Psychological Reports, 2010, 107, 3, 983-1009. © Psychological Reports 2010

VIOLENCE RISK APPRAISAL OF MALE AND FEMALE YOUTH, ADULTS, AND INDIVIDUALS[1]

ROBERT JOHN ZAGAR
Private Practice, Chicago, Illinois

WILLIAM M. GROVE
Private Practice, Woodbury, Minnesota

Summary.—Data from youth (n = 1,127), adults (n = 1,595), and individuals (N = 2,722, combined youth and adults) were followed 3 to 12 years in records to develop items to predict abuse, violence, and homicide in these 3 relevant groups for risk appraisal or safety scales. Shao's bootstrapped logistic regression yielded 14 predictors for youth (AUC = .91), 11 for adults (AUC = .99), and 13 for individuals (AUC = .96). Three regression equations were cross-validated with in-bag and out-of-bag techniques. Pearson coefficients were computed with intelligence, achievement, adaptive behavior, and perception tests. Test-retest reliability was acceptable. Using case-control quasi-experimental design, this study extends probation-parole decision-making tests to infants and children as young as 3 years, with convergent and divergent validity and reliability with other tests. Sensitivity and specificity were high and minimized over- or under-identification challenges in identifying potentially violent persons in the general population.

In the courts, Burgess (1928) was the first researcher to track prisoners' behavior over time, evaluating those who returned to court with the goal of predicting recidivism. His parole-probation decision-making test was followed by eight decades of continuously improving risk appraisal in the courts all over the Western world. Currently, British Commonwealth, European Union, and U.S. federal prisons use these tools in determinations of release into the community, with the dual purpose of allowing rehabilitation of criminals and safeguarding society. Courts have affirmed the reliability and the validity of parole-probation decision-making tests in the U.S., the British Commonwealth, the European Union, and the United Nations (Zagar, Zagar, Bartikowski, Busch, & Stark, 2009).

There are currently two widely used risk appraisal guides for violence and sexual offense (Quinsey, Rice, Harris, & Cormier, 1998). Zagar, Busch, Grove, and Hughes (2009b) sought to improve upon the interpretive power of these measures by randomly collecting case-control samples of abused infants and children who were later homicidal, homicidal

[1]Address correspondence to Robert John Zagar, 445 East Ohio Street, Suite 303, Chicago, IL or e-mail (drzagar@msn.com). The authors wish to express their gratitude to Professor Michael Marasco, the Robert Farley Entrepreneurship Center, the McCormick School of Engineering, and the Kellogg Graduate School of Management at Northwestern University for providing strategic and tactical support in making possible the use of these tests on the internet as part of a battery of assessments. Original research collection funded from the Edith Schiller Neurology Fund, Departments of Neurology, Psychiatry, and Behavior Sciences, Northwestern University Medical School, Vogelbach Computing Center, Northwestern University, and the Juvenile Division, Circuit Court of Cook County, while continuing research was funded by the senior author.

DOI 10.2466/02.03.16.PR0.107.6.983-1009 ISSN 0033-2941

youth, and assaulters, rapists, and molesters who later were homicidal. The intent was to gather pertinent longitudinal data for a description of the developmental pattern of violent behavior. Each participant was followed in legal, medical, and school records for three to 12 years after the initial court contact, and robust statistics[2] were used to strengthen accuracy, reliability, and validity. This approach was applied to static demographic data from physical, psychiatric, psychological, social, and educational exams. Accuracy, measured as sensitivity and specificity with area under the curve (*AUC*, similar to receiver operation characteristic *ROC*), was .81–.98 (Zagar, Busch, Grove, & Hughes, 2009a). This research was replicated among teens in the Chicago Public Schools for 350 ultra high-risk students in 38 high-homicide areas; the high schools following treatment programs based on the identified variables, reported a 44% reduction in shootings.[3] This on-the-ground research promises to provide a wealth of new information about the practical validity of using prediction models and following up with empirical interventions (Zagar, Busch, & Hughes, 2009).

The importance of increased accuracy in risk appraisal cannot be overstated. Since the first risk appraisal was developed in 1928, there are more U.S. homicides than total battlefield casualties from the Revolutionary War of 1776 through the Afghanistan War (Zagar, Busch, Grove, & Hughes, 2009b). Predicting and preventing violence and homicide in workplaces, schools, and universities, let alone urban high-risk areas, are important economic concerns affecting human resources, school admissions, the need for tax dollars to support prisons, and insurance costs for private citizens as well as government and corporate organizations. Direct and indirect cost for a single homicide was estimated at $3.9M in 2006 U.S. dollars (Zagar, Zagar, Bartikowski, & Busch, 2009). Tens of billions of dollars are spent in the U.S. and the European Union on the consequences of homicide. Intercepting at-risk individuals before fatal incidents is most effectively achieved when scarce resources are targeted to "hot spots" within urban areas, and to the most at-risk individuals: dropouts, alcoholics, addicts, career delinquents-criminals, and those prone to committing homicide. Probability models can be used with a pragmatic, empirical focus on costs, given the limited effectiveness of single preventive treatments for delinquency (10–40% reduction in recidivism; Lipsey, 1999; Zagar, Zagar, Bartikowski, Busch, & Stark, 2009).

The purpose of this study was to demonstrate the reliability and the

[2]Shao's bootstrapped logistic regression (Shao, 1996).

[3]Under a U.S. Justice Department grant to the Chicago Public Schools, Zagar, Busch, Grove, and Hughes' prediction model for youth (2009a) was used in an empirical approach to provide jobs, anger management training, and mentoring to the most at-risk youth in the high schools in 38 high-homicide areas as part of the senior author's participation in Mayor Daley's Youth Violence Task Force (Saulny, 2009; Shelton & Banchero, 2009; Ahmed, 2010; Rossi, 2010).

validity of three risk-appraisal measures derived from the work of Zagar and colleagues: one for male and female youth, another for male and female adults, and the third for males and females from infancy through adulthood. Probability models, using risk factors found previously, should allow differentiation of potentially violent or homicidal male and female youth (infants, children, youth) compared to controls. Similarly, violent homicidal male and female adults should be differentiated by the model from adult controls. Cross validating the probability models will add to the predictive validity of all three instruments.

Method

Participants and Selection

From five published studies of infants, children, and adolescents, three groups were selected: youth (Sample 1, $n = 1{,}127$), adults (Sample 2, $n = 1{,}595$), and individuals (Sample 3, combined youth and adults; $N = 2{,}722$). The age, sex, race, family composition, and socioeconomic status of the three samples are presented in Table 1.

TABLE 1

Age, Sex, Race, Family Composition, and Socioeconomic Status For Youth, Adults, and Individuals

	Youth ($n = 1{,}127$)	Adults ($n = 1{,}595$)	Individuals ($N = 2{,}722$)
Age, *M*	10.29	14.1	12.5
Sex			
Female	301	235	536
Male	826	1,360	2,186
Race			
African American	711	941	1,652
Euro-American	239	476	715
Hispanic	177	178	355
Family			
Orphan	73	42	115
Single-parent	507	1,048	1,555
Stepparent	224	247	471
Mother + father	323	258	581
Socioeconomic status			
Low	638	977	1,615
Middle	489	618	1,107

Population and samples.—The population was approximately 1 million total school-aged children in Cook County, Illinois, from 1980 to 1988 (Fig. 1). This population included 47,987 abused and delinquent youth with records in Juvenile Court. Of these, five samples (2,408 abused infants, 2,455 abused children, and 2,507, 3,455, and 3,555 delinquent youth) were selected. These samples comprised approximately 0.04% of the population

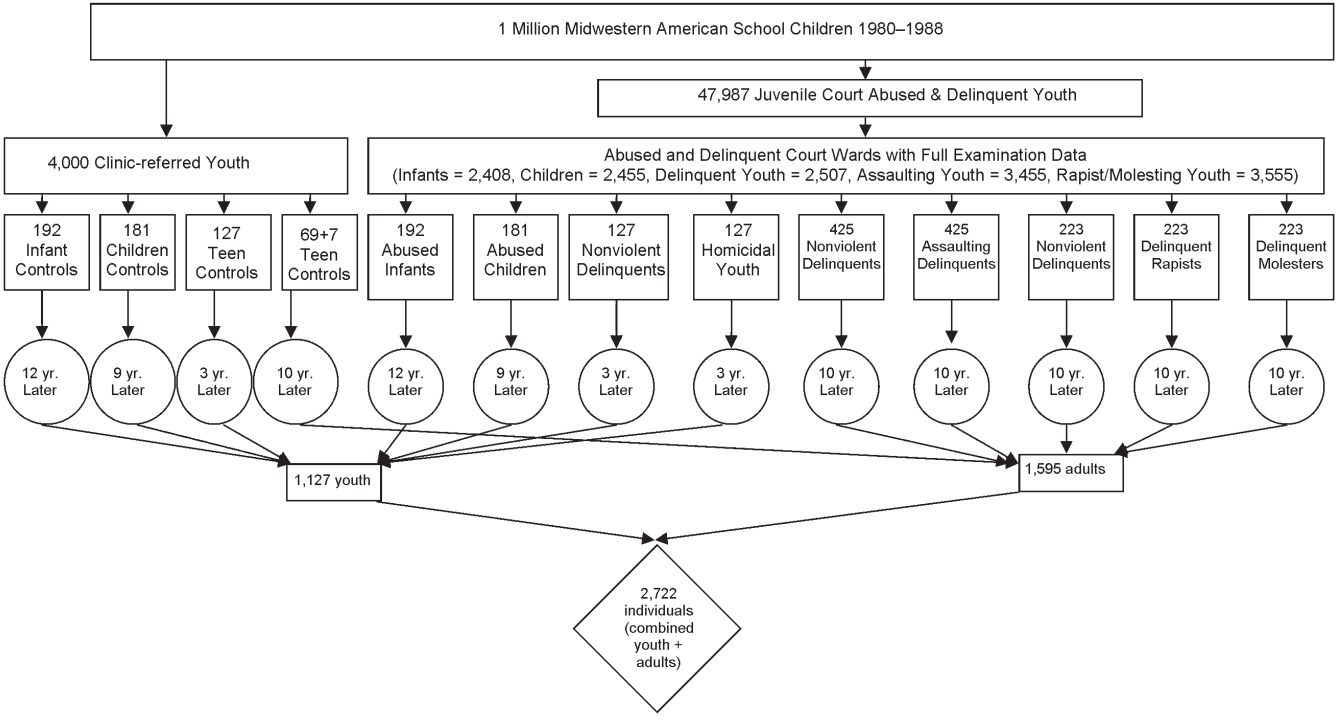

FIG. 1. Sample Selection of study of infants, children, youth, and adults to make up youth, adults, and combined individuals

and were completely unique samples of court wards exclusive of cases used in prior studies, consisting of all not-as-yet sampled referred court wards with full records and complete examinations from 1980 to 1988. The five samples represented 5% of the adjudicated wards. The Abused Infant and Abused Children groups were randomly selected from infants and children whose parents were adjudicated for abuse. The Homicidal Youth were not randomly selected. Delinquent Assaulters, Rapists, and Molesters, and Nonviolent Delinquents were randomly selected from children and youth reported and adjudicated as delinquent whose parents also may have been adjudicated for abuse. Of course, some youth who would have met these criteria were not detected or convicted or had known records of abuse by parents or caretakers or later delinquent offense.[4]

Abused infant and control groups.—The sample of 2,408 abused infants was assigned consecutive numbers and selected with a random number table to obtain the group of 192 Abused Infants. This group included 192 infants who were judged as abused within a court and whose parents were convicted between 1980 and 1988 who had full records from court-referred physical, psychiatric, psychological, educational, and social exams including measures of executive function, social maturity, and sexual abuse. The group included only those abused infants who could be completely matched with Controls. Infants were 1 to 71 months old (*M* age = 3.12 yr., *SD* = 1.48) when parents were convicted of abuse. There were 82 girls and 110 boys. The Control group comprised 192 inpatient and outpatient infants referred to hospital and university clinics from 1980 to 1988 by nurses, parents, physicians, and others in clinics, hospitals, schools, and universities, pairs who had full records from physical, psychiatric, psychological, educational, and social exams including measures of executive function, social maturity, and sexual abuse. Referrals were for development, hearing, learning, motor, physical, speech, visual, or other issues to assess level of function for interventions, schooling, and/or treatment. Court, police, protection agency, and school records were checked so that

[4]Collecting the sample and the groups from the population undoubtedly reflected some selection bias in convictions and referrals for full exams and trial. Obtaining a random sample completely free of selection bias is impossible, so the question is always which type of selection bias could be present in this study. Within a population of school-aged children, some undetermined number of youth would not have been detected as abused or convicted of a crime, or had known court, hospital, jail, police, or school records of abuse by parents or caretakers or of delinquent offenses. Others avoided court through psychiatric hospitalization or therapeutic interventions with caretakers. This particular selection bias is not avoidable when court records are used. However, the advantage of court documentation of the abuse and delinquency is that it is not affected by experimenter biases, a potential major source of bias in other studies of abuse and delinquency. The examiners' knowledge of conviction and referral was a source of expectancy bias in this prospective study because examiners expect to find more risks among convicted court wards based on experience acquired over time. However, there were many different examiners for the court over time and settings, hopefully reducing any consistent bias.

no Control case had a record of being abused or later being delinquent or criminal. Abused Infants were matched to the Control Infants on age and race to obtain individually matched groups. Controls were the best individual matches with the Abused Infants, selected from the sample of Controls until there were 192. First, Abused Infants were matched to Controls on age within 6 numerical months, e.g., matching an age of 18 months with a Control 12 to 24 months of age. Second, the cases were matched on race, African-American, Euro-American, Hispanic, and other (Asian, American Indian, or Filipino). Cases were not matched on sex or socioeconomic status (SES) because it was theorized that these might influence the phenomena under study: abuse, delinquency, and homicide. SES was measured by family annual income in 2006 dollars with a national median of $50,233 (U.S. Census Bureau, 2006). Low SES consisted of incomes from $0 to $25,000 and middle from $25,001 to $50,000. Lower and middle class were also defined by area of residence and parental occupation, similar to the procedure used in the standardization of the Wechsler Intelligence Scales (Wechsler, 1974). There were no partial matches.

Abused children and control groups.—The sample of 2,455 abused wards of the court was assigned consecutive numbers and selected with a random number table to obtain the sample of 181 Abused Children, adjudicated by a court as abused, whose parents were convicted between 1980 and 1988 who had full records from court-referred physical, psychiatric, psychological, educational, and social exams including measures of executive function, social maturity, and sexual abuse. The real number of abused children was larger, but only 181 could be rigorously matched with clinic-referred controls. Children were from 6 to 16 years old (M age = 12.8 yr., SD = 2.7) when parents were convicted of abuse. There were 58 girls and 123 boys. The Control group comprised 181 inpatient and outpatient infants referred to hospital and university clinics from 1980 to 1988 by nurses, parents, physicians, and others in clinics, hospitals, schools, and universities, who had full records from court-referred physical, psychiatric, psychological, educational, and social exams including measures of executive function, social maturity, and sexual abuse. Referrals were for development, hearing, learning, motor, physical, speech, visual, or other issues to assess level of function for interventions, schooling, and/or treatment. Court, police, protection agency, and school records were checked so that no Control case had a record of being abused or later being delinquent or criminal. Abused Children were matched to the Control Children on age and race to obtain individually matched groups. Controls were the best individual matches with the Abused Children, selected from the sample of Controls until there were 181 pairs. First, children were matched on age

within 12 numerical months, e.g., 8.0 years with 7.0 to 9.0 years. Second, children were matched on race.

Homicidal youth and control groups.—Youth were from 10 to 17 years of age (*M* age = 14.9 yr., *SD* = 1.4). The group consisted of 127 youth convicted of homicide within the court between 1980 and 1988 who had full records from court-referred physical, psychiatric, psychological, educational, and social exams including measures of executive function, social maturity, and sexual abuse. There were 7 girls and 120 boys. Because homicide is a rare event, occurring at U.S. population annual rates of 1 to 15 per 100,000 over the past century (Fox, 1992), truly random selection of homicidal youth would be difficult. Thus, best individual matching with randomly selected Nonviolent Delinquents and clinic-referred Controls was used. Nonviolent Delinquents and Homicidal Youth were matched on age and race to obtain 127 pairs. Homicidal Youth were the best individual matches with Nonviolent Delinquents from the sample of 2,507 nonviolent delinquent court wards in the sample, continuing the matching procedure until there were 127 pairs. Nonviolent Delinquents were those judged within a court, some of whose parents were convicted of abusing the youth from 1980 to 1988; they were assigned consecutive numbers and selected with a random number table to obtain a sample of 127 that had full records from court-referred physical, psychiatric, psychological, educational, and social exams including measures of executive function, social maturity, and sexual abuse. Youth were first matched on age within 12 numerical months and second on race. Controls comprised 127 inpatient and outpatient youth referred to hospital and university clinics from 1980 to 1988 by nurses, parents, physicians, and others in clinics, hospitals, schools, and universities, randomly selected from 4,000 cases in the overall referred sample that had full records from court-referred physical, psychiatric, psychological, educational, and social exams including measures of executive function, social maturity, and sexual abuse. Referrals were for developmental, hearing, learning, motor, physical, speech, visual, or other issues to determine level of function for interventions, schooling, and/or treatment. Court, police, protection agency, and school records were checked so that no Control case had a record of being abused or being later delinquent or criminal. Controls were the best individual matches with the Homicidal Youth, selected from the sample of Controls until there were 127 pairs. Youth were matched first on age within 12 numerical months, e.g., 8.0 years with 7.0 to 9.0 years and second on race.

Delinquent assaulters and control groups.—The sample of 3,455 was assigned consecutive numbers and selected with a random number table to obtain a sample of 425 Delinquent Assaulters. The group included 425 delinquents convicted in court from 1980 to 1988 of aggravated battery and

assault or homicide who had full records from court-referred physical, psychiatric, psychological, educational, and social exams including measures of executive function, social maturity, and sexual abuse. Some of the parents of these court wards also had been convicted of physical abuse. These children and youth were from 8 to 17 years of age (M age = 14.1 yr., SD = 1.7). There were 77 girls and 348 boys. Both abused and delinquent children and youth had an age range from 6 to 16 years with the overlap of age for abuse and delinquency. This occurred because children and youth were abused earlier or later in life and because there were early- and later-onset delinquents. A sample of 3,425 nonviolent delinquents was assigned consecutive numbers and selected with a random number table to obtain a sample of 425 Nonviolent Delinquents. Nonviolent Delinquents were convicted on nonviolent offenses and matched on age and race to obtain a sample that had full records from court-referred physical, psychiatric, psychological, educational, and social exams including measures of executive function, social maturity, and sexual abuse. Youth were matched with Delinquent Assaulters on age within 12 numerical months, e.g., 8.0 years with 7.0 to 9.0 years, and then on race. No cases were discarded because of a failure to find an adequate match. Controls (n = 69) were randomly selected from 4,000 clinic-referred youth and individually matched with 69 Homicidal Assaulters. Referrals were for developmental, hearing, learning, motor, physical, speech, visual, or other issues to assess level of function for interventions, schooling, and/or treatment. These were youth who were referred for pediatric neurological or psychological workups, or vocational guidance, and should be considered somewhat at risk but without known delinquency or history of abuse.

Delinquent rapists and molesters and control groups.—The delinquents in this sample of 3,555 were assigned consecutive numbers and were selected with a random number table to obtain a sample of 223 Rapists and 223 Molesters. Although the total number of rapists and molesters was larger, only for 223 of the rapists were the matched triads of Rapists, Molesters, and Nonviolent Delinquents possible. The first group included 223 delinquent Rapists convicted in court from 1980 to 1988 of rape and homicide as part of a rape, some of whose parents had also been convicted of physical abuse, who had full records from court-referred physical, psychiatric, psychological, educational, and social exams including measures of executive function, social maturity, and sexual abuse. Youth were from 7 to 17 years old (M age = 14.2 yr., SD = 1.5) with 25 girls and 198 boys in each group. The sample of 3,555 was assigned consecutive numbers and selected with a random number table to obtain a sample of 223 Nonviolent Delinquents. Nonviolent Delinquents were a comparison group of 223 delinquents convicted of nonviolent offenses and matched on age and

race with the Rapists and Molesters to obtain a sample of 223. Youth were matched on age within 12 numerical months, e.g., 8.0 years with 7.0 to 9.0 years, and then on race. No cases were discarded because of a failure to find an adequate match. Controls ($n = 7$) were randomly selected from 4,000 clinic-referred youth and individually matched on age and sex with Rapists and Molesters. Referrals were for developmental, hearing, learning, motor, physical, speech, visual, or other issues to assess level of function for interventions, schooling, and/or treatment. These were youth who were referred for pediatric neurological or psychological workups, or vocational guidance, and should be considered somewhat at risk but without known delinquency or history of abuse.

Samples Used in Regressions

Sample 1.—The original samples of 192 Abused Infants matched with 192 Controls tested at a mean age of 3.1 yr. (Zagar, Busch, Grove, Hughes, & Arbit, 2009c), the 181 Abused Children matched with 181 Controls test-

TABLE 2

Descriptive Time Line of Examinations and Court Contacts and Mean Ages at Examinations and Records Follow-up For All Groups

Time Line
First court contact
Police or protective agency contact, either:
Parent contact for child abuse, or
Youth contact for delinquency
One or more court contacts, either
Finding of child abuse (Parent conviction), or
Finding of delinquency (Youth conviction)
First data collection (Time 1)
Court contact for abused status and/or delinquency
Court order for physical, psychiatric, psychological, educational, and social examinations
Gathered data from examinations
Second (follow-up) data collection (Time 2)
Records check: court, hospital, police, protective agencies, school
Finding of child abuse (Parent conviction) or
Finding of delinquency (Youth or adult conviction)
Gathered data from examinations, records

Mean Ages (yr.) at Examinations and Records Follow-up

Group	*N*	Examination	Records Follow-up	*N* Homicides
Abused infants	192	3.1	15.2	21
Abused children	181	12.8	20.6	10
Homicidal youth	127	14.9	7.2–18.0	127
Assaulting youth	425	14.0	10.0–24.1	69
Molesters, rapists	446	14.2	9.8–24.2	7
Total				234

ed at a mean age of 12.8 yr. (Hughes, Zagar, Busch, Grove, & Arbit, 2009), and the original sample of 127 Homicidal Youth matched with 127 Nonviolent Delinquents and 127 Controls tested at a mean age of 14.9 yr. (Zagar, Busch, Grove, Hughes, & Arbit, 2009a) comprised the sample of 1,127 youth with a mean age of 10.29 yr.

Sample 2.—The original samples of 425 Delinquent Assaulters matched with 425 Nonviolent Delinquents and 69 Controls tested at a mean age of 14.0 yr. (Zagar, Busch, Grove, Hughes & Arbit, 2009b) were combined with 223 Molesters, 223 Rapists, with 223 matched Nonviolent Delinquents and 7 Controls tested at a mean age of 14.2 yr. (Busch, Zagar, Grove, Hughes, Arbit, & Bussell, 2009) for a total sample of youth ($n = 1,595$) with a mean age of 14.1 yr.

Sample 3.—For the total sample of youth and adults, the 1,127 youth tested at a mean age of 10.29 yr. were combined with 1,595 adults to total 2,722 individuals. It was already established that these groups of adults, youth, and individuals were representative of the larger samples of school children and delinquents from a Midwestern American city, and the U.S. and European Union populations on demographics (see cited studies for comparisons on demographics for each group).

Measures

Record follow-up.—Infants, children, youth, delinquents, and controls were followed through records for information about risk factors and future offenses,[5] namely abuse by parents or caretakers and for violent and nonviolent delinquent and criminal offenses. No cases were lost through attrition because of failure to find records. See Table 2 for time line.

General notes on measures.—In order to assess thoroughly the overall pattern of risks, data for both the parents and the infants, children, and

[5]"Abuse" is defined by the federal Child Abuse and Preventive Treatment Act and by Illinois State statute as an act of a caretaker which resulted in death or severe physical or emotional harm, or sexual abuse or exploitation, or an act or failure to act which presents an imminent risk of serious harm. These harms include evidence of documented bruises, lacerations, contusions and injuries to the infants, children, or youth, including molestation and rape. A Molester was an identified, detected, and convicted individual who molested a minor, supported by evidence presented in court. Molestation was consistent with the definition of pedophilia in the Diagnostic Statistical Manual-III–Revised (DSM-III–R; American Psychiatric Association, 1987). A Rapist was an identified, detected, and convicted individual who had raped a female or male, with evidence presented in court, and met the criterion for rape or aggravated criminal sexual assault as defined in Illinois criminal statutes. "Delinquency" is defined as conviction before a judge. "Nonviolent delinquency" includes offenses of truancy, disorderly conduct, solicitation, telephone harassment, forgery, violation of court order, drug use, property damage, auto theft, theft, and burglary. "Violent delinquency" is unlawful restraint, kidnapping, robbery, unlawful weapon use, battery/assault, rape, and aggravated sexual abuse toward own sex. "Homicide" is defined as adjudication and conviction before a judge for killing another individual(s). Age range is "infant" (0–5 yr.), "child" (6–12 yr.), and "youth" (13–17 yr.).

youth were collected from exam and court records created in response to the parents' initial court contact for abuse or the youths' initial court contact for delinquency. The data collected from each type of record are described below. All were gathered from examinations and court records. Nominal variables included parental risks (violent family, the parent having been physically abused by a caretaker in youth, gang membership, alcohol abuse, substance abuse, alcohol plus substance abuse), infant risks (family composition, injury, burn, poisoning, or fetal substance exposure, three or more home/school moves, later court contacts, executive function, hyperactivity, sex, socioeconomic status, family composition, and sexual abuse), child risks (underachievement, respiratory, infectious, neurological, genitourinary, pregnancy, childbirth, or perinatal complications, truancy, suspension or expulsion, injury, burn or poisoning, epilepsy, psychiatric hospital, later court contacts, executive function, social maturity, and family composition), youth risks (prior court contacts, underachievement, truancy, suspension or expulsion, gang membership, physically abused, alcohol abuse, substance abuse, alcohol plus substance abuse, violent family, sex, socioeconomic status, illnesses, executive function, social maturity, weapon possession/conviction, family composition, head injury, asthma, epilepsy, neurological disorder, jaundice, sensory and speech disorders, special education services, and psychiatric hospitalization), and adult-adolescent risks (personality disorder, unemployment, not with parents until age 14, single relationship status, violent family, gang membership, convicted before age 12, truancy, suspension or expulsion, socioeconomic status, sex, family composition, victim, victim injury, prior court contact, executive function, social maturity, underachievement, prior court contact for sex offenses). Records were obtained because of interest in the youth's welfare. See the Appendix (pp. 1008-1009).

Standardized tests.—The dependent variable, violence, represented a continuum from no court contact to homicide (see Table 3). Physical examinations were completed by pediatricians and coded and defined using the International Classification of Diseases (ICD–9, World Health Organization, 1977; see the Appendix, pp. 1008-1009). In the psychiatric examinations, physicians assessed infants, children, and youth along with their caretakers for one hour or longer using a nonstandardized developmental review of defined psychiatric symptoms of youth and guardians (DSM-III–R, American Psychiatric Association, 1987; ICD–9, World Health Organization, 1977; see the Appendix). Psychological and school examinations included the Bayley Scales of Infant Development (Bayley, 1969), the Stanford Binet Intelligence Scale, Third Edition or L-M Version (Terman & Merrill, 1973), Wechsler Intelligence Scale for Children–Revised (WISC–R;

TABLE 3
Dependent Variables for Logistic Regressions in Youth, Adults, and Individuals

Code	Variable
0	No court contact
1	Neglect
2	Dependency
3	Abuse (physical, sexual)
4	Minor in need of supervision
5	Truancy
6	Disorderly conduct, solicitation, phone harassment, forgery
7	Violating court order
8	Drug possession/sales
9	Property damage
10	Auto theft
11	Theft
12	Burglary
13	Robbery
14	Unlawful weapon possession [firearm(s)]
15	Arson
16	Assault/battery
17	Molestation (aggravated criminal sexual assault to minor)
18	Rape (aggravated criminal sexual assault to adult)
19	Homicide

Wechsler, 1974) which was scored by two independent psychologist raters, the Bender Visual-Motor Gestalt Test (Bender, 1938), a measure of age-appropriate visual and motor skills scored by two independent raters with the Koppitz (1964) method, obtaining a Pearson correlation coefficient of interobserver agreement of .91 ($p < .01$), the Gates-MacGinitie Reading Tests Survey D, Form 1 (Gates & MacGinitie, 1965) with subscales Reading Speed and Accuracy, Vocabulary, and Comprehension to assess general reading skills, the Stanford Achievement Test, Form W, Intermediate I (Kelly, Madden, Gardner, & Ruderman, 1964)—specifically, the Arithmetic Computation subscale to assess mathematics achievement, the Vineland Social Maturity Scale (Doll, 1965) used to assess maturity, i.e., social competence including communication, locomotion, self-help, dressing, eating, and general self-direction, socialization, and vocation, the Zagar Executive Function Scale for Infants (Zagar, Busch, Grove, Hughes, & Arbit, 2009c), and the Zagar Executive Function Scale for Children-Youth (Hughes, *et al.*, 2009). Using data from records, the Diagnostic Statistical Manual-III–Revised criteria (DSM-III–R, American Psychiatric Association, 1987), and an algorithm defined earlier by Zagar, Arbit, Hughes, Bussell, and Busch (1989), two independent psychologists scored IQs and two classified psychopathology; Pearson correlation coefficients of inter-observer agreement were .87 to .94 for each of the scores and categories ($p < .01$).

Social examinations.—Probation officers, child protection personnel, and/or social workers collected the data listed in the Appendix (pp. 1008-

1009). These were recorded on nominal, ordinal, or interval scales. Two independent psychologist raters coded parent and infant, child, or youth nominal data (0: No, 1: Yes). Inter-rater correlation coefficients for parental data were .75 to .98, and for infant, children, and youth data were .76 to .97 ($ps < .01$). Child nominal data were childhood respiratory, infectious, neurological, or genitourinary complications or epilepsy; pregnancy, childbirth, or perinatal complications; truancy, suspension, or expulsion from school, and/or underachievement (defined earlier in Zagar, *et al*., 1989); psychiatric hospitalization not necessarily involved with any court order; sexual abuse; and family composition (coded 1: Orphan, 2: Single-parent, 3: Stepparents, 4: Biological mother and father, "mother + father"). Infant, child, and youth illnesses were collapsed into an ordinal variable because the organ systems involved were not different when the groups were compared—presumably because the sample was small. Details were described in Zagar, *et al.* (1989). Ordinal data were Bayley, Stanford Binet, and Wechsler, Bender errors, Gates-MacGinitie, Stanford Achievement, Vineland, and Zagar Executive Function Checklist raw scores. See the Appendix.

Analyses

Since some variables were not collected for infants, children and youth, or adults, these were not included in the analyses. Likewise, since some variables were not collected across youth and adults, these were not included in the analyses for individuals. The sample sizes were sufficient for a 20-variable, multiple group, multivariate design with expected small effect sizes (Lachenbruch, 1968; Bock, 1975; Cohen & Cohen, 1982; Kirk, 1982; Stevens, 1986; Hosmer & Lemeshow, 2000). Three logistic regressions with Shao's bootstrapping (1996) yielded multiple R, R^2, and area under the curve (AUC) for the three groups: youth ($n = 1{,}127$), adults ($n = 1{,}595$), and combined youth and adults ($n = 2{,}722$).

Logistic regression gives a linear equation showing which variables are most strongly associated with violent (homicidal) behavior. Therefore, in the common sense of the term, it is a classification of group membership (status). Analyses are not only statistical analyses predicting outcomes; since the data are longitudinal, the resulting equations are truly "predictive" in the common sense of the word. Prior risk factors are used to predict whether the individual would later commit homicide. Comparisons between violent and nonviolent or non-delinquent groups were conducted to identify and quantify risks of violent behavior in youth and adults separately, and for an individual in the general population (combined youth and adults) at risk for later committing a violent act.

In statistical bootstrapping, there is a large-sample unbiased estimate

of accuracy using the same cases to develop the logistic regression and also to test the regression. In Shao's bootstrapping (1996),[6] the sampling distribution of a statistic of interest is approximated by repeatedly sampling from within the sample at hand. In logistic regression, it is the area under the receiver operating characteristic (*ROC*) curve, the *AUC*, which captures the overall accuracy of prediction outcomes of interest such as homicide by comparing "hits" versus "misses." The total group is treated as if it is a population, and samples are drawn without replacement from it, one after another, until 1,000 or more have been drawn. Logistic regressions are calculated for each sample and a corresponding *AUC* is estimated. The overall analysis yields an estimate of *AUC* that is the mean of all 1,000 samples' *AUC*s. The *AUC* is a desirable measure of prediction performance, because it is not influenced by the base rate of the phenomenon being predicted and of the cutting scores on predictors used to make predictions.[7]

The *AUC*, or area under the binormal receiver operating characteristic (*ROC*) curve, is the proportion of the area to the range of the area index, plotted on linear probability scales, ranging from 0.5 to 1.0. This area under the curve is equal to the probability of a correct response in a two-alternative, forced-choice test (such as membership or nonmembership in a group) that accounts for both the true positives and the false positives, i.e., sensitivity and specificity. Logistic regressions with resulting *AUC*s provide an easily understood, quantitative measure of the risk factors that in-

[6]Shao's bootstrapping method relies on repeatedly sampling with replacement from the study sample (of size N) with each new sample being of size $m = N^{3/4}$ with the exponent chosen so that $m << N$ and $m/N \rightarrow 0$ as $N \rightarrow \infty$. Shao has proven that this method of bootstrapping produces the correct population model, in terms of the set of variables selected with $P \rightarrow 1$ as $N \rightarrow \infty$. Other analytic methods include leaving-one-out and 10-fold cross-validation, and produce the correct model with $P \rightarrow 0$ as $N \rightarrow \infty$. Shao's method also provides unbiased estimates of the regression weight for each variable and of the *AUC*. In application, slightly more complicated estimators and confidence intervals are used to decrease bias and improve confidence interval coverage. If the variables entered into the logistic regression are specified *a priori* and are not selected in or dropped out based on significance tests, then this simple type of bootstrapping is known to have attractive, optimal properties.

[7]However, if forward selection, backward elimination, or stepwise selection of variables occurs, then an element of capitalization on chance comes into the picture, which the simple bootstrap theory does not take into account. Three things in an analysis occur when variable selection is going on: (a) the correct model (the right variables included and the right variables excluded) with as high a probability P as possible; (b) a correct measure of the prediction model's accuracy as is obtainable; and (c) correct measures of uncertainty of aspects of the model, the model's coefficients (or, more easily interpreted, the model's predictive accuracy). Shao has shown that traditional methods of single-sample cross-validation do not satisfy (a), (b), or (c). These methods include leaving-one-out and 10-fold cross-validation. With this N, each logistic regression is done on just a quarter of the data, and variables are selected (in this study, by forward selection). The resulting regression equation, with the selected variables and their assigned weights, is then cross-validated on all N observations, and its *AUC* calculated as an index of model performance. The 1,000 models and the *AUC*s are collected to compute the final model, namely the one with the highest *AUC*, and its estimated performance, which is the average of all 1,000 *AUC*s.

crease the probability of homicide. In this set of three samples, the regressions with their weights comprise the "Safety Scales," measures predictive of homicide in each of three groups.

Bagging, which stands for "bootstrap aggregations," is a way of generating many replicas of the data set by randomly selecting N observations out of N without replacement, where N is the data-set size. Bagging works by reducing the variance of an unbiased set. Drawing N out of N observations with replacement omits, on average, 37% of observations. These are "out-of-bag" observations. One uses them to estimate the predictive power and feature importance. For each observation, one can estimate the out-of-bag prediction by averaging predictions for which this observation is out-of-bag. One can compare the computed prediction against true responses for this observation. By doing this comparison, one can estimate the average out-of-bag error, which is an unbiased estimator of the true error (Breiman, 1996, 2001; Martinez-Munoz & Suarez, 2010).[8]

There were 11 significant variables or items identified for Sample 1 (youth), 14 for Sample 2 (adults), and 13 for Sample 3 (combined group). Test-retest reliabilities (r_{tt}) were computed for the youth (children and teens; $n = 40$), adults ($n = 83$), and individuals ($n = 123$) approximately two years apart. Mean age at the first testing for the total retest group ($n = 123$) was 12.0 yr. ($SD = 2.1$) and at the second testing 14 yr. ($SD = 2.1$). Pearson product-moment correlations were computed between the Safety Scale scores for each of the three samples and the Bayley (1969) Mental and Motor Developmental Index, Stanford Binet L-M (Terman & Merrill, 1973) IQ, Wechsler Intelligence scales (1974) Verbal, Performance, and Full Scale IQ, Vineland Social Maturity Scale (Doll, 1965), Bender (1938) Visual-Motor Gestalt Test Koppitz (1964) error score, Gates and MacGinitie (1965) Reading Test Survey D Form 1 Speed, Comprehension, and Vocabulary scores, Stanford Achievement Test Form W Intermediate I (Kelly, *et al.*, 1964), and Executive Function Checklist (Hughes, *et al.*, 2009; Zagar, Busch, Grove, Hughes, & Arbit, 2009c). Pearson product-moment item-total correlations for the youth, adult, and individual Safety Scales were also completed.

Results

Sample 1, Youth

In Table 3, the dependent (predicted) variables are listed from normal or (no court contact) to violence (homicide) along a continuum. In Table 4, the independent variables (predictors) for youth violence are list-

[8]Cronbach's α (1951) and other measures of internal consistency, e.g., Revelle's β (1979) and McDonald's Ω (1970), that would give dimensionality were not considered because it was known that two primary measures (prior court contacts, executive function) and all other static descriptors were the three main dimensions of the youth, adult, and individual Safety Scales.

TABLE 4
Independent Variables for Logistic Regressions of Youth, Adults, and Individuals

	Youth (n = 1,127) $t_{14,1112}$	Adults (n = 1,595) $t_{11,1583}$	Combined (N = 2,722) $t_{14,2697}$
Intercept	†	†	†
Executive function	†	†	†
Prior court contacts	†	†	†
Sex	†	†	†
Alcohol & substance abuse	†	†	†
Violent family	†	†	†
Unemployment		†	†
Underachievement	†	†	†
Antisocial personality disorder		*	
Hyperactivity		†	†
SES		†	
Illnesses	†	†	†
Family composition	†		†
Alcohol abuse	†		
Substance abuse	†		
Social maturity	†		
Physical abuse	†		
Truancy, suspension, expulsion	†		†
Epilepsy	†		†
Gang membership			†
Multiple R	.86	.66	.78
R^2	.85	.66	.78
Area Under the Curve (AUC)	.91	.99	.96

*$p > .05$. †$p < .01$ for beta coefficients (not provided to maintain test integrity and security).

ed in order of importance. There were 14 significant predictors of violence for Sample 1: Youth: executive function, sex, court contact, violent family member, alcohol and substance abuse, underachievement, illness, epilepsy, alcohol abuse, substance abuse, truancy/suspension/expulsion, family composition, physical abuse, Vineland Social Maturity, and gang participation. The multiple R, R^2, and AUC were all significant. It should be noted that for the subgroups of infants and children, many of these variables refer to *parental* characteristics. To show divergent validity, Pearson correlations were calculated between the Youth Safety Scale and ability, achievement, adaptive behavior, and perception test scores. These are provided in Table 5 with the two-year test-retest reliabilities. The distribution of scores is shown in Fig. 2**.**

Sample 2, Adults

In Table 4 are listed the 11 significant predictors of violence for adults: executive function, sex, court contact, violent family member, unemployment, antisocial personality disorder, alcohol and substance abuse, hyper-

TABLE 5
Two-year Test-retest (r_{rtr}), and Test Inter-correlations
For Youth (n = 1,127) Safety Scale

Infants (n = 384)	1	2	3	4	5
1. Bayley Mental Scale of Infant Development					
2. Bayley Motor Scale of Infant Development	.10				
3. Stanford Binet L-M	−.22	−.29			
4. Vineland Social Maturity Scale	.10	.08	−.01		
5. Zagar Executive Function Checklist	.20	.12	−.26	−.15	
6. Zagar Youth Safety Scale	.03	−.03	.00	.99	.08

Children and Teens (n = 743)	1	2	3	4	5	6	7	8	9	10
1. Wechsler Verbal IQ Scale										
2. Wechsler Performance IQ Scale	.64									
3. Wechsler Full IQ Scale	.87	.85								
4. Vineland Social Maturity Scale	.54	.81	.73							
5. Bender Visual-Motor Gestalt Test	−.12	−.10	−.10	−.20						
6. Gates-MacGinitie Speed Test	.17	.10	.17	.15	−.24					
7. Gates-MacGinitie Vocabulary Test	.22	.13	.22	.14	.16	.76				
8. Gates-MacGinitie Comprehension Test	.23	.15	.22	.17	−.20	.77	.88			
9. Stanford Arithmetic Test	.17	.06	.15	.09	−.30	.71	.80	.87		
10. Zagar Executive Function Checklist	.03	.06	.07	−.16	.41	.01	.08	.05	.03	
11. Zagar Youth Safety Scale	.32	.47	.43	−.44	.38	.15	.10	.12	.08	.01

Note.—Test-retest (n = 40) r_{rtr} = .75.

activity, lower socioeconomic status, underachievement, and illness. The multiple *R*, R^2, and *AUC* were statistically significant. The unavoidable dilemma of decision making with an actuarial tool is the "sensitivity-specificity tradeoff": a higher hit rate is nearly always obtained at the expense of a higher alarm rate. The tradeoff is observed when the cutoff point or decision threshold of the test is moved, i.e., changing the cutoff score for an established test. With an *AUC* of .99, the false negative and false positive rates are essentially equal to 1%, which is within the acceptable error of measurement of any test. Since the false negative and false positive rates are low, i.e., the possibility of not detecting a homicidal individual or of identifying someone as homicidal when the individual is not, the excellent accuracy of this test provides a way out of the dilemma (Zagar, Busch, Grove, & Hughes, 2009a). To show divergent validity, Pearson correlations were calculated between the Adult Safety Scale and ability, achievement, adaptive behavior, and perception test scores (Table 6) with the two-year test-retest reliability. See Fig. 3 for distributions of Safety Scale scores.

Sample 3, Combined Youth and Adults

In Table 4 are shown the 15 significant predictors of violence in Sample 3, youth and adults: executive function, sex, court contact, violent

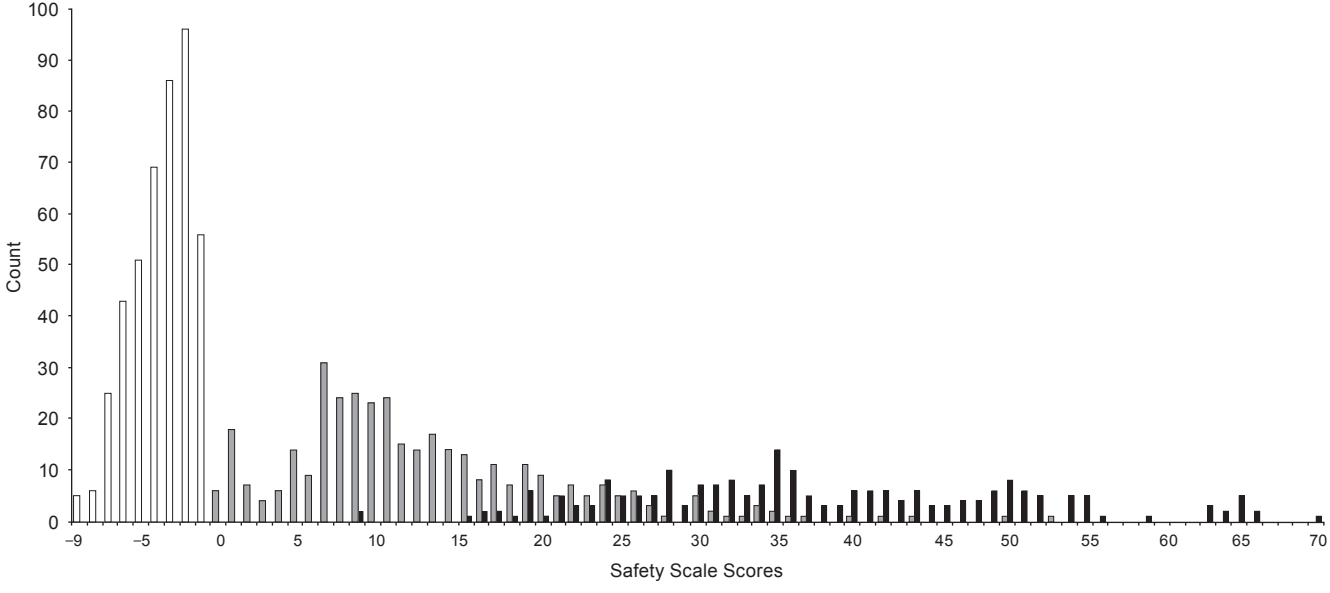

FIG. 2. 1,127 youth ($M = 11.73$, Variance = 17.7); Infant Nonviolent (□); Child–Teen Nonviolent (▩); Infant–Teen Violent (■).

TABLE 6
Two-year Test-retest (r_{rtr}), and Test Inter-correlations For Adult (n = 1,595) Safety Scales

Adults (n = 1,595)	1	2	3	4	5	6	7	8	9	10
1. Wechsler Verbal IQ Scale										
2. Wechsler Performance IQ Scale	.62									
3. Wechsler Full IQ Scale	.89	.89								
4. Vineland Social Maturity	.58	.77	.75							
5. Bender Visual Motor Gestalt Test	−.05	−.09	−.08	−.29						
6. Gates MacGinitie Speed Test	.32	.26	.30	.36	−.35					
7. Gates MacGinitie Vocabulary Test	.23	.27	.32	.29	−.39	.90				
8. Gates MacGinitie Comprehension Test	.21	.26	.30	.27	−.37	.90	.93			
9. Stanford Arithmetic Test	.24	.22	.25	.36	−.34	.91	.93	.96		
10. Zagar Executive Function	−.13	−.04	−.08	−.46	−.29	−.36	−.38	−.37	−.38	
11. Zagar Adult Safety Scale	−.10	−.01	−.06	−.46	−.27	−.38	−.40	−.40	−.40	.94

Note.—Test-retest (n = 83) r_{rtr} = .76.

family member, alcohol and substance abuse, socioeconomic status, illness, epilepsy, truancy/suspension/expulsion, family composition, Vineland Social Maturity, and gang participation. The multiple R, R^2, and AUC were statistically significant. Two-year test-retest reliability and the Pearson correlations with the Bayley Scale of Infant Development, Stanford Binet Intelligence Scale L-M or third edition, Wechlser Intelligence Scale–Revised, Bender Test, Gates-MacGinitie Reading Achievement Test, Stanford Arithmetic Test, and Executive Function Checklist are presented in Table 7. See Fig. 4 for a comparison of distributions of Safety Scale scores in the groups.

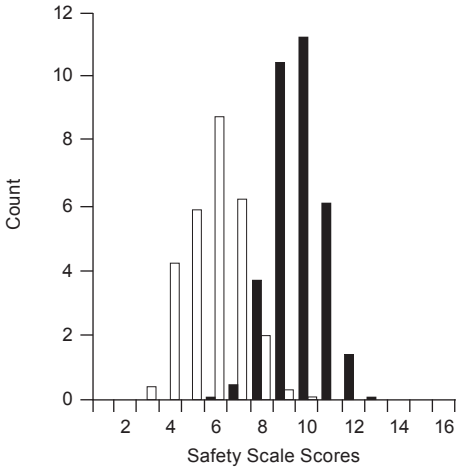

Fig. 3. 1,595 adults: Violent (■); Nonviolent (□)

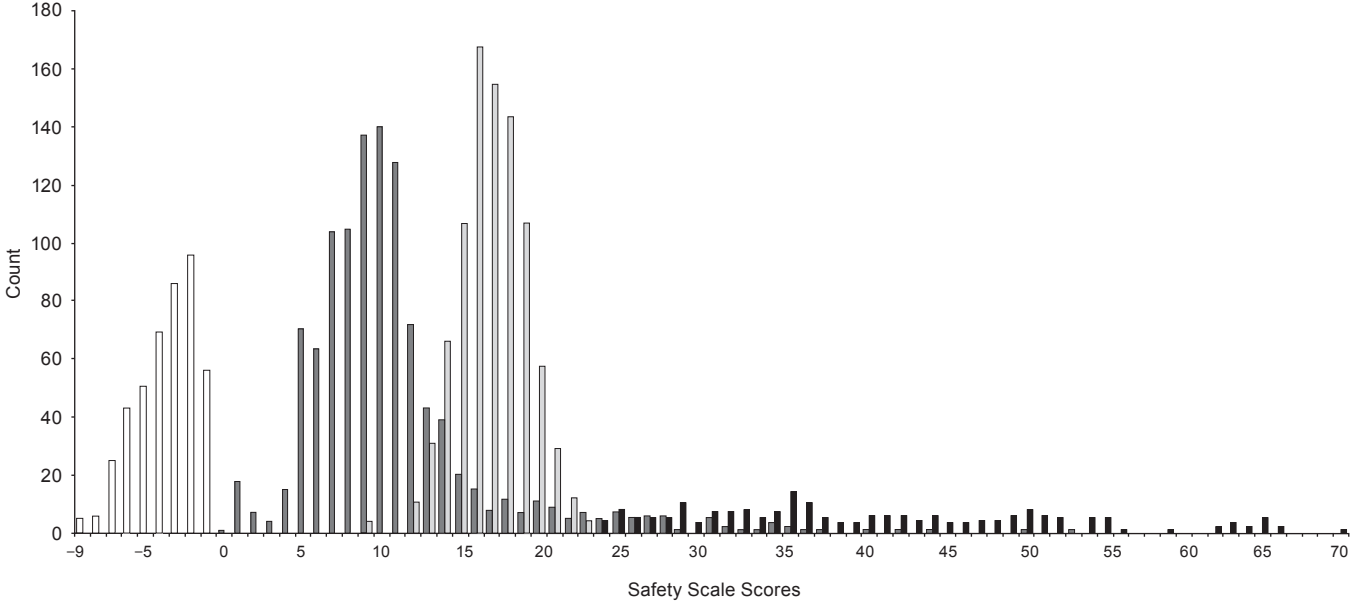

FIG. 4. 2,722 individuals ($M = 12.43$, Variance = 11.62); Infant Nonviolent (□); Child–Teen Nonviolent (■); Adult Nonviolent (▩); Infant–Adult Violent (■);

TABLE 7
Two-year Test-Rretest (r_{rtr}), and Test Inter-correlations
For Individual (N = 2,722) Safety Scale

Infants (n = 384)	1	2	3	4	5
1. Bayley Mental Scale of Infant Development					
2. Bayley Motor Scale of Infant Development	.10				
3. Stanford Binet L-M	−.22	−.29			
4. Vineland Social Maturity Scale	.10	−.08	−.01		
5. Zagar Executive Function Checklist	.20	.12	−.26	−.15	
6. Zagar Youth Safety Scale	.32	.21	−.01	.52	.09

Children,Teens, and Adults (n = 2,338)	1	2	3	4	5	6	7	8	9	10
1. Wechsler Verbal IQ Scale										
2. Wechsler Performance IQ Scale	.61									
3. Wechsler Full IQ Scale	.08	.65								
4. Vineland Social Maturity Scale	.67	.84	.84							
5. Bender Visual-Motor Gestalt Test	−.19	−.10	−.08	−.10						
6. Gates-MacGinitie Speed Test	.23	.19	.16	.10	−.26					
7. Gates-MacGinitie Vocabulary Test	.24	.19	.18	.21	−.28	.90				
8. Gates-MacGinitie Comprehension Test	.23	.18	.17	.10	−.27	.90	.93			
9. Stanford Arithmetic Test	.21	.16	.14	.17	−.28	.91	.93	.96		
10. Zagar Executive Function Checklist	−.40	−.15	−.08	−.06	.29	−.28	−.29	.29	−.30	
11. Zagar Youth Safety Scale	−.31	−.11	−.06	−.05	.30	−.24	−.24	−.25	−.27	.76

Note.—Test-retest (n = 123) r_{rtr} = .755.

Discussion

Issues related to overmodeling of data may be a source of criticism. Nunnally's formal logic (1978) of analysis first calls for specification of domains of interest, then application of statistical analyses to assess the extent to which observables measure one, several, or many outcomes. The logic is tested by performing studies of individual differences or by controlled experiments to evaluate the extent to which supposed measures of a construct produce results are predictable from theoretically derived and thoroughly tested hypotheses. In this study, there are a variety of observables (medical, psychiatric, psychological, social, and educational exams and records), with several variables representing different domains that are related to the constructs of interests: functioning (e.g., executive function, social maturity, family, health, or illness, school achievement, alcohol-substance abuse) and personal variables (e.g., sex, socioeconomic status). The outcome of interest is the violence (homicide) status at time of follow-up in records. The issue of over-modeling of data is much less relevant, because the domains of interest were chosen, robust quasi-experimental design was used in collection of data, and strong statistical analyses were chosen to assess which observables measure the outcome (violence or homicide). The results fit predictable, theoretically derived, and empirically

supported hypotheses of Lipsey and Derzon (1998) and Bonta, Hanson, and Law (1998)—also, see Zagar, Busch, Grove, and Hughes (2009a, pp. 253-257).

Researchers have debated the accuracy of decision-making tests for violence and homicide, typically citing the issues of false negatives and positives and their associated costs. If the accuracy of the three measures presented here for youth (*AUC* = .91), adults (*AUC* = .99), and individuals (combined youth and adults; *AUC* = .96) is replicated in independent population samples, the concern about over- and under-identification will be minimal. In addition, the tests have excellent psychometric qualities and stability: acceptable internal consistency, high test-retest reliability, and divergent coefficients with ability, achievement, and perception tests. Representation of females and youth in these samples supports validity of the Safety Scales, mitigating long-term controversies about risk appraisal in these less well characterized subgroups. Particularly considering the fact that parole and probation decision-making tests have been developed over decades and are based on enormous and diverse histories of tens of thousands of delinquents and criminals in many geographic areas, these new data indicate that prediction of violence and homicide is not only practical, but reliable, valid, and probably generalizable to the general population. In contrast, there are a litany of studies showing that professionals' ability to predict violence is very poor: opinions of psychiatrists and psychologists were only about 30–53% accurate in prediction of violence among patients and parolees (Sepajak,1983; Lidz, 1993; Monahan, 1996). Rice, Harris, and Quinsey (1996) found that forensic clinicians seemed most ready to release, without supervision, the most likely to re-offend violently, while retaining or recommending intense supervision for the least dangerous patients. Meehl (1954) and Grove and Meehl (1996) proved conclusively the usefulness of statistical probation and parole decision-making tests: actuarial judgment was superior in 128 of 136 empirical studies comparing clinical versus objective prediction. Quinsey, *et al.* (1998) reviewed violence risk assessments for parolees, estimating accuracy at 69 to 75%.

The goal of the current line of research is to continue the development of an accurate, inexpensive screening battery for assessment of ability, personality, alcohol/substance abuse, abuse, and violence potential to reduce the incidence of violence. The cost-effective approach would allow practical decision-makers to overcome the current inconsistent and subjective approach that costs lives every day. Obviously, it is not reasonable to proactively identify homicide-prone behavior and intervene by limiting constitutional freedoms, but, by applying empirical treatments and interven-

tions to youth through the public schools, or through employee assistance and other health interventions for adults, the likelihood of violence could be reduced. The legality and practical utility of this approach is indicated by the use of actuarial identification combined with empirical treatments in six high-risk high schools in Chicago, resulting in a 44% reduction in shootings and 77% lowering of violent acts (Saulny, 2009; Shelton & Banchero, 2009; Ahmed, 2010; Rossi, 2010).

REFERENCES

Ahmed, A. (2010) Chicago Public Schools antiviolence program takes off slowly. *Chicago Tribune*, June 3.

American Psychiatric Association. (1987) *Diagnostic and statistical manual of mental disorders–revised.* (3rd ed.) Washington, DC: Author.

Bayley, N. (1969) *Manual for Bayley Scales of Infant Development*. New York: Psychological Corp.

Bender, L. (1938) A visual motor gestalt test and its clinical use. *American Orthopsychiatric Association Research Monographs*, No. 3.

Bock, R. D. (1975) *Multivariate statistical methods in behavioral research.* New York: Academic Press.

Bonta, J., Hanson, K., & Law, M. (1998) The prediction of criminal and violent recidivism among mentally disordered offenders. *Psychological Bulletin*, 123, 123-142.

Breiman, L. (1996) Bagging predictors. *Machine Learning*, 24, 123-140.

Breiman, L. (2001) Using iterated bagging to debias regressions. *Machine Learning*, 45, 261-277.

Burgess, E. W. (1928) Factors making for success or failure on parole. In A. A. Bruce, E. W. Burgess, A. J. Harno, & J. Landesco (Eds.), *The workings of the intermediate sentence law and the parole system in Illinois*. Springfield, IL: State Board of Parole.

Busch, K. G., Zagar, R. J., Grove, W. M., Hughes, J. R., Arbit, J., & Bussell, R. E. (2009) Looking forward in records of young adults convicted of sexual homicide, rape, or molestation as youth: risks for reoffending. *Psychological Reports*, 104, 155-184.

Cohen, J., & Cohen, C. (1982) *Statistical power analysis for the behavioral sciences.* Hillsdale, NJ: Erlbaum.

Cronbach, L. J. (1951) Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297-334.

Doll, E. A. (1965) *Manual for Vineland Social Maturity Scale*. Circle Pines, MN: American Guidance Service.

Fox, J. A. (1992) *Sourcebook on criminal justice statistics*. Washington, DC: U.S. Department of Justice.

Gates, A. I., & MacGinitie, W. H. (1965) *Gates-MacGinitie Reading Test technical manual*. New York: Teacher's College Press.

Grove, W. M., & Meehl, P. E. (1996) Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: the clinical–statistical controversy. *Psychology, Public Policy, and Law*, 2, 293-299.

Hosmer, D. W., & Lemeshow, S. (2000) *Applied logistic regression.* New York: Wiley.

Hughes, J. R., Zagar, R. J., Busch, K. G., Grove, W. M., & Arbit, J. (2009) Looking forward in records of youth abused as children: risk for homicidal, violent, and delinquent offenses. *Psychological Reports*, 104, 77-101.

Kelly, T. L., Madden, R., Gardner, E. F., & Ruderman, H. C. (1964) *Stanford Achievement Test manual*. New York: Harcourt, Brace and World.

Kirk, K. (1982) *Experimental design*. New York: Brooks-Cole.

Koppitz, E. M. (1964) *The Bender Gestalt Test for Young Children*. New York: Grune and Stratton.

Lachenbruch, P. A. (1968) On expected values of probabilities of misclassification in discriminant analysis. *Biometrics*, 23, 639-645.

Lidz, C. (1993) The accuracy of predictions of violence to others. *Journal of the American Medical Association*, 269, 1007-1009.

Lipsey, M. W. (1999) Can intervention rehabilitate serious delinquents? *Annals of the American Academy of Political and Social Science*, 564, 142-166.

Lipsey, M. W., & Derzon, J. H. (1998) Prediction of serious or violent delinquency in adolescence and early adulthood. In R. Loeber & D. P. Farrington (Eds.), *Serious and violent juvenile offenders: risk factors and successful interventions*. Thousand Oaks, CA: Sage. Pp. 86-105.

McDonald, R. P. (1970) The theoretical foundations of principal factor analysis, canonical factor analysis, and alpha factor analysis. *British Journal of Mathematical Psychology,* 23, 1-21.

Martinez-Munoz, G., & Suarez, A. (2010) Out-of-bag estimation of the optimal sample size in bagging. *Machine Learning*, 43, 143-152.

Meehl, P. (1954) *Clinical versus statistical prediction*. Minneapolis, MN: Univer. of Minnesota Press.

Monahan, J. (1996) Violence prediction. *Criminal Justice and Behavior*, 23, 107-120.

Nunnally, J. (1978) *Psychometric theory*. New York: McGraw-Hill.

Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. A. (1998) *Violent offenders: appraising and managing risk*. Washington, DC: American Psychological Association.

Revelle, W. (1979) Hierarchical cluster analysis and the internal structure of tests. *Multivariate Behavioral Research*, 13, 57-74.

Rice, M. E., Harris, G. T., & Quinsey, V. L. (1996) Treatment for forensic patients. In B. D. Sales & S. A. Shah (Eds.), *Mental health and law: research, policy and services.* New York: Academic Press. Pp. 141-189.

Rossi, R. (2010) Behavior issues drop at 6 schools. *Chicago Sun Times*, May 27.

Saulny, S. (2009) Focus in Chicago: students at risk of violence. *New York Times*, October 7, p. 1.

Sepajak, D. S. (1983) Clinical predictions of dangerousness: two-year follow-up of 408 pretrial forensic cases. *Bulletin of the American Academy of Psychiatry and the Law*, 11, 171-180.

Shao, J. (1996) Bootstrap model selection. *Journal of the American Statistical Association*, 91, 655-665.

Shelton, D. L., & Banchero, S. (2009) Seeking safe passage: curing a public epidemic. *Chicago Tribune*, October 7, p. 1.

Stevens, J. (1986) *Applied multivariate analysis for the social sciences*. Hillsdale, NJ: Erlbaum.

Terman, L. M., & Merrill, M. A. (1973) *Stanford Binet Intelligence Scale: manual for third edition*. Boston, MA: Houghton-Mifflin.

U.S. Census Bureau. (2006) *Household income in the United States, income quintiles*. Washington, DC: U.S. Government Printing Office.

Wechsler, D. (1974) *Wechsler Intelligence Scale–Revised manual*. New York: Psychological Corp.

World Health Organization. (1977) *Manual for the international classification of diseases, injuries, and causes of death.* (9th ed.) Geneva, Switzerland: Author.

Zagar, R. J., Arbit, J., Hughes, J. R., Bussell, R. E., & Busch, K. G. (1989) Developmental and disruptive behavior disorders among delinquents. *Journal of the American Academy of Child and Adolescent Psychiatry*, 28, 437-440.

Zagar, R. J., Busch, K. G., Grove, W. M., & Hughes, J. R. (2009a) Can violent (re)offense be predicted? A review of the role of the clinician and use of actuarial tests in light of new data. *Psychological Reports*, 104, 247-277.

Zagar, R. J., Busch, K. G., Grove, W. M., & Hughes, J. R. (2009b) Summary of studies of abused infants and children later homicidal, and homicidal, assaulting later homicidal, and sexual homicidal youth and adults. *Psychological Reports*, 104, 17-45.

Zagar, R. J., Busch, K. G., Grove, W. M., Hughes, J. R., & Arbit, J. (2009a) Looking forward and backward in records for risks among homicidal youth. *Psychological Reports,* 104, 103-127.

Zagar, R. J., Busch, K. G., Grove, W. M., Hughes, J. R., & Arbit, J. (2009b) Looking forward in records of young adults who were convicted of homicide or assault as youth: risks for reoffending. *Psychological Reports*, 104, 129-154.

Zagar, R. J., Busch, K. G., Grove, W. M., Hughes, J. R., & Arbit, J. (2009c) Looking forward in records of youth abused as infants: risks for homicidal, violent, and delinquent offenses. *Psychological Reports*, 104, 47-75.

Zagar, R. J., Busch, K. G., & Hughes, J. R. (2009) Empirical risk factors for delinquency and best treatments: where do we go from here? *Psychological Reports*, 104, 279-308.

Zagar, R. J., Zagar, A. K., Bartikowski, B., & Busch, K. G. (2009) Cost comparisons of raising a child from birth to 17 years among samples of abused, delinquent, violent, and homicidal youth using victimization and justice system estimates. *Psychological Reports*, 104, 309-338.

Zagar, R. J., Zagar, A. K., Bartikowski, B., Busch, K. G., & Stark, R. (2009) Accepted legal applications of actuarial testing and delinquency intervention: examples of savings in real-life situations. *Psychological Reports*, 104, 339-362.

Accepted December 8, 2010.

APPENDIX

Experimental and Control Examinations Gathered From Court Records and Psychologists' Reviews

Type of Record[a]	Information Reviewed[a,b]
Pediatric and medical: pregnancy, neonatal, hospital	Central nervous system (CNS) problems; postpartum and general health of mother and infant; substance abuse by parents influencing pregnancy or fetus; psychiatric problems
	(IC): Neurological, respiratory, renal, hepatic, musculoskeletal, endocrine, circulatory, gastrointestinal, and genito-urinary systems (ICD–9)
	(IC): Visual, hearing, and speech disorders, headaches, blackouts, tremors, epilepsy, head and neck injuries, skull fractures, cerebral palsy, and hydrocephalus
	(IC): Anoxia, prematurity, fetal alcohol, cocaine, heroin, nicotine, marijuana, transient neonatal diabetes, apnea, jaundice, and breech or Caesarian delivery
	(IC): Hospitalization, enuresis, hyperactivity, developmental delay, overdose, and posttraumatic stress
	(CP): Weekly use of one or more substances: alcohol, cannabis, cocaine, hallucinogens, heroin, amphetamines, and/or aromatics (paint or glue)
	(ICP): Organic brain syndrome, malingering, brief reactive psychosis, major affective, schizoaffective, paranoid disorders, and schizophrenia
Psychological, intelligence, school	Review and DSM-III–R coding by psychologist
	(I): Bayley Scales of Infant Development; Stanford-Binet Intelligence Scale, Alternate Forms L-M, Third Edition; Vineland Social Maturity Scale
	(C): Wechsler Intelligence Scale for Children–Revised (WISC–R); Vineland Social Maturity Scale; Bender Visual-Motor Gestalt Test Koppitz scores; Gates-MacGinitie Achievement Test, Survey D, Form 1, Reading Speed and Accuracy, Vocabulary, and Comprehension; Stanford Achievement Test, Form W, Intermediate I, Arithmetic
	(IC): Expressive or receptive language, ADD, ADHD, pervasive developmental delay, learning or conduct disorder, IQ less than 70, and executive function
Psychiatric	Infant, child, teen, parent, and relatives' and guardians' mental status
	(IC): Fainting, blackouts, seizures, paralysis, local weakness, numbness, tingling, tremors, memory, nervousness, tension, mood, depression, medications, allergies, substance use, number and type of hallucinations, overdoses, and hospitalizations
	(ICP): History of developmental, family, and medical conditions
	(ICP): Education, housing, employment, finances, future, insurance, habits, routines, exercise, and general health of family
	(ICP): Number and type of offenses and convictions

(continued on next page)

APPENDIX (CONT'D)

EXPERIMENTAL AND CONTROL EXAMINATIONS GATHERED FROM COURT RECORDS AND PSYCHOLOGISTS' REVIEWS

Type of Record[a]	Information Reviewed[a,b]
Social worker, probation officer, child protection worker	General demographic and family information, health, antisocial or criminal activity, school, substance abuse
	(ICP): Residences, schools, family income, parents' work, family composition, home or school moves, community, and school history
	(ICP): Complications of pregnancy, childbirth, puerperium, and certain conditions originating in the perinatal period; injury, burn, poisoning, and fetal substance exposure; respiratory, infectious, neurological, and genitourinary conditions; epilepsy, IQ less than 70, ADHD; and general medical history
	(ICP): Family offenses; child, teen, and adult convictions; criminally violent family; physical abuse of parent or youth; gang membership; prior arrests; and weapon use or possession
	(IC): Underachievement, use of special education services, truancy, suspension, and expulsion
	(ICP): Alcohol abuse, substance abuse, and alcohol and substance abuse

[a]Information from records of infant, child (also youth, assaulter, rapist), or parents, indicated as (I), (C), or (P), respectively, or all (ICP). "Infant" refers to all records obtained for the Abused Infant and Control group members, at any age. "Child" refers to all records obtained at any age for the Abused Children and Control groups, Homicidal Youth, Nonviolent Delinquent, and Control, Assaulter and Control groups, and Rapist-Molester and Control groups. (See four other studies in this series.) "Parent" refers to the infants' parents or guardians. Left-hand column entries indicate focus of the record review; subentries indicate specific data recorded.

[b]The mean age at which the records were reviewed for the youth that comprised the Abused Infants and Control groups was 3.12 yr. ± 1.48. This is provided as general information only, since the analyses within these groups were not dependent on the age at record review nor, for the most part, the age at which testing was done for court purposes. See text for details. Child age is the mean age at which the records were reviewed for those who comprised the Abused Children and Control groups, 12.85 yr. ± 2.74. Youth age is the mean age at which the records were reviewed from Homicidal Youth and Control groups, 14.9 yr. ± 1.4. Assaulter age is the mean age at which the records were reviewed for Assaulter and Control groups, 14.1 yr. ± 1.7. Rapist and Molester age is the mean age at which the records were reviewed for the Rapist, Molester, and Control groups, 14.2 yr. ± 1.5. (See four other studies in this series.)