TECHNICAL NOTE: IDENTIFYING HISTORIC EEO-1 STANDARD INDUSTRIAL CLASSIFICATION (SIC) VARIABLES
Prepared by EEOC ORIP

PROBLEM EXAMINED:

The historic EEO-1 files contain multiple Standard Industrial Classification (SIC) variables (primarily SICX and SIC). They also represent different versions of the official SIC coding categories (signified by the SIC Manuals for 1972 and 1987). We examined selected SIC frequencies from the historic EEO-1 statistics files to determine:

(a) The SIC variables that were used to create the annual SIC Job Patterns tables. Presumably the SIC variable matching the published Job Patterns counts for a given year is the most accurate and reliable SIC variable for that year.

(a) The SIC manual used to classify establishments in each year. To ascertain the appropriate version of the Standard Industrial Classification manual, we identified SIC categories present in the 1972 manual but not the 1987 manual; and SIC categories present in the 1987 manual but not the 1972 manual. Presumably EEO-1 files containing the former SIC values relied on the 1972 manual, and EEO-1 files containing the latter SIC values relied on the 1987 manual. We selected SIC categories with at least 100 EEO-1 observations to minimize the effects of growing and declining industries (i.e., industries which confound fluctuations in the actual number of respondents with potential changes in the SIC coding categories).

We concluded that SAS variable SICX should be used for years 1991 and earlier and that SAS variable SIC should be used for years 1992 and later. SAS variable SICX appears to represent pre-1987 SIC coding categories. SAS variable SIC appears to represent the 1987 SIC coding categories. There is some evidence to suggest that the published data in Job Patterns for 1991 contains both omissions and errors in the reported counts for selected industries.

It should be noted that the standardized EEO-1 SAS files from 1990 forward do not contain SICX values so no adjustment is necessary. Researchers unsure of whether or not they are using a standardized data set should check for a variable named "YEAR." The presence of this variable indicates that the data set has been standardized (see the last paragraph of this memorandum for more details).

MISSING VALUES:

The EEO-1 SIC variables have at least two kinds of missing values, zero and 999. A shift in the

presence or absence of missing values provides clues to changes in the SIC variables.

SIC Value Zero: In 1991, variable SICX had no zero values, and variable SIC had 139,410 zero values. In 1992, variable SICX had 139,678 zero values, and variable SIC had no zero values. This suggests a shift in SIC variables between 1991 and 1992 such that industry was coded using the SICX variable in 1991 and in the SIC variable in 1992. Any SIC values appearing in the SIC variable before 1992 are probably pre-test values, but this has not been investigated.

SIC Value 999: Variable SIC has 999 values from 1989 through 1991 but no 999 values in 1992. Conversely, the SICX variable has no SIC 999 values from 1989 through 1991, but 1,520 SIC 999 values in 1992. This pattern reinforces the impression that the SIC variables switched from SICX to SIC after 1991.

SIC CATEGORIES AVAILABLE IN BOTH VERSIONS OF THE SIC MANUAL (1972 AND 1987):

SIC 201: The SIC category 201 is present in both the 1972 and the 1987 SIC Manuals. We compared the frequencies in the EEO-1 data for SIC value 201 to the published frequencies for SIC value 201 in Job Patterns for Minorities and Women in Private Industry. The Job Patterns information matches the SIC variable in 1992 and the SICX variable in 1991. There were 731 establishments present in the 1992 SIC variable and 731 establishments reported in the 1992 Job Patterns. Likewise, there were 729 establishments present in the 1991 SICX variable and 729 establishments reported in the 1991 Job Patterns. Thus, the switch from variable SICX to variable SIC in 1992 is accurately reflected in Job Patterns for 1991 and 1992.

SIC CATEGORIES NOT FOUND IN THE 1987 SIC MANUAL

The SIC codes 264, 307, and 739 could not be found in the 1987 SIC manual but are listed in the 1972 manual. Presumably variable SICX represents the 1972 classification scheme, and variable SIC represents the 1987 classification scheme. Since SIC codes 264, 307, and 739 are non-existent in the 1987 classification scheme, the switch to variable SIC in 1992 should result in a comparable shift in published SIC codes - - they should be present in Job Patterns before 1992 but absent after 1991. The attached spreadsheet confirms this pattern for all the years except 1991. Job Patterns contains frequency counts for SIC codes 264, 307, and 739 in 1989 and 1990 but no entries for SIC codes 264, 307, and 739 in 1991 and 1992. The puzzle is 1991. The SICX variable shows a substantial number of observations for these industries in 1991 ranging from 595 establishments for SIC 264 to 3,413 establishments for SIC 739. Since the

SICX variable appears to reflect the 1972 classification scheme, Job Patterns in 1991 should have reported data for industries 264, 307, and 739 based on variable SICX. The most likely explanation for omitting these industries is that a decision was made to exclude some or all of the industries dropped from the 1972 classification scheme in compiling the 1991 Job Patterns even though the data was available in 1991 EEO-1 statistics file. It is also should be noticed that the SIC variable has 178 observations for industry 264 in 1989 and 188 observations for industry 264 in 1990. There is no apparent reason why industry 264 from the 1972 Manual should appear under variable SIC in 1989 and 1990 (since, presumably, variable SIC is based on the 1987 Manual). The remaining pre-1987 industries, 307 and 739, have zero values in variable SIC.

SIC CATEGORIES NOT FOUND IN THE 1972 SIC MANUAL:

The SIC codes 267 (Converted Paper and Paperboard Products, Except Containers), and 308 (Miscellaneous Plastics) could not be found in the 1972 SIC Manual, but were listed in the 1987 SIC Manual. Since SIC codes 267 and 308 are non-existent in the 1972 classification scheme, the switch to variable SIC in 1992 should result in a comparable shift in published SIC codes - - they should be absent in Job Patterns before 1992 but present after 1991. The attached spreadsheet confirms this pattern for all the years except 1991. Job Patterns contains no entries for SIC codes 267 and 308 in 1989 and 1990, but it does have frequencies counts for SIC codes 267 and 308 in 1991 and 1992. The puzzle again is 1991. To be consistent, these industries should be reported for 1992 but not 1991. Yet the 1991 Job Patterns does have entries for SIC codes 267 and 308 based on variable SICX. Notice that published counts for these industries are substantially lower in 1991 than in 1992 - - the number of SIC 267 establishments increases from 75 establishments in 1991 to 708 establishments in 1992; the number of SIC 308 establishments increases from 139 establishments in 1991 to 1,539 establishments in 1992. This suggests that the 1991 variable SICX mistakenly includes, perhaps as a pretest, some of the observations that were latter transferred to the 1992 variable SIC. Whatever the explanation, it appears that industry table in the 1991 Job Patterns should be treated with caution since it contains implausible entries for several industries that should have been omitted from the report.

STANDARDIZED FILES STARTING IN 1990

The SAS EEO-1 files for 1990 through the present contain a standardized set of SAS variable names (i.e., a given variable has the same SAS variable name from 1990 through the present). Therefore, starting in 1990, there is a single group of SIC variables named SIC2 (two-digit values), SIC3 (three-digit values) and SIC4 (four-digit values) with the SAS label "GE92 v87/LT 92 V72." This SAS label is a shorthand for saying that the values for variables SIC2-SIC4 use the1987 SIC coding categories from 1992 to the present and the pre-1987 SIC coding categories

before 1992. Therefore, although the variable names remain the same, the meaning of the SIC values in the standardized files changes from the pre-1987 industry categories in 1991 to the 1987 industry categories in 1992. In particular, the standardized values SIC2-SIC4 for 1990 and 1991 are based on variable SICX in the historic EEO-1 files. We have not researched the SIC variables before 1989 and do not know under what circumstances researchers may discover multiple versions of the industry variable (such as SIC and SICX) in the historic EEO-1 files. We also do not know what version of the industry categories were used before 1989, although it is probably safe to assume that most, if not all, of the industry values in the period 1980 through 1988 are based on the 1972 SIC coding categories as modified in 1977. However, researchers should be cautious in interpreting the historic EEO-1 files because the industry codes may have been implemented with leads and lags (as well as errors) in the years surrounding the publication of SIC revisions in 1972, 1977, and 1987. We welcome comments from researchers and encourage you to send us any comments or data that could be used to refine this memorandum as we learn more about the historic EEO-1 files.

Resource Materials:

Attached table: Comparative Frequencies for Historical EEO-1 SIC Values

Last Revised: November 3, 2003

COMPARATIVE FREQUENCIES FOR HISTORICAL EEO-1 SIC VALUES

NOT FOUND IN 1987

1972	1989			1990			1991			1992		
SIC	JOB			JOB			JOB			JOB		
ONLY	PATT	SICX	SIC	PATT	SICX	SIC	PATT	SICX	SIC	PATT	SICX	SIC
264	621	621	178	664	664	188	0	595	2	0	563	0
307	1379	1380	0	1486	1487	0	0	1411	0	0	1262	0
739	14537	14619	0	7889	7930	0	0	3413	0	0	2968	0

NOT FOUND IN 1972

1987	1989			1990			1991			1992		
SIC	JOB			JOB			JOB			JOB		
ONLY	PATT	SICX	SIC	PATT	SICX	SIC	PATT	SICX	SIC	PATT	SICX	SIC
267	0	0	0	0	0	0	75	75	0	708	14	708
308	0	0	154	0	0	157	139	139	0	1569	20	1569

BOTH TIME PERIODS

1972	1989			1990			1991			1992		
AND	JOB			JOB			JOB			JOB		
1987	PATT	SICX	SIC	PATT	SICX	SIC	PATT	SICX	SIC	PATT	SICX	SIC
201	664	666	0	720	720	0	729	729	0	731	109	731

	1989			1990			1991			1992		
MISSING												
VALUES		SICX	SIC		SICX	SIC		SICX	SIC		SICX	SIC
0		0	0		0	0		0	139410		139678	0
999		0	31337		0	32194		0	1609		1520	0