

ARTICLE OPEN



Estimating the incidence of cocaine use and mortality with music lyrics about cocaine

Yulin Hswen^{1,2,3,4}, Amanda Zhang^{3,5} and John S. Brownstein^{1,3,4}

In the United States, cocaine use and mortality have surged in the past 5 years. Considering cocaine's reputation as a fashionable social drug, the rise of cocaine mentions in popular music may provide a signal of epidemiological trends of cocaine use. We characterized the relationship between mentions of cocaine in song lyrics and incidence of cocaine use and mortality in the US. Incidence of cocaine use from 2002 to 2017 was obtained from the National Survey on Drug Use and Health and cocaine overdose mortality rate from 2000 to 2017 was obtained from the Centers for Disease Control. Distributed lag models were fit using ordinary least squares on the first difference to identify associations between changes in cocaine lyric mentions and changes in incidence of cocaine use and mortality. A total of 5955 song lyrics with cocaine mentions were obtained from Lyrics.com. Cocaine mentions in song lyrics were stable from 2000 to 2010 then increased by 190% from 2010 to 2017. The first-order distributed lag model estimated that a 0.01 increase in mentions of cocaine in song lyrics is associated with an 11% increase in incidence of cocaine use within the same year and a 14% increase in cocaine mortality with a 2-year lag. Lag-times were confirmed with cross-correlation analyses and the association remained after accounting for street pricing of cocaine. Mentions of cocaine in song lyrics are associated with the rise of incidence of cocaine use and cocaine overdose mortality. Popular music trends are a potentially valuable tool for understanding cocaine epidemiology trends.

npj Digital Medicine (2021)4:100; <https://doi.org/10.1038/s41746-021-00448-x>

INTRODUCTION

Cocaine popularity in the United States has rebounded in the past 5 years after many years of decline. Annual deaths by cocaine increased 200% from 2012 to 2017, making cocaine the leading non-opioid cause of drug overdose death in the US¹. At the same time, cocaine's user base has expanded with more people trying cocaine for the first time. From 2012 to 2017, the number of people using cocaine for the first time in the past year increased by 57%^{2,3}. In 2017, 6.2% of people in the US aged 18–25 reported using cocaine in the past year, compared to 0.6% for heroin³.

Historically, patterns of cocaine use differed from those of other illicit drugs in that cocaine has been used almost exclusively in social settings, especially to indicate high social status. Cocaine possesses the unique reputation of being the “rich man's speed,” a symbol of affluence flouted by celebrities and artists and the object of aspiration for others⁴. The glamorization of cocaine's bodily effects and an underestimation of its dangers likely also influenced public perception of cocaine. A 1977 Newsweek story on cocaine reported “Among hostesses in the smart sets of Los Angeles and New York, a little cocaine, like Dom Perignon and Beluga caviar, is now de rigueur at dinners”⁵. Although the impression of cocaine is that it gives the user strength and vigor, its actual use is harmful. In a study ranking the danger of drugs⁶, cocaine received the highest harm ratings and notably scored highest in the “social harm” category. Furthermore, epidemiological evidence on cocaine has shown it to be a highly addictive and widely abused drug specifically among young people^{6,7}.

*“In its pharmacological action, cocaine—
perhaps more than any other of the*

*recognized psychoactive drugs—reinforces
and boosts what we recognize as the highest
aspirations of American initiative, energy,
frenetic achievement, and ebullient opti-
mism...”—Stimmel, 1984⁴*

Its attractiveness among the public has been arbitrated by popular media, including coverage of celebrity drug activities associating the drug with high society². In these ways, references to cocaine in the media and popular culture portrayed real-world behaviors of cocaine use.

There is reason to believe that popular culture today is again capturing real-world trends around a renewed interest in cocaine. Although cocaine's reputation waned during the 1980s War on Drugs government crackdown, other forms of popular media such as the music industry continue to relay information about cocaine. Major artists today frequently mention cocaine and other drugs extensively in their songs, particularly in the context of glamour, wealth, and sociability^{8,9}. For instance, studies analyzing popular music and videos have found drugs to be a dominant theme^{8,10–12} with one study reporting that 33.3% of top-charting songs portrayed substance use with an average of 32.5 drug references per hour⁹. Considering that 90 percent of Americans regularly listen to music with an average listening time of 32.1 h per week¹¹, this means Americans on average are exposed to 54 explicit references to drugs every day.

A growing body of research supports that drug mentions in popular media may be linked with substance use, including for

¹Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, CA, USA. ²Bakar Computational Health Sciences Institute, University of California San Francisco, San Francisco, CA, USA. ³Innovation Program, Boston Children's Hospital, Boston, MA, USA. ⁴Computational Epidemiology Lab, Harvard Medical School, Boston, MA, USA. ⁵Applied Mathematics, Harvard University, Cambridge, MA, USA. [✉]email: yulin.hswen@ucsf.edu

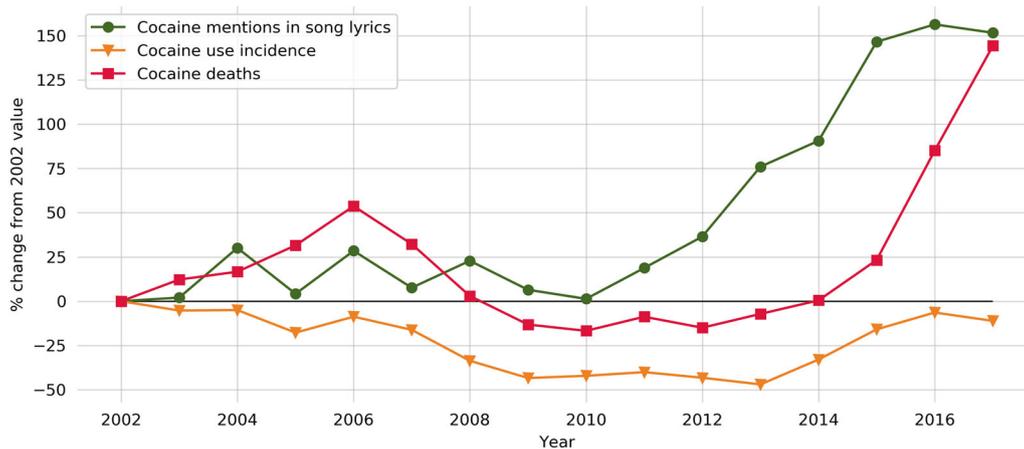


Fig. 1 Cocaine mentions in song lyrics, incidence of cocaine use, and cocaine death rate. Summarizes the percent change from the baseline year of 2002 for 2002–2016 for cocaine mentions in song lyrics, incidence of cocaine use, and cocaine death rate.

smoking^{13–15}, alcohol¹⁶, and cannabis¹⁷. A qualitative study found that an underground form of hip-hop music called “screw” is a strong reinforcer of codeine syrup use; respondents named “media modeling” as the foremost reason for the popularity for syrup usage¹⁸. A content analysis of another genre of music, popular country music, found that lyrics were more likely to describe women in association with alcohol use and sex in the 2010’s compared to earlier decades, confirming the known association between alcohol use and sexual assault in this genre¹⁹. However, the majority of these studies on substance use in song lyrics have been descriptive and limited in the scope of songs analyzed. For instance, Primack et al.⁹ conducted a content analysis of 279 songs for drug-related content in the single year of 2005 to determine degree of drug exposure for music listeners. Hall, West, and Neeley²⁰ analyzed time trends of alcohol, tobacco, and other drugs in 1100 songs over 50 years, an average of only 22 songs per year. Herd performed a qualitative analysis of the social context surrounding 341 rap songs⁸. While these studies describe important findings regarding drug exposure in song lyrics, trends over time, and qualitative content analysis, none were specific to cocaine nor quantified the effect size of drug exposure in song lyrics with drug use behavior. Additionally, all studies compiled song titles from ranked lists, most commonly the Billboard Top-100, thereby skipping over thousands of popular songs that did not make the list.

Given cocaine’s fashionable reputation and highly social patterns of use²¹, descriptions of cocaine in music are potentially an indication of increased cocaine initiation and use. However, the epidemiological trends between popular media and cocaine have yet to be studied empirically. This study investigated the relationship between prevalence of drug mentions in music lyrics and epidemiological trends for cocaine. We sought to offer insights on the recent rise in cocaine use incidence and cocaine overdose mortality observed during the first decades of the 21st century with respect to the influence of contemporary music trends on population-based health behaviors.

RESULTS

Descriptive results

A total of 6923 unique mentions of cocaine from 5955 unique songs and 2418 unique artists were obtained from Lyrics.com. The mean number of cocaine mentions per year was 364 (SD = 73). Cocaine mentions in lyrics was stable from 2000 to 2010, then increased by 190% from 2010 to 2017. Data on cocaine overdose mortality, obtained from the Centers for Disease Control, also exhibited a recent surge. Deaths increased every year starting in

Table 1. Parameter estimates for distributed lag models.

	Death model	Death model	Use model	Use model
	β	p	β	p
Intercept	0.42	0.142	0.08	0.752
Lyrics (same year)	–	–	11.16	0.039
Lyrics (1 year ago)	9.48	0.162	6.22	0.193
Lyrics (2 years ago)	14.39	0.037	–	–
Lyrics (3 years ago)	7.35	0.331	–	–
Price (same year)	–0.00	0.118	–0.00	0.573

2012 with a total growth of 200% from 2012 to 2017. The percent change in cocaine mentions in song lyrics, incidence of cocaine use from 2002 for 2002–2016 is visualized in Fig. 1.

The top three cocaine slang words used from 2000 to 2018 were “coke”, “cocaine”, and “powder” and made up over 60% of drug mentions in song lyrics. “Coke” led with 2358 song counts (34%), then “cocaine” with 1469 song counts (21%), “powder” with 810 song counts (12%), “coca” with 579 song counts (8.4%) and “kilo” with 470 song counts (6.8%). Slang terms with the greatest growth from 2000 to 2018 were “kilo” at 1950%, “yayo” at 750%, and “coco” at 640%, followed by “baking soda” at 300%.

Distributed lag regression

Results from the time distributed-lag model are presented in Table 1 and show that a relative 0.01 increase in mentions of cocaine relative to “love” mentions in song lyrics was associated with a 100 * (e^(0.01*11.1609) – 1) = 11.8% increase in incidence of cocaine use during the same year and a 100 * (e^(0.01*14.398) – 1) = 15.5% increase in mortality by cocaine in 2 years. These associations were robust to changes in street price. A decrease in cocaine price was associated with an increase in cocaine mortality with a lag of 2 years.

Lag-time results were robust to sensitivity analyses that removed the term “8-ball” (death model: $p = 0.04$, cocaine use model: $p = 0.03$). No significant lag-times were seen for the distributed lag regression for the control models of codeine lyrics and heroin lyrics with cocaine epidemiology.

Regression parameter estimates were used to project future incidence of cocaine use and cocaine-related death rates for 2018,

2019, and 2020 and these results are presented in Fig. 2. Based on the model, we estimate the incidence of cocaine use and cocaine overdose mortality will continue to increase. We project incidence of cocaine use to be 344 new users per 100,000 in 2018, 385 new users per 100,000 in 2019, and 416 new users per 100,000 in 2020. We project overdose deaths by cocaine to be 5.16 deaths per 100,000 in 2018, 5.66 deaths per 100,000 in 2019, and 7.12 deaths per 100,000 in 2020.

Cross-correlation

Cross-correlation analyses are displayed in cross-correlation plots in Fig. 3. These analyses confirmed a significant cross-correlation between cocaine lyrics and incidence of cocaine use in the same year ($r=0.5713$, $p<0.05$) and a significant correlation between cocaine lyrics and mortality 2 years later ($r=0.5747$, $p<0.05$). Sensitivity analyses did not yield significant cross-correlations in the lags in the opposite direction. Cocaine mortality was not associated with future cocaine mentions in lyrics at any lag time.

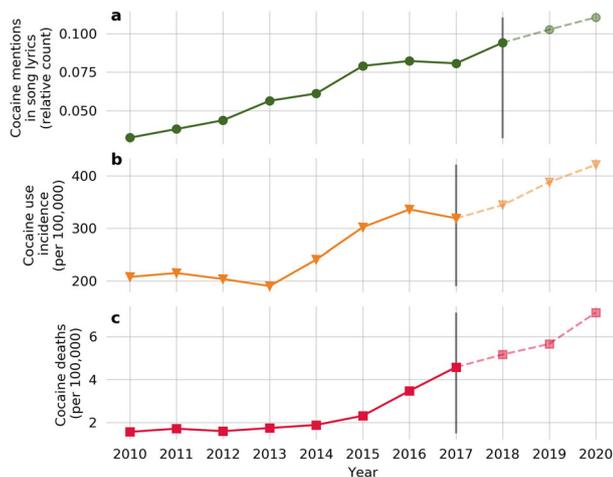


Fig. 2 Cocaine use and cocaine-related death rates for 2018, 2019, and 2020. **a** Projection of cocaine mentions in song lyrics for 2019 and 2020. **b** Projection of cocaine use incidence for 2018, 2019, and 2020. **c** Projection of cocaine mortality for 2018, 2019, and 2020. Predictive models were fitted with lyrics and epidemiological data from 2000 to 2018 and assumed linear growth of cocaine mentions in song lyrics.

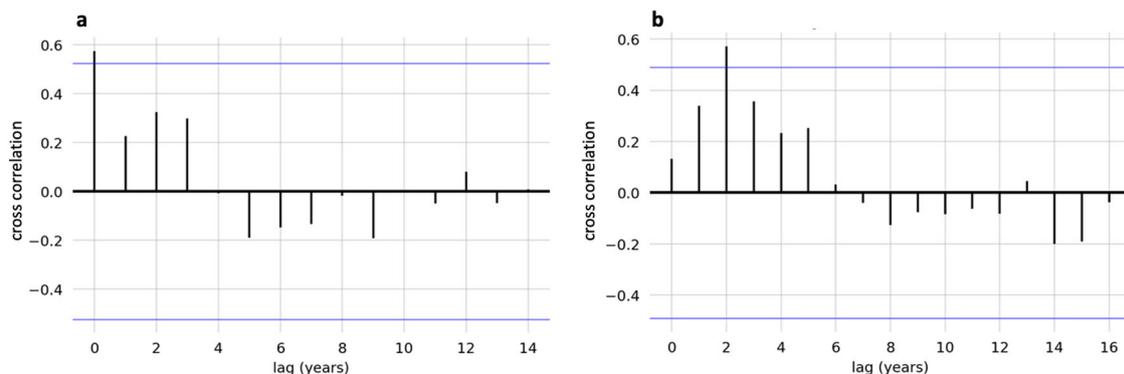


Fig. 3 Cross-correlation between cocaine lyrics and incidence of cocaine use. **a** Cross correlation for cocaine mentions in lyrics leading cocaine use incidence at different lag times. **b** Cross correlation for cocaine mentions in lyrics leading cocaine deaths at different lag times. Blue lines indicate the 95% confidence interval around cross-correlation value of 0.

DISCUSSION

Our study explores the ability of song lyrics to signal epidemiological trends in incidence and mortality of cocaine use. Specifically, our results found that an increase in cocaine mentions in song lyrics is associated with increased incidence of cocaine use in the same year and death by cocaine 2 years later. The temporal relationship between cocaine lyrics and incidence of cocaine use and related deaths was confirmed with significant cross-correlation in the same year for cocaine use and 2 years for cocaine overdose mortality. The lead time period of 2 years found in this study between lyrics mentioning cocaine and incidence of cocaine and cocaine overdose mortality is supported by epidemiological evidence that the time between initiation of cocaine use and seeking help for addiction is between 1 and 3 years^{22,23}. Musical trends that depict cocaine in lyrics may be an early signal of a rising interest and use of cocaine in the same year. The lag-time period of 2 years between cocaine lyrics and cocaine mortality may indicate that there is an incubation period of 2 years between addiction and fatality. Treatment intervention is critical in this time period to prevent the onset of drug dependence and death.

Our results also provide insights into the pathways by which media may influence cocaine use behaviors. In addition to being a measurement tool to estimate cocaine epidemiology, media itself can influence public perceptions of drug use and lead to increased drug use at the population level. Therefore, music about cocaine may not only provide an early signal of cocaine-related behaviors, but it could also act as an exposure that encourages the use of cocaine. Our study did not test the potential dynamic and cyclical relationship between cocaine lyrics and patterns of cocaine use. While cocaine mortality did not appear to estimate future cocaine mentions in song lyrics, further investigation into the influence of cocaine use on song lyrics is warranted.

Our study showed growth in slang terminology such as “kilo”, “yayo”, “coco”, and “baking soda” instead of more explicit terms for cocaine such as “coke” and “cocaine” used in older song lyrics. These results may indicate that the current generation has inherited an attachment to cocaine expressed through the development of new slang terms for cocaine. Evidence of this shifting use of cocaine by younger generations is provided by a 2018 United Nations study reporting that settings of cocaine use have branched out from exclusive clubs to include more accessible environments including college parties, concerts, and bars²². Furthermore, in the past decade, first time users of cocaine have moved from high income areas to lower income and under resourced neighborhoods²⁴. The popularity of new slang terms in our study may signal a different population and a reemergence of a black market for cocaine as a rebranding of street terminology is being used to disguise discussion of this illegal drug. Finally, our

argument about epidemiological ties with cocaine mentions in song lyrics is strengthened by our results that showed that popular music lyrics related to codeine and heroin did not follow the same trends with cocaine mortality. If this trend in cocaine lyrics continues, we estimate that future incidence of cocaine use may increase by close to 40% and mortality by cocaine use will increase by almost 80% for 2020, making it important to monitor this growth in lyrics about cocaine over time.

An important limitation with our study is that our results depict associations and cannot confirm causality. Analytical steps were included in the methodology to remove potential spurious correlations but this is not enough to deem our results as causal. Therefore, our findings should be interpreted with caution. Cocaine street price was included in the model to control for economic fluctuations that would impact purchasing behaviors and initiation of drug use²⁵. Spurious correlations were controlled through first-order differentiation. Further tests on the differenced series confirmed autocorrelations were no longer present after first-order differentiation.

Another limitation of our study is the potential of not having captured all the cocaine slang terminology to identify song lyrics describing cocaine. Although we queried a number of top websites for the most referenced slang terms for cocaine, it is possible that certain ambiguous slang terms were excluded. Generating a comprehensive list of slang drug terminology is exceedingly difficult because the vocabulary around cocaine evolves to conceal its discussion²⁶. However, sensitivity analyses revealed that our results were robust to the removal of ambiguous terms such as “8-ball”. Inclusion of additional terms like “blow”, which was not included in the model because of its possible description of guns rather than drugs, increased the degree of association between mentions in lyrics and cocaine mortality. Because of this, we believe that genuinely uncertain references to cocaine are rare and have limited potential to significantly alter study results.

In this study, we were only able to conduct analyses at the yearly level and not more granularly. Data derived from Lyrics.com is only available at the annual level. We explored several song lyrics engines including Spotify, Genius Lyrics, Metro lyrics, and Billboard; however, these other platforms either do not allow their data to be available or it is also at the yearly level. We selected Lyrics.com because of its word query capabilities within lyrics that other song lyric engines do not provide. This restricted us to have 18 effective observations from 2000 to 2017 for song lyrics data. However, in our analysis, another limiting factor was the incidence of cocaine data from the National Survey on Drug Use and Health published by the Substance Abuse and Mental Health Services Administration and cocaine-related deaths were obtained from the Center for Disease Control Multiple Causes of Death WONDER database, which only are available aggregated at the yearly level. Therefore, our analysis still would have been limited to this level of granularity. Additionally, since our definition of cocaine-related mortality includes deaths using the ICD-10 code T40.5, it may include underlying causes of death other than cocaine. However, other ICD-10 codes in the mortality dataset do not specify the actual drug involved. While this is a limitation in the use of death certificates for identifying cause-of-death, this is a consistently used measure for cocaine-related deaths. Furthermore, the quality of testing for drug overdose has improved over time²⁷. Thus the rise in deaths by cocaine may be in part due to the more accurate determination of deaths attributed to cocaine.

Lastly, it is possible that mortality by cocaine could occur in the same year of initiation, which was not seen in our results. However, addiction to cocaine is often characterized by repeated use that changes brain and psychological function that promote transitions to problematic patterns of use²⁸. Thus, the accumulation of conditional use of cocaine overtime that leads to mortality by cocaine is likely to occur after years of addictive use of cocaine.

In conclusion, these associations underscore the importance of monitoring trends in music to understand drug patterns over time. Media are a powerful indicator of social norms and our study offers initial epidemiological evidence that music lyrics about cocaine may provide an early signal to incidence of cocaine use and mortality. Additionally, new slang terminology for drugs in music lyrics could indicate a new generation of cocaine users and a surge in under-detected use. Future studies should conduct a deeper investigation into lyrics about cocaine and other substances, with particular focus on how these messages may shape cultural perceptions and behaviors toward drug use. Given the wide audience of popular music, popular music artists should consider the potential influence of their lyrics on the drug epidemic impacting today's youth.

METHODS

Data

The use of computers to record, distribute, store and play music in the early 2000's shifted the way people accessed and listened to music²⁹. The digital transformation of the music industry beginning in the early 2000's has allowed peer-to-peer sharing of digital audio files, as well as decreased the barriers to entry for artists³⁰. For instance, artists can now self-release their music to various online platforms, thereby decreasing the time required for a song to be written, performed, and released to the public, and making it possible for emerging artists to expand the reach of their audience^{31,32}. Additionally, the most recent surge of cocaine use and deaths occurred in the 2010s³³. Therefore, our data was restricted to 2000-onward to focus on contemporary trends in music and the most recent resurgence of cocaine use and mortality in the United States.

National incidence of cocaine use

Data on incidence of cocaine use from 2002 to 2017, defined as the nationwide number of new cocaine initiates aged 12 and older in the past year, was obtained from Table 7.28A of the 2017 National Survey on Drug Use and Health published by the Substance Abuse and Mental Health Services Administration³. National incidence was calculated by dividing initiation counts by midyear population estimates obtained from the United States Census Bureau³⁴.

National drug overdose deaths involving cocaine

National drug overdose deaths involving cocaine were collected as yearly counts of cocaine-related deaths for 2000–2017, obtained from the Center for Disease Control Multiple Causes of Death WONDER database¹. Query criteria for drug overdose involving cocaine were all conditions under the ICD-10 code T40.5 (“Poisoning by, adverse effect of and underdosing of cocaine”). National death count by cocaine was divided by midyear population estimates obtained from the United States Census Bureau³⁴.

Cocaine mentions in song lyrics

To measure cocaine mentions in song lyrics, first a comprehensive list of informal words that describe cocaine (slang) was compiled by analyzing top websites that have the primary purpose of deciphering slang terminology such as Urban Dictionary³⁵ as well as the Drug Enforcement Association Slang Terms and Code Words reference guide³⁶. Two independent researchers coded a random sample of 1000 to remove ambiguous terms that did not refer to cocaine. Slang terms such as “snow”, “blow”, and “dust” were found to be used in reference to connotations other than cocaine in the lyrics' dataset, and were therefore excluded from the set of terms. The final list used for this analysis was composed of 18 terms: “cocaine”, “coke”, “coca”, “coco”, “powder”, “baking soda”, “bakin soda”, “arm and hammer”, “kilo”, “8 ball”, “eight ball”, “llemo”, “yayo”, “yeyo”, “nose candy”, “white horse”, “white line”, and “doing lines”.

Cocaine time trends

Yearly counts of cocaine mentions in song lyrics were obtained from Lyrics.com, the largest known database for references to song lyrical information with keyword search³⁷. A programmatic code was written to query each individual slang term to obtain the count of songs that mentioned that term each year. Slang term counts were summed to obtain the total

mentions of unique slang terms per year. To account for variation in song volume over time, we transformed counts of cocaine mentions into a ratio of relative prevalence by dividing mentions of cocaine terms by mentions of “love”. Previous studies have shown that love is consistently the dominant theme in song lyrics for every decade since the 1960s^{38,39}, therefore “love” was used in this study to adjust for changes in song volume over time and provide a relativistic interpretation for the time trends.

Cocaine street price

Cocaine street price data in the US from 2000 to 2016, adjusted for purity and inflation, was obtained from the United Nations Office on Drugs and Crime²⁶.

Analysis

To examine the association of cocaine mentions in lyrics with incidence of cocaine and mortality, we used a first-order finite distributed lag model with unrestricted coefficients. In this model, the first differences of the time series of national counts of incidence of cocaine and mortality were regressed on first differences of the time series of cocaine mentions in lyrics earlier in time^{27,40–43}.

First-order differencing

To regress time-series data, the data must be stationary whereby the mean and variance are independent of time. To test the stationarity of our data, we conducted Augmented Dickey–Fuller (ADF) tests on the lyrics, price, incidence, and mortality time series. ADF tests deemed the data was nonstationary and a time trend in the data existed. To transform our data to be stationary, we differenced the data on the first order. Inspection of autocorrelation function plots and partial autocorrelation function plots before and after first order differencing confirmed autocorrelation was removed after differencing. Subsequent analyses were performed using the differenced series.

Distributed lag regression

A standard epidemiological method for identifying relationships between time series is distributed lag regression^{43,44}. The general format of this model is presented in Eq. (1). In this framework, the dependent variable is regressed on delayed (or lagged) instances of the independent variable. This practice is commonly used in economics⁴² and has been applied to epidemiological research⁴¹ in order to express associations that change over time⁴³. In our study, the dependent variable was cocaine mortality or cocaine use incidence and the independent variable was cocaine mentions in song lyrics. We chose regression terms using a stepwise process, beginning with contemporary values and adding previous year lag terms one at a time until further additions were no longer significant⁴⁰. Epidemiological research and theory support a 1–3 year delay between media exposure and behavior^{22,23}. Thus only lag years were included in the model that were consistent with a potential epidemiological relationship. For instance, we did not include year 1 in the distributed lag model because of the lack of biological plausibility that music about cocaine would estimate cocaine overdose mortality in the same year. The final regression model for cocaine mortality is presented as Eq. (2) and the final regression model for cocaine use incidence is presented as Eq. (3).

Sensitivity analyses were conducted whereby the term “8-ball” was excluded—although most frequently used to reference cocaine in music lyrics³¹, “8-ball” can also be used as a reference to crack cocaine. Additionally, we employed a negative control by collecting slang terms for codeine: ‘codeine’, ‘purple drank’, ‘sizzurp’, ‘syrup’, ‘lean’; and for heroin: ‘heroin’, ‘skag’, ‘china white’, ‘white horse’, ‘black tar’, ‘black eagle’, ‘brown sugar.’ These sets of terms were tested in the distributed lag model in the same procedure to determine if there was a relationship between cocaine epidemiology and drug lyrics about non-cocaine drugs.

The models were fit using ordinary least squares and included variables that could act as “proxy” confounders—those that may have distorted the relationship between lyrics and mortality⁴⁵. The response variable was logged to interpret the regression coefficients as percent changes. The fitted model and 2018 lyrics data about cocaine were used to project 2018, 2019, and 2020 incidence of cocaine and cocaine overdose mortality assuming a linear rate of growth of cocaine prevalence in song lyrics.

Model.

$$\Delta \ln(y_t) = \alpha + \beta_0 \Delta x_t + \beta_1 \Delta x_{t-1} + \dots + \beta_p \Delta x_{t-p} + \epsilon_t \quad (1)$$

where $\Delta x = x_t - x_{t-1}$

$$\Delta \ln(\text{deaths}_t) = \alpha + \beta_1 \Delta \text{lyrics}_{t-1} + \beta_2 \Delta \text{lyrics}_{t-2} + \beta_3 \Delta \text{lyrics}_{t-3} + \beta_4 \text{price}_t + \epsilon_t \quad (2)$$

$$\Delta \ln(\text{use}_t) = \alpha + \beta_0 \Delta \text{lyrics}_t + \beta_1 \Delta \text{lyrics}_{t-1} + \beta_2 \text{price}_t + \epsilon_t \quad (3)$$

Cross-correlation

To provide further validation of our lag-time results, we performed cross-correlation tests between lyrics and cocaine use and lyrics and deaths at varying lags to evaluate time-delayed associations. Cross-correlation analyzes the relationship between two signals and calculates the correlation coefficient at designated lags (displacements) between the two series. The correlation coefficient is maximal at the time point that the two series correlate most closely and the coefficient reveals the strength of the correlation. Cross-correlation analyses were conducted in both directions to examine whether the data supported cocaine epidemiology preceding lyrics or lyrics preceding cocaine epidemiology.

The cross-correlation between two discrete sequences x and y at lag k is defined as:

$$(x * y)[k] \triangleq \sum_n x[n+k] \cdot y^*[n],$$

where $y^*[n]$ denotes the conjugate of y .

All analyses were conducted using the statsmodels package in Python⁴⁶.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

DATA AVAILABILITY

The data that support the finding from this study are publicly available from The Substance Abuse and Mental Health Services Administration and The Centers for Disease Control and Lyrics.com. Aggregated datasets are available on reasonable request.

CODE AVAILABILITY

Statistical analyses were conducted using the statsmodels package in Python.

Received: 14 December 2019; Accepted: 16 February 2021;

Published online: 30 June 2021

REFERENCES

- Centers for Disease Control and Prevention. *Multiple Cause of Death 1999–2017 on CDC WONDER Online Database*, <https://wonder.cdc.gov/mcd.html> (2018). Accessed 1 Oct 2019.
- Jonnes, J. *Hep-cats, narcs, and pipe dreams: A history of America's romance with illegal drugs*. (Scribner, 1996).
- Substance Abuse and Mental Health Services Administration. *Results from the 2017 National Survey on Drug Use and Health*, <https://www.samhsa.gov/data/report/2017-nsduh-annual-national-report> (2018). Accessed 15 Nov 2019.
- Stimmel, B. *Alcohol and Drug Abuse in the Affluent*. (The Haworth Press, 1984).
- Marion, N. E., & Oliver, W. M. (Eds.) *Drugs in American Society: An Encyclopedia of History, Politics, Culture, and the Law*. (Vol. 3. ABC-CLIO. Chicago, 2014).
- Nutt, D., King, L. A., Saulsbury, W. & Blakemore, C. Development of a rational scale to assess the harm of drugs of potential misuse. *Lancet* **369**, 1047–1053 (2007).
- Hser, Y.-I., Huang, D., Brecht, M.-L., Li, L. & Evans, E. Contrasting trajectories of heroin, cocaine, and methamphetamine use. *J. Addict. Dis.* **27**, 13–21 (2008).
- Herd, D. Changes in drug use prevalence in rap music songs, 1979–1997. *Addict. Res. Theory* **16**, 167–180 (2008).
- Primack, B. A., Dalton, M. A., Carroll, M. V., Agarwal, A. A. & Fine, M. J. Content analysis of tobacco, alcohol, and other drugs in popular music. *Arch. Pediatrics Adolesc. Med.* **162**, 169–175 (2008).
- Gruber, E. L., Thau, H. M., Hill, D. L., Fisher, D. A. & Grube, J. W. Alcohol, tobacco and illicit substances in music videos: a content analysis of prevalence and genre. *J. Adolesc. Health* **37**, 81–83 (2005).

11. Markert, J. Sing a song of drug Use-Abuse: Four decades of drug lyrics in popular Music-From the sixties through the nineties. *Sociological Inq.* **71**, 194–220 (2001).
12. Christenson, P., Roberts, D. F. & Bjork, N. Booze, drugs, and pop music: trends in substance portrayals in the Billboard top 100—1968–2008. *Subst. Use Misuse* **47**, 121–129 (2012).
13. Gidwani, P. P., Sobol, A., DeJong, W., Perrin, J. M. & Gortmaker, S. L. Television viewing and initiation of smoking among youth. *Pediatrics* **110**, 505–508 (2002).
14. Dalton, M. A. et al. Effect of viewing smoking in movies on adolescent smoking initiation: a cohort study. *Lancet* **362**, 281–285 (2003).
15. Morgenstern, M. et al. Smoking in movies and adolescent smoking initiation: longitudinal study in six European countries. *Am. J. Prevent. Med.* **44**, 339–344 (2013).
16. Primack, B. A., McClure, A. C., Li, Z. & Sargent, J. D. Receptivity to and recall of alcohol brand appearances in US popular music and alcohol-related behaviors. *Alcohol: Clin. Exp. Res.* **38**, 1737–1744 (2014).
17. Primack, B. A., Douglas, E. L. & Kraemer, K. L. Exposure to cannabis in popular music and cannabis use among adolescents. *Addiction* **105**, 515–523 (2010).
18. Peters, R. J. Jr et al. Beliefs and social norms about codeine and promethazine hydrochloride cough syrup (CPHCS) onset and perceived addiction among urban Houstonian adolescents: an addiction trend in the city of lean. *J. Drug Educ.* **33**, 415–425 (2003).
19. Lin, R. & Rasmussen, E. Why don't we get drunk and screw? a content analysis of women, sex and alcohol in country music. *J. Pop. Music Stud.* **30**, 115–128 (2018).
20. Hall, P. C., West, J. H. & Neeley, S. Alcohol, tobacco, and other drug references in lyrics of popular music from 1959 to 2009. *Addiction Res. Theory* **21**, 207–215 (2013).
21. Green, A., Pickering, H., Foster, R., Power, R. & Stimson, G. V. Who uses cocaine? Social profiles of cocaine users. *Addiction Res.* **2**, 141–154 (1994).
22. Tucker, J. S., Miles, J. N. & D'Amico, E. J. Cross-lagged associations between substance use-related media exposure and alcohol use during middle school. *J. Adolesc. Health* **53**, 460–464 (2013).
23. Miller, N. S. & Gold, M. S. Criminal activity and crack addiction. *Int. J. Addict.* **29**, 1069–1078 (1994).
24. Petronis, K. & Anthony, J. A different kind of contextual effect: geographical clustering of cocaine incidence in the USA. *J. Epidemiol. Community Health* **57**, 893–900 (2003).
25. Caulkins, J. P. Drug prices and emergency department mentions for cocaine and heroin. *Am. J. Public Health* **91**, 1446–1448 (2001).
26. Buntain, C. & Golbeck, J. This is your Twitter on drugs: Any questions? *Proceedings of the 24th international conference on World Wide Web.* 777–782 <https://doi.org/10.1145/2740908.2742469> (2014).
27. Substance Abuse and Mental Health Services Administration. *Clinical Drug Testing in Primary Care*. (U.S. Department of Health and Human Services, Substance Abuse and Mental Health Services Administration, Center for Substance Abuse Treatment, 2012).
28. Keramati, M., Durand, A., Girardeau, P., Gutkin, B. & Ahmed, S. H. Cocaine addiction as a homeostatic reinforcement learning disorder. *Psychological Rev.* **124**, 130–153 (2017).
29. Dogan, S. L. Is Napster a VCR—The Implications of Sony for Napster and Other Internet Technologies. *Hastings Law J.* **52**, 939–959 (2000).
30. Ena, A. R. *What Other Businesses Could Have Learnt From It, And The Convergence (or Not) of Spotify and Netflix*, <https://medium.com/@alvarorebolloena/digitaltransformationinthemusicbusiness-e7e92ced125> (2019). Accessed 15 Aug 2020.
31. Perez, S. *Spotify launches an app for artists with real-time streaming data, audience demographics*, <https://techcrunch.com/2017/10/12/spotify-launches-an-app-for-artists-with-real-time-streaming-data-audience-demographics/> (2017). Accessed 15 Aug 2020.
32. Media Entertainment Tech Outlook. *How Streaming Technology is Changing the Music Industry*, <https://www.mediaentertainmenttechoutlook.com/news/how-technology-is-transforming-online-entertainment-nwid-629.html> (2019). Accessed 15 Aug 2020.
33. Hedegaard, H., Bastian, B. A., Trinidad, J. P., Spencer, M. & Warner, M. Drugs most frequently involved in drug overdose deaths: United States, 2011–2016. *Natl Vital. Stat. Rep.* **67**, 1–14 (2018).
34. U.S. Census Bureau. *National Population Totals and Components of Change*, https://www.census.gov/data/tables/time-series/demo/popest/2010s-national-total.html-par_textimage_2011805803 (2018). Accessed 15 Nov 2019.
35. Urban Dictionary. *Cocaine*, <https://www.urbandictionary.com/> (2019). Accessed 1 Mar 2019.
36. Drug Enforcement Administration. *Slang Terms and Code Words: A Reference for Law Enforcement Personnel*. (Drug Enforcement Administration, 2018).
37. Lyrics, *The Web's Largest Resource for Music, Songs & Lyrics*. <https://www.lyrics.com> (2019). Accessed 1 Mar 2019.
38. Ostlund, D. R. & Kinnier, R. T. Values of youth: Messages from the most popular songs of four decades. *J. Humanist. Educ. Dev.* **36**, 83–91 (1997).
39. Christenson, P. G., de Haan-Rietdijk, S., Roberts, D. F. & ter Bogt, T. F. What has America been singing about? Trends in themes in the US top-40 songs: 1960–2010. *Psychol. Music* **47**, 194–212 (2019).
40. Hswen, Y., Brownstein, J. S., Liu, J. & Hawkins, J. B. Use of a digital health application for influenza surveillance in China. *Am. J. Public Health* **107**, 1130–1136 (2017).
41. Schwartz, J. The distributed lag between air pollution and daily deaths. *Epidemiology* **11**, 320–326 (2000).
42. Almon, S. The distributed lag between capital appropriations and expenditures. *Econometrica: J. Econom. Soc.* **33**, 178–196 (1965).
43. Gasparrini, A. Modeling exposure–lag–response associations with distributed lag non-linear models. *Stat. Med.* **33**, 881–899 (2014).
44. Zeger, S. L., Irizarry, R. & Peng, R. D. On time series analysis of public health and biomedical data. *Annu. Rev. Public Health* **27**, 57–79 (2006).
45. VanderWeele, T. J. Principles of confounder selection. *Eur. J. Epidemiol.* **34**, 211–219 (2019).
46. Seabold, S. & Perktold, J. Statsmodels: Econometric and statistical modeling with python. *Proceedings of the 9th Python in Science Conference.* 92–96 <http://conference.scipy.org/proceedings/scipy2010/pdfs/seabold.pdf> (2010).

ACKNOWLEDGEMENTS

Y.H. thanks Malcolm Mitchell for the introduction to understanding the deeper meaning of music lyrics. J.S.B. and Y.H. report funding by the National Library of Medicine of the National Institutes of Health (R01LM010812).

AUTHOR CONTRIBUTIONS

Y.H. secured funding for this study, conceived the idea, designed the study and analyses, and wrote the manuscript. A.Z. conducted the analysis and performed the computations, and helped write the manuscript. J.S.B. helped secure funding for this study, assisted in the design, and supervised the project and supported the interpretation of the study findings. All authors contributed substantially to the acquisition, analysis, interpretation, critical revision, and are accountable for all accuracy and integrity of the work.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41746-021-00448-x>.

Correspondence and requests for materials should be addressed to Y.H.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021