

D6b Final Validation Report Part 1

Prepared exclusively for the U.S. Food and Drug Administration

Table of Contents

[List of Tables ii](#_Toc52396433)

[List of Figures ii](#_Toc52396434)

[1 Executive Summary 3](#_Toc52396435)

[2 Introduction 3](#_Toc52396436)

[3 Constraints and Limitations 4](#_Toc52396437)

[4 Review of Harvard Grant Team Responses to Model Verification #2 5](#_Toc52396438)

[5 Validation by Section 6](#_Toc52396439)

[5.1 Materials Validation 6](#_Toc52396440)

[5.1.1 Documentation 6](#_Toc52396441)

[5.1.2 Major Findings 7](#_Toc52396442)

[5.2 General Feedback 9](#_Toc52396443)

[5.3 Prescribing Practices 12](#_Toc52396444)

[5.3.1 Structural Findings 14](#_Toc52396445)

[5.3.2 Behavioral Findings 15](#_Toc52396446)

[5.4 Prescription Availability and Price 16](#_Toc52396447)

[5.4.1 Structural Findings 17](#_Toc52396448)

[5.4.2 Behavioral Findings 17](#_Toc52396449)

[5.5 Medical Use and Misuse 18](#_Toc52396450)

[5.5.1 Structural Findings 23](#_Toc52396451)

[5.5.2 Behavioral Findings 25](#_Toc52396452)

[5.6 Use Disorder 30](#_Toc52396453)

[5.6.1 Structural Findings 32](#_Toc52396454)

[5.6.2 Behavioral Findings 33](#_Toc52396455)

[5.7 Treatment 34](#_Toc52396456)

[5.7.1 Structural Findings 37](#_Toc52396457)

[5.7.2 Behavioral Findings 37](#_Toc52396458)

[5.8 Effects of Abuse-Deterrent Formulations (ADF) 43](#_Toc52396459)

[5.8.1 Structural Findings 43](#_Toc52396460)

[5.8.2 Behavioral Findings 44](#_Toc52396461)

[6 Conclusion 45](#_Toc52396462)

[6.1 Summary of Findings 45](#_Toc52396463)

[6.2 Next Steps 45](#_Toc52396464)

[7 References 46](#_Toc52396465)

[8 Appendix 47](#_Toc52396466)

[8.1 Replication of Goodness-of-Fit Results 47](#_Toc52396467)

[8.2 Buprenorphine Treatment (Bup) Capacity Startup Delay 50](#_Toc52396468)

[8.3 Statement of Work (SOW) Checklist 51](#_Toc52396469)

List of Tables

[Table 4‑1: Review of Response to Model Verification #2 5](#_Toc52396470)

[Table 5‑1: Organization of Findings 6](#_Toc52396471)

[Table 5‑2: Material Validation Findings 8](#_Toc52396472)

[Table 5‑3: General Feedback Findings 10](#_Toc52396473)

[Table 5‑4: Model Successes for Prescribing Practices 13](#_Toc52396474)

[Table 5‑5: Prescribing Practices Structural Findings 14](#_Toc52396475)

[Table 5‑6: Prescribing Practices Behavioral Findings 15](#_Toc52396476)

[Table 5‑7: Prescription Availability and Price Successes 17](#_Toc52396477)

[Table 5‑8: Prescription Availability and Price Structural Findings 17](#_Toc52396478)

[Table 5‑9: Model Successes for Medical Use and Misuse 22](#_Toc52396479)

[Table 5‑10: Replication of Grant Team Results 22](#_Toc52396480)

[Table 5‑11: Medical Use and Misuse Structural Findings 23](#_Toc52396481)

[Table 5‑12: Medical Use and Misuse Behavioral Findings 25](#_Toc52396482)

[Table 5‑13: Model Successes for Use Disorder 32](#_Toc52396483)

[Table 5‑14: Use Disorder Structural Findings 32](#_Toc52396484)

[Table 5‑15: Use Disorder Behavioral Findings 33](#_Toc52396485)

[Table 5‑16: Treatment Successes 36](#_Toc52396486)

[Table 5‑17: Treatment Structural Findings 37](#_Toc52396487)

[Table 5‑18: Treatment Behavioral Findings 37](#_Toc52396488)

[Table 5‑19: ADF Structural Findings 44](#_Toc52396489)

[Table 8‑1: SOW Checklist 51](#_Toc52396490)

List of Figures

[Figure 5‑1: Sensitivity Test of Initiation rate heroin with Rx misuse on Total overdose deaths synth heroin 19](#_Toc52396491)

[Figure 5‑2: Sensitivity Test of Initiation rate heroin with Rx misuse on Overdose death Rx OUD no H in Tx 20](#_Toc52396492)

[Figure 5‑3: Rx misuse no PY heroin 21](#_Toc52396493)

[Figure 5‑4: Sensitivity Test of Initiation rate heroin with Rx OUD relative to Rx misuse on Overdose death Rx OUD with H 30](#_Toc52396494)

[Figure 5‑5: Sensitivity Test of Initiation rate heroin with Rx OUD relative to Rx misuse on Overdose death Rx OUD no H in Tx 31](#_Toc52396495)

[Figure 5‑6: Sensitivity Test of Initiation rate heroin with Rx OUD relative to Rx misuse on Tx intake capacity utilization 31](#_Toc52396496)

[Figure 5‑7: Sensitivity of Tx seeking fraction Bup Rx OUD on overdose death Rx OUD no H in Tx 35](#_Toc52396497)

[Figure 8‑1: Initiating Rx misuse own Rx Goodness-of-Fit 47](#_Toc52396498)

[Figure 8‑2: Rx misuse no PY heroin Goodness-of-Fit 47](#_Toc52396499)

[Figure 8‑3: Total Rx misuse initiation Goodness-of-Fit 48](#_Toc52396500)

[Figure 8‑4: Rx OUD no PY heroin total Goodness-of-Fit 48](#_Toc52396501)

[Figure 8‑5: Total overdose deaths base heroin Goodness-of-Fit 48](#_Toc52396502)

[Figure 8‑6: Total overdose death base Rx Goodness-of-Fit 49](#_Toc52396503)

[Figure 8‑7: Initiating Rx misuse diverted Goodness-of Fit 49](#_Toc52396504)

[Figure 8‑8: HUD total Goodness-of Fit 49](#_Toc52396505)

[Figure 8‑9: Initiating heroin no Rx Goodness-of-Fit 50](#_Toc52396506)

[Figure 8‑10: Simulation Result for Decreasing Bup capacity startup delay 50](#_Toc52396507)

# Executive Summary

The D6b Final Validation Part 1 details our high-level findings and recommendations for the U.S. Food and Drug Administration’s (FDA) opioid systems dynamics model version “OSM\_Master\_0831” (hereinafter “model v831”) submitted by the Harvard Medical School Grant Team (hereinafter “Grant Team”) on September 1, 2020. The objectives of Part 1 of the Final Validation were to: 1) review and verify updates resulting from the Model Verification #2 feedback; 2) verify technical documentation is complete and accurate; 3) certify the model adheres to sound modeling principals; and 4) provide an unbiased evaluation of the model’s readiness to be used in FDA decision-making. For Part 1, we focused our activities around structural and behavioral validation. We will conduct activities related to policy validation and the model comparator during Part 2 of the Final Validation. The following is a list of key findings and recommendations from this report:

1. We found that, in general, model v831 is well-developed and there are no major behavioral or structural issues.
2. We found that for many model assumptions, more thorough descriptions and enhanced justifications are necessary for model transparency. We recommend providing additional supporting evidence regarding assumptions in the documentation to improve model confidence.
3. We performed sensitivity testing on a select group of variables that were identified as relevant to FDA’s mission. The tests identified that some output variables (e.g., **Total overdose deaths** and **Total overdose deaths synth Rx[[1]](#footnote-2)**) are highly sensitive to change in certain parameters. If the accuracy of certain parameter estimates is well- documented and supported, these variables can become “high leverage policy interventions.” These interventions can be considered opportunities for further policy analysis as they allow decision-makers to focus on elements that lead to significant change. We recommend estimates for these variables be well-documented to increase trust in the model (see [Section 5.5](#_Medical_Use_and_1) for more details).

# Introduction

Booz Allen and a system dynamics subject matter expert (SME) previously reviewed versions 406b, 511, and 622 of FDA’s model and documented findings and recommendations in the [Initial Validation Report](http://sharepoint.fda.gov/orgs/CDER-OPSA/PEIS/opioidspolicytool/Shared%20Documents/Opioids%20Policy%20Tool/D06ab%20-%20Initial%20Validation%20Report%20+%20Final%20Validation%20Report/D6a_Initial%20Validation_Report_202003270110.docx), [Model Verification #1](http://sharepoint.fda.gov/orgs/CDER-OPSA/PEIS/opioidspolicytool/Shared%20Documents/Opioids%20Policy%20Tool/D06ab%20-%20Initial%20Validation%20Report%20+%20Final%20Validation%20Report/D6b_ModelVerification1_202005290841.docx), and [Model Verification #2](http://sharepoint.fda.gov/orgs/CDER-OPSA/PEIS/opioidspolicytool/Shared%20Documents/Opioids%20Policy%20Tool/D06ab%20-%20Initial%20Validation%20Report%20+%20Final%20Validation%20Report/D6b_ModelVerification2_20207101030.docx) respectively. For Final Validation Part 1, a Validation Team comprised of Booz Allen modelers and analysts along with another system dynamics SME reviewed model v831. We evaluated the model to: 1) review and verify updates resulting from the Model Verification #2 feedback; 2) verify technical documentation is complete and accurate; 3) certify the model adheres to sound modeling principals; and 4) provide an unbiased evaluation of the model’s readiness to be used in FDA decision-making. We present the following sections in this document:

* [Constraints and Limitations](#_Constraints_and_Limitations)
* [Review of Harvard Medical School Grant Team Responses to Model Verification #2](#_Validation_by_Section)
* [Validation by Section](#_Constraints_and_Limitations)
* [Conclusion](#_Conclusion_1)

# Constraints and Limitations

System dynamics model validation can be limited by the nature of the modeling itself which allows for a range of approaches (Homer 2019). The absence of a standardized system dynamics modeling approach and prescriptive validation process can make it challenging to compare validation findings between different versions of a model. We attempted to address this by following our [Validation Plan](http://sharepoint.fda.gov/orgs/CDER-OPSA/PEIS/opioidspolicytool/Shared%20Documents/Opioids%20Policy%20Tool/D05%20-%20Validation%20Plan/D5_Validation%20Plan_Updated_202001170446.docx), which provided a systematic, reproducible, and adaptable model validation framework.

While there are many ways to validate a model, the Validation Plan focuses on the most relevant validation activities and tests as highlighted by the literature (Department of Defense 2018; Sterman 2000). Select tests and standards require judgment from SMEs. SME judgement is necessary where there is no standard test or metric described in the literature to assess appropriateness or reasonableness for a component of the model.

To make efficient use of resources while providing an impactful validation, Part 1 of the Final Validation is not exhaustive (i.e., we did not review every single variable in the model). However, the validation was thorough, and there are sufficient findings to substantiate a strong base model.

We outline the constraints and limitations of the report below:

* **Validation findings** - The Final Validation Report Part 1 is only relevant to model v831. If model development has progressed (e.g., addition of new components, changing the structure of the current model) since the Validation Team began validation, findings and recommendations may no longer be relevant.
* **Replication of historical data** - We selected a subset of variables most relevant to potential FDA policies to compare the historical data with the simulation results.
* **Review of data processing files** - This validation period focused on reviewing the reasonableness of parameter values in the model rather than reviewing supporting documentation of how constants in the model were calculated. The data processing files and related data documentation sent by the Grant Team will be assessed during the next model maintenance cycle.
* **Two-part validation process** -The primary focus of Final Validation Part 1 was structural and behavioral validation, whereas the focus of Part 2 will be policy validation and the model comparator. Therefore, certain model elements were not fully explored during Part 1, but will be further evaluated as part of the policy validation and/or the model comparator.

# Review of Harvard Grant Team Responses to Model Verification #2

This section summarizes the review of model v831 against Model Verification #2 feedback related to structural and behavioral validation activities. We will assess responses to feedback related to policy validation in Part 2 of the Final Validation. Table 4‑1 summarizes our review of Model Verification #2 findings.

Table ‑: Review of Response to Model Verification #2

|  |  |
| --- | --- |
| Verification #2 Finding based on v622 | Review of Finding in v831 |
| Assumptions regarding street Rx consumption that lead to unexpected behavior in the model | Limiting treatment (i.e., decreasing **Bup capacity effective fraction[[2]](#footnote-3)** by 50%) did not result in a doubling of heroin initiation as documented in the Model Verification #2. There was a slight increase in heroin initiation for both Rx misusers and Rx opioid use disorder (OUD) patients but, as mentioned in Model Verification #2, this is as expected  Reducing treatment capacity appears to have a minimal impact on Rx demand. This behavior is partially related to the formulation of **Rx supply relative** (ID 49, [Section 5.4](#_Prescription_Availability_and)), for which we have recommended a modification. We will be able to more effectively assess behavior of this sector once the equations for **Rx supply relative** are modified. |
| The exclusion of the positive effects from abuse-deterrent formulations (ADF) and of routes of administration (ROA) from the model | We found that parts of ADFs are still structurally under-developed in the model. Please see [Section 5.8](#_Effects_of_Abuse-Deterrent) for more details |
| The model did not reasonably follow historical values of: 1) buprenorphine (Bup) treatment from 2014-2018; and 2) prices of illicit prescriptions (Rx) on the streets | We will conduct this review during Part 2 of the Final Validation as it is part of the policy validation for medication-assisted treatment (MAT) and ADFs |
| We found strong assumptions in the baseline run of the model may lead to improbable baseline run estimates for the “business-as-usual” scenario | We will perform these checks as part of the policy validation in Part 2 of the Final Validation during the evaluation of reasonableness of the policy effect vs. baseline |

# Validation by Section

The Validation Team separated findings into different sections: materials validation ([Section 5.1](#_Materials_Validation)); general feedback ([Section 5.2](#_General_Feedback)); structural and behavioral validation by model component (i.e., Prescribing Practices ([Section 5.3](#_Prescribing_Practices)) Prescription Availability and Price ([Section 5.4](#_Prescription_Availability_and)), Medical Use and Misuse ([Section 5.5](#_Medical_Use_and_1)), Use Disorder ([Section 5.6](#_Use_Disorder)), Treatment ([Section 5.7](#_Treatment)), and Effects of ADFs ([Section 5.8](#_Effects_of_Abuse-Deterrent)).

For clarity, names of equations and parameters in the model are **bolded**.

The material, structural, and behavioral validation findings are organized in tables using the fields described in Table 5‑1.

Table ‑: Organization of Findings

|  |  |
| --- | --- |
| Column Header | Description |
| ID | Unique ID, used to refer to and relate findings, not necessarily in sequential order |
| Validation Type | Structural or Behavioral |
| Finding | Summary of the finding from the validation test |
| Test Conditions | Description of test steps or scenario |
| Discussion of Finding | Detailed explanation of the finding; includes relevant references that support the suggested recommendations |
| Recommended Actions | Suggested solutions to address the finding |

## Materials Validation

The Validation Team performed the materials validation to verify documentation and technical materials (e.g., model files) necessary for model validation were available, complete, and accurate. We used these materials to understand the scope of the model including assumptions, capabilities, and limitations of the model.

### Documentation

On September 1, 2020, the Validation Team received model v831 for validation. This package included:

* **Cover Letter -** This PDF provides a summary of recent model development, current model scope, and some limitations.
* **Vensim model file**
* **Model Overview.ppsx** - This PowerPoint provides a high-level description of the model and some assumptions.
* **Documentation Master File.xlsx –** This document is hereinafter referred to as “documentation.” It contains: information on model variables and assumptions; a list of model equations; and a hard copy of input time series and validation time series.
* **Five supplementary Excel files required for data transformation** - 1) Medical Use, Rx, and Mortality.xlsx; 2) Price Data.xlsx; 3) Supply and Naloxone.xlsx; 4) Treatment.xlsx; and 5) Use, Disorder, and Transitions.xlsx. We will assess these files during the next model maintenance cycle.
* **NSDUH for Documentation.sas** - SAS code required for transformation of NSDUH data. We will assess these files during the next model maintenance cycle.
* **Input time series and validation series** - The time series are included in .vdfx files delivered alongside the Vensim model.
* **Three supplementary files for validation -** 1) Parameter Sensitivity – a series of graphs that show the sensitivity of outcomes with respect to value changes of model parameters, 2) Policy Lever Tests – a series of graphs that show the behavior of the system (e.g., changes in the outputs) as a result of changes in policy levers; and 3) Goodness-of-fit results – a comparison of historical data versus model simulation results.
* **List of Policy levers -** This word document contains a list of policy levers and provides descriptions of potential use cases.
* **References -** This Word document contains a list of the supporting literature used in the model.

The [Final Validation Part 1 Findings](http://sharepoint.fda.gov/orgs/CDER-OPSA/PEIS/opioidspolicytool/Shared%20Documents/Opioids%20Policy%20Tool/D06ab%20-%20Initial%20Validation%20Report%20+%20Final%20Validation%20Report/D6b_Final%20Validation%20Part%201_Findings_202009231715.xlsx) punch list’s “Material Validation” tab provides a summary evaluation of the validation documents.

### Major Findings

We have two main findings from the materials validation. First, many assumptions lack the thorough description and justification necessary for model transparency. Second, the “Documentation Master File.xlsx” lack ranges for constants which could have significant impacts to model outputs. Please see [Section 5.5](#_Medical_Use_and_1) regarding sensitivity testing for more detail.

Assumptions are a necessary part of modeling. Modelers make assumptions to simplify and incorporate real-life processes in models and/or speed up model run time. Modelers also make assumptions to correctly incorporate data into the models. Most assumptions related to data are included in the documentation, but we did not find justifications for many key modeling assumptions; these are noted in the behavioral and structural findings.

Multiple model stakeholders including the Validation Team, modeling analysts, and policy analysts need to understand assumptions made about data and relationships in the model. Thorough documentation of assumptions is necessary for the Validation Team to determine their appropriateness according to the scope of the model during validation. Modeling and policy analysts need to understand and review key model assumptions to accurately conduct simulations and interpret the results. See Table 5‑2 for documentation-related findings.

Table ‑: Material Validation Findings

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ID | Validation Type | Test Condition | Finding | Discussion of Finding | Recommended Actions |
| 32 | Materials Validation | Model Overview.ppsx | Slide 9 of the Model Overview PowerPoint shows an arrow from Total Opioid RX to People with Opioid Rx. This arrow does not describe a physical causal relationship; rather, it describes how data were estimated. Everywhere else in the document, arrows depict causal relations that can be understood by the reader. In this case, “prescriptions” do not "cause" people | Especially since this is the first arrow to appear in the presentation, it is not intuitive, and it is not representative of the generally sound causal logic elsewhere in the model | Reserve the use of arrows in this presentation to represent intuitive causal relationships. Adjust the diagram accordingly |
| 33 | Materials Validation | Model Overview.ppsx | Slide 13 of the Model Overview states that people who "initiate RX misuse with their own prescription," estimated at 38%, have been excluded | As written, it sounds like a significant portion of opioid users are excluded from consideration, although it does not appear so in the model structure | Clarify if this sub-population is excluded and explain the consequences. If this sub-population is not excluded, re-write for more clarity |

## General Feedback

This section includes findings for the model as a whole rather than focusing on specific parameters. In summary, we find the physics of the model, overall unit consistency, and high-level structure are reasonable.

As part of our inspection, we checked the net population with net inflows and outflows to the system. Our calculation shows that the *population in the stocks + deaths and quits -inflow of people coming into the model = 15.20M* for the duration of the model*.* This test confirms that the physics of the system work well (e.g., no artifacts in the system erroneously add or subtract users).

We also inspected the dimensional consistency in the model. All core model equations pass the dimensional consistency test. Vensim's automated units checker flagged some units’ errors, but these are in peripheral equations such as those used for calibration or model parsimony (e.g., subscripting). The result of this test increases confidence in the model. However, we recommend addressing the minor unit errors identified because it could allow the unit checker to discover other potentially consequential unit errors.

The high-level structure of the model rests on a stock and flow diagram representing various categories (stocks) of opioid users, classified according to severity, heroin use, and treatment status. The flows into, out of, and between these categories represent transitions between the stocks. The transitions (flows) are appropriately modeled as base rates influenced by several feedback loops. The base rates are fractions per year of the relevant stocks of candidates for transition. These base fractions are increased or decreased by feedback influences from risk perception, social influence, and street prices (which are in turn influenced by opioid availability). The clarity and sensibility of this stock and flow and feedback structure are strengths of the model that enhance usability, foster ease of interpretation, and form a strong foundation for future model development.

A notable exception to these feedback influences is the set of flows modeling users seeking treatment. The various fractional rates for seeking treatment (**Tx seeking rate[[3]](#footnote-4)** of various populations) are not influenced by any endogenous feedback; rather they are held constant over the time horizon of the simulation. We find this curious, particularly since the flows for quitting from the various populations are influenced by feedback from risk perception and in some cases availability. It appears likely that some combination of risk perception, social influence, and availability would indeed influence the likelihood of users seeking treatment. We recommend the Grant Team consider incorporating feedback influences on the flows for seeking treatment.

It also appears likely that some combination of risk perception, social influence, and availability would indeed influence the likelihood of success in the various treatment programs, although such feedback is absent in the model. We suggest considering incorporating feedback influences on the flows for treatment success.

See Table 5‑3 for the remaining general feedback findings.

Table ‑: General Feedback Findings

| ID | Validation Type | Test Condition | Finding | Discussion of Finding | Recommended Actions |
| --- | --- | --- | --- | --- | --- |
| 37 | General Feedback | Boundary Adequacy | The current model includes a variable for **Policy Activation Time,** used as the start time for all policy tests | Does not allow for a policy that uses two or more levers commencing at different times | Consider some enhancement to the dashboard or user interface that will facilitate testing of multi-pronged policies with multiple start dates. This enhancement will provide a user-friendly basis for testing multiple policies with different starting times |
| 38 | General Feedback | Boundary Adequacy | Model does not appear to contemplate any "early" intervention policy lever that would increase quits of RxMisuse before advancing to RxOUD | Upstream interventions can sometimes have higher success because of less advanced disease and can be cost-effective as well | Note in commentary the absence of any consideration of such possible interventions and Identify what such an intervention would be |
| 43 | General Feedback | Structure Assessment | Model includes several variables or constants that are not used, e.g., **Table for Tx Success Fraction**, **Average opioid Rx length chronic**, **Average opioid Rx length acute** | Potentially confusing to future users | Remove unused variables or add documentation |
| 44 | General Feedback | Model boundary | Model does not include social influence feedback effects that influence treatment demand | If social influence impacts quitting, as in the model, we might expect influences on demand for treatment from various stock populations. Addressing this point has the potential to improve model accuracy and produce more realistic predictions | Investigate with appropriate experts |
| 45 | General Feedback | Boundary Adequacy | As with Findings in ID 44, model does not include risk perception or price feedback effects that influence treatment demand | See Discussion of Findings ID 44 | Investigate with appropriate experts |
| 46 | General Feedback | Boundary Adequacy | As with Findings in ID 44, model does not include feedback effects that influence success of treatment | See Discussion of Findings ID 44 | Investigate with appropriate experts |

## Prescribing Practices

The Prescribing Practices model component comprises the number of people who received opioid prescription, number of morphine milligram equivalents (MME) that have been dispensed, and the average MME per prescription.

In this section, the Validation Team examined prescription factors (e.g., prescription duration) and their impacts on user populations. We tested a decrease in the average duration of opioid prescriptions by 50% and saw an expected reduction in overdose deaths and total overdose deaths. This scenario can be informative for prescribers and policy-makers to understand how changes in prescription duration can impact overdose mortalities in medical users and those with an opioid prescription. Overall, we found the results of this test to be reasonable, for more information please see ID 1 in Table 5‑4.

We documented behavioral findings related to the unintuitive effects of **Sensitivity of Rx supply to MME per Rx**. Our inspection regarding the **Sensitivity of Rx supply to MME per Rx** shows that after setting the value of this variable at 0.5 (100% increase) from the beginning of simulation (2002), the **Total overdose deaths Rx** will not start increasing until after 2018. Also, by doubling **Sensitivity of Rx supply to MME per Rx** we see a decrease in **Initiating Rx misuse own Rx.** It is plausible that this decrease could represent Rx misusers selling their own pills, but it is not clear why they would sell their own pills if those actions resulted in a need to find an illicit supply source (e.g., indicated by the increased **Initiating Rx misuse diverted**). If these findings result from model assumptions, we recommend the Grant Team provide supporting documentation in this regard (ID 5, Table 5‑6).

Table ‑: Model Successes for Prescribing Practices

| ID | Validation Type | Test Condition | Finding | Discussion of Finding | Recommended Actions |
| --- | --- | --- | --- | --- | --- |
| 1 | Behavioral | **Avg duration per opioid Rx** = 0.02 (around 50% reduction) | Reducing **Avg duration per opioid Rx** by 50% caused a decrease in **Overdose death MU**,[[4]](#footnote-5) **Effective patients with opioid RX, Total overdose deaths Rx,** and **Total overdose deaths base Rx** | These results are as expected | NA |
| 6\* | Behavioral | **Effect of MOUD Tx on Rx consumption[[5]](#footnote-6)** = 1 (set it at its max) | There were no observable changes in **Rx demand for misuse**; **Rx price endogenous**; and **Rx street supply disruption** when **Effect of MOUD Tx on Rx consumption** was set to the max value of 1 | When we change **Effect of MOUD Tx on Rx consumption** before running the simulation, we will be potentially increasing the strength of a feedback loop, but we note that this parameter is used only in the equation for **Rx demand for misuse**, which is in turn normalized in **Rx demand for misuse relative**. Since, it is normalized with the initial value of **Rx demand for misuse**, so any change we make will be “cancelled” in the normalization. Therefore, these results make sense. | NA |

\* ID 6 was initially reported as a “model action” (i.e., finding associated with a recommended action) in the preliminary Final Validation Part 1 Findings punch list. Upon further review, we have revised our discussion of the finding and reclassified ID 6 as a “model success.”

### Structural Findings

See Table 5‑5 for detailed structural findings from the Prescribing Practices component.

Table ‑: Prescribing Practices Structural Findings

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ID | Validation Type | Test Condition | Finding | Discussion of Finding | Recommended Actions |
| 39 | Structural | Boundary Adequacy | Prescribing behavior of providers is exogenous in the current model | Appears like a reasonable choice for this version of the model and might be a useful domain for future enhancement. We can capture the effect of other factors on prescribing behavior if it is included in the model endogenously and as part of a feedback loop. Also, it provides opportunities for further policy testing and introducing related policy levers | Acknowledge limitation |

### Behavioral Findings

See Table 5‑6 for detailed behavioral findings from the Prescribing Practices component.

Table ‑: Prescribing Practices Behavioral Findings

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ID | Validation Type | Test Condition | Finding | Discussion of Finding | Recommended Actions |
| 5 | Behavioral | **Sensitivity of Rx supply to MME per Rx** = 0.5 (100% increase) | When we double the initial value of **Sensitivity of** **Rx supply to MME per Rx** to .5 we noted the following: 1. **Rx supply relative** and **Rx demand for misuse** increases after 2014  2. There is a delayed increase in **Total overdose deaths Rx** which appeared after 2018 3. Increased **Initiating Rx misuse diverted** 4. Decreased **Initiating Rx misuse own Rx** | We question the delayed increase in **Total overdose deaths Rx**. We changed the **sensitivity of Rx supply to MME per Rx** from 2002 but only saw its effect on overdose deaths after 2018  Regarding item 4, it is plausible that decreased **Initiating Rx misuse own Rx** could represent Rx misusers selling their own pills, but it is unclear why they would want to sell their own pills if they would need to find an illicit supply source (e.g., indicated by the increased **Initiating Rx misuse diverted**) | Check this variable and other downstream variables  If these findings are intentional, provide supporting documentation |

## Prescription Availability and Price

Prescribing Availability and Price component comprises opioid prescription demand, price, and availability for misuse in the market.

We evaluated **Sensitivity of Rx price to Rx demand** to see if Rx and illicit drug prices in the model have expected impacts on demand; the outcome was as anticipated. For more information on the test, please see Table 5‑7.

We find the modeling of Rx supply a bit concerning, particularly the equation for **Rx supply relative**. This equation attempts to model not the “true” supply of Rx opioids available for diversion, but instead a proxy for the supply of Rx opioids available for diversion. The proxy is used to model effects on price, on initiating or quitting Rx misuse and on developing OUD.

Whereas we can conceptualize supply as the product of people with prescriptions times the average MME of those prescriptions, the equation for **Rx supply relative** includes the idea that the effect of a change (e.g., a 10% increase) in the number of people with prescriptions may not be the same as the effect of a change (e.g., a 10% increase) in the average MME per prescription. This idea makes sense, but we question the way it is operationalized in the formulation.

The equation includes two exponents that govern the strength of these two effects. One exponent is **Sensitivity of Rx supply to MME per Rx** and the other is (1- **Sensitivity of Rx supply to MME per Rx**). The problem is that by constraining these two exponents to sum to 1, the overall behavior may become distorted. The constraint that the two exponents must sum to 1 yields an equation that generates “constant returns to scale.” To exemplify the concern, consider what happens if we double both the number of people with prescriptions and the average amount of MME per prescription. The unsophisticated expectation is that such a change would increase supply by four-fold. However, the equation used here would result in only a two-fold increase in **Rx supply relative**, regardless of the value of **Sensitivity of Rx supply to MME per Rx.** This is a mathematical consequence of the form of the equation, and we do not think it is reasonable.

We recommend the Grant Team either reformulate this equation to use two exponents that can vary independently, thus allowing for **Rx supply relative** to increase in a way that is closer to linear in the two factors, or to choose a different form of equation altogether (ID 49, Table 5‑8).

All behavioral tests evaluating prescription availability and price resulted in reasonable or expected results.

Table ‑: Prescription Availability and Price Successes

| ID | Validation Type | Test Condition | Finding | Discussion of Finding | Recommended Actions |
| --- | --- | --- | --- | --- | --- |
| 7 | Behavioral | **Sensitivity of Rx price to Rx demand** = 4 (original value = 2.224) | We saw an increase in **Initiating Rx misuse own Rx** and **Total overdose deaths**, but a decrease in **Total overdose deaths Rx** | It implies that **Total overdose deaths RX** decrease because fewer people seek early treatment. The stock of **Rx OUD no heroin in MOUD Tx** is lower, and thus the death outflow from this stock is lower. Therefore, the results are intuitive | NA |

### Structural Findings

Please see Table 5‑8 for detailed structural findings from the Prescription Availability and Price component.

Table ‑: Prescription Availability and Price Structural Findings

| ID | Validation Type | Test Condition | Finding | Discussion of Finding | Recommended Actions |
| --- | --- | --- | --- | --- | --- |
| 49 | Structural | Structure Assessment | **Rx Supply Relative** uses exponents α and (1-α) | This creates constant returns to scale function. Supply is likely approximated by **Average MME per opioid Rx** times **Effective patients with opioid Rx** (or relative supply approximated by the product of the relative values). A doubling of **Average MME per opioid Rx** and a doubling of **Effective patients with opioid Rx** would give an almost quadruple value of supply, but this equation would result in only a doubling of Rx supply relative | Consider using two separate exponents so an increase in one factor while the other is constant can yield a close-to-linear increase in **Rx Supply Relative** or choose an entirely different functional relationship |

### Behavioral Findings

There are no behavioral findings related to the Prescription Availability and Price component that require follow-up. Please see Table 5‑7.

## Medical Use and Misuse

The Medical Use component comprises people using Rx opioids for therapeutic purposes, whether in accordance with a doctor’s instructions or not (i.e., misuse) and excludes casual abusers of opioids or people with OUD.

The behavioral tests for the Medical Use section are intuitive and the results are as expected (ID 13 in Table 5‑9). We see that reducing **Overdose death rate base MU** decreases overdose deaths in medical users and those with opioid prescription which confirms that there is no anomalous behavior in this part of the model.

The structural findings for the Medical Use and Misuse section relate to recommendations for equations and questions about calculations for parameters. The equation for **NonOD death rate misuse[[6]](#footnote-7)** is:

NonOD death rate misuse = (NonOD death rate nonuser + 0.0145) / 2

There is no documentation or explanation of the number 0.0145, so it is not clear what this value represents or what the logic of this equation is. Moreover, standard modeling practice eschews using numerical values equations, because: a) any such numerical value must have some real-world meaning; and b) representing that number as an explicit constant to be used in the equation brings greater transparency and facilitates analysis by enabling visibility and changes using the Vensim toolset. We suggest adding a constant for this value, naming it appropriately to signal its meaning, and using it in the equation for **NonOD death rate misuse** instead of the number itself (ID 34, Table 5‑11).

Prescriptions per person are appropriately calculated as the quotient of opioid prescriptions divided by patients. The data streams used are Total prescription opioid Rx IQVIA for the numerator and Patients with opioid prescription IQVIA SH[[7]](#footnote-8) for the denominator. The comments for Total prescription opioid Rx IQVIA explain that the values exclude prescriptions for patients in hospitals “because interest is in people who have an opioid prescription that is taken at home and has potential for misuse or diversion,” a reasonable choice. To operationalize the idea of excluding prescriptions for patients in hospital, it is important to also exclude patients in hospitals from the denominator. We recommend reviewing that data source and confirm that patients in hospitals are excluded from the denominator (ID 35, Table 5‑11).

Most of the findings in the Misuse component relate to the sensitivity of model’s outputs (e.g., **Total overdose deaths, Total overdose deaths synth Rx, Total overdose Rx**)to any changes in specific constants or inputs (ID 21, 23, and 24 in Table 5‑12).

Sensitivity testing involves changing the values of model inputs/constants while performing multiple simulations, then examining the uncertainty in selected output variables. Sensitivity testing can be very helpful in understanding the behavioral boundaries of a model and testing the robustness of model-based policies.

We performed all sensitivity tests using a Random Uniform distribution for 1,000 simulation runs. Examples of our findings are presented in Figure 5‑1 and Figure 5‑2. These figures represent the sensitivity of some model outputs to changes in the value of **Initiation rate heroin with Rx misuse** as a constant or input. In these graphs, the grey area represents the 90% confidence interval (CI); therefore, wider CIs mean more sensitivity of outputs to any changes in the input. As captured in Figure 5‑1, the 90% CI for **Total overdose deaths synth heroin** widens greatly after 2016 in response to changes in the **Initiation rate heroin with Rx misuse**. Also, Figure 5‑2 shows that **Overdose death Rx OUD no H in Tx[[8]](#footnote-9)** for Bup, methadone maintenance treatment (MMT), and Vivitrol (Viv) are sensitive to the constant named **Initiation rate heroin with Rx misuse**.

The results of these sensitivity tests have two implications. First, they underscore the need for reliable estimates of the (range of plausible) values for the parameters to which behavior is highly sensitive. It is important that estimates for these variables are well-documented so future users can build trust in the model. Second, the sensitivity test results highlight thatif the accuracy of baseline values of constants is well- documented and supported, these "high leverage policy intervention" variables can be considered as an opportunity for further policy analysis.

Having highly sensitive outputs is not necessarily a cause for concern. However, it is important that confidence intervals and estimates for variables that significantly impact the model are documented so future users can build trust in the model. Therefore, we recommend emphasizing in the documentation that these inputs/constants could be potential "high leverage policy interventions" so that analysts are aware prior to scoping scenarios to test in the model.



Figure ‑: Sensitivity Test of Initiation rate heroin with Rx misuse on Total overdose deaths synth heroin



Figure ‑: Sensitivity Test of Initiation rate heroin with Rx misuse on Overdose death Rx OUD no H in Tx

In another test we set all inflows to stock of **Rx misuse no PY heroin[[9]](#footnote-10)** to zero to evaluate anomalous behavior in this part of the model. We observed that it takes more than 20 years to drain most of the people out of the **Rx misuse no PY heroin** stock when no new individuals are incoming into the system (see Figure 5‑3). When we set all the inflows at zero at time 2002, at time 2010 we still have two million people in the stock, which predicts that after eight years of cutting all the Rx misuse sources, we still have almost 25% of the initial stock of Rx misusers (ID 2 in Table 5‑12). We request the Grant Team consider whether this behavior is reasonable and provide more explanation.

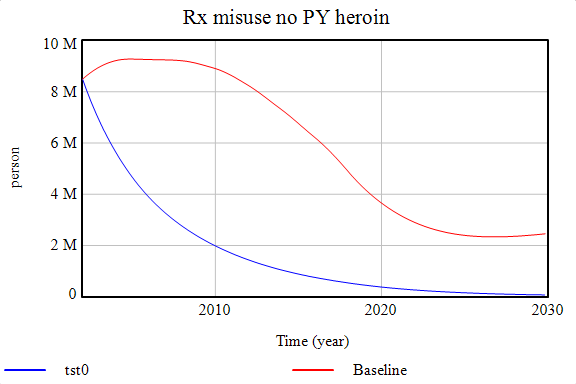


Figure ‑: Rx misuse no PY heroin

Additionally, we replicated the results that the Grant Team provided in the Goodness-of-fit document for several variables (see Table 5‑10). Our findings confirm the provided results in most cases. However, for **Initiating Rx misuse own Rx** and **Total Rx misuse initiation** root-mean-square error (RMSE), R^2, and mean absolute percentage error (MAPE), there was a more than 5% difference between our results and the values provided in the documentation. For **Initiating Rx misuse diverted** we also observed a more that 5% difference in the RMSE and MAPE between our findings and those reported by the Grant Team (ID 29 in Table 5‑12). For more detail on the replicated results see Table 5‑10; in this table, differences more than 5% are highlighted in red. Finally, we replicated a selection of the Goodness-of-fit graphs provided in the Grant Team’s documentation. In this document they compared the data versus the simulation results. All representative graphs are shown in the Appendix ([Section 8.1](#_Replication_of_Goodness-of-Fit)).

Table ‑: Model Successes for Medical Use and Misuse

| ID | Validation Type | Test Condition | Findings | Discussion of Finding | Recommended Actions |
| --- | --- | --- | --- | --- | --- |
| 13 | Behavioral | Test 1: **Overdose death rate base MU =** 0.000136 (-20%) Test 2: **Overdose death rate base MU** = 0 | Reducing **Overdose death rate base MU** by 20% reduced of **Overdose death MU, Total overdose deaths** and **Total overdose deaths Rx** as expected. There were no **Overdose deaths MU** when **Overdose death rate base MU** was set to 0 | These results are as expected | NA |

Table ‑: Replication of Grant Team Results

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | Validation Team | | | From the Goodness-of-fit Documentation | | | % Difference | | |
| Variable name | **Stock or Flow** | **RMSE** | **R^2** | **MAPE** | **RMSE** | **R^2** | **MAPE** | **RMSE** | **R^2** | **MAPE** |
| Initiating Rx misuse own Rx | Flow | 62516 | 0.03 | 0.16 | 57,449 | 0.02 | 0.15 | 8.11 | 33.33 | 6.25 |
| Rx misuse no PY heroin | Stock | 354034 | 0.94 | 0.04 | 362,208 | 0.92 | 0.04 | -2.31 | 2.13 | 0.00 |
| Total Rx misuse initiation | Flow | 137004 | 0.89 | 0.06 | 51,591 | 0.55 | 0.17 | 62.34 | 38.20 | -183.33 |
| Rx OUD no PY heroin total | Stock | 178479 | 0.43 | 0.06 | 180,824 | 0.43 | 0.06 | -1.31 | 0.00 | 0.00 |
| Total overdose deaths base heroin | Flow | 1782 | 0.86 | 0.44 | 1782 | 0.86 | 0.3 | 0.00 | 0.00 | 31.82 |
| Total overdose deaths base Rx | Flow | 809 | 0.85 | 0.06 | 809 | 0.85 | 0.06 | 0.00 | 0.00 | 0.00 |
| Initiating Rx misuse diverted | Flow | 88317 | 0.95 | 0.05 | 73,195 | 0.95 | 0.04 | 17.12 | 0.00 | 20.00 |
| HUD total[[10]](#footnote-11) | Stock | 204041 | 0.91 | 0.11 | 201,754 | 0.94 | 0.11 | 1.12 | -3.30 | 0.00 |
| Initiating heroin no Rx | Flow | 21118 | 0.01 | 0.31 | 21,565 | 0.01 | 0.3 | -2.12 | 0.00 | 3.23 |

### Structural Findings

See Table 5‑11 for detailed structural findings from the Medical and Misuse components.

Table ‑: Medical Use and Misuse Structural Findings

| ID | Validation Type | Test Condition | Findings | Discussion of Finding | Recommended Actions |
| --- | --- | --- | --- | --- | --- |
| 31 | Structural | Parameter Assessment | In the "Opioid Prescribing and Supply" view for the **Overdose death rate base MU**, a comment mentions chronic users were considered here. However, there is no mention of acute users and it is unclear whether they were included in this number | Both chronic users and acute users are considered medical users. Therefore, if we only consider chronic users in the **Overdose death rate base MU**, we might underestimate the **Overdose death MU** population | Check data sets and adjust as necessary or provide reason on why acute users were excluded |
| 34 | Structural | Structure Assessment | Equation for NonOD death rate misuse in the "User stocks & flows" view includes a numerical value (0.0145) | Recommended modeling practice avoids embedding numerical quantities in a model equation | Define a constant that specifies the meaning of the value 0.0145, document the source, and use the new constant in the equation |
| 35 | Structural | Parameter Assessment | Calculation of **Prescriptions per Person** excludes prescriptions for hospitalized patients in the numerator. Does the denominator (**Patient with opioid prescriptions IQVIA SH**) exclude hospitalized patients as well? | Numerator and denominator should be consistently defined | Check data sets to confirm consistency; adjust as necessary |
| 40 | Structural | Structure Assessment | **Net quitting heroin with Rx misuse** is diagrammed as an inflow to **Nondisordered heroin use** | The diagram is inconsistent with equation. Net quitting is an outflow, but the diagram represents it as an inflow. It can be confusing for viewers of the model if the diagram representation is different from the equation | Remove incorrect arrowhead |
| 41 | Structural | Structure Assessment | **Initiating heroin with Rx misuse** is diagrammed as an inflow to **Rx misuse no PY heroin** | Inconsistent with equation. Initiating heroin is an outflow from **Rx misuse no PY heroin**, but the diagram represents it as an inflow. It can be confusing for viewers of the model if the diagram representation is different than the equation | Adjust pipe and valve structure by adding another outflow pipe so that arrows are not overlapping |

### Behavioral Findings

See Table 5‑12 for detailed behavioral findings from the Medical Use and Misuse components.

Table ‑: Medical Use and Misuse Behavioral Findings

| ID | Validation Type | Test Condition | Findings | Discussion of Finding | Recommended Actions |
| --- | --- | --- | --- | --- | --- |
| 2 | Behavioral | Set all inflows to stock of **Rx misuse no PY heroin** to zero | It takes more than 20 years to drain most of the people out of the **Rx misuse no PY heroin** stock when we zero out all the inflows | We question this finding because when we set all the inflows at zero at time 2002, at time 2010 we still have 2 million people in the stock, which shows after 8 years of cutting all the Rx misuse sources we still have almost 25% of the initial stock of Rx misusers | Consider whether this is reasonable and explain |
| 21 | Behavioral | We performed sensitivity testing in Vensim for **Initiation rate heroin with Rx misuse:** Uniform [0.003, 0.009] | **Total overdose deaths, Total overdose deaths synth Rx, Total overdose deaths synth heroin, and Overdose death Rx OUD no H in Tx[Viv]** are very sensitive to this constant at 90% CI after 2015.  **Total overdose Rx, Total overdose heroin, Overdose death Rx OUD with H, and Tx intake capacity utilization[MMT]** are sensitive to this constant at 90% CI starting after 2002 to the end of simulation time (2030).  **Overdose death Rx OUD no H in Tx[MMT]** and **Overdose death Rx OUD no H in Tx[Bup]** are both sensitive to this constant especially after 2009. MMT is more sensitive than Bup. **Total overdose deaths base heroin and Tx intake capacity utilization[Viv]** were not sensitive at all. **Tx intake capacity utilization[Bup]** was sensitive only between 2009 to 2018 | Total overdose deaths from Rx and heroin and treatment capacity utilization for MMT are highly sensitive to **Initiation rate heroin with Rx misuse**. This implies that any policy effect that causes slight changes to **Initiation rate heroin with Rx misuse** would cause drastic changes to overdose deaths and have significant policy implications  Having outputs that are highly sensitive to changes in certain model inputs is not necessarily a cause for concern, but worth noting. It is important that confidence intervals and estimates for variables that significantly impact the model are documented so future users can build trust in the model. If the accuracy of these values is well documented/supported, these "high leverage policy intervention" variables are also an opportunity for further policy analysis | 1. Check the values and provide context on whether ACAL or LIT was chosen as the final source.  2. Identify actions to tighten parameter estimates. 3. Provide a range of plausible values for selected parameters and information on level of confidence in estimate.  4. Emphasize in the documentation that this input is a potential "high leverage policy intervention" so that analysts are aware prior to scoping scenarios to test in the model |
| 23 | Behavioral | We performed sensitivity testing in Vensim for **Initiation rate Rx misuse own Rx:** Uniform [0.002, 0.006] | **Total overdose heroin, Total overdose deaths synth heroin,** and **Tx intake capacity utilization[MMT]** are partially sensitive to this constant after 2018 at 90% CI.  However, None of the following variables are sensitive to this constant at 90% CI: **Total overdose deaths,  Total overdose deaths synth Rx,  Total overdose Rx,  Overdose death Rx OUD with H,  Overdose death Rx OUD no H in Tx Total overdose deaths base Rx  Total overdose deaths base heroin Tx intake capacity utilization[Bup] Tx intake capacity utilization[Viv]** | See ID 21 for more context on the results of sensitivity tests | See ID 21 |
| 24 | Behavioral | We performed sensitivity testing in Vensim for **Developing HUD rate no Rx OUD:** Uniform [0.27, 0.67] | **Overdose death Rx OUD with H** is very sensitive to this parameter at 90% CI especially after 2016. **Tx intake capacity utilization[MMT]** shows sensitivity from 2002 to around 2014 and also from 2018 to 2030. We see no sensitivity between 2014 to 2018  **Total overdose deaths synth heroin,** and **Total overdose heroin**, are partially sensitive at 90% CI after 2018.  None of the following variables showed any sensitivity to this constant: **Total overdose deaths,  Total overdose deaths synth Rx,  Total overdose Rx,  Overdose death Rx OUD no H in Tx Total overdose deaths base Rx  Total overdose deaths base heroin Tx intake capacity utilization** | See ID 21 for more context on the results of sensitivity tests | See ID 21 |
| 29 | Behavioral | We replicated the results provided in the Goodness-of-fit documentation and then calculated RMSE, R^2, and MAPE for the following variables: **Initiating Rx misuse own Rx Total Rx misuse initiation  Initiating Rx misuse diverted Total overdose deaths base heroin** | We replicated some of the results provided in the "Goodness-of-fit" documentation. The reported values for RMSE, R^2, and MAPE for **Initiating Rx misuse own Rx** and **Total Rx misuse initiation** had more that 5% difference between our results and the reported values in the documentation. We also, observed more that 5% difference in the RMSE and MAPE for **Initiating Rx misuse diverted** between our finding and the reported ones in the documentations. The same story is true for MAPE for **Total overdose deaths base heroin** | A larger difference between our calculated results versus the results provided in the "Goodness-of-fit" documentation suggests differences in replicability | Check the results of RMSE, R^2, and MAPE regarding the replication of historical data vs the simulation values for the following variables: **Initiating Rx misuse own Rx Total Rx misuse initiation  Initiating Rx misuse diverted Total overdose deaths base heroin** |

## Use Disorder

The Use Disorder component comprises people with opioid use disorder (OUD) and heroin use disorder (HUD) as well as casual abuse of opioids and heroin.

For the replication of the results that were provided by the Grant team in the Goodness-of-fit document, our calculation for the RMSE, R^2, and MAPE are the same as what is reported (see ID 31 in Table 5‑13). Findings of this test bring more confidence in the results of the model. For more detail see Table 5‑10 and Table 5‑13. The Appendix ([Section 8.1](#_Replication_of_Goodness-of-Fit)) includes the associated graphs which show the simulated results versus the real data.

The structural finding related to Use Disorder is minor and we believe it may be due to a transcription error. Please refer to ID 47 in Table 5‑14 for more information.

While conducting the sensitivity tests for the Use Disorder component we found that many of the output variables that we selected to track downstream (see ID 22 in Table 5‑15) are sensitive to changes in **Initiation rate heroin with Rx OUD relative to Rx misuse** at 90% CI.

For example, we can see in Figure 5‑4 and Figure 5‑5 that **Overdose death Rx OUD with H** and **Overdose death Rx OUD no H in Tx** for Bup, MMT, and Viv are sensitive to changes in **Initiation rate heroin with Rx OUD relative to Rx misuse** especially after 2016. We also observe that only **Tx intake capacity utilization[MMT]** is sensitive to the value of this input (see Figure 5‑6).

These sensitivity results highlight opportunities to explore potential policy interventions. However, it is also important to have accurate estimates for such constants so that future users can build trust in the model.Therefore, we recommend the Grant Team to: 1. Check the values and provide context on whether ACAL or LIT was chosen as the final source. 2. Identify actions to tighten parameter estimates. 3. Provide a range of plausible values for selected parameters and information on level of confidence in estimate. 4. Emphasize in the documentation that this input is a potential "high leverage policy intervention" so that analysts are aware prior to scoping scenarios to test in the model.



Figure ‑: Sensitivity Test of Initiation rate heroin with Rx OUD relative to Rx misuse on Overdose death Rx OUD with H



Figure ‑: Sensitivity Test of Initiation rate heroin with Rx OUD relative to Rx misuse on Overdose death Rx OUD no H in Tx



Figure ‑: Sensitivity Test of Initiation rate heroin with Rx OUD relative to Rx misuse on Tx intake capacity utilization

Table ‑: Model Successes for Use Disorder

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ID | Validation Type | Test Condition | Findings | Discussion of Finding | Recommended Actions |
| 30 | Behavioral | We replicated the results provided in the Goodness-of-fit documentation and then calculated RMSE, R^2, and MAPE for the following variables**: Rx misuse no PY heroin Rx OUD no PY heroin total Total overdose deaths base Rx HUD total Initiating heroin no Rx** | Our findings are the same as what is reported in the documentation | This increases confidence in the model | NA |

### Structural Findings

See Table 5‑14 for detailed structural findings from the Use Disorder component.

Table ‑: Use Disorder Structural Findings

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ID | Validation Type | Test Condition | Findings | Discussion of Finding | Recommended Actions |
| 47 | Structural | Structure Assessment | **Relapsing to HUD** = **HUD in remission** \* **Relapse rate to HUD** | Not based on Relapse rate to HUD net, which means the policy change parameter would not directly affect this flow. Is that correct? | Consider using Relapse rate to HUD net |

### Behavioral Findings

See Table 5‑15 for detailed behavioral findings from the Use Disorder component.

Table ‑: Use Disorder Behavioral Findings

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ID | Validation Type | Test Condition | Findings | Discussion of Finding | Recommended Actions |
| 22 | Behavioral/ Sensitivity test | We performed sensitivity testing in Vensim for **Initiation rate heroin with Rx OUD relative to Rx misuse:** Uniform [1.5, 5.5] | **Total overdose deaths, Total overdose deaths synth Rx, Total overdose deaths synth heroin, and Overdose death Rx OUD no H in Tx[Viv]** are very sensitive to this constant at 90% CI after 2015.  **Total overdose Rx, Total overdose heroin, Overdose death Rx OUD with H, and Tx intake capacity utilization[MMT]** are sensitive to this constant at 90% CI starting after 2002 to the end of simulation time (2030).  **Overdose death Rx OUD no H in Tx[MMT]** and **Overdose death Rx OUD no H in Tx[Bup]** are both sensitive to this constant especially after 2009. MMT is more sensitive than Bup. **Total overdose deaths base heroin and Tx intake capacity utilization[Viv]** were not sensitive at all. **Tx intake capacity utilization[Bup]** was sensitive only between 2009 to 2018. | See ID 21 for more context on the results of sensitivity tests. | See ID 21 |

## Treatment

The Treatment component encompasses people who are in any kind of systematic treatment for opioid-related substance use disorder, whether for OUD or HUD. Three types of opioid agonist treatment (i.e., OAT) are addressed in the model: Bup, MMT, and Viv.

The Validation Team completed tests in the Treatment component to understand if users engaged in treatment and treatment factors (e.g., capacity and duration) are reasonably represented in the model. Overall, when the Validation Team made changes to duration of treatment for each of the three possible MAT types, the model produced expected results. More information about the successful tests can be found in Table 5‑16.

The major structural findings in this section are related to the representation of fentanyl and Bup capacity in the model.

The variable **Switch for fentanyl shift**, part of the model controls, is used correctly as a switch in the equation for **Fentanyl penetration H supply**. However, its use in the equations for **Excess OD deaths synth H user estimated** and for **OD deaths synth H user baseline estimated** is questionable. In these latter two equations, **Switch for fentanyl shift** is used to specify the amount of time of (i.e., the value of) the fentanyl shift. While this shortcut has practically no mathematical consequence when the value is 1 or 0, it is stylistically weak, and it introduces dimensional inconstancy in the latter two equations (where a dimensionless constant is used to represent years). We suggest adding a new constant for **Length of fentanyl shift**, or another name of their choosing, and using it in the equations for **Excess OD deaths synth H user estimated** and for **OD deaths synth H user baseline estimated** (ID 36, Table 5‑17).

In examining equations, we note that **Bup capacity effective** is modeled as a material delay (using DELAY1). The phenomenon modeled with this delay is that it takes time, as described in the comments for **Bup capacity startup delay,** for “newly waivered providers to effectively 'activate' their capacity by building a practice with clients, administrative support, etc.” We expect that this ramping up of capacity would behave as an effective capacity value moving towards a theoretical capacity value as providers close the gap between what they are currently doing and what they are waivered to do. This process would exhibit a goal-seeking pattern of behavior, with the goal to bring effective capacity up to theoretical capacity.

We question the choice of delay type for this formulation. There are two types of delays, known in system dynamics as material delays and information delays. The current model uses a material delay for **Bup capacity effective.** A material delay (which is a stock) accumulates what flows in and generates an outflow as a fraction per unit time of what is in the stock. What goes in eventually comes out. With this formulation, **Bup capacity effective** is the outflow (not a stock). Conversely, an information delay is constructed as a stock for the current value (e.g., **Bup capacity effective**) with an inflow that closes the gap between the current value and the target value (**e.g., Bup total theoretical capacity**). We believe typical system dynamics modeling practice is to model an adjustment process for ramping up of capacity using an information delay.

The behaviors of material delays and information delays differ in one important way. Material delays are said to conserve the flows: what flows in from the inflows exits through an outflow or remains in the stock, so there is a “conservation” of the cumulative quantity of the inflow. An information delay does not conserve flows; rather, new information overwrites old information. To understand the difference, the standard test is to introduce a change in the value of the delay time. A decrease in the delay time for a material delay causes a (temporary) increase in the outflow (i. e., **Bup capacity effective),** because the “material” in the stock flows out faster. See Appendix [Section 8.2](#_Buprenorphine_Treatment_(Bup)) for the associated graph. We do not believe such behavior is consistent with expected behavior in the real-world system. There is no such burst of extra capacity when this test is performed with an information delay (SMOOTH).

We recommend the Grant Team either justify the choice of a material delay or change the formulation to an information delay, for example using the SMOOTH function. Table 5‑17 is a summary of the Treatment structural findings.

The tests related to user behavior in Treatment identified some potential areas that would benefit from further investigation from the Grant Team. For example, the team found that **Overdose death Rx OUD no H in Tx** for Bup, MMT, and Viv are sensitive to changes in **Tx seeking fraction Bup Rx OUD** at 90% CI especially for MMT (see Figure 5‑7). This sensitivity result highlights the opportunities for introducing appropriate policy interventions regarding seeking treatment in people with OUD to be able to reduce overdose deaths. However, it is important to have an accurate estimation for such constants so that the future users can build trust in the model. For more details on sensitivity tests see ID 25-28.



Figure ‑7: Sensitivity of Tx seeking fraction Bup Rx OUD on overdose death Rx OUD no H in Tx

Table ‑: Treatment Successes

| ID | Validation Type | Test Condition | Finding | Discussion of Finding | Recommendation |
| --- | --- | --- | --- | --- | --- |
| 14 | Behavioral | **Tx average duration MMT** = 1.092 (+20%) | When **Tx average duration MMT** increased by 20%, **Tx exit in remission total** and **Tx success fraction** also increased over baseline. | These results are intuitive and reasonable | The documentation shows several values from several different studies. Please provide a range for this constant in the documentation |
| 15 | Behavioral | **Tx average duration Viv** = 0.372 (+20%) | We increased **Tx average duration Viv** by 20% which reduced **Tx exit in remission total** and increased **Tx success fraction** compared to baseline. | These results are intuitive and reasonable | NA |
| 16 | Behavioral | **Tx average duration Bup** = 0.912 (+20%) | A 20% increase in the duration of treatment for Bup increased the **Tx exit in remission total** and **Tx success fraction** compared to baseline | These results are intuitive and reasonable | The documentation shows several values from several different studies. Provide a range for this constant in the documentation |
| 18 | Behavioral | **Bup capacity effective fraction** = .3 | When **Bup capacity effective fraction** decreases from .59 to .3, **Tx demand total by type** increases for all variables. | These results are as expected | NA |

### Structural Findings

See Table 5‑17 for detailed structural findings from the Treatment component.

Table ‑: Treatment Structural Findings

| ID | Validation Type | Test Condition | Finding | Discussion of Finding | Recommendation |
| --- | --- | --- | --- | --- | --- |
| 36 | Structural | Structure Assessment | The variable **Switch for fentanyl shift** appears as a switch in some equations and as a length of time in others | While a convenient shortcut, this double use of **Switch for fentanyl shift** introduces unit errors and may be misleading | Introduce a constant for time period of the fentanyl shift (currently = 1 year) |
| 42 | Structural | Change Bup capacity startup delay from 0.5 to 0.25 at time 2012 | **Bup capacity effective** exhibits a sharp peaked increase followed by a somewhat sharp decline and then remains permanently higher than in the baseline run | **Bup capacity effective** is modeled as a material delay (using DELAY1). This delay choice will conserve flows, which explains the observed behavior. Because this is modeling a ramp up towards providers' full capacities, an information delay is likely more appropriate | Change formulation to an information delay and test behavior |

### Behavioral Findings

See Table 5‑18 for detailed behavioral findings on the Treatment component.

Table ‑: Treatment Behavioral Findings

| ID | Validation Type | Test Condition | Finding | Discussion of Finding | Recommendation |
| --- | --- | --- | --- | --- | --- |
| 8 | Behavioral | Tx intake delay Bup is increased from .04 to 4 | See ID 9. (**Tx intake delay MMT** increased from .09 to 2) | See ID 9 | See ID 9 |
| 9 | Behavioral | **Tx intake delay MMT** is increased from .09 to 2 | When **Tx intake delay MMT** was increased from .09 to 2, we saw **Total in Tx by type[Bup]**, **Tx exit in remission total[Bup]**, **Tx exit with UD total[Bup]** all increase after 2015  **Total in Tx by type[MMT]**, **Tx exit in remission total[MMT]**, and **Tx exit with UD total[MMT]** all decrease by a little over half  However, no changes were seen in any of the three variables for the Viv subscript  When **Tx intake delay for Bup** was adjusted there were no impacts to **Total in Tx by type**, **Tx exit in remission total**, and **Tx exit with UD total for MMT** and Viv subscripts. Likewise, when **Tx intake delay for Viv** was adjusted, there were no impacts seen in the other subscripts | It is difficult to tell if there is a relationship between the methods of treatment (e.g., if decreases in accessing one treatment type should cause a change in accessibility for another treatment type). If it is believed that access to different MOUD Tx is related and dependent, then that should be reflected in all of the subscripts. There is no explicit stock for people seeking and awaiting entry to a treatment program, although there is explicit representation of the intake delay. There is a need for clarity about assumptions for possible switching to alternative treatments while waiting. Also, with no explicit backlog, intake delays are currently exogenous, and thus do not respond to the supply/demand balance. One would expect longer intake delays as the system reaches higher capacity utilizations  For example, the current model shows that changes in MMT availability impacts Bup availability and treatment enrollment. For consistency, changes in Bup or Viv availability/enrollment should also impact MMT availability/enrollment. If this is not the case, that rationale/ underlying assumptions should be documented | Please confirm that model behavior is reasonable. Review treatment structure and include information regarding assumptions for users engaging in different methods of treatment. (e.g., if it is a model assumption that access to one method of treatment should impact access to another) |
| 10 | Behavioral | **Tx intake delay Viv** is increased from .04 to 4 | See ID 9. (**Tx intake delay MMT** increased from .09 to 2) | See ID 9 | See ID 9 |
| 25 | Behavioral/ Sensitivity test | We performed sensitivity testing in Vensim for **Tx seeking fraction Bup HUD**: Uniform [0.3, 0.7] | **Overdose death Rx OUD no H in Tx[MMT]** and **Tx intake capacity utilization[MMT]** are sensitive to this constant at 90% CI. The original value for **Tx intake capacity utilization[MMT]** is closer to its upper bound. **Overdose death Rx OUD no H in Tx[Viv]** shows sensitivity after 2018. **Overdose death Rx OUD no H in Tx[MMT]** and **Tx intake capacity utilization[MMT]** show some sensitivity between 2005 to 2016 and do not show any sensitivity after 2018  None of the following variables showed any sensitivity to this constant: **Total overdose deaths,  Total overdose deaths synth Rx,  Total overdose deaths synth heroin,  Total overdose Rx,  Total overdose heroin,  Overdose death Rx OUD with H,  Total overdose deaths base Rx  Total overdose deaths base heroin** | See ID 21 for more context on the results of sensitivity tests. ID 21 is located in the Medical Use and Misuse section ([Section 5.5](#_Medical_Use_and_1)), but the discussion of finding is still relevant | See ID 21 |
| 26 | Behavioral/ Sensitivity test | We performed sensitivity testing in Vensim for **Tx seeking fraction MMT HUD relative**: Uniform [0.65, 0.95] | **Overdose death Rx OUD no H in Tx[MMT]** is sensitive to this constant during the whole simulation period. **Overdose death Rx OUD no H in Tx[Viv]** is only sensitive after 2016 but it is more sensitive to this constant compared to the MMT subscript. **Overdose death Rx OUD no H in Tx[Bup]** does not show any sensitivity  **Tx intake capacity utilization[MMT]** is sensitive to this constant during the whole simulation period. **Tx intake capacity utilization[Viv]** is only sensitive after 2016 but it is less sensitive to this constant compare to the MMT subscript. **Tx intake capacity utilization[Bup]** does not show any sensitivity  None of the following variables were sensitive to changes in this constant:  **Total overdose deaths**,  **Total overdose deaths synth Rx**,  **Total overdose deaths synth heroin**,  **Total overdose Rx**,  **Total overdose heroin**,  **Overdose death Rx OUD with H**,  **Total overdose deaths base Rx**  **Total overdose deaths base heroin** | **Overdose death Rx OUD no H in Tx** and **Tx intake capacity utilization for Viv** was sensitive to changes in **Tx seeking fraction MMT HUD relative** but **Tx intake capacity utilization[Bup]** was not sensitive  Also, we observed that OD deaths and overdoses from heroin are not sensitive to Tx seeking fraction MMT HUD relative.  See ID 21 for more context on the results of sensitivity tests | See ID 21 |
| 27 | Behavioral/ Sensitivity test | We performed sensitivity testing in Vensim for **Tx seeking fraction Bup Rx OUD:** Uniform [0.425, 0.825] | **Overdose death Rx OUD no H in Tx** for Bup, MMT, and Viv are very sensitive to this constant. However, the sensitivity of **Tx intake capacity utilization** for Bup, MMT, and Viv is very minor  We saw no sensitivity for the following variables: **Total overdose deaths,  Total overdose deaths synth Rx,  Total overdose deaths synth heroin,  Total overdose Rx,  Total overdose heroin,  Overdose death Rx OUD with H,  Total overdose deaths base Rx  Total overdose deaths base heroin** | **Tx intake capacity utilization[MMT]** shows sensitivity to changes in **Tx seeking fraction Bup Rx OUD** during the whole simulation time except for around 2014 to 2018  See ID 21 for more context on the results of sensitivity tests | See ID 21 |
| 28 | Behavioral/ Sensitivity test | We performed sensitivity testing in Vensim for **Tx seeking fraction MMT Rx OUD relative:** Uniform [0.1, 0.3] | **Overdose death Rx OUD no H in Tx[MMT]** was highly sensitive to this constant. **Tx intake capacity utilization[MMT]** was partially sensitive from 2002 to 2014, and then after 2023. No sensitivity were observed regarding the following variables: **Total overdose deaths,  Total overdose deaths synth Rx,  Total overdose deaths synth heroin,  Total overdose Rx,  Total overdose heroin,  Overdose death Rx OUD with H,  Total overdose deaths base Rx  Total overdose deaths base heroin** | **Tx intake capacity utilization[MMT]** shows sensitivity to changes in **Tx seeking fraction MMT Rx OUD relative** during the whole simulation time excluding ~2014 to 2023  See ID 21 for more context on the results of sensitivity tests | See ID 21 |

## Effects of Abuse-Deterrent Formulations (ADF)

The Validation Team focused on the structural assessment of ADFs for Part 1 of the Final Validation. We will complete a behavioral assessment as part of the policy validation in Part 2 of the Final Validation.

### Structural Findings

In Model Verifications #1 and #2, we suggested the adding structure to incorporate both helpful and unhelpful effects of ADFs (ID 48, Table 5‑19). This sector of the model still appears underdeveloped.

One might expect high percentages of ADFs to reduce availability of opioids for non-oral use and a concomitant increase in price, triggering various feedbacks in the model, and perhaps others. Currently, the variable for **ADF fraction of** **Rx street supply net** affects the rate of initiating heroin use, but it does not appear to have other effects. We would expect, for example, that a high fraction of ADF in the street supply might lead to an increase in prices of non-ADF Rx because those who are seeking an opioid suitable for non-oral use, perhaps the least price-sensitive population, would face lower availability than otherwise and thus drive up prices. The model does not distinguish between oral and non-oral use of Rx opioids, so it is difficult to incorporate the possible effects of ADFs. We suggest that the Grant Team, in lieu of modifying the model to distinguish oral and non-oral use, clearly document this limitation of the model, with a particular emphasis on how this limitation does or does not limit its use for policy investigation. For example, are there some policy interventions related to ADFs for which the model can be useful? Are there others where the model is incomplete and potentially misleading?

See Table 5‑19 for detailed structural findings from the ADF component.

Table ‑: ADF Structural Findings

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ID | Validation Type | Test Condition | Findings | Discussion of Finding | Recommended Actions |
| 48 | Structural | Boundary Adequacy | Model does not include any direct effect of ADF on street supply | One might expect high percentages of ADF to reduce availability of opioids for non-oral use and a concomitant increase in price, triggering various feedbacks in the model, and perhaps others. This sector seems underdeveloped | Consider enhancing the endogenous structure |

### Behavioral Findings

Part 2 of the Final Validation will include an assessment of ADFs as part of the policy validation. Since behavioral testing methods are similar to policy validation methods and an ADF scenario will be evaluated during the policy validation, it is logical to include ADF behavioral findings in Part 2 of the Final Validation.

# Conclusion

Below are the high-level summary of findings from Part 1 of the Final Validation ([Section 6.1](#_Summary_of_Findings)) and next steps ([Section 6.2](#_Next_Steps))

## Summary of Findings

Overall, the validation of model v831 identified no major behavioral or structural concerns. Our tests suggest the model is well-developed and the physics, overall unity, consistency, and high-level structure of the model are reasonable. However, the technical documentation requires additional detail to enhance future usability and enable continued maintenance.

The Validation Team will make a final determination of the model’s readiness for use in FDA decision-making following completion of our Final Validation efforts (Part 1 and Part 2). However, based on our finding to date, we remain optimistic that the model will be ready to proceed to the next phase of operationalization following completion of the Final Validation.

## Next Steps

Part 1 of the Final Validation focused on the general behavior and structural assessment of the v831 of the model. Part 2 of the Final Validation will address the following, non-exhaustive, list of objectives.

* **Policy Validation** - The Validation Team will test a pre-approved list of scenarios and assess the reasonableness of the outputs reported by the model.
* **Model Comparator** - The Validation Team will compare the latest version of the model with an appropriate published model (e.g. Pitt 2018 model).
* **Usability Assessment** - The Validation Team will assess the usability of the model and propose applicable levels of sharing based on the model’s maturity.

For Part 2 of the Final Validation to effectively build upon the work of Part 1, we recommend that model development between Parts 1 and 2 of the Final Validation should be limited to addressing Part 1 findings.

# References

Department of Defense. “DoD Standard Practice Documentation of Verification, Validation, and Accreditation (VV&A) for Models and Simulations,” 2018.

Homer, Jack. “Best Practices in System Dynamics Modeling, Revisited: a Practitioners View.” *System Dynamics Review* 35, no. 2 (2019): 177–81. <https://doi.org/10.1002/sdr.1630>.

Pitt, Allison L., Keith Humphreys, and Margaret L. Brandeau. “Modeling Health Benefits and Harms of Public Policy Responses to the US Opioid Epidemic.” American Journal of Public Health 108, no. 10 (2018): 1394–1400. <https://doi.org/10.2105/ajph.2018.304590>.

Sterman, John D. Business Dynamics: Systems Thinking and Modeling for a Complex World. Boston: Irwin, 2000.

# Appendix

The Appendix includes the Validation Team’s replications of Goodness-of-fit ([Section 8.1](#_Replication_of_Goodness-of-Fit)), the simulation result for decreasing **Bup capacity startup delay** ([Section 8.2](#_Bup_Capacity_Startup)), and a crosswalk of the statement of work (SOW) with Part 1 of the Final Validation Report ([Section 8.3](#_SOW_Checklist)).

## Replication of Goodness-of-Fit Results

See Figure 8‑1 through Figure 8‑9 for the Validation Team’s replications of Goodness-of-fit.

Figure ‑: Initiating Rx misuse own Rx Goodness-of-Fit

Figure ‑: Rx misuse no PY heroin Goodness-of-Fit

Figure ‑: Total Rx misuse initiation Goodness-of-Fit

Figure ‑: Rx OUD no PY heroin total Goodness-of-Fit

Figure ‑: Total overdose deaths base heroin Goodness-of-Fit

Figure ‑: Total overdose death base Rx Goodness-of-Fit

Figure ‑: Initiating Rx misuse diverted Goodness-of Fit

Figure ‑: HUD total Goodness-of Fit

Figure ‑: Initiating heroin no Rx Goodness-of-Fit

## Buprenorphine Treatment (Bup) Capacity Startup Delay

In Figure 8‑10, the lines are associated with the results of the following simulations:

* *ChangeStartupDelay* - decrease **Bup capacity startup delay** by 50% in 2012.
* *ChangeStartupDelaySmooth* - decrease **Bup capacity startup delay** by 50% in 2012 and modify **Bup capacity effective** from a DELAY1 (i.e., material delay) to a SMOOTH (i.e., information delay).
* *Base* - behavior with baseline conditions, for comparison.

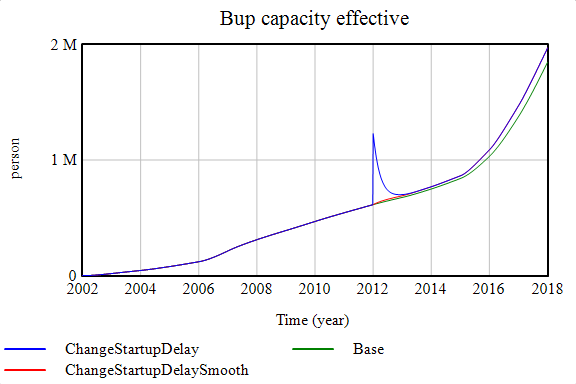


Figure ‑: Simulation Result for Decreasing Bup capacity startup delay

## Statement of Work (SOW) Checklist

Table 8‑1 crosswalks the SOW with Part 1 of the Final Validation Report.

Table ‑: SOW Checklist

| SOW | Location |
| --- | --- |
| i. Ensuring the structure of the model aligns and is consistent with documented assumptions and dynamic system modeling principles | [Section 5: Validation by Section](#_Validation_by_Section_1) |
| ii. Ensuring that all documented assumptions have been correctly incorporated into the model | [Section 5: Validation by Section](#_Validation_by_Section_1) |
| iii. Validating and assessing the appropriateness of the methodology by which the model data was obtained, generated and/or manipulated | [Section 5: Validation by Section](#_Validation_by_Section_1) |
| iv. Ensuring alignment between the stated definitions of the variable used in the model and the definitions of the variables in the source documentation | [Section 5: Validation by Section](#_Validation_by_Section_1) |
| v. Conducting an extensive review of all technical documentation provided by FDA | [Section 5.1: Materials Validation](#_Materials_Validation) |
| vi. Replicating at least two policy scenario analyses conducted by FDA and assessing any discrepancies | This content will be included in Part 2 of the Final Validation |
| vii. Comparing at least 2 policy scenario analyses conducted by FDA’s model with similar analyses in least one comparable published opioids policy analysis model. FDA will work with the contractor to confirm the comparator model and analyses. | This content will be included in Part 2 of the Final Validation |

1. Note: “synth” stands for “synthetic”; “Rx” stands for “prescription.” [↑](#footnote-ref-2)
2. Note: “Bup” stands for “buprenorphine treatment.” [↑](#footnote-ref-3)
3. Note “Tx” stands for “treatment.” [↑](#footnote-ref-4)
4. Note: “MU” stands for “medical user.” [↑](#footnote-ref-5)
5. Note: “MOUD” stands for “medication for opioid use disorder.” [↑](#footnote-ref-6)
6. Note: “OD” stands for “overdose.” [↑](#footnote-ref-7)
7. Note: “SH” stands for “Symphony Health.” [↑](#footnote-ref-8)
8. Note: “H” stands for “heroin.” [↑](#footnote-ref-9)
9. Note: “PY” stands for “past year.” [↑](#footnote-ref-10)
10. Note: “HUD” stands for “heroin use disorder.” [↑](#footnote-ref-11)