Long-Term Effects of Increasing Buprenorphine Treatment Seeking, Duration, and Capacity on Opioid Overdose Fatalities: A Model-Based Analysis

Erin J. Stringfellow, PhD, Tse Yang Lim, PhD, Catherine DiGennaro, BA, Ziyuan Zhang, Pritika Paramasivam, Benjamin Bearnot, MD, Keith Humphreys, PhD, and Mohammad S. Jalali, PhD

Objectives: Because buprenorphine treatment of opioid use disorder reduces opioid overdose deaths (OODs), expanding access to care is an important policy and clinical care goal. Policymakers must choose within capacity limitations whether to expand the number of people with opioid use disorder who are treated or extend duration for existing patients. This inherent tradeoff could be made less acute with expanded buprenorphine treatment capacity.

Methods: To inform such decisions, we used a validated simulation model to project the effects of increasing buprenorphine treatment-seeking, average episode duration, and capacity (patients per provider) on OODs in the United States from 2023 to 2033, varying the start time to assess the effects of implementation delays.

Results: Results show that increasing treatment duration alone could cost lives in the short term by reducing capacity for new admissions yet save more lives in the long term than accomplished by only increasing treatment seeking. Increasing provider capacity had negligible effects. The most effective 2-policy combination was increasing capacity and duration simultaneously, which would reduce OODs up to 18.6% over a decade. By 2033, the greatest reduction in OODs (≥20%) was achieved when capacity was doubled and average duration reached 2 years, but only if the policy changes started in 2023. Delaying even a year diminishes the benefits. Treatment-seeking increases were equally beneficial whether they began in 2023 or 2025 but of only marginal benefit beyond what capacity and duration achieved.

Conclusions: If policymakers only target 2 policies to reduce OODs, they should be to increase capacity and duration, enacted quickly and aggressively.

Keywords: buprenorphine treatment, opioid overdose deaths, modeling analysis

(J Addict Med 2023:00: 00–00)

Opioid overdose deaths (OODs) in the United States reached their highest levels ever recorded in 2021, surpassing 75,000—a 13% rise from 2020. Buprenorphine is a Food and Drug Administration–approved medication for opioid use disorder (OUD) that reduces OODs. After multiple years and billions of dollars of federal investment, buprenorphine providers have doubled since 2018 and the number of programs offering buprenorphine has increased 15% since 2017, while receipt has steadily increased. Nonetheless, too few people with OUD receive any treatment, due to barriers ranging from stigma to inaccessibility and lack of insurance coverage. Once buprenorphine treatment has begun, people often receive it for much less time than is needed. Buprenorphine treatment episodes average 8 to 9 months, with only half of patients making it to 4 months and a third making it to 6 months. However, at least 15 months is associated with the greatest improvements and reductions in adverse events, although the optimal duration of treatment remains unknown.

Because all forms of health care have finite capacity, there is an inherent tension between how many individuals can receive care and how intensive that care can be. This principle applies to buprenorphine treatment of OUD, which has inspired an important debate about whether public health would be furthered more by increased treatment entry (ie, greater population coverage) or increased duration for those who access treatment. Currently, the buprenorphine treatment system has inadequate capacity to meet even existing demand. At most, 62% of callers seeking buprenorphine treatment obtain an appointment. In the 10 states with the highest OOD rates, three quarters of buprenorphine providers did not have available appointments, and the remainder had an average waitlist of over 2 weeks. This is concerning because appointments scheduled as little as 2 days after first contact are associated with significantly higher no-show risk than same- or next-day appointments. If existing patients are retained in treatment longer, these limitations on enrolling new patients would become more acute. Outreach efforts to increase treatment initiation could...
create additional demand and make successfully accessing treatment even harder.

At state and federal levels, capacity-building efforts are underway. Concern that the DATA 2000 (Drug Addiction Treatment Action of 2000) waiver requirement posed a significant barrier to prescribing buprenorphine has led to incentivizing waiver training and, more recently, relaxing training requirements. However, the waiver is not, in fact, a significant barrier to offering treatment. Once waivered, most providers treat very few patients; in 2017–2018, they treated a median of 1.5 patients at any given time. Consequently, half of buprenorphine prescriptions are written by 5% of prescribers. A recent analysis suggested that an increase in waivered providers is unlikely to result in meaningful OODs reductions at the national level. Rather, increasing existing providers’ capacity to treat more patients reduced OOD more. Thus, it is essential for policymakers to consider federal strategies that support providers to increase the number of patients they treat, and not simply increase waivered providers.

The present analysis examines changes in OODs that result from increases in treatment seeking (necessary for treatment initiation), average episode duration, and buprenorphine provider capacity (patients per provider) from 2023 to 2033. We further examine how the timing of these increases changes outcomes. Pursuing increased entry, duration, and capacity simultaneously and with equal intensity is not possible because these initiatives would compete for available federal resources. Even when competition not present, some efforts could yield diminishing returns. We examine this problem using dynamic simulation modeling, an essential tool to guide decision making when unintended consequences are a real risk. Dynamic simulation models account for complexities in systems, such as competing goals, to project short- and long-term intervention outcomes, analyze tradeoffs, and identify the optimal timing for interventions.

METHODS

Overview of SOURCE

We used SOURCE (Simulation of Opioid Use, Response, Consequences, and Effects), a US population-level model calibrated to data (1999–2020). SOURCE tracks the population through opioid misuse, OUD, medication for OUD treatment, remission, and OOD (Fig. S1, http://links.lww.com/JAM/A414). We estimated initiation, misuse, and OUD prevalence using the National Survey on Drug Use and Health. We assumed that National Survey on Drug Use and Health’s trends are accurate but that the absolute numbers are not, so we adjusted for underreporting of heroine use (S2a, http://links.lww.com/JAM/A414). Because OUD prevalence data underpin our results, we tested SOURCE’s sensitivity to potential inaccuracies in those data (S2ai, http://links.lww.com/JAM/A414). The sensitivity analysis showed that historically estimated or future projected trends are not meaningfully sensitive to these potential inaccuracies (S2ai, http://links.lww.com/JAM/A414).

SOURCE uses a feedback perspective: changes in one part of the system reverberate throughout the system, eventually feeding back to the original source. The feedback perspective supports making more realistic projections of OUD and OOD and can identify potential unintended consequences of interventions. For this analysis, there is a treatment capacity balancing feedback effect, described hereafter.

Capacity Balancing Feedback Effect

SOURCE explicitly acknowledges finite treatment capacity. As more people enter treatment, capacity is increasingly used, limiting treatment entry for new patients. Thus, treatment entry cannot be artificially increased in SOURCE. Rather, entry arises endogenously from the interaction between treatment seeking and buprenorphine capacity (Fig. 1). To increase treatment entry in SOURCE, either capacity or treatment seeking, or both, must be increased, depending on which is the limiting factor. However, increasing the average treatment episode duration slows treatment exit, thus using capacity; hence, increasing average duration indirectly reduces treatment entry.

Model Parameters Modified for This Analysis

Treatment seeking includes any actions taken to receive treatment, such as phone calls to providers or Internet searches, even when these actions do not result in treatment receipt. We tested increased treatment-seeking rates of 5% to 20%, based on the proportion of people with OUD who report a perceived need for treatment but do not seek it. In SOURCE, overdose fatality risk decreases while in buprenorphine treatment, so a longer treatment duration translates into more sustained protection. The longer people are retained in buprenorphine treatment, the more likely they are to exit treatment in remission from OUD. Therefore, longer durations reduce repeated episodes. We tested treatment duration at magnitudes ranging from a 10% to a 2.3-fold increase, equivalent to 2 years.

The average number of patients that can be treated per provider ("capacity") is multiplied by “Total buprenorphine providers” in SOURCE to arrive at Buprenorphine capacity (Fig. 1)—the total number of patients that can be treated. In SOURCE’s baseline, capacity from 2023–2033 is between 5 and 6—much fewer than the 30 or more patients that providers can legally treat. In SOURCE, overdose fatality risk decreases while in buprenorphine treatment, so a longer treatment duration translates into more sustained protection.

Model Analysis and Outcomes

Our analysis projected fractional changes in cumulative OODs resulting from increases in treatment seeking, duration, and capacity. Accumulation of OODs begins in 2023. The fractional changes are relative to baseline projections, which reflect the status quo. In the status quo, we assumed that the number of buprenorphine providers will continue to increase, albeit at a decelerating pace. We made no other assumptions regarding...
time-varying inputs relevant to this analysis. All model inputs and their sources have been previously reported.25,27

See Table S1, http://links.lww.com/JAM/A414, for our model analysis and outcomes. We refer to increases as “policies.” We assumed all policies came into effect over 5 years rather than instantaneously. We varied the policy start times from the years 2023–2025 to assess the potential effects of delays in enacting policy decisions. This yielded implementation time frames of 2023–2028, 2024–2029, and 2025–2030.

We first tested increases in treatment seeking, treatment duration, and capacity, alone and in combinations of 2 and 3, with all policies implemented simultaneously. We report the short- and long-term reduction in OODs, defined respectively as 3 and 8 years after the policy start year (ie, in 2026 and 2031 for the implementation time frame of 2023–2028). Comparing outcomes at fixed points after policy start (rather than in specific years) allows direct assessment of differences in outcomes arising from the timing of implementation, rather than how long policies have been in effect. We assess outcomes at 2 time points to identify worse-before-better and better-before-worse scenarios.

We then repeated the 3-policy combination but varied the policy start times. Unlike the first set of analyses, we assessed cumulative reductions in OODs in 2033 (a fixed point) regardless of policy start time. This was to identify where delayed implementation does not have a significant detrimental impact, which could provide guidance on how to prioritize allocation of limited resources.

Once the policy reaches its full effect, it remains there, representing sustained change and not short-lived interventions. We also tested a subset of the analyses where direct policy effects cease after the 5-year implementation period, representing short-lived interventions. We tested all policies at their maximum value starting in 2023 and assessed the reduction in OODs 8 years after implementation (2031) and as of 2033.

RESULTS

Baseline

As detailed elsewhere, in the baseline, scenario buprenorphine treatment demand and thus receipt are projected to start declining by 202525 due to falling OUD prevalence.27 Changing assumptions about OUD prevalence data (S2ai, http://links.lww.com/JAM/A414) delays this projected fall by 1 year. Decreasing prevalence and thus demand reduces capacity constraints, allowing greater entry for those who seek buprenorphine. Until then, the system is projected to remain capacity limited, such that on average nationally, new patients cannot enter without existing patients leaving.

All results hereinafter remain qualitatively the same when using different assumptions about OUD prevalence (S2ai, http://links.lww.com/JAM/A414).
Short- and Long-Term Effects of Increasing Treatment Seeking, Duration, or Capacity Alone

Figure 2 shows the results of single policy changes in the short and long term.

Increasing treatment seeking reduces OODs by less than 1% in the short term, regardless of what year the policy starts. At most, OODs reduce in the long term by 2.1%, if started in 2023. Later implementation erodes these gains.

The effect of increasing duration differs depending on magnitude and timing. Increases of at least 70%, beginning in 2023, create a worse-before-better outcome; deaths are projected to rise in the short term but more lives are saved in the long term (almost 3%) than achieved by treatment seeking or capacity increases. Delaying the policy start time to 2025 yields even greater reductions in OODs—up to 4.2%, which is also more than any other single policy change. This only occurs with an increase to approximately 14 months on average. Increases beyond 14 months result in fewer lives saved because keeping patients in treatment longer exacerbates capacity constraints. With fewer people entering treatment, more people experience overdose than would have otherwise.

Increasing capacity starting in 2023 has a better-before-worse outcome, as OODs fall more in the beginning than they do later. In the short term, OODs are reduced no more than 0.5%, falling to no more than 0.2% in the long term. Starting in 2024 or 2025 yields no benefit.

Short- and Long-Term Effects of 2-Policy Combinations

Figure 3 shows the results of 2-policy combinations implemented simultaneously.

Simultaneously increasing treatment seeking and duration had mixed results depending on the magnitude and timing of the duration increase. Starting in 2023 resulted in more deaths in the short term (up to 1.2%) unless implementing the smallest increases in duration. A 3.6% decrease in OODs is the greatest possible reduction in the long term, if implementation begins in 2023, but duration can only be increased by 30% before OODs begin to rise because of unmet demand. Delaying implementation to 2025 did not increase OODs in the short term and slightly increased the long-term benefits, reducing deaths by at most 4.5% in the long term. However, if duration exceeds a year, these benefits decline.

Of the 2-policy combinations, simultaneous increases in duration and capacity produced the largest reductions in OODs—up to 18.6% fewer deaths in the long term, if implemented in 2023 and with average duration reaching 2 years. Although this required increasing capacity to 28 patients/provider, OOD reductions almost as great (−17.4%) were achieved with only an additional 7 patients/provider. An increase of only one patient/provider combined with duration increases to a year reduced deaths by 4.6%, more than the other 2-policy combinations.

Finally, increasing capacity and treatment seeking simultaneously achieved, at most, a short-term 1.4% decline in OODs and a long-term 2.5% decrease, if implemented in 2023. In the short term, this required as few as 2 additional patients/provider. However, even these small reductions required the greatest increase in treatment seeking we tested (20%). These benefits quickly drop off with later implementation times.

Three- and 8-Year Effects of 3-Policy Combinations

Simultaneously increasing treatment seeking, duration, and capacity in 2023 reduced OODs in the short term by up to 8.6% (the upper right corner of the upper right cell of Fig. S3, http://links.lww.com/JAM/A414). In the long term (2031), this same policy combination reduced OODs by up to 20.9% (Fig. S4, http://links.lww.com/JAM/A414). This represents a 2.3
percentage point gain over the 2-policy combination that did not increase treatment seeking. In the long term, similar reductions were seen in OODs regardless of the magnitude of treatment-seeking increases. These short- and long-term reductions fell to 6.5% and 18% with a policy start in 2024 and 4.9% and 15.5% with a policy start in 2025.

Even with all 3 policies, it is possible to create a worse-before-better outcome. Increasing average duration to at or near 2 years without a concomitant increase in provider capacity of at least 5 more patients/provider increases OODs in the short term before falling in the long term.

Three-Policy Combinations Achieving 20% Reduction in OODs by 2033, With Varying Start Times

Finally, we tested increases in treatment seeking, duration, and capacity together at varying magnitudes and policy start times. This helped identify whether delaying implementation of one or more policies could achieve equivalent or even greater effects than starting them all simultaneously. The result was 11,664 possible combinations, 66 of which reduced OODs by at least 20% by 2033.

All scenarios achieving at least a 20% reduction in OODs required a minimum 210% increase in average treatment duration (to 1.9 years). Reductions in OOD were seen with as little as 5% treatment-seeking increases. Provider capacity would require, at minimum, just over a doubling (to 12 patients/provider, from an estimated 5 at baseline in 2028).

To identify more feasible options, we excluded scenarios involving the largest increases in all 3 policies. The 9 remaining scenarios are in Table 1; for each of the 3 policies, we report the final value after the increase rather than the percentage increase. All resulting scenarios required starting capacity and duration increases in 2023, whereas increases in treatment seeking could be delayed to 2025.

<table>
<thead>
<tr>
<th>Treatment-Seeking Rate for Rx OUD and HUD* (Attempts per Person-Year)</th>
<th>Latest Start Time to Increase Treatment Seeking</th>
<th>Average Buprenorphine Episode Duration, yr†</th>
<th>Latest Start Time to Increase Duration</th>
<th>Provider Capacity (Patients per Provider That Can Be Treated)‡</th>
<th>Latest Start Time to Increase Capacity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 0.32–1.38</td>
<td>2023</td>
<td>2</td>
<td>2023</td>
<td>15</td>
<td>2023</td>
</tr>
<tr>
<td>2 0.32–1.38</td>
<td>2024</td>
<td>2</td>
<td>2023</td>
<td>15</td>
<td>2023</td>
</tr>
<tr>
<td>3 0.32–1.38</td>
<td>2025</td>
<td>2</td>
<td>2023</td>
<td>15</td>
<td>2023</td>
</tr>
<tr>
<td>4 0.34–1.44</td>
<td>2023</td>
<td>2</td>
<td>2023</td>
<td>12</td>
<td>2023</td>
</tr>
<tr>
<td>5 0.34–1.44</td>
<td>2023</td>
<td>2</td>
<td>2023</td>
<td>15</td>
<td>2023</td>
</tr>
<tr>
<td>6 0.34–1.44</td>
<td>2024</td>
<td>2</td>
<td>2023</td>
<td>12</td>
<td>2023</td>
</tr>
<tr>
<td>7 0.34–1.44</td>
<td>2024</td>
<td>2</td>
<td>2023</td>
<td>15</td>
<td>2023</td>
</tr>
<tr>
<td>8 0.34–1.44</td>
<td>2025</td>
<td>2</td>
<td>2023</td>
<td>12</td>
<td>2023</td>
</tr>
<tr>
<td>9 0.34–1.44</td>
<td>2025</td>
<td>2</td>
<td>2023</td>
<td>15</td>
<td>2023</td>
</tr>
</tbody>
</table>

The largest reduction observed was 23%, so we do not make a distinction once more than 20%.

*The baseline treatment-seeking rate among people with OUD involving prescription opioids (Rx OUD) or OUD involving heroin (HUD) is 0.31 and 1.31 treatment-seeking attempts per person-year, respectively.

†The baseline average buprenorphine episode duration is 0.61 years.

‡The baseline average provider capacity as of 2023 is 6 patients per provider, decreasing to 5 by 2028.

© 2023 American Society of Addiction Medicine
Maximum OOD Reductions When Stopping the Policy After 5 Years

Short-lived policies have lower impacts than sustained ones, although the same policies or policy combinations remain most effective at reducing OODs: for single policies delaying an increase in duration (2.6% reduction vs 4.2% when sustained); for 2-policy combinations, capacity and duration (12.5% reduction vs 18.6%); and 14% OOD reduction for all 3 policies combined versus 20.9% when sustained. In all cases, and in contrast to when policies are sustained, the cumulative reduction in OODs was smaller in 2033 than in 2031.

DISCUSSION

In an ideal world, the healthcare system could treat an infinite number of patients with OUD for an infinite length of time. However, in real-world settings, clinicians and policymakers must balance increasing access to and duration of medication for OUD treatment with practical considerations of capacity constraints. Our results show that increasing duration without increasing capacity—specifically, the average number of patients that providers can treat—may inadvertently lead to more deaths. In contrast, increased treatment seeking does not require more capacity, if duration is not increased. However, churning more people through brief episodes of treatment will likely benefit each patient less than would retaining them in care for an extended period.14,15 All 3-policy improvements—in treatment seeking, duration, and capacity—are thus desirable, and this study attempted to identify which type and timing of improvements would ideally be part of the strategy to meaningfully reduce OODs.

In our analysis, maximizing average treatment duration was critical to achieving the largest reductions in OODs, whereas increases in provider capacity and treatment seeking can vary much more and still achieve large (≥20%) reductions in OOD. Opioid overdose death reductions of 20% or more required increasing duration and capacity soon—as early as 2023, although treatment-seeking increases can be delayed somewhat. Keeping more people in treatment now, when OUD prevalence is still near its peak, is critical to reduce OUD and thus OODs. If the policies start later, when we project OUD prevalence will be lower,25,27 there is less room for improvement. Finally, maximum reductions in OODs are smaller and decline rather than grow over time, if the policies are not sustained.

Many people with OUD receive treatment in systems not designed to support them long term.12 Efforts are already underway to increase duration, with greater focus on chronic care models that de-emphasize punitive policies for nonabstinence or nonattendance.33–35 Programmatic efforts to increase retention include ensuring continuity of care for people in jails and prisons36 and those experiencing homelessness.37 Telehealth shows great promise in sustaining patients in the long term, perhaps for as long as 2 years.38

In our analyses, increases in duration required concomitant increases in capacity to avoid short-term increases in OODs. We estimate that the average provider would need to be able to prescribe to 12 to 15 patients (up from 5 to 6) to achieve 20% reductions in OODs by 2033. Buprenorphine providers’ capacity is constrained by numerous barriers, including stigma, insufficient reimbursement, lack of confidence in treating OUD, and inadequate patient access to more intensive levels of care.11 Only one study, of Vermont’s hub-and-spoke model, has tracked patients per provider as an outcome.33 That study found that over 5 years, Vermont achieved a 50% increase in capacity. In SOURCE, this translates to 7 to 8 patients/provider. By 2033, the greatest reduction in deaths achieved at this level was 15%, when combined with 20% treatment-seeking increases and 2-year average duration.

Treatment-seeking increases conferred the least additional benefit when tested together with increases in duration and capacity. This is because of the underlying secular trend of decreasing OUD prevalence,10 even when accounting for heroin use underestimates.37 There are diminishing returns on outreach to a shrinking population. Our tested range of treatment-seeking increases (5%–20%) included the 8.2% we defined as an anchor (S2b, http://links.lww.com/JAM/A414). This reflects the maximum possible fractional increase achieved if outreach were successful to those who perceive a need for treatment but do not seek it. Further increases would correspond, for instance, to outreach to increase the fraction of people with OUD who perceive a need for treatment and who also go on to seek it. Emergency department induction might be the closest real-world comparator to our test. Up to 50% increases in treatment engagement have been observed relative to controls,39 although effects on retention (ie, duration) are weak.40

Finally, our findings highlight policies that work better in the short term than the long term, and vice versa. Increasing duration without capacity created worse-before-better effects while the effect of increasing capacity, in contrast, weakens over time. These tradeoffs disappear when the 2 policies are combined, suggesting that these policies should be implemented together.

Limitations

Our finding regarding average treatment duration was limited to a 2-year window. We cannot say with certainty what longer durations might achieve, although they generally confer better outcomes.14,15 Another limitation is that because of limited evidence regarding long term (ie, at least 1 year) remission during and after treatment,38 SOURCE uses parameters based on expert judgment. Regardless, the relative impact of increasing duration compared with capacity would not be expected to change because any short-term increases in duration will require a capacity increase in our current capacity-constrained system.

We do not account for the possibility that as people stay in treatment longer, their treatment needs might become less complex, reducing the per-patient time needed to treat them and freeing up provider capacity to treat new patients. We could be overestimating the negative impact that increased duration has in the short term on limiting new patient engagement. On the other hand, we do not account for the possibility that the perceived urgency of the crisis could dissipate as OUD or OOD fall, which would shift resources and priorities away from buprenorphine prescribing. This could mean capacity falls below our baseline estimated level. Thus, our results are most valid assuming that providers and systems maintain, at minimum, their current dedication to treating OUD.
SOURCE is a national simulation model that was not built to account for local variation in treatment seeking, capacity, or duration. Therefore, strategic timing might need to shift regarding these 3 policies. For instance, in areas where buprenorphine capacity is lower than the national average, policymakers might anticipate that near-term capacity-building efforts will have a stronger and longer impact on OODs than shown in our results; however, for maximum impact, they, too, will need to enact policies that increase duration.

Finally, we are aware of the benefits of methadone and extended-release injectable naltrexone, which many patients might prefer. We focus only on buprenorphine because of a lack of national capacity data on these other medications.

CONCLUSIONS

To reduce OODs, our modeling analysis finds that a focus on buprenorphine treatment entry is less urgent than quickly and aggressively implementing policies that increase capacity and duration. Although there are significant challenges to achieving these goals, failure to do so only serves to maintain the status quo.

ACKNOWLEDGMENTS

The authors thank Zeynep Hasgül for her assistance in conducting sensitivity analysis calibrations.

REFERENCES

10. LaBelle CT, Han SC, Bergeron A, et al. Office-based opioid treatment with buprenorphine (OBOT-B): Statewide implementation of the Massachusetts Opioid Overdose Fatalities


