Commentary

The triple wave epidemic: Supply and demand drivers of the US opioid overdose crisis

Daniel Ciccarone

Department of Family and Community Medicine, University of California San Francisco, United States

ARTICLE INFO

Keywords:
Fentanyl
Heroin
Opioid
Overdose
Injection drug use

Introduction

The US mortality rate has gone up three years in a row from 2014 to 2017 (Xu, Murphy, Kochanek, & Arias, 2016; Murphy, Xu, Kochanek, & Arias, 2018; Kochanek, Murphy, Xu, & Arias, 2017). Correspondingly, life expectancy at birth has declined; the first triple year decline since World War One and the devastating influenza pandemic one hundred years ago (Tejada Vera, Bastian, & Arias, 2017). Most of the top ten causes of death are declining year over year; however, the third leading cause of death, unintentional injuries, has climbed in rate and rank since 2014 (Xu et al., 2016). Driving this are deaths due to drug poisoning which exceeded 70,000 in 2017 (Hedegaard, Miniño, & Warner, 2018). Annual deaths due to drug overdoses now exceed those from motor vehicle deaths, gun violence and even HIV at the height of the 1990s HIV epidemic (Katz, 2017).

The triple wave epidemic: opioid pills, heroin and synthetic opioids

The US is suffering a triple wave epidemic of overdose deaths from three classes of opioids: prescription opioid pills (“semi-synthetic opioids” in Fig. 1), heroin and synthetic opioids other than methadone (Ciccarone, 2017). Fig. 1 shows three waves of opioid mortality, each wave cresting on top of the one before it. In the first wave, overdoses related to opioid pills, started rising in the year 2000 and have steadily grown through 2016. The second wave saw overdose deaths due to heroin, which started increasing clearly in 2007, surpassing the number of deaths due to opioid pills in 2015. The third wave mortality has arisen from fentanyl, fentanyl analogues and other synthetic opioids of illicit supply, climbing slowly at first, but dramatically after 2013. Data from 2017 show synthetic opioid deaths continuing to rise, reaching a peak of over 28,000, while opioid pill and heroin overdose deaths leveled off, albeit at very high levels of approximately 15,000 deaths in each category (Hedegaard et al., 2018).

Supply and demand drivers

To address this crisis, its supply and demand drivers need to be better understood; both forces are needed to create the immense waves of consumption and their consequences that are occurring. Historians have observed the implications of supply in several opioid misuse cycles beginning with morphine in the latter half of the 19th century (Courtwright, 2001b; Musto, 1999). Isolated in 1805 by the German pharmacist Friedrich Sertturner, morphine was mostly dispensed by physicians in America, particularly to women (Courtwright, 2001a, 2001b). Thus the subsequent misuse problem was iatrogenic, exacerbated by the technological advance of the hypodermic syringe (Courtwright, 2001b, 2001b).

Heroin (diacetylmorphine) had a short life as a licit medication. There have been a number of illicit heroin waves beginning around the 1920s, the first of which may have been due to restriction on licit supplies driving its use underground. Consumption of all kinds occurs within an economic and cultural framework and both demand-side and supply-side forces can bring innovation to established consumption patterns. The upswings in American heroin use in the 1940s and ‘70s were in part stimulated by strong social and cultural elements, where heroin use often conferred an outsider status, signaling the rejection of mainstream values (Courtwright, 2001b). Thus the subsequent misuse problem was iatrogenic, exacerbated by the technological advance of the hypodermic syringe (Courtwright, 2001b, 2001b).

Heroin (diacetylmorphine) had a short life as a licit medication. There have been a number of illicit heroin waves beginning around the 1920s, the first of which may have been due to restriction on licit supplies driving its use underground. Consumption of all kinds occurs within an economic and cultural framework and both demand-side and supply-side forces can bring innovation to established consumption patterns. The upswings in American heroin use in the 1940s and ‘70s were in part stimulated by strong social and cultural elements, where heroin use often conferred an outsider status, signaling the rejection of mainstream values (Courtwright, 2001b). They also had strong supply-side forces with the post-WW2 emergence of the Italian and French ‘connection’ supplying heroin to the US (Courtwright, 2001b) and new sources of heroin imported from Southeast and Southwest Asia in the 1970s (McCoy, 2003). In the 1990s, a new form of heroin, produced by Colombian transnational criminal organizations (TCOs), was brought into the United States resulting in increased heroin use and adverse
In sum, since the uptake of morphine and subsequent licit and illicit opioids, the US has experienced multiple waves of opioid misuse and their medical consequences caused by the forces of supply (iatrogenic and new illicit sources) and demand (social, cultural and new technologies for use). In the current triple wave epidemic, we see both forces at work again. On the supply side, we have witnessed the iatrogenic sourcing of opioid pills, a new source-form of refined heroin and an illicit opioid sub-class, fentanyl, resurfacing from a new source. On the demand side, there are social and structural root causes of opioid use that have led to population dependency on opioids; starting with pills, yet leading to spill-over effects driving heroin and subsequently fentanyl demand.

Wave one: prescription opioid pills

The supply side drivers underlying the first wave of prescription opioid overdose have been extensively discussed (Madras, 2017; Van Zee, 2009). Wave one is often considered to have been driven iatrogenically with a tripling of opioid prescriptions starting in the 1990s and peaking around 2011 (Kolodny et al., 2015). This increase in prescriptions has been correlated to rising adverse consequences, particularly opioid overdose (Centers for Disease Control & Prevention, 2011). The introduction of extended release long-acting (ERLA) opioid formulations support both supply-side and demand-side pressures. ERLAs are a source of opioid in a novel form with a technological advance that allowed higher, longer lasting doses in a single capsule. However, the ease with which their delayed release mechanisms could by bypassed and the whole dose discharged at once, for instance by crushing and insufflating (nasal snorting) or injecting, led to a wave of misuse (Cicero, Ellis, & Surratt, 2012; Mars, Bourgois, Karandinos, Montero, & Ciccarone, 2014).

A demand-side argument has been introduced examining the structural factors that might be driving the epidemic. The most compelling structural determinants include an aging population with rises in reported pain and disability, economic distress, declining social cohesion and rising psychological malaise that may have led an at-risk population to seek opioids in the first place (Dasgupta, Beletsky, & Ciccarone, 2018).

Wave two: heroin

Coincident with rising heroin-related deaths, the number of heroin users, especially young heroin users, has been increasing since the mid-2000s (Center for Behavioral Health Statistics and Quality, 2017; Kilmer et al., 2014). The first two overdose waves, from opioid pills and heroin, have been termed ‘intertwined epidemics’ (Unick, Rosenblum, Mars, & Ciccarone, 2013). Young and new heroin users have described transitioning to heroin from opioid pills as their growing dependence required larger and more consistent pill supplies than they could obtain either by prescription or on the street. The more ready availability of high purity, low cost heroin made the switch to heroin economically logical and difficult to resist (Mars et al., 2014).

Drug treatment data show that in successive cohorts from the 1960s through the 2000s, patients admitted to treatment with heroin use disorder increasingly reported starting their opioid dependency with opioid pills (Cicero, Ellis, Surratt, & Kurtz, 2014). However, this has begun to change as an increasing proportion of heroin use disorder patients entering treatment report heroin as their first experience of an opioid (Cicero, Kasper, & Ellis, 2018). Overdoses due to heroin began to accelerate in 2011. In late 2010 OxyContin, a brand name ERLA formulation of oxycodone, was reformulated to be abuse-deterrent. The reformulation of this popular diverted opioid pill may have had the unintended consequence of driving a small proportion of the at-risk population to heroin (Cicero et al., 2012).

Although overlapping and related, the first and second opioid overdose waves show some contrasts in age and regional distribution that require explanation. Examining the years 2012–2014, the age distribution of patients hospitalized for opioid pill overdose had its largest peaks in the 50 to 64 year old group. Meanwhile, the peak age group for heroin overdose admissions was 20 to 34 year-olds. In this data we see possible evidence of population level transitions from opioid pills to heroin use as the rates for overdose among 20–34 year-olds declined for opioid pills while, in the same time period, increased for heroin (Unick & Ciccarone, 2017). In sum, we have seen elements of demand-side drive in wave two, with rising numbers of heroin users transitioning from opioid pill dependency followed more recently by younger persons initiating first with heroin.

Geographic disparities are also evident: opioid pill overdose is
relatively even across the country whereas heroin overdose is much higher in the US Northeast and Midwest regions, along with higher rates of increase (Unick & Ciccarone, 2017). Some of this regional disparity may be endemic, stemming from the 1970s, but there are also significant new supply-side forces shaping it. A dramatic transformation in the US heroin supply, including changes in its country of origin, has occurred in the last 10 years. Prior to 2000, heroin was imported from four source regions/countries in the world, including Southeast Asia, Southwest Asia, Mexico, and South America (Colombia). The 2000s began an era in which most US heroin was transshipped by TCOs from two countries: Colombia and Mexico (US Drug Enforcement Administration, 2015). Regional heroin distribution became starkly divided with Colombian-sourced heroin predominant in the eastern US while ‘black tar’ heroin from Mexico the major source-form in the western US (Ciccarone, 2009; Ciccarone et al., 2009). Accelerating this trend from oligopoly to monopoly, Mexican TCOs have increasingly dominated the US heroin market with their market share increasing from 50% in 2005 to 90% in 2016 (US Drug Enforcement Administration, 2017b).

Mexican-sourced heroin is also becoming more refined. From 2005 to 2012, a growing and substantial proportion of analyzed heroin samples obtained in eastern US cities by the US Drug Enforcement Administration (DEA) for its Heroin Domestic Monitor Program were from an unknown source and of unknown quality (Drug Enforcement Administration, 2016). Subsequent DEA analyses have led to the conclusion that a more refined heroin has emerged from Mexican sources. This so-called ‘Mexican White’ is a mimic of Colombian-sourced powder heroin, replacing it in its traditional retail outlets of the Northeast and Midwest (US Drug Enforcement Administration, 2015).

Wave three: synthetic opioids

Synthetic opioids in the heroin supply, chiefly illicitly produced fentanyl, are responsible for the third wave of overdose mortality (Ciccarone, 2017; NIDA, 2019). ‘Heroin’, particularly in the Northeast and Midwest, the regions with the greatest increases in wave two overdose, currently exists as fentanyl-adulterated and/or fentanyl-substituted heroin (FASH) (Ciccarone, Ondocsin, & Mars, 2017). Illicitly manufactured fentanyl is integrated into the illicit drug supply and sold as ‘heroin’ in powder form, or as counterfeit opioid or benzodiazepine pills (Gladden, Martinez, & Seth, 2016).

Fentanyl is the main chemical in a growing family of chemical analogues. These analogues come in a range of morphine-equivalent potencies with some such as butyryl-fentanyl being less potent than fentanyl by weight while others have much greater potency (Suzuki & El-Haddad, 2017). In addition, there are other novel synthetic opioids in circulation including U47700 and U48800. The greatest concern arises from a branch of the fentanyl family that includes some extremely potent opioids including carfentanil, sufentanil and re-mifentanil. It is unclear whether these extremely potent fentanyl will become established elements in the opioid marketplace or if they are just accidents or experiments in the rapidly evolving illicit opioid supply.

According to the US Drug Enforcement Administration (DEA), the main source of illicitly manufactured fentanyl is China. Fentanyl sourced from China take a number of routes on their way into the US including internet purchases, routing through Canada (typically pill form), or through Mexico in powder or pill forms (US Drug Enforcement Administration, 2016). Perhaps the most revealing aspect of the supply that fuels the US fentanyl epidemic is its regional discreetness. Comparing drug seizure data with overdose death data one finds remarkable geographical correlation between fentanyl seizures and synthetic opioid overdoses (Gladden et al., 2016). These fentanyl events overlap in the same regions as wave two heroin overdose: the Northeast and Midwest. The reasons for this regional disparity are unclear. One possibility is that fentanyl distribution is regionally orchestrated by a branch of the Sinaloa TCO (US Drug Enforcement Administration, 2017b). Another hypothesis is that source-forms of powder heroin, predominant in the Northeast and Midwest, are more easily adulterated with powder fentanyl than solid ‘black tar’ heroin, which predominates in the western US (Carroll, Marshall, Rich, & Green, 2017; Ciccarone, 2017). Such stark regional disparities support the notion that the third wave is a supply-side event (Mars, Rosenblum, & Ciccarone, 2018). A demand or culturally driven event, such as through entrepreneurial or individual internet purchases, would more likely have led to a more even geographic spread of fentanyl-related overdose or one that reflected similar social conditions in separate geographical locations.

Ethnographic research with persons who use heroin confirm that the introduction of FASH has been unexpected and unsettling, that fentanyl was not a demand-driven phenomenon and that there is a range of desirability for FASH from abhorrence and avoidance through acceptance to enthusiasm (Carroll et al., 2017; Ciccarone et al., 2017; Mars, Ondocsin, & Ciccarone, 2018). Those who favor fentanyl are nevertheless hampered from choosing it in the market place by its concealed identity as ‘heroin’ or counterfeit brand name pills (Ciccarone et al., 2017; Mars, Ondocsin et al., 2018; Mars, Rosenblum et al., 2018). Importantly, cultural idioms for fentanyl have been slow to emerge despite four years of steady supply; slang terms have arisen for most desired illicit drugs and their absence is evidence for a lack of strong demand. Consequentially, emergence of slang can be seen as a marker for growing acceptance of fentanyl.

In addition to the dangerous potency of fentanyl, ethnographic observations support the notion of a possibly greater danger: rapid changes in potency and purity, as well as varying mixtures of heroin, fentanyl and its analogues (Ciccarone, 2017; Ciccarone et al., 2017; Mars, Ondocsin et al., 2018; Mars, Rosenblum et al., 2018). Australian research on heroin overdose has shown that fluctuations within a wider range of street heroin purity, particularly when around a higher mean purity level, are an independent predictor of fatal overdose (Darke, Hall, Weatherburn, & Lind, 1999). Vicissitudes in potency/purity/mixture in the fentanyl street market may be discovered to have profound effects on the overdose rate in a given location.

In summary, all three waves have impressive supply-side drivers including excessive prescribing of medication, a new form of highly refined Mexican-sourced heroin and a new illicit source of synthetic opioids adulterating heroin and counterfeit pills. Demand for opioid pills partially drove demand for heroin while demand for heroin unsurprisingly feeds demand for synthetics-as-substitute. What is driving increases in opioid mortality now are deaths due to FASH. The second and third waves are regional, with the Northeast (including Mid-Atlantic) and Midwest (including Appalachia), the most affected regions.

There are other medical consequences, in addition to overdose, that are growing in concern. The change from opioid pill misuse to heroin involved, for many, a change in route of administration from oral ingestion to intravenous injection. While heroin can be smoked or inhaled, it tends, in the US, to be injected (Ciccarone, 2019). This raises concerns about the transmission of blood-borne viruses such as hepatitis C and HIV. The US has recently had two documented injection-drug-related HIV outbreaks in Scott County, Indiana and eastern Massachusetts (Peters et al., 2016; US Centers for Disease Control & Prevention, 2018).

Addressing the fentanyl crisis

The triple-wave opioid overdose epidemic is an intertwined, three drug sub-class epidemic. To comprehend and address it fully the drivers of each wave need to be elucidated. Positive supply shocks have historically led to drug epidemics and the same can be seen in the current opioid crisis. Fentanyl, in particular, comes as a positive supply shock leading to disastrous consequences. It is thus tempting to focus efforts on controlling supply. There is evidence that supply-side interventions...
can work if part of a comprehensive program that also includes demand reduction (Caulkins, Reuter, Iuigi, & Chiesa, 2005). Unipolar supply-side interventions however, may cause paradoxical unwanted results; for example see (Ciccarone et al., 2009). These phenomena may be occurring within the current crisis. Downward pressure on opioid prescribing may be driving a portion of the at-risk population from opioid pill misuse to heroin, thus exposing them to the even more dangerous family of fentanyl. Another driver of unintended consequences may have been the reformulation of ERLA opioids to abuse-deterrent formulations, examples of which include OxyContin and Opana, with misuse of the latter implicated in the Scott County Indiana HIV outbreak (Broz et al., 2018). The goals of curtailing excessive prescribing practices or creating abuse-deterrent formulations may be reasonable, but the untoward consequences must be recognized, monitored and responded to accordingly.

The end of interdiction?

Regarding fentanyl, supply-side interventions include source control and interdiction in the drug supply chain. Considering source control, it is crucial to find areas of cooperation between the United States and the Government of China to improve monitoring, regulation and enforcement of pharmaceutical and chemical manufacturing in order to deter illicit suppliers (Pardo, 2018). Bilateral efforts by governmental agencies in both countries have led to an expanding list of controlled psychoactive substances, including opioids and precursors, leading to discussions regarding scheduling fentanyl as a class (Knierim, 2018). Interdiction will be challenging given the size of illicit fentanyl flows. In 2016, a mere 668 kg of fentanyl was seized in the US (Baum, 2017), a fraction of the estimated 11 metric tons of cocaine seized in 2016 at the US Southwest Border alone (US Drug Enforcement Administration, 2017a). Fentanyl’s high potency allows shipment in small volumes. Considering a seizure to importation ratio of 1:4, a total of 2.6 metric tons of fentanyl may have been distributed in the US in 2016. This would fit into approximately 10 industrial drum barrels – a tiny volume that if divided up over the huge trade that occurs across the Pacific Rim constitutes a proverbial needle in a haystack.

The “Iron Law of Prohibition” suggests that highly potent-by-volume drugs like fentanyl are expected due to the strong effects of interdiction (Beletsky & Davis, 2017). Following this is the concern that constraining the mother chemical fentanyl too robustly and too rapidly will foster the supply of fentanyl analogues. The number of known fentanyl analogues exceeds 60; the number of potential fentanyl analogues could exceed 600. Care must be taken not to foster the ingenuity and creativity of the illicit drug manufacturers to push in even more dangerous directions. Based on this concern, the DEA has imposed a first-ever class restriction on the family of fentanyl, the utility of which is uncertain.

Surveillance of the drug supply

One supply-side intervention with potentially positive impact is drug surveillance. Investments in drug monitoring, identification and data collection could assist in interdicting supply (Pacula & Powell, 2018). US government officials have called for greater public safety and public health collaboration to address this crisis. One way for these two domains to work together is by increasing local drug surveillance – with sharing of the data. Local criminal justice systems have a ready pool of geo-located analyzable drug samples in their crime labs. Analysis of drug samples and dissemination of the findings could be achieved in rapid cycles and enhance our understanding of the drugs in circulation and especially how rapidly heroin, FASH and fentanyl analogue mixtures are changing.

Improved surveillance would benefit not only the interdiction and public safety side but also the public health side including first responders, emergency and hospital clinicians as well as those who work in community based programs serving the affected population (Ciccarone, 2017). Point of use testing or drug checking is an intimate form of surveillance that is emergent in the US because of the FASH crisis. Evidence for its utilization is growing (Peiper et al., 2018), although concerns have also been raised (McGowan, Harris, Platt, Hope, & Rhodes, 2018).

Harm reduction

Considering the inadequate and paradoxical effects of current opioid supply interventions, supply-side policies must be combined with sufficient investments in and expansion of effective prevention, substance use treatment and harm reduction (Paucala & Powell, 2018). Opioid use disorder has a number of medical treatment options that have been shown to be medically efficacious as well as cost-effective (Volkow, Frieden, Hyde, & Cha, 2014). The US Surgeon General has called for greater distribution of naloxone, the opioid antagonist used to treat an opioid pill, heroin or fentanyl overdose (US Surgeon General, 2018). Getting wider distribution of naloxone into the community is an essential strategy in the current epidemic (Fairbairn, Coffin, & Walley, 2017; Wheeler, Jones, Gilbert, & Davidson, 2015). Sterile syringe provision must be greatly expanded to meet the increasing population at risk. Drug surveillance, discussed above, can be utilized as a harm reduction strategy. Supervised consumption spaces can aid in prevention of overdose and reduce HIV and HCV transmission risks (Dolan et al., 2000; Marshall, Milloy, Wood, Montaner, & Kerr, 2011; Potier, Laprévote, Dubois-Arber, Cottencin, & Rolland, 2014). These services can act as a safety net or hub for persons at risk and provide them with necessary resources and referrals to services (Kerr, Small, Moore, & Wood, 2007).

Conclusion

A crisis level response is needed to address the triple wave epidemic. In the 1990s, US government intervention, albeit slow to start, was essential in curtailing the HIV epidemic. The Ryan White Care Act led to a dramatic increase in funding for HIV prevention and treatment. That coupled with medical progress led to a dramatic decrease in HIV incidence, morbidity and mortality. The economic and social costs of the opioid crisis are enormous, with estimates approaching $80 billion per year. Estimates to address it range from $60 billion for treatment over the next 5 years (Simmons, 2017) to $100 billion for a multi-pronged approach to prevention, treatment and community resilience efforts (Katz, 2018). The opioid epidemic is the latest phase of a multi-decade increase in drug-related mortality (Jalal et al., 2018). With overdose deaths increasing steadily since 1979, regardless of the primacy of any particular drug, more resolve is needed to address the structural determinants of this relentless epidemic.

Conflicts of interest

Dr Ciccarone reports receiving consulting fees from Mallinckrodt Pharmaceuticals and Nektar Therapeutics.

Acknowledgements

I gratefully acknowledge research funding from US National Institutes of Health, National Institute of Drug Abuse Grant DA037820. I thank Dr. Sarah Mars for extensive feedback and editing and Dr. Jay Unick for creating Fig. 1.

References


