PROJECT ABSTRACT

The age crime curve and its implications for effect size change over the life course: Establishing meaningful estimates of expected change

E. C. Hedberg, Ph.D.
Senior Data Scientist, Academic Research Centers, NORC at the University of Chicago

Charles Katz, Ph.D.
Professor, School of Criminology & Criminal Justice at Arizona State University
Director, Center for Violence Prevention at Arizona State University

Our project seeks to provide context for the effect sizes estimated from evaluations of interventions seeking to reduce criminal and other deviant behavior. Many interventions seek to influence the rate of criminal or deviant activity of individuals at various ages. [CHUCK—EXAMPLES]. Such projects estimate statistical models which can yield scale-free effect size metrics for use in meta-analysis or more informal cross-study comparisons. However, a common but difficult to answer question about these effect sizes is "is it large?"

We propose to use publicly available Uniform Crime Reports and Census Data to estimate risk over the life course for a variety of ages typically targeted by evaluations. Year-over-year changes in these risk estimates are then the basis of effect sizes that provide standardized metrics of the growth and desistance of criminal behavior over the life course.

These empirical benchmarks can then be used by evaluation researchers to contextualize their achieved impacts. This project, in effect, helps to answer the question of "what is a large effect size" by establishing natural change. Once the parameters are estimated, we will produce an online tool where researchers can examine these tables interactively to easily access our findings.

Significance

In this section we detail the purpose of our study by establishing the need for empirical benchmarks.

What do Quantitative Evaluations Estimate?

Quantitative evaluations then seek to estimate a difference between groups exposed to a program or intervention ("treatment") and another group without exposure, comparison or "control." Typically, impacts are estimated by taking differences in means or rates (of some outcome) between these groups through descriptive or model-based statistical techniques. The estimated difference is then a quantification of the intervention's "impact."

The impact is then presented in the units of the outcome (e.g., rate or likelihood of offence, mean on a risk factor scale, etc.) and the ratio of that estimate to its standard error (a statistical measure of uncertainty) is then converted into a probability of the result (or one more extreme) given the assumption of "no effect" (aka, the null hypothesis). This probability is the "p-value" and typically values less than .05 are considered "statistically significant."

Is the p-value a Good Measure of Impact?

The use of p-values and phrases such as "statistically significant" have long come under intense scrutiny. The primary reason is that p-values are misused. The two prime misuses are 1) interpreting the value to be the probably of the null hypothesis (which it is not) or 2) interpreting the value to be a measure of the *strength* of the association (which it is not). The latter is a problem that has caused considerable damage to policy and even human life (Ziliak and McCloskey 2008). The American Statistical Association (ASA) has repeatedly sought to provide guidance and clarify the meaning of p-values (Wasserstein and Lazar 2017), but the ASA has recently proposed removing their usage entirely (Wasserstein, Schirm, and Lazar 2019).

The Effect Size as a Better Measure of Impact

A key statistic that *does* summarize the strength of an association is the effect size. Effect sizes take many forms, such as correlation coefficients or ratios, but the most common is the standardized difference between means (SDBM), such as Cohen's d (Cohen 1992) or Hedges' g. The SDBM statistic is simple, given two means, μ_0 and μ_1, for, say, control and treatment groups, the SDBM effect size is simply the difference in the means divided by the population standard deviation (σ, which is the variation in the outcome without the introduction of treatment effects)

$$SDBM = \frac{\mu_1 - \mu_0}{\sigma} = \frac{\Delta}{\sigma}$$

This quantity is simply the difference between treatment and control in standard deviation units. Thus, the effect size is the only measure of the strength of the association, and its estimate does not depend on sample size.[1] Effect sizes can (and should) be computed from such studies, but the magnitudes are difficult to interpret without context (Hedberg 2018).

Problem: What is a Big Effect Size?

In applied research, p-values have been the preference because there are near universal conventions (e.g., $p < .05$ is statistically significant). Unfortunately, there are no usefully broad conventions for effect sizes. Cohen introduce some conventions for psychological research (e.g.,

[1] Effect sizes only relate to p-values though sample size. Holding the sample size as fixed, larger effect sizes are associated with smaller p-values, but unfortunately the reverse is also true, holding the effect size constant, larger sample sizes are associated with smaller p-values.

a difference in .5 standard deviations is a “medium” effect size), but these are not preferable for other fields because the context is very important (e.g, .5 standard deviations may be very large in criminology outcomes, but not meaningful in cancer research).

Contextual Benchmarks in Other Fields

In Health, effect sizes in the form of risk or odds ratios have prima-facia value. In the context of disease or death, knowing that a treatment reduces the odds of death by 50 percent is straightforward to understand. These effect sizes are simple to calculate from statistical models as they are based on simple dichotomous outcomes, say death or not, and can be aggregated into a probability, percent chance, or odds. Then, either descriptive or model-based statistics can produce ratios between treatment and control. If the chance of death for treatment is 1 percent, and the chance for control is 3 percent, then the effect size is a risk ratio of 1:3.

In many social sciences, the outcomes are less defined. In psychological fields, such as education, outcomes are typically based on standardized scales like the ACT or SAT standardized test. Thus, a common effect size is not a ratio of chances, but instead a SDBM statistic. This solves the problem of varying scale metrics, allowing studies to be compared to each other, but it does not solve the inherent problem of knowing “what is big?”
In education research, one solution has been to derive year-over-year effect sizes that summarize typical changes in outcomes for various grades (Hedberg 2016; Hill et al. 2008). For example, using the mean fall score $\bar{Y}_{Fall}$, the mean spring score $\bar{Y}_{Spring}$ and the spring standard deviation, σ_{Spring}, the year-over year gains in SDBM effect size units can be computed as
$$\delta = \frac{\bar{Y}_{Spring} - \bar{Y}_{Fall}}{\sigma_{Spring}}.$$

These statistics allow researchers to establish the context of an impact study. For example, typical growth in reading for Kindergartners is 1.5 standard deviations (Hill et al. 2008). If an intervention achieves an effect size of .5 standard deviations, this impact represents a 1/3 of typical growth, which is important. If the same study achieved.05 standard deviations, even if statistically significant, the result should not be met with excitement. However, typical growth for High School students is about .1 standard deviations, making an impact of .05 standard deviations worth 1/2 of typical growth, which is impressive. Thus, even within a field of study, context matters.

The Proposed Study

We propose to produce age-based empirical benchmarks of changes in criminal behavior. We propose to do this by using aggregate age-crime statistics to estimate age-crime polynomial curves. Such curves have long been the subject of criminological analysis, e.g., Farrington (1986). From those curves, we can manipulate the parameters produce age-based changes in log odds-ratio and SDBM effect size units. This will produce a similar “context” that can be used by researchers and evaluators to better understand their achieved effect sizes.

This project will produce the following products

1) Tables of age-based effect sizes for a variety of outcomes and demographic groups

2) An online tool where researchers can look up subsets of the many hundreds of estimates produced, and to also enter data to produce their own age-based effect sizes.
3) Academic papers and presentations

This project has the advantage is that it does not require individual data collection. Our method, detailed below, is based on using aggregate rates that can be derived from Uniform Crime Reports and Census Data, for a variety of outcomes, demographic groups, geographies, and cross-products thereof.

Methodology

Our proposed methodology is to utilize aggregate data of rates by age (and race and gender) for a variety of outcomes. With the aggregate data, we can represent the age-crime curve mathematically, then use calculus to find the instantaneous effect size for each age.

The first step is to convert the typical rate as presented in criminology, the rate of a crime per 100,000 individuals, into an individual probability. The probability metric can further be transformed in log-odds using the logit function suitable for linear modeling. If variable r is a rate per 100,000 persons, then we can convert r into the logit y with

$$y = \Lambda(r \times 10^{-5})$$

where $r \times 10^{-5}$ converts the rate to an individual chance, p, and Λ is the logit transformation

$$\Lambda(p) = \ln\left(\frac{p}{1-p}\right)$$

We then fit these log-odds to a polynomial linear model based on age; e.g., a 5th polynomial model would be

$y_{age} = \alpha + \beta_1 age + \beta_2 age^2 + \beta_3 age^3 + \beta_4 age^4 + \beta_5 age^5 = \alpha + \sum \beta_q age^q$,

for example. This creates an algebraic function of the log-odds based on age. We then then use basic calculus to find the slope of this curve at each value of age. If we note a linear function as $f(x)$, then the slope at any point of x is noted as $f'(x)$ or $\frac{\Delta y}{\Delta x}$, for example.

The rules of calculus that are applied are

the constant rule:	the power rule:	the constant product rule:	and the sum rule:
if $f(x) = c$, then $f'(x) = 0$;	if $f(x) = x^p$, then $f'(x) = px^{p-1}$;	if $f(x) = bg(x)$, then $f'(x) = bg'(x)$;	if $f(x) = g(x) + h(x)$, then $f'(x) = g'(x) + h'(x)$.

These rules combine to state that if the age function is defined as

$f(age) = \alpha + \sum \beta_q age^q$,

then its first derivative is

$f'(age) = \sum_q q\beta_q age^{q-1}$.

Thus, the slope of the age-crime curve for a, say, 5-degree polynomial function is

$$\frac{\Delta y}{\Delta age} = \beta_1 + 2\beta_2 age + 3\beta_3 age^2 + 4\beta_4 age^3 + 5\beta_5 age^4.$$

This function, given that rate is transformed into a logit metric, gives the difference in log-odds for a unit increase in age for a particular age.

One the model is fit to achieve an algebraic expression, and we calculate the first derivative, we then convert the difference in log-odds into a Cohen's d effect size. The meta-analysis literature (Borenstein et al. 2009) provides a formula to turn a difference in log-odds into a SDBM effect size. If a slope is a difference in means between x and $x+1$, then we can create a SDBM statistic by dividing the slope by the variance. In the case of log-odds, the variance is

$$V\{\Lambda\} = \frac{\pi^2}{3}$$

Thus, the effect size at each age is equal to (in the case of 5th degree polynomial)

$$d_{age} = \frac{\frac{\Delta y}{\Delta age}}{\sqrt{\frac{\pi^2}{3}}} = \frac{\sqrt{3}}{\pi}(\beta_1 + 2\beta_2 age + 3\beta_3 age^2 + 4\beta_4 age^3 + 5\beta_5 age^4)$$

Example Analysis

In this section we employ data collected by Cohen and Rosenfeld (2005)[2] to create a table of effect sizes for robbery. These data are from decades ago (1965 to 1985) but serve the purpose of being an example. The data points for robbery, by race and year, are shown in Figure 1. Each year-race panel includes 7 data points for age categories for midpoints 12.5, 16.5, 18.5, 22, 29.5, 39.5, and 60. Note that our effect sizes can be computed for any age once the curve is fitted. We used a 5-degree polynomial to allow 1 degree of freedom (5 age variables plus the constant is 6, with 7 total observations). A separate model was fit for each race-age combination, and an average model was fit for each year. Each model explained nearly 100 percent of the variation.

We then calculated the SDBM effect sizes for a 1-year age increase for the ages 12, 15, 18, and 25. The results are presented in Table 1. From this table, we make three major observations. First, the effect sizes increase as a function of time. For example, 12-year-olds are increasing their rate of robbery by about .25 standard deviations in 1965, but this increases to .4 by 1985. This indicates that these curves are not stable over time. Second, the rates for white and non-whites are very similar. Third, of course, these effects differ by age, with a decrease evident between 18 and 25.

Implications

[2] https://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/9589

As a tool for context, these numbers can provide information on natural change by age. For example, if an intervention in 1985 found that robberies decreased as a result of a program by -.06 standard deviations for 15 year holds, many in the research community would consider that effect small. However, -.06 accounts for 50 percent of the increase in robbery incidences at that age, meaning that the intervention was quite effective.

This also has implications for power. Many studies are powered to detect effects around --.3???, which this study shows is actually a very large effect for most ages. In other words, studies are powered for effects they are unlikely to achieve. This means that the studies are under-powered and unlikely to detect effects.

Figure 1: Data of Age-by-Race Crime Rates for Robbery by Year

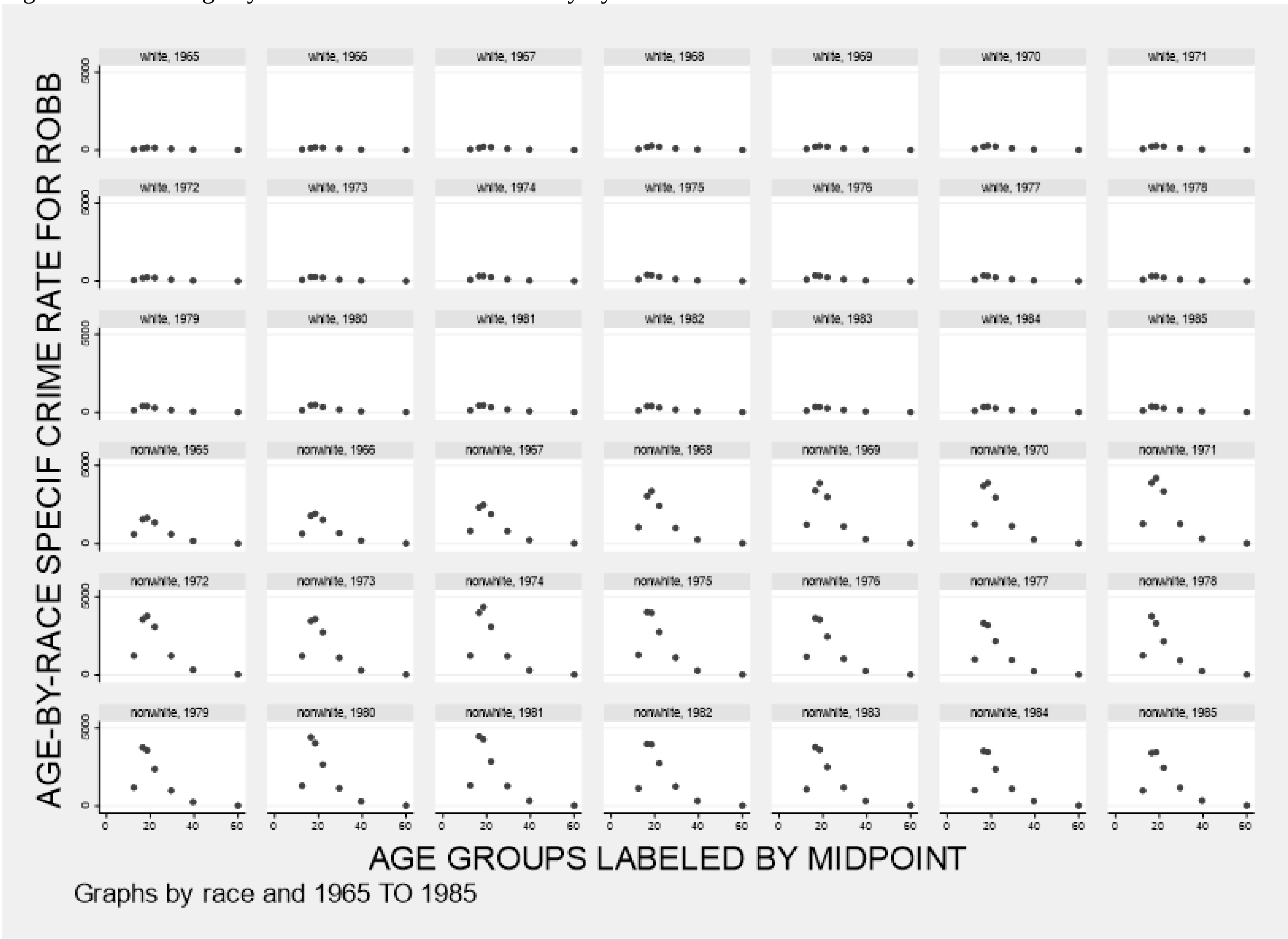

Table 1: Effect sizes												
	Age 12			Age 15			Age 18			Age 25		
Yr.	All	White	Non-White	All	White	Non-White	All	White	Non-White	All	White	Non-White
1985	0.40	0.40	0.40	0.12	0.12	0.13	-0.01	-0.01	-0.01	-0.05	-0.05	-0.06
1984	0.41	0.40	0.42	0.13	0.13	0.13	-0.02	-0.01	-0.02	-0.06	-0.05	-0.06
1983	0.42	0.41	0.43	0.12	0.13	0.12	-0.02	-0.01	-0.02	-0.05	-0.05	-0.06
1982	0.41	0.40	0.42	0.13	0.13	0.13	-0.01	-0.01	-0.02	-0.06	-0.05	-0.06
1981	0.41	0.40	0.42	0.12	0.13	0.12	-0.02	-0.01	-0.03	-0.06	-0.05	-0.07
1980	0.41	0.40	0.43	0.12	0.13	0.12	-0.02	-0.01	-0.03	-0.06	-0.06	-0.07
1979	0.39	0.38	0.40	0.11	0.12	0.11	-0.02	-0.02	-0.03	-0.06	-0.06	-0.07
1978	0.38	0.37	0.40	0.11	0.11	0.10	-0.03	-0.02	-0.04	-0.06	-0.06	-0.07
1977	0.41	0.40	0.41	0.12	0.12	0.12	-0.02	-0.02	-0.02	-0.06	-0.05	-0.07
1976	0.38	0.39	0.38	0.11	0.11	0.12	-0.02	-0.02	-0.02	-0.06	-0.05	-0.07
1975	0.38	0.38	0.38	0.12	0.11	0.12	-0.02	-0.02	-0.02	-0.06	-0.06	-0.07
1974	0.37	0.37	0.37	0.13	0.12	0.13	-0.01	-0.01	0.00	-0.07	-0.06	-0.08
1973	0.32	0.31	0.32	0.11	0.11	0.11	0.00	0.00	0.00	-0.06	-0.06	-0.07
1972	0.29	0.28	0.30	0.12	0.12	0.12	0.02	0.03	0.01	-0.06	-0.06	-0.07
1971	0.31	0.31	0.32	0.13	0.13	0.12	0.01	0.02	0.01	-0.07	-0.06	-0.07
1970	0.32	0.31	0.33	0.13	0.13	0.12	0.01	0.02	0.00	-0.07	-0.06	-0.07
1969	0.29	0.29	0.29	0.12	0.12	0.12	0.01	0.02	0.01	-0.07	-0.06	-0.08
1968	0.32	0.31	0.33	0.12	0.13	0.12	0.01	0.03	0.00	-0.06	-0.06	-0.07
1967	0.30	0.28	0.32	0.13	0.14	0.11	0.02	0.05	0.00	-0.06	-0.06	-0.07
1966	0.29	0.28	0.30	0.12	0.13	0.11	0.02	0.04	0.00	-0.06	-0.05	-0.06
1965	0.25	0.23	0.27	0.12	0.13	0.10	0.03	0.05	0.00	-0.05	-0.05	-0.06

References

Borenstein, Michael, Larry V. Hedges, Julian P. T. Higgins, and Hannah R. Rothstein. 2009. *Introduction to Meta-Analysis*. Chichester, UK: John Wiley & Sons, Ltd.

Cohen, Jacob. 1992. "A Power Primer." *Psychological Bulletin*.

Cohen, Jacqueline and Richard Rosenfeld. 2005. "Age-by-Race Specific Crime Rates, 1965-1985: [United States]."

Farrington, David P. 1986. "Age and Crime." *Crime and Justice* 7:189–250.

Hedberg, E. C. 2016. "Academic and Behavioral Design Parameters for Cluster Randomized Trials in Kindergarten: An Analysis of the Early Childhood Longitudinal Study 2011 Kindergarten Cohort (ECLS-K 2011)." *Evaluation Review* 40(4).

Hedberg, E. C. 2018. *Introduction to Power Analysis: Two-Group Studies*.

Hill, Carolyn J., Howard S. Bloom, Alison Rebeck Black, and Mark W. Lipsey. 2008. "Empirical Benchmarks for Interpreting Effect Sizes in Research." *Child Development Perspectives*.

Wasserstein, Ronald L. and Nicole A. Lazar. 2017. "The ASA's Statement on p-Values: Context, Process, and Purpose." *The American Statistician*.

Wasserstein, Ronald L., Allen L. Schirm, and Nicole A. Lazar. 2019. "Moving to a World Beyond " P." *The American Statistician* 73(sup1):1–19.

Ziliak, Steve and Deirdre Nansen McCloskey. 2008. *The Cult of Statistical Significance: How the Standard Error Costs Us Jobs, Justice, and Lives.* University of Michigan Press.