	7.8%	20.7%	33.6%	43.3%	70.5%	78.8%

Analyses of Factors Affecting Condition of Items and Likelihood of Being Missing: Study data were also analyzed to identify factors which affect the likelihood of monographs being missing or in poor condition (or alternatively, being in excellent condition). Table 3 provides results of a multivariate logistic regression on being missing using as independent variables the factors of item age, frequency of circulation, number of duplicates among other EAST libraries, and subject area (as represented by a two-character call number). Individual indicators for each EAST library were also included in the model specification.

Results of these regression models show that after controlling for library, significant predictors for an item being missing included age, frequency of circulation, level of duplication across EAST libraries, and having call numbers denoting Mathematics (call number 'QA') or US Law (call number 'KF'). In particular, US law monographs were 1.7% more likely to be missing (for example, a 10% likelihood would become a 10.17% likelihood), mathematical monographs were .3% more likely to be missing, and each increment of twenty additional checkouts of a monograph related to a 1.6% relative decrease in likelihood. Smaller effects include a .17% decrease in likelihood for every five additional copies of the item among EAST libraries, and a .07% decrease in likelihood for every ten years increase in the age of the item. While all these predictors represent statistically significant factors, none would appear to be large enough to require any special action or attention.

Table 3: Estimated Effects on Likelihood of Being Missing

Variable	Estimate	t-value	p-value
Age of monograph	.07%	4.73	<.0001
Frequency of Circulation	1.6%	8.64	<.0001
Number of duplicates within EAST	-.17%	-7.05	<.0001
Call number for US Law ('KF')	1.7%	4.41	<.0001
Call number for Mathematics ('QA')	.56%	2.06	.0396

[1] estimate of effect for 10 year increase in age of item
[2] estimate of effect for increase of 20 in frequency of use
[3] estimate of effect for increase in 5 in number of duplicate items in EAST libraries

Logistic models with the same specification were also used to identify factors predicting poor condition (and then alternatively, excellent condition) among the library monographs. As provided in Table 4, results of these models show that many of the same factors significant in predicting missing items were also significant in predicting their condition (at least among examined items). For example, monographs on Paintings and African History (call numbers ND and DT respectively) were 3.2% and 1.9% more likely to be in poor condition. More importantly, each increment of twenty additional checkouts of a monograph and each increment of ten years in the age of the item increased the likelihood of being in poor condition by 5.1% and 3.0% respectively. Since some items have recorded uses in the hundreds and some items are greater than 50 years old, both of these factors could predict much higher likelihoods of poor condition.

Table 4: Estimated Effects Predicting Condition of Monographs

Effects Predicting Poor Condition in Monographs			
Variable	Estimate	t-value	p-value
Age of monograph[1]	3.0%	124.1	<.0001
Frequency of Circulation[2]	5.1%	14.61	<.0001
Number of duplicates within EAST[3]	-.39%	-9.32	<.0001
Call number for African History ('DT')	1.9%	2.72	.007
Call number for Paintings ('ND')	3.2%	5.92	<.0001
Effects Predicting Excellent Condition in Monographs			
Variable	Estimate	t-value	p-value
Age of monograph [1]	-4.9%	-135.0	<.0001
Frequency of Circulation [2]	-12.1%	-23.3	<.0001
Number of duplicates within EAST [3]	-.76%	-12.2	<.0001
Call number for Architecture ('NA')	-4.89	-4.70	<.0001
Call number for Geology ('QE')	-.4.89	-3.60	.0003

[1] estimate of effect for 10 year increase in age of item
[2] estimate of effect for increase of 20 times in frequency of use
[3] estimate of effect for increase in 5 in number of duplicate items in EAST libraries

For the sake of comparison, Table 4 also provides results of predicting excellent condition in monographs. As might be expected, the model found that newer items and items with lower recorded uses were more likely to be in excellent condition. Each ten year decrease in the age and each twenty fewer checkouts of an item increased the likelihoods for being in excellent condition by 4.8% and 12.2% respectively.

===

Cohort 2 - <u>From report dated January 11, 2018</u>

Description Results: Table 1 provides summary results on rates of missing for the 12 additional EAST libraries (note: Appendix 1 provides the individual rates of missing among these 12 libraries). Table 1 results, which are calculated based on equal weights for all participating libraries, show a mean estimated rate of missing of 2.18%, accurate to within 1.0% with 99% likelihood. The table also shows that the distribution of estimated rates of missing has a 90th percentile of 4.40% (indicating that only two libraries had missing rates of 4.40% or more), a 75th percentile of 2.89% (indicating that three quarters, or 9 libraries, had a missing rate of 2.89% or less), and a median of 1.85% (indicate that half of the libraries, or 6 of them, had estimated missing rates of 1.85% or less). In addition, all participating libraries had estimated missing rates under 5.42%, which means there is 99% likelihood that all estimates are accurate to within 1.0%.

Table 1: Distribution in Rates of Missing Monographs among 12 EAST Libraries

	mean	std[1]	5th pctl	10th pctl	25th pctl	50th pctl	75th pctl	90th pctl	95th pctl
Rate	2.18%	1.55%	.25%	.77%	.86%	1.85%	2.89%	4.40%	5.42%

[1] The standard deviation (std) gives some idea of the spread on the values. Roughly 80% of all library rates should be with one std of the mean, and roughly 95% of all rates should be within two std of the mean.

Table 2 provides summary results on the distributions for the conditions of the monographs at the 12 study libraries (note: Appendix 2 provides the individual condition rates for these 12 EAST libraries). Table 2, again based on equal weights, indicates that the mean poor condition rate among titles was 7.1% (note: the mean acceptable condition rate among the libraries was 46.1%, and the mean rate for excellent condition was 41.3%). Estimates of poor condition monographs at all libraries were accurate to within .3% with 99% likelihood. Table 2 also provides the distributions for various percentile values, including that the 10th and 90th percentiles for poor condition titles were .8% and 14.8% respectively (indicating that all but four libraries had between .8% and 14.8% of their titles in poor condition). It can be noted that the variation in poor condition monographs among the 12 libraries is much greater than the variation in missing. This is due in part because the mean rate of poor condition is higher than the mean rate of missing, but it is likely also due to the fact that being in poor condition is a judgement call and there could be significant variation among reviewers in what constituted poor condition.

Table 2: Distribution of Monographs in Poor Condition among additional 12 EAST Libraries

	mean	std	5th pctl	10th pctl	25th pctl	50th pctl	75th pctl	90th pctl	95th pctl
Poor	7.1%	5.9%	.68%	.83%	2.20%	5.55%	11.2%	14.8%	18.8%

Analyses of Factors Affecting Likelihood of Being Missing or Being in Poor Condition: Study data were also analyzed to identify factors which affect the likelihood of monographs being missing or in poor condition. Table 3 provides results of a multivariate logistic regression on being missing using as independent variables the factors of item age, frequency of circulation, number of titles of the monograph in the US, and subject area (as represented by a two-character call number). Individual indicators for each EAST library were also included in the model specification. Results of these regression models show that after controlling for library, significant predictors for an item being missing included age, frequency of circulation, and having call numbers denoting Religion (call number 'BL') or US Law (call number 'KF'). In particular, US Law monographs were 53% more likely to be missing (for example, a 2% likelihood would become a 3% likelihood) and Philosophy and Religion monographs were 44% more likely to be missing. In addition, each increment of twenty additional checkouts of a monograph related to a .9% relative decrease in likelihood of being missing and every ten years increase in the age of the item associates with .6% increase in likelihood. Since some items have recorded uses in the hundreds and some items are greater than 50 years old, both of these factors could predict much different likelihoods of being missing than the overall mean.

Table 3: Estimated Effects on Likelihood of Being Missing

Variable	Estimate	t-value	p-value
Age of monograph (per 10 years)	.6%	6.72	<.0001
Frequency of Circulation (per 20 times)	-.9%	2.37	.0179
Call number for US Law ('KF')	53%	9.93	<.0001
Call number for Religion ('BL')	44%	2.47	.0137

Logistic models with the same specification were also used to identify factors predicting poor condition among the library monographs. As provided in Table 4, results of these models show many factors are significant in predicting their condition (at least among examined items). For example, each increment of twenty additional checkouts of a monograph, each increment of ten years in the age of the item and each increment of 5 in the number of US holdings increased the likelihood of being in poor condition by 4.5%, 3.6%, and .13% respectively. In addition, certain call numbers are indicative of monographs with much higher likelihoods of being in poor condition. For example, monographs in the areas of Psychology, Asian History, Economic History, Family and Marriage, Welfare and Criminology, Theory and Practice of Education, Painting, French and Spanish Literature, and English Literature all have increased likelihood of being in poor condition.

Table 4: Estimated Effects Predicting Condition of Monographs

Effects Predicting Poor Condition in Monographs			
Variable	Estimate	t-value	p-value
Age of monograph[1]	3.6%	65.8	<.0001
Frequency of Circulation[2]	4.5%	23.1	<.0001
Number of US holdings[3]	.13%	4.64	<.0001
Call number for Psychology ('BF')	33%	2.78	.0055
Call number for Asian History ('DS')	26%	2.74	<.0062

Call number for Economic History ('HC')	30%	2.42	.0156
Call number for Family and Marriage ('HQ')	48%	4.43	<.0001
Call number for Welfare and Criminology ('HV')	79%	7.22	<.0001
Call number for Theory and Practice of Education ('LB')	39%	2.92	.0035
Call number for Paintings ('ND')	41%	3.21	.0013
Call number for French and Spanish Literature ('PQ')	18%	2.39	.0166
Call number for English Literature ('PS')	16%	2.40	.0163

[1] estimate of effect for 10 year increase in age of item
[2] estimate of effect for increase of 20 times in frequency of use
[3] estimate of effect for increase in 5 in number of US holdings

===

Cominded C1 & C2 - From report dated June 9 2018

Description Results: Table 1 provides percentile values on the rates of missing among library cohorts and then for the 52 EAST libraries as a single group (note: Appendix 1 provides the individual rates of missing among the 52 library samples). Table 1 results, which are calculated based on equal weights for all participating libraries, show mean estimated rates of missing equal to 3.0% for cohort 1 libraries, 2.18% for cohort 2 libraries, and 2.79% for the combined group. All estimates are accurate to within 1.0% with 99% likelihood. The table also shows that the distribution of estimated rates of missing has a 90th percentile value of 7.4% for cohort 1 libraries, 4.4% for cohort 2 libraries, and 5.22% across all 52 EAST libraries (indicating that only five libraries had missing rates of 5.22% or more). Similarly, the 75th percentile rate of missing was 4.2% for cohort 1 libraries, 2.9% for cohort 2 libraries, and 3.6% for the combined group (indicating that three quarters, or 39 libraries, had a rate of missing of 3.6% or less). Notably, all participating libraries had estimated rates of missing under 10.0%, indicating strong likelihood that all estimated rates of missing were accurate to within 1.0%.

Table 1: Rates of Missing Monographs – by Cohort and Overall Among all 52 EAST Libraries

	mean	std[1]	5th pctl	10th pctl	25th pctl	50th pctl	75th pctl	90th pctl	95th pctl
Cohort 1	3.0%	2.1%	.3%	.6%	1.3%	2.8%	4.2%	7.4%	9.7%
Cohort 2	2.18%	1.55%	.25%	.77%	.86%	1.85%	2.89%	4.40%	5.42%
Combined	2.79%	2.03%	.3%	.63%	1.18%	2.44%	3.60%	5.22%	6.47%

[1] The standard deviation (std) gives some idea of the spread on the values. Roughly 80% of all library rates should be with one std of the mean, and roughly 95% of all rates should be within two std of the mean.

Table 2 provides summary results on the distributions for the conditions of the monographs at the two library cohorts and then as a combined group (note: Appendix 2 provides the individual condition proportions for all 52 EAST libraries). Table 2, again based on equal weights, indicates that the mean poor condition proportion was 10.4% among cohort 1 volumes, 7.1% among cohort 2 volumes, and 9.94% for the combined group. Similarly, the mean acceptable condition proportion was 55.7% among cohort 1 volumes, 46.9% among cohort 2 volumes, and 54.1% for the combined group, while the mean

excellent condition proportion was 36.6% among cohort 1 volumes, 42.2% among cohort 2 volumes, and 36.0% for the combined group. Estimates of poor condition monographs at all libraries were accurate to within 1% with 99% likelihood. Table 2 also provides the distributions for various percentile values, including that the 10th and 90th percentiles for poor condition titles were 1.09% and 21.5% respectively (indicating that all but 11 libraries had between 1.09% and 21.5% of their titles in poor condition). It can be noted that the variation in poor condition monographs among the 52 libraries is much greater than the variation in missing. This is due in part because the mean rate of poor condition is higher than the mean rate of missing, but it is likely also due to the fact that judging a monograph to be in poor condition is subjective and there could be significant variation among reviewers in what constituted poor condition.

Table 2: Distribution of Condition of Monographs - by Cohort and Overall Among all 52 EAST Libraries

	mean	std	5th pctl	10th pctl	25th pctl	50th pctl	75th pctl	90th pctl	95th pctl
Cohort 1									
Poor	10.7%	9.8%	1.31%	1.54%	3.04%	8.19%	15.2%	22.8%	29.3%
Acceptable	55.7%	17.3%	28.2%	34.8%	45.5%	54.1%	65.9%	84.1%	91.0%
Excellent	33.6%	17.7%	7.84%	12.7%	20.7%	33.6%	43.3%	57.1%	70.5%
Cohort 2									
Poor	7.1%	5.9%	.68%	.83%	2.20%	5.55%	11.2%	14.8%	18.8%
Acceptable	46.9%	18.4%	3.2%	38.2%	42.5%	45.6%	53.7%	61.2%	84.1%
Excellent	42.2%	20.0%	7.9%	30.8%	33.2%	41.0%	47.0%	50.9%	94.6%
Overall									
Poor	9.94%	9.09%	1.09%	1.54%	2.89%	7.96%	14.5%	21.5%	26.6%
Acceptable	54.1%	17.9%	27.4%	35.2%	45.4%	52.2%	60.4%	82.4%	89.9%
Excellent	36.0%	18.6%	7.88%	13.7%	23.0%	34.8%	45.8%	52.7%	71.0%

Factors Affecting Likelihood of Being Missing: The study also attempted to determine factors which predicted the likelihood of monographs being missing. Table 3 provides results of such multivariate logistic regressions using variables for age, frequency of circulation, number of titles of the monograph in the US, and subject area (as represented by a two-character call number). Individual indicators for each EAST library were also included in the model specification. These models were conducted using the cohort 2 sample only (n=72,000) and then using the combined sample (n=312,000 volumes). Results of the two regression models were very similar and showed that after controlling for library, the only consistently significant predictors for an item being missing were the age of the monograph and having its subject matter classified as Religion ('BL') or US Law ('KF'). In particular, US Law monographs had an odds ratio of approximately 4.5 for being missing (for example, a 2% likelihood would become a 9% likelihood) and Philosophy and Religion monographs demonstrated an odds ratio of approximately 1.8 for being missing. In addition, every 10 year increase in the age of an item associates with a 4% to 5% increase in its likelihood of being missing. Since some monographs are greater than 50 years old, this

factor could predict a much higher likelihood (e.g., 20%-25% higher) compared with fairly new items. However, both models were consistent in noting the strongest characteristic increasing the likelihood of a monograph being missing was the library itself.

Table 3: Factors Predicting Likelihood of Being Missing

Variable	Cohort 2 Only			Combined Sample (1 and 2)		
	Odds Ratio	t-value	p-value	Odds Ratio	t-value	p-value
Age of monograph (per year)[1]	1.05	6.12	<.0001	1.04	6.07	<.0001
Call number for Religion ('BL')	1.82	2.46	.0137	1.83	2.35	.0185
Call number for US Law ('KF')	4.43	9.74	<.0001	4.65	9.96	<.0001

[1] estimate of effect per 10 year increase in age of item

Factors Affecting Likelihood of Being in Poor Condition: Similar logistic models were used to determine factors which predict monographs in poor condition. Table 4 provides the results of these models, first based on Cohort 2 only and then using the combined sample. These logistic results are somewhat different from missing, in that they have a much larger number of significant predictors. Based on the Cohort 2 sample, each ten year increase in the age of the item, each increment of twenty additional checkouts of a monograph, and each increment of 5 in the number of US holdings leads to an odds ratio of being in poor condition of 1.036, 1.05, and 1.038 respectively. In addition, a larger set of call numbers are indicative of monographs being in poor condition. For example, monographs in the areas of Psychology (OR=1.48), Asian History (OR=1.38), Economic History (OR=1.53), Family and Marriage (OR=1.70), Welfare and Criminology (OR=2.31), Theory and Practice of Education (OR=1.63), Painting (1.61), and French and Spanish Literature (1.28) all are more likely to have monographs in poor condition. Almost all of these factors were also significantly in the logistic models using the combined sample of Cohorts 1 and 2 and had similar odds ratios. Only two factors, number of US holdings and being concerned with French or Spanish literature (I.e., call number 'PQ'), did not retain their significance with the combined sample. At the same time, only one factor, being concerned with English literature (i.e., call number 'PR') gained significance in the larger combined sample.

Just as with the outcome of being missing, library indicators were among the strongest predictors for monographs being in poor condition. ... As noted in the first EAST study, one caveat to all of the analyses of monograph conditions is that the assessments at different libraries were done by different reviewers. Although reviewers were given training, it is likely they employed somewhat different standards for determining monograph condition.

Table 4: Estimated Effects Predicting Monographs in Poor Condition

Variable	Cohort 2 Only			Combined Sample		
	Odds Ratio	t-value	p-value	Odds Ratio	t-value	p-value
Age of monograph[1]	1.036	65.8	<.0001	1.035	65.8	<.0001
Frequency of Circulation[2]	1.05	23.1	<.0001	1.038	33.0	<.0001
Number of US holdings[3]	1.038	4.64	<.0001	NS	-	-
Call number for Psychology ('BF')	1.48	2.61	.009	1.60	2.98	.003
Call number for Asian History ('DS')	1.38	2.50	.013	1.43	2.48	.013

Call number for Economic History ('HC')	1.53	2.78	.005	1.53	2.51	.012
Call number for Family and Marriage ('HQ')	1.70	4.16	<.0001	1.85	4.65	<.0001
Call number for Welfare and Criminology ('HV')	2.31	6.90	<.0001	2.48	7.29	<.0001
Call number for Theory and Practice of Education ('LB')	1.63	3.05	.002	1.65	2.90	.004
Call number for Paintings ('ND')	1.61	3.07	.002	1.68	3.12	.002
Call number for French and Spanish Literature ('PQ')	1.28	2.14	.032	NS	-	-
Call number for English Literature ('PS')	NS	-	-	1.36	2.81	.005

[1] estimate of effect per 10 year increase in age of item
[2] estimate of effect for increase in frequency of use
[3] estimate of effect for increase in US holdings by 100