Predicting Recidivism in North Carolina, 1978 and 1980

Peter Schmidt
Ann D. Witte

ICPSR 8987

June 2002

Predicting Recidivism in North Carolina, 1978 and 1980

Peter Schmidt
and Ann D. Witte

PREDICTING RECIDIVISM IN NORTH CAROLINA, 1978 AND 1980

(ICPSR 8987)

Principal Investigators

Peter Schmidt and Ann D. Witte

Acknowledgement of Assistance

All manuscripts utilizing data made available through the Consortium

should acknowledge that fact as well as identify the original collector

of the data. The ICPSR Council urges all users of the ICPSR Data

facilities to follow some adaptation of this statement with the

parentheses indicating items to be filled in appropriately or deleted

by the individual user.

In order to provide funding agencies with essential information

about the use of archival resources and to facilitate the exchange

of information about ICPSR participants' research activities, each

user of the ICPSR data facilities is expected to send two copies

of each completed manuscript or thesis abstract to the Consortium

Please indicate in the cover letter which data were used.

Peter Schmidt and Ann D. Witte
    PREDICTING RECIDIVISM IN NORTH CAROLINA, 1978 AND 1980 (ICPSR 8987)

SUMMARY: This data collection examines the relationship between individual characteristics and recidivism for two cohorts of inmates released from North Carolina prisons in 1978 and 1980. The survey contains questions on the background of the offenders, including their involvement in drugs or alcohol, level of schooling, nature of the crime resulting in the sample conviction, number of prior incarcerations and recidivism following release from the sample incarceration. The data collection also contains information on the length of time until recidivism occurs. CLASS IV

UNIVERSE: All individuals released from North Carolina prisons during the periods of July 1, 1977, through June 30, 1978, and July 1, 1979, through June 30, 1980.
NOTE: For 1978 data, 4,709 individual records were missing one or more variables and placed in a missing data file. The other 4,618 observations, which contained complete information, were randomly split into an estimation sample of 1,540 observations and a validation sample of 3,078. For 1980 data, 3,810 individual records were missing information on one or more variables and were placed in a missing data file. The other 5,739 observations, which contained complete information, were randomly split into an estimation sample of 1,435 observations and a validation sample of 4,304 observations.

EXTENT OF COLLECTION: 2 data files + machine-readable documentation
DATA FORMAT: Card Image

PART 1: 1978 Data   PART 2: 1980 Data
FILE STRUCTURE: rectangular   FILE STRUCTURE: rectangular
CASES: 9,327   CASES: 9,549
VARIABLES: 19   VARIABLES: 19
RECORD LENGTH: 80   RECORD LENGTH: 80
RECORDS PER CASE: 1   RECORDS PER CASE: 1

RELATED PUBLICATION:
    Schmidt, P., and A.D. Witte. PREDICTING RECIDIVISM USING SURVIVAL MODELS. New York: Springer-Verlag, 1988.

ABSTRACT

Peter Schmidt and Ann D. Witte

Predicting Recidivism: North Carolina, 1978 and 1980

Michigan State University

84-IJ-CX-0021

Purpose of the Study

This study examines individual characteristics and recidivism (measured as length of time until a released prisoner returns to prison) for two cohorts of North Carolina prison releasees.

Methodology

Sources of information:

The North Carolina Department of Corrrections provided data tapes which contained information on all individuals released from North Carolina prisons during the periods July 1, 1977 through June 30, 1978 and July 1, 1979 through June 30, 1980.

Sample:

1978: After deletions for obvious data defects, there were 9327 individual records on the tape. Of these, 4709 were missing information on one or more variables and these observations constitute a "missing data" file. The other 4618 observations, which contained complete information, were randomly split into an "estimation sample" of 1540 observations and a "validation sample" of 3078.

1980: After deletions for obvious data defects, there were 9549 individual records on the tape. Of these, 3810 were missing information on one or more variables and these observations constitute a "missing data" file. The other 5739 observations, which contained complete information, were randomly split into an "estimation sample" of 1435 observations and a "validation sample" of 4304 observations.

Dates of data collection:

April, 1984

Summary of Contents

Special characteristics of the study:

The data set is particularly useful for the application of survival models because it contains information on the length of time until recidivism occurs.

Description of variables:

Variables include the sex, race, age, and marital status of the inmate, involvement in drugs or alcohol, level of schooling, the nature of the crime which resulted in the "sample conviction," (e.g. felon vs. misdemeanor, against person vs. against property), participation in work release, number of rules broken during the "sample sentence," amount of time served in "sample sentence," number of prior incarcerations, the nature of the inmate's release (e.g. supervised), recidivism following release from the "sample incarceration," the length of time from release from the "sample incarceration" until return to prison in North Carolina, and the amount of time in the followup period (from release until North Carolina Department of Correction records were searched). A variable called FILE indicates to which data sample the individual record belongs--analysis sample, validation sample, or missing data sample.

Unit of observation:

Released inmates

Geographic Coverage

North Carolina

File Structure

Data files:	2 (1978 data and 1980 data)
Unit:	Released inmates
Variables:	19
Cases:	9,327 (1978)
	9,549 (1980)

Reports and Publications

Schmidt, P. & Witte, A. D. (1988). *Predicting Recidivism Using Survival Models.* New York: Springer-Verlag.

Predicting Recidivism

Peter Schmidt and Ann Witte,
principal investigators

1. General Description of the 1978 Data

This is a description of the data on the file, DATA1978. The description was prepared by Peter Schmidt, Department of Economics, Michigan State University, East Lansing, Michigan 48824. The data were gathered as part of a grant from the National Institute of Justice to Peter Schmidt and Ann Witte, "Improving Predictions of Recidivism by Use of Individual Characteristics," 84-IJ-CX-0021. A more complete description of the data, and of the uses to which they were put, can be found in the final report for this grant. Another similar dataset, contained in a file DATA1980 on a separate diskette, is also described in that report.

The North Carolina Department of Correction furnished a data tape which was to contain information on all individuals released from a North Carolina prison during the period from July 1, 1977 through June 30, 1978. There were 9457 individual records on this tape. However, 130 records were deleted because of obvious defects. In almost all cases, the reason for deletion is that the individual's date of release was in fact not during the time period which defined the data set. This left a total of 9327 individual records, and accordingly there are 9327 records on DATA1978.

The basic sample of 9327 observations contained many observations for which one or more of the variables used in our analyses were missing. Specifically, 4709 observations were missing information on one or more such variables, and these 4709 observations constitute the "missing data" file. The other 4618 observations which contained complete information were randomly split into an "analysis file" of 1540 observations and a "validation file" of 3078 observations.

2. Locations of Variables

DATA1978 contains 9327 individual records. Each individual record contains 28 columns of data, representing the following 19 variables. The variables will be defined in the next section.

VARIABLE	POSITION
WHITE	1
ALCHY	2
JUNKY	3
SUPER	4
MARRIED	5
FELON	6
WORKREL	7
PROPTY	8
PERSON	9
MALE	10
PRIORS	11-12

SCHOOL	13-14
RULE	15-16
AGE	17-19
TSERVD	20-22
FOLLOW	23-24
RECID	25
TIME	26-27
FILE	28

3. Definitions of Variables

Recall that individuals in the 1978 data set were released from prison from July, 1977 through June, 1978. The sentence from which they were released will be called the sample sentence, and the conviction which resulted in the sample sentence will correspondingly be called the sample conviction. The records of the North Carolina Department of Correction were searched in April, 1984 to determine the values of those variables representing post-release outcomes. Thus the length of the followup period ranged from 70 to 81 months.
We now define the variables in the data set, as listed above.

WHITE is a dummy (indicator) variable equal to zero if the individual is black, and equal to one otherwise. Basically, WHITE equals one for whites and zero for blacks. However, the North Carolina prison population also contains a small number of Native Americans, Hispanics, Orientals, and individuals of "other" race. They are treated as whites, by the above definition.

ALCHY is a dummy variable equal to one if the individual's record indicates a serious problem with alcohol, and equal to zero otherwise. It is important to note that for individuals in the missing data sample (FILE = 3), the value of ALCHY is recorded as zero, but is meaningless.

JUNKY is a dummy variable equal to one if the individual's record indicates use of hard drugs, and equal to zero otherwise. It is important to note that for individuals in the missing data sample (FILE = 3), the value of JUNKY is recorded as zero, but is meaningless.

SUPER is a dummy variable equal to one if the individual's release from the sample sentence was supervised (e.g., parole), and equal to zero otherwise.

MARRIED is a dummy variable equal to one if the individual was married at the time of release from the sample sentence, and equal to zero otherwise.

FELON is a dummy variable equal to one if the sample conviction was for a felony, and equal to zero if it was for a misdemeanor.

WORKREL is a dummy variable equal to one if the individual participated in the North Carolina prisoner work release program during the sample sentence, and equal to zero otherwise.

PROPTY is a dummy variable equal to one if the sample conviction was for a crime against property, and equal to zero otherwise. A detailed listing of the crime codes which define this variable (and PERSON below) can be found in A. Witte, Work Release in North Carolina: An Evaluation of Its Post Release Effects, Chapel Hill, North Carolina: Institute for Research in Social Science.

PERSON is a dummy variable equal to one if if the sample conviction was for a crime against a person, and equal to zero otherwise. (Incidentally, note that PROPTY plus PERSON is

not necessarily equal to one, because there is an additional miscellaneous category of offenses which are neither offenses against property nor offenses against a person.)

MALE is a dummy variable equal to one if the individual is male, and equal to zero if the individual is female.

PRIORS is the number of previous incarcerations, not including the sample sentence. The value -9 indicates that this information is missing.

SCHOOL is the number of years of formal schooling completed. The value zero indicates that this information is missing.

RULE is the number of prison rule violations reported during the sample sentence.

AGE is age (in months) at time of release.

TSERVD is the time served (in months) for the sample sentence.

FOLLOW is the length of the followup period, in months. (The followup period is the time from relase until the North Carolina Department of Correction records were searched, in April, 1984.)

RECID is a dummy variable equal to one if the individual returned to a North Carolina prison during the followup period, and equal to zero otherwise.

TIME is the length of time from release from the sample sentence until return to prison in North Carolina, for individuals for whom RECID equals one. TIME is rounded to the nearest month. (In particular, note that TIME equals zero for individuals who return to prison in North Carolina within the first half month after release.) For individuals for whom RECID equals zero, the value of TIME is meaningless. For such individuals, TIME is usually recorded as zero, but it is occasionally recorded as the length of the followup period. We emphasize again that neither value is meaningful, for those individuals for whom RECID equals zero.

FILE is a variable indicating to which data sample the individual record belongs. The value 1 indicates the analysis sample; 2 indicates the validation sample; and 3 indicates the missing data sample.

Predicting Recidivism

Peter Schmidt and Ann Witte,
principal investigators

1. General Description of the 1980 Data

This is a description of the data on the file, DATA1980. The description was prepared by Peter Schmidt, Department of Economics, Michigan State University, East Lansing, Michigan 48824. The data were gathered as part of a grant from the National Institute of Justice to Peter Schmidt and Ann Witte, "Improving Predictions of Recidivism by Use of Individual Characteristics," 84-IJ-CX-0021. A more complete description of the data, and of the uses to which they were put, can be found in the final report for this grant. Another similar dataset, contained in a file DATA1978 on a separate diskette, is also described in that report.

The North Carolina Department of Correction furnished a data tape which was to contain information on all individuals released from a North Carolina prison during the period from July 1, 1979 through June 30, 1980. There were 9679 individual records on this tape. However, 130 records were deleted because of obvious defects. In almost all cases, the reason for deletion is that the individual's date of release was in fact not during the time period which defined the data set. This left a total of 9549 individual records, and accordingly there are 9549 records on DATA1980.

The basic sample of 9549 observations contained many observations for which one or more of the variables used in our analyses were missing. Specifically, 3810 observations were missing information on one or more such variables, and these 3810 observations constitute the "missing data" file. The other 5739 observations which contained complete information were randomly split into an "analysis file" of 1435 observations and a "validation file" of 4304 observations.

2. Locations of Variables

DATA1980 contains 9549 individual records. Each individual record contains 28 columns of data, representing the following 19 variables. The variables will be defined in the next section.

VARIABLE	POSITION
WHITE	1
ALCHY	2
JUNKY	3
SUPER	4
MARRIED	5
FELON	6
WORKREL	7
PROPTY	8
PERSON	9
MALE	10
PRIORS	11-12

SCHOOL	13-14
RULE	15-16
AGE	17-19
TSERVD	20-22
FOLLOW	23-24
RECID	25
TIME	26-27
FILE	28

3. Definitions of Variables

Recall that individuals in the 1980 data set were released from prison from July, 1979 through June, 1980. The sentence from which they were released will be called the sample sentence, and the conviction which resulted in the sample sentence will correspondingly be called the sample conviction. The records of the North Carolina Department of Correction were searched in April, 1984 to determine the values of those variables representing post-release outcomes. Thus the length of the followup period ranged from 46 to 57 months. We now define the variables in the data set, as listed above.

WHITE is a dummy (indicator) variable equal to zero if the individual is black, and equal to one otherwise. Basically, WHITE equals one for whites and zero for blacks. However, the North Carolina prison population also contains a small number of Native Americans, Hispanics, Orientals, and individuals of "other" race. They are treated as whites, by the above definition.

ALCHY is a dummy variable equal to one if the individual's record indicates a serious problem with alcohol, and equal to zero otherwise. It is important to note that for individuals in the missing data sample (FILE = 3), the value of ALCHY is recorded as zero, but is meaningless.

JUNKY is a dummy variable equal to one if the individual's record indicates use of hard drugs, and equal to zero otherwise. It is important to note that for individuals in the missing data sample (FILE = 3), the value of JUNKY is recorded as zero, but is meaningless.

SUPER is a dummy variable equal to one if the individual's release from the sample sentence was supervised (e.g., parole), and equal to zero otherwise.

MARRIED is a dummy variable equal to one if the individual was married at the time of release from the sample sentence, and equal to zero otherwise.

FELON is a dummy variable equal to one if the sample conviction was for a felony, and equal to zero if it was for a misdemeanor.

WORKREL is a dummy variable equal to one if the individual participated in the North Carolina prisoner work release program during the sample sentence, and equal to zero otherwise.

PROPTY is a dummy variable equal to one if the sample conviction was for a crime against property, and equal to zero otherwise. A detailed listing of the crime codes which define this variable (and PERSON below) can be found in A. Witte, Work Release in North Carolina: An Evaluation of Its Post Release Effects, Chapel Hill, North Carolina: Institute for Research in Social Science.

PERSON is a dummy variable equal to one if if the sample conviction was for a crime against a person, and equal to zero otherwise. (Incidentally, note that PROPTY plus PERSON is

not necessarily equal to one, because there is an additional miscellaneous category of offenses which are neither offenses against property nor offenses against a person.)

MALE is a dummy variable equal to cne if the individual is male, and equal to zero if the individual is female.

PRIORS is the number of previous incarcerations, not including the sample sentence. The value -9 indicates that this information is missing.

SCHOOL is the number of years of formal schooling completed. The value zero indicates that this information is missing.

RULE is the number of prison rule violations reported during the sample sentence.

AGE is age (in months) at time of release.

TSERVD is the time served (in months) for the sample sentence.

FOLLOW is the length of the followup period, in months. (The followup period is the time from relase until the North Carolina Department of Correction records were searched, in April, 1984.)

RECID is a dummy variable equal to one if the individual returned to a North Carolina prison during the followup period, and equal to zero otherwise.

TIME is the length of time from release from the sample sentence until return to prison in North Carolina, for individuals for whom RECID equals one. TIME is rounded to the nearest month. (In particular, note that TIME equals zero for individuals who return to prison in North Carolina within the first half month after release.) For individuals for whom RECID equals zero, the value of TIME is meaningless. For such individuals, TIME is usually recorded as zero, but it is occasionally recorded as the length of the followup period. We emphasize again that neither value is meaningful, for those individuals for whom RECID equals zero.

FILE is a variable indicating to which data sample the individual record belongs. The value 1 indicates the analysis sample; 2 indicates the validation sample; and 3 indicates the missing data sample.