



The big data potential of epidemiological studies for criminology and forensics



Matt DeLisi

Iowa State University, 203A East Hall, Ames, IA 50011-1070, United States

ARTICLE INFO

Article history:

Received 3 May 2016

Received in revised form

29 August 2016

Accepted 5 September 2016

Available online 6 September 2016

Keywords:

Epidemiology

Criminology

Big data

Forensics

Criminal justice

ABSTRACT

Big data, the analysis of original datasets with large samples ranging from ~30,000 to one million participants to mine unexplored data, has been under-utilized in criminology. However, there have been recent calls for greater synthesis between epidemiology and criminology and a small number of scholars have utilized epidemiological studies that were designed to measure alcohol and substance use to harvest behavioral and psychiatric measures that relate to the study of crime. These studies have been helpful in producing knowledge about the most serious, violent, and chronic offenders, but applications to more pathological forensic populations is lagging. Unfortunately, big data relating to crime and justice are restricted and limited to criminal justice purposes and not easily available to the research community. Thus, the study of criminal and forensic populations is limited in terms of data volume, velocity, and variety. Additional forays into epidemiology, increased use of available online judicial and correctional data, and unknown new frontiers are needed to bring criminology up to speed in the big data arena.

© 2016 Elsevier Ltd and Faculty of Forensic and Legal Medicine. All rights reserved.

1. Introduction

For most of its existence, academic criminology has been largely devoid of any reference to epidemiology except for occasional works. For instance, Cressey noted that an epidemiological understanding of the statistical distribution of crime in time and space was an important mission for criminology. And while many studies repeatedly found that crime was disproportionately committed by males compared to females, African Americans compared to whites, and youth compared to older adults, there was nevertheless a lack of follow-up on Cressey's suggestion.¹ Indeed, it was not until the late 1970s that theoretical and empirical research began to take seriously the idea of epidemiology in terms of understanding the distribution of crime and victimization across various social statuses (e.g., age, sex, race, social class, personality).^{2–4} Once it was understood that crime and victimization disproportionately occurred among similarly situated young males congregating in specific types of contexts, enforcement and prevention efforts could tailor their modalities accordingly.

In recent years, Akers and Lanier formally called for epidemiological criminology to bring the methods and concepts from epidemiology—primarily concerned with health and illness in the

population in the interests of public health and preventive medicine—to the study of criminal offenders who often display serious health-compromising behaviors and impose a substantial public health burden.^{5,6} Similar to the muted reaction to Cressey decades earlier, there has not been a broad response to the more recent calls for epidemiological criminology. However, some criminologists have followed the suggestion and incorporated epidemiological datasets into criminal justice theory and research. Fundamentally, recent research has shown the promise of mining big data for answering an array of questions relating to criminal and forensic populations.

2. Big data epidemiology

To date, scholars have primarily utilized two large-scale data sources to mine epidemiological data for criminological and forensic purposes. The first dataset is the National Epidemiologic Survey on Alcohol and Related Conditions or NESARC. The NESARC contains data on 43,093 participants and includes scores of questions on alcohol and substance use disorders, psychiatric disorders, personality disorders, and a bevy of behavioral indicators. The NESARC oversampled young adults between the ages of 18–24, Hispanics, and African Americans and has a response rate of 81%. The second dataset is the National Survey on Drug Use and Health or NSDUH. The 2010 NSDUH is a nationwide survey of 68,487

E-mail address: delisi@iastate.edu.

randomly selected participants aged 12 or older. It is designed primarily to measure prevalence, patterns, and trends in alcohol, tobacco, and illicit substance use and abuse, but like the NESARC, contains dozens of items relating to psychological and psychiatric features and antisocial conduct. Some studies employ multiple years of the NSDUH to produce samples between 200,000 and 300,000 participants and allow for trends analyses.^{7,8}

3. Criminal and forensic content areas informed by big data

3.1. Prevalence and correlates of assorted typologies

Study groups have utilized the richness of big data to create behavioral profiles of rare offender types that were heretofore studied via case studies or very small clinical samples from a single jurisdiction. Using NESARC data, Vaughn and colleagues reported that the prevalence of fire setting in the United States was 1% and that fire setters were characterized by behavioral disorders, substance use disorders, and a family history of antisocial behavior.⁹ Subsequent studies of individuals who engage in cruelty to animals (lifetime prevalence of 1.8%), bullying (lifetime prevalence of 6%), criminal victimization (lifetime prevalence 4.1%), delinquency abstinence (lifetime prevalence of 11%), reckless driving (lifetime prevalence of 25.2%), being a drifter (lifetime prevalence of 3.4%), deliberate self-harm (lifetime prevalence of 2.9%), handgun carrying (past year prevalence of 3.1%), and truancy (past month prevalence of 11%) have also revealed that offender groups have distinct demographic, behavioral, and psychiatric profiles while also displaying a general liability for externalizing behaviors that are significantly more severe than unaffected individuals.^{10–18}

3.2. Pathological offenders

Big data has also been useful to reinforce findings on a small group of pernicious offenders who denote lifelong antisocial conduct. Data from the NESARC and NSDUH converged on an isomorphic group of severe offenders who comprised 5% (5.3% and 4.7%, respectively) of the population and who had the most violent, serious, and chronic antisocial careers.^{19,20} These severe offenders are noteworthy for antisocial personality features and a psychological profile that favors risk-taking, poor self-regulation, and limited emotional regulation.

3.3. Subtypes

Big data permits disaggregated analyses to explore the heterogeneity of the population vis-à-vis specific forms of crime, substance use, or psychiatric condition. Disaggregation of data permits nuanced analyses that can inform prevention and policy efforts. For instance, a study of 1226 individuals with Antisocial Personality Disorder derived from the NESARC found that about 70% of affected individuals nevertheless have higher socioeconomic functioning and there was little intergenerational continuity in antisocial conduct. In contrast, about 9.4% of affected individuals had parents and children with conduct problems and 20.3% of affected individuals have multigenerational histories of problem behaviors. The latter group was characterized by clinical and personality disorders, alcohol abuse, versatile criminal behaviors, and diverse acts of physical violence.²¹ Prevention programs should target antisocial individuals for whom crime runs in their family in order to effect the greatest potential crime reduction. In contrast, most individuals with Antisocial Personality Disorder nevertheless are able to achieve a modicum of prosocial functioning.

Drawing on data from more than 283,000 participants in the 2008 to 2012 NSDUH, DeLisi and colleagues discovered three latent

classes of substance use among those charges with a serious violent offense in the United States. They found that 61.2% of the sample had limited substance use morbidity and 28.2% had comorbid alcohol and marijuana use disorders. About 10.7% of the sample displayed polydrug abuse and dependence and had severe criminal careers. For many persons charged with a serious violent offense, the behavior is not generally part and parcel of a broader drug lifestyle and criminal involvement. But for the most severe offenders, violence and substance use were commingled.²² Additional studies of subtypes of sexually-impulsive adults, juvenile detainees, drug sellers, criminal justice system clients, immigrants, and individuals with personality disorders similarly reveal the importance of broadband temperamental and personality deficits relating to low self-control and poor emotional regulation as central to antisocial conduct and collateral health and social burdens.^{23–31} These findings are also consonant with leading theories in the criminal justice sciences.^{32–35}

4. Discussion

Recent criminological interest in epidemiology has occurred on two fronts, one conceptual and the other empirical that took advantage of the big data yield that epidemiological samples offer. This research has been successful at elucidating—on a panoramic scale—the empirical existence and cascade of behavioral and health problems that unfold from the small core of the most severe offenders. These offenders are disproportionately responsible for the bulk of criminal activity in a population, disproportionately responsible for most of the murders, rapes, armed robberies, and kidnappings occurring in a population, disproportionately responsible for drug activity, drug sales, drug trafficking and the corresponding emergency medical resources that are used in their wake, and disproportionately responsible for the social burdens associated with crime. Big data research has also shown that even among persons with clinical conditions or who display severe conduct problems, there is also heterogeneity within the seemingly narrow typologies, and this heterogeneity reveals usually the existence of a large, normative group, one or more moderate groups, and a small pathological group.^{15,19–23,32} These grouping within groupings need different interventions for behavioral change with the former group (of normative offenders) requiring the least time and resources and the latter group (of pathological offenders) requiring the most intensive interventions and/or most punitive correctional response.

There are still pressing research needs that big data can serve to enhance understanding of truly pathological offenders that are usually only found in forensic samples. In this sense, epidemiology has evolved from the general population description of delinquency discussed nearly 60 years ago by Cressey to the severe offenders of the research described here to offenders who we know less about, those who commit mass homicide, serial homicide, sexual homicide, and other rare offenses. Although the NESARC and NSDUH are extraordinary datasets, even they cannot encompass forensic inmates. For this reason, it is imperative that criminologists partner with practitioners and juvenile justice and criminal justice agencies in order to access their precious, but often, analytically-untouched data. Many criminal justice organizations have an outreach or research component that conducts analyses of their own data for presentations to the community and other criminal justice organizations. Agencies often desire more analytically sophisticated analyses, which is something that social scientists can contribute. A typical criminal justice agency-researcher partnership involves a contract or memorandum of understanding where the researcher develops a proposal to obtain de-identified data, describes the research questions that will be explored, and delineates

the products or deliverables to the agency. In turn, the agency completes an IRB-like approval process that usually requires the ultimate approval of the local chief judicial officer (e.g., chief district judge). In the best scholar-practitioner collaborations, the academic and practitioner jointly conduct research, present their findings at practitioner and academic conferences, and publish their work as coauthors. In stronger collaborative relationships, there is greater likelihood that data are being analyzed and interpreted correctly and practitioners are critical in providing qualitative insights into the quantitative data that are being analyzed.

Fortunately, some of these partnerships are already in place and others are emerging. Barrett and his colleagues conducted a matched-control study of 99,602 delinquent youth and matched controls using data from South Carolina and found that prior diagnoses for conduct problems were the strongest predictors of delinquency.³⁶ Baglivio and his colleagues have utilized population data of between 60,000 and 222,640 juvenile offenders encompassing upwards of 363,617 records in Florida to provide large-scale understanding of the developmental sequelae for serious, violent, and chronic juvenile delinquents. A particularly pernicious issue is the presence of adverse childhood experiences. Each additional adverse childhood experience increases the likelihood of pathological offending by 35% and youth with extensive victimization histories overwhelmingly develop into career offenders.^{37–39} Bales and colleagues have analyzed Florida prisoner data samples ranging up to 80,000 offenders to examine the effects of incarceration on recidivism.⁴⁰ Using data from 297,600 offenders from Washington, Hamilton and his colleagues examined the predictive validity of a static risk assessment instrument, and interestingly, found that a big data approach to correctional research was not superior to traditional sampling approaches.⁴¹ It is critical for scholar-practitioners to employ their considerable data access to the most severe and often rare offenders to enhance understanding of their forensic and criminological characteristics.

There are nationally representative datasets that the United States government employs, such as the National Crime Information Center (NCIC) which contains more than 12 million active criminal records and handles more than 12 million data transactions each day. However, access to NCIC is restricted and not permitted to researchers. The Combined DNA Index System (CODIS) is similarly restricted and not permitted to researchers. There are scores of local, county, and state criminal justice organizations that make their data freely available. These include online access to courts records and online access to correctional records that often contain information on instant conviction offense, sentence, misconduct, prior commitments to prison, and other information. These types of online data have been used in prior research, but the data collection is labor intensive and expensive.^{42,43} In other words, there is limited data volume, velocity, and variety in criminology. Although researchers can be creative with data collection, it is clear that “outside” data sources are helpful. And since epidemiological samples contain so many measures relating to behavioral health, they are a perfect way for criminologists to bring a big data sensibility to their research.

5. Conclusion

Big data *per se* has not been utilized in criminology. However, there have been recent calls for greater synthesis between epidemiology and criminology and a small number of scholars have utilized epidemiological studies that were designed to measure alcohol and substance use to harvest behavioral and psychiatric measures that relate to the study of crime. Such creativity is needed because true big data initiatives that the government owns are limited to criminal justice purposes and restricted to researchers. In

this sense, the study of criminal and forensic populations is limited in terms of data volume, velocity, and variety. Additional forays into epidemiology, increased use of available online judicial and correctional data, and unknown new frontiers are needed to bring criminology up to speed in the big data arena.

Conflicts of interest

The current study was not sponsored by any funder and there are no conflicts of interest.

References

- Cressey DR. Epidemiology and individual conduct: a case from criminology. *Pac Sociol Rev.* 1960;3:47–58.
- Cohen LE, Felson M. Social change and crime rate trends: a routine activity approach. *Am Sociol Rev.* 1979;44:588–608.
- Hindelang MJ. Variations in sex-race-age-specific incidence rates of offending. *Am Sociol Rev.* 1981;46:461–474.
- Hindelang MJ, Gottfredson MR, Garofalo J. *Victims of Personal Crime: An Empirical Foundation for a Theory of Personal Victimization.* Cambridge, MA: Ballinger; 1978.
- Akers TA, Lanier MM. Epidemiological criminology: coming full circle. *Am J Public Health.* 2009;99:397–402.
- Lanier MM. Epidemiological criminology (EpiCrim): definition and application. *J Theor Philos Criminol.* 2010;2:63–103.
- Fears NE, Vaughn MG, Nelson EJ, Salas-Wright CP, DeLisi M, Qian Z. Trends and correlates of substance use disorders among probationers and parolees in the United States 2002–2014. *Drug Alcohol Depend.* 2016. <http://dx.doi.org/10.1016/j.drugalcdep.2016.08.003>.
- Vaughn MG, Nelson EJ, Salas-Wright CP, DeLisi M, Qian Z. Handgun carrying among white youth increasing in the United States: new evidence from the National Survey on Drug Use and Health 2002–2013. *Prev Med.* 2016;88:127–133.
- Vaughn MG, Fu Q, DeLisi M, et al. Prevalence and correlates of fire-setting in the United States: results from the national epidemiological survey on alcohol and related conditions. *Compr Psychiatry.* 2010;51:217–223.
- Vaughn MG, Fu Q, DeLisi M, et al. Correlates of cruelty to animals in the United States: results from the national epidemiologic survey on alcohol and related conditions. *J Psychiatr Res.* 2009;43:1213–1218.
- Vaughn MG, Fu Q, Bender K, et al. Psychiatric correlates of bullying in the United States: findings from a national sample. *Psychiatr Q.* 2010;81:183–195.
- Vaughn MG, Fu Q, DeLisi M, Beaver KM, Perron BE, Howard MO. Criminal victimization and comorbid substance use and psychiatric disorders in the United States: results from the NESARC. *Ann Epidemiol.* 2010;20:281–288.
- Vaughn MG, Fu Q, Wernet SJ, et al. Characteristics of abstainers from substance use and antisocial behavior in the United States. *J Crim Justice.* 2011;39:212–217.
- Vaughn MG, Define RS, DeLisi M, et al. Sociodemographic, behavioral, and substance use correlates of reckless driving in the United States: findings from a national sample. *J Psychiatr Res.* 2011;45:347–353.
- Vaughn MG, Salas-Wright CP, DeLisi M. Drifter: an exploration into a classical criminological construct. *Am J Crim Justice.* 2015;40:560–578.
- Vaughn MG, Salas-Wright CP, DeLisi M, Larson M. Deliberate self-harm and the nexus of violence, victimization, and mental health problems in the United States. *Psychiatry Res.* 2015;225:588–595.
- Vaughn MG, Perron BE, Abdon A, Olate R, Groom R, Wu L-T. Correlates of handgun carrying among adolescents in the United States. *J Interpers Violence.* 2012;27:2003–2021.
- Vaughn MG, Maynard BR, Salas-Wright CP, Perron BE, Abdon A. Prevalence and correlates of truancy in the US: results from a national sample. *J Adolesc.* 2013;36:767–776.
- Vaughn MG, DeLisi M, Gunter T, et al. The severe 5%: a latent class analysis of the externalizing behavior spectrum in the United States. *J Crim Justice.* 2011;39:75–80.
- Vaughn MG, Salas-Wright CP, DeLisi M, Maynard BR. Violence and externalizing behavior among youth in the United States: is there a severe 5%? *Youth Violence Juv Justice.* 2014;12:3–21.
- Vaughn MG, Salas-Wright CP, DeLisi M, Qian Z. The antisocial family tree: family histories of behavior problems in antisocial personality in the United States. *Soc Psychiatry Psychiatr Epidemiol.* 2015;50:821–831.
- DeLisi M, Vaughn MG, Salas-Wright CP, Jennings WG. Drugged and dangerous: prevalence and variants of substance use comorbidity among seriously violent offenders in the United States. *J Drug Issues.* 2015;45:232–248.
- Behnken M, Vaughn MG, Salas-Wright CP, DeLisi M. Risky business: behavioral profiles of sexually-impulsive adults in the United States. *Am J Crim Justice.* 2016;41:185–201.
- Vaughn MG, Salas-Wright CP, DeLisi M, Maynard BR, Boutwell B. Prevalence and correlates of psychiatric disorders among former juvenile detainees in the United States. *Compr Psychiatry.* 2015;59:107–116.
- Vaughn MG, Salas-Wright CP, DeLisi M, Shook JJ, Terzis L. A typology of drug

- selling among young adults in the United States. *Subst Use Misuse*. 2015;50:403–413.
26. Shook JJ, Vaughn MG, Salas-Wright CP. Exploring the variation in drug selling among adolescents in the United States. *J Crim Justice*. 2013;41:365–374.
 27. Vaughn MG, DeLisi M, Beaver KM, Perron BE, Abdon A. Toward a criminal justice epidemiology: behavioral and physical health of probationers and parolees in the United States. *J Crim Justice*. 2012;40:165–173.
 28. Vaughn MG, Salas-Wright CP, DeLisi M, Piquero AR. Health associations of drug-involved and criminal-justice-involved adults in the United States. *Crim Justice Behav*. 2014;41:318–336.
 29. Vaughn MG, Salas-Wright CP, Maynard BR, et al. Criminal epidemiology and the immigrant paradox: intergenerational discontinuity in violence and antisocial behavior among immigrants. *J Crim Justice*. 2014;42:483–490.
 30. Vaughn MG, Salas-Wright CP, DeLisi M, Maynard BR. The immigrant paradox: immigrants are less antisocial than native-born Americans. *Soc Psychiatry Psychiatr Epidemiol*. 2014;49:1129–1137.
 31. Vaughn MG, Fu Q, Beaver K, DeLisi M, Perron B, Howard M. Are personality disorders associated with social welfare burden in the United States? *J Personal Disord*. 2010;24:709–720.
 32. Moffitt TE. Adolescence-limited and life-course-persistent antisocial behavior: a developmental taxonomy. *Psychol Rev*. 1993;100:674–701.
 33. Caspi A, Houts RM, Belsky DW, et al. The p factor: one general psychopathology factor in the structure of psychiatric disorders? *Clin Psychol Sci*. 2014;2:119–137.
 34. Gottfredson MR, Hirschi T. *A General Theory of Crime*. Stanford, CA: Stanford University Press; 1990.
 35. DeLisi M, Vaughn MG. Foundation for a temperament-based theory of anti-social behavior and criminal justice system involvement. *J Crim Justice*. 2014;42:10–25.
 36. Barrett DE, Katsiyannis A, Zhang D, Zhang D. Delinquency and recidivism: a multicohort, matched-control study of the role of early adverse experiences, mental health problems, and disabilities. *J Emot Behav Disord*. 2014;22:3–15.
 37. Baglivio MT, Jackowski K, Greenwald MA, Howell JC. Serious, violent, and chronic juvenile offenders: a statewide analysis of prevalence and prediction of subsequent recidivism using risk and protective factors. *Criminol Public Policy*. 2014;13:83–116.
 38. Baglivio MT, Epps N, Swartz K, Huq MS, Sheer A, Hardt NS. The prevalence of adverse childhood experiences (ACE) in the lives of juvenile offenders. *J Juv Justice*. 2014;3:1–20.
 39. Baglivio MT, Wolff KT, Piquero AR, Epps N. The relationship between adverse childhood experiences (ACE) and juvenile offending trajectories in a juvenile offender sample. *J Crim Justice*. 2015;43:229–241.
 40. Bales WD, Piquero AR. Assessing the impact of imprisonment on recidivism. *J Exp Criminol*. 2012;8:71–101.
 41. Hamilton Z, Neuilly M-A, Lee S, Barnoski R. Isolating modeling effects in offender risk assessment. *J Exp Criminol*. 2015;11:299–318.
 42. DeLisi M. Criminal careers behind bars. *Behav Sci Law*. 2003;21:653–669.
 43. DeLisi M, Scherer AM. Multiple homicide offenders: offense characteristics, social correlates, and criminal careers. *Crim Justice Behav*. 2006;33:367–391.