Development and Application of a Data-Driven Signal Detection Method for Surveillance of Adverse Event Variability Across Manufacturing Lots of Biologics

Development and Application of a Data-Driven Signal Detection Method for Surveillance of Adverse Event Variability Across Manufacturing Lots of Biologics

Running Head: A Signal Detection Method for Adverse Event Variability Across Manufacturing Lots of Biologics

Joshua T. Wilde

Massachusetts Institute of Technology, Operations Research Center, Cambridge, MA, United States

josh.t.wilde@gmail.com

ORCID: 0000-0002-3251-0211

Stacy Springs

Massachusetts Institute of Technology, Center for Biomedical Innovation, Cambridge, MA, United States

ssprings@mit.edu

ORCID: 0000-0003-2133-5689

Jacqueline M. Wolfrum

Massachusetts Institute of Technology, Center for Biomedical Innovation, Cambridge, MA, United States

jwolfrum@mit.edu

ORCID: 0000-0002-9125-828X

Retsef Levi

Massachusetts Institute of Technology, Sloan School of Management, Cambridge, MA, United States

retsef@mit.edu (Corresponding Author)

Building E62, 100 Main Street Cambridge, MA, 02142

ORCID: 0000-0002-1994-4875

Abstract

Introduction and Objective: Post-marketing drug safety surveillance research has focused on the product-patient interaction as the primary source of variability in clinical outcomes. However, the inherent complexity of pharmaceutical manufacturing and distribution, especially of biologic drugs, also underscores the importance of risks related to variability in manufacturing and supply chain conditions that could potentially impact clinical outcomes. We propose a data-driven signal detection method called HMMScan to monitor for manufacturing lot-dependent changes in adverse event (AE) rates and herein apply it to a biologic drug.

Methods: The HMMScan method chooses the best-fitting candidate from a family of probabilistic Hidden Markov models to detect temporal correlations in per lot AE rates that could signal clinically relevant variability in manufacturing and supply chain conditions. Additionally, HMMScan indicates the particular lots most likely to be related to risky states of the manufacturing or supply chain condition. The HMMScan method was validated on extensive simulated data and applied to three actual lot sequences of a major biologic drug by combining lot metadata from the manufacturer with AE reports from the FDA Adverse Event Reporting System (FAERS).

Results: Extensive method validation on simulated data indicated that HMMScan is able to correctly detect the presence or absence of variable manufacturing and supply chain conditions for contiguous sequences of 100 lots or more when changes in these conditions have a meaningful impact on AE rates. Applying the HMMScan method to FAERS data, two of the three actual lot sequences examined exhibited evidence of potential manufacturing or supply chain related variability.

Conclusions: HMMScan could be utilized by both manufacturers and regulators to automate lot variability monitoring and inform targeted root cause analysis. Broad application of HMMScan would rely on a well-developed data input pipeline. The proposed method is implemented in an open-source GitHub repository.

Key Points

- The pharmacovigilance community lacks methods for systematically detecting variation in adverse event (AE) rates due to manufacturing and supply chain factors.
- Our HMMScan method uses probabilistic Hidden Markov models to screen sequences of per manufacturing lot AE rates, separating patient and manufacturing-related variability and proposing lot ranges for targeted root cause analysis.
- The method is extensively validated and applied to three actual lot sequences of a major biologic, detecting regions of potential manufacturing or supply chain variability in two of these sequences.

1 Introduction

Methods for detecting post-marketing safety signals have long been the subject of active pharmacovigilance academic research, as well as regulatory and industrial work. These efforts have primarily focused on uncovering novel drug-adverse event combinations [1–4], and specifically on the product-patient interaction as the primary source of variability in clinical outcomes.

However, there are also known examples of serious adverse events (AEs), including fatalities of patients, caused by pharmaceutical products with root causes linked to manufacturing and supply chain sources [5]. The manufacturing process and related control mechanisms of drugs are specified in detail during the regulatory assessment and approval, and are mandated during the post-approval phase. Nevertheless, in 2019 the U.S. Food and Drug Administration (FDA) stated that monitoring the impact of manufacturing and supply chain variability on patients remains an open challenge for the pharmacovigilance community [6].

The risk of temporal variability in manufacturing and supply chain conditions, including raw materials sourcing, with potential impact on clinical outcomes is particularly relevant to biologic drugs because of the inherent complexity of the respective processes. Biologics are structurally much more complex than small molecule drugs. In addition to the increased complexity of a biological manufacturing process compared to chemical synthesis, the number of critical quality attributes is higher and their type is also more complex. Previously designed statistical approaches take as input a time series of monthly AE reports and aim to identify points in time where either temporary or systematic changes in

the rate of AE reports occur [7,8]. However, these approaches operate on aggregated monthly AE data and are not designed to specifically identify sequences of lots with unusually high AE rates. Additionally, in practice multiple lots may be used in parallel to treat patients, and the overall AE rates capture the aggregated number of AEs across all lots that are on the market.

Mahaux *et al*. [9] apply a hierarchical statistical scanning method to simulated batch genealogy data, which links inputs and outputs of material through a sequence of manufacturing process steps, to identify steps that are associated with excess adverse events. The method relies on data that capture relationships between final product batches that share bulk intermediate product batches. Whereas their method could be used by manufacturers, particularly for detailed root cause analysis, it would likely be impractical for use by a regulator that does not often have access to such granular data consistently across multiple different products.

This paper aims to augment existing post-marketing surveillance frameworks, specifically by addressing this challenge. The paper describes a new data-driven signal detection method, called HMMScan, inspired by the well-known family of Hidden Markov models (HMMs) [10]. Relying on standard reported clinical outcomes and manufacturing attributes, it is designed to monitor for manufacturing and supply chain lot-dependent changes. Specifically, the newly proposed method relies on the rate of reported AEs per final product lot to flag potential safety signals that could be related to variability in manufacturing and supply chain conditions.

Section 2 details the HMMScan method, as well as the design of the validation experiments on simulated data, as well as a use case based on industry data. The results of these experiments and an application to FDA Adverse Event Reporting System (FAERS) data are presented in Section 3. Section 4 provides a discussion of the aspects related to the application of the method and its limitations, as well as outlines directions for future data gathering and analysis, and Section 5 concludes.

2 Methods

2.1 High Level Approach

The goal of the HMMScan method is to provide an alert when the pattern of per lot AE rates in a time-ordered series of lots suggests that there might exist serial correlation in consecutive lots. The HMMScan method is applied to a single product at a time and takes as input a sequence of final product lots with their respective reported AE rates.

The HMMScan method aims to detect signals that suggest the potential presence of clinically meaningful variability in manufacturing and supply chain-related processes. Under the hypothesis that no such variability exists, differences in per lot AE rates are expected to be driven solely by random variation in aggregate patient covariates, such as age, race, the average level of polypharmacy, and other societal factors with no consistent dependence of the sequence of the lots. Thus, the per lot AE rate should be statistically independent across lots. In contrast, if temporal variability of the manufacturing and supply chain conditions impacts patient outcomes, then it is expected that the per lot AE rates will exhibit

temporal, or serial, correlation between lots manufactured at similar times. These differences in manufacturing and supply chain between lots encompass both acceptable variation in measured quality attributes within the approved operating limits, as well as variation in unmeasured (or undetected) quality attributes that could impact patient outcomes and safety.

The newly proposed method relies on a probabilistic modeling framework that can detect serial correlation in a series of lots ordered by packaging date, which is used as a proxy for the manufacturing timing of the respective lots. However, more generally, the specific order of lots could be further refined using information about the source of the intermediate materials for each lot.

One of the main challenges in identifying the manufacturing or supply chain related impact on the per lot AE rates is the fact that conditions of the manufacturing or supply chain systems may not be fully observable. This motivates the use of candidate models that fall into the broad category of HMMs, each consisting of two major elements. The first element is the number of underlying hidden (unobserved) states, and the respective transition probabilities from each state to all other states. The model assumes that each lot is manufactured and handled under a hidden state that corresponds to a certain state of the underlying manufacturing or supply chain conditions. The second element is a state-dependent binomial mixture distribution that captures the probabilistic pattern of the AEs per lot manufactured under the respective state. The dynamic transition between hidden states in the HMMs captures the potential variability in the underlying manufacturing or supply chain conditions. The corresponding impact of these states on the number of reported AEs per lot is captured through

the respective state-dependent binomial mixture distribution. The 'best' model is selected using Bayesian Information Criterion (BIC) [11], which weighs the explanatory power of the model with respect to the observed data against the complexity of the model (number of parameters). If the best model has only one state, this corresponds to the scenario with no clinically meaningful variability in the underlying manufacturing and supply chain conditions. Alternatively, if the best model exhibit multiple states, then there is a signal for potential impact of changing conditions on patient outcomes.

2.2 Detailed Method Description

2.2.1 Method Inputs

The input data to the HMMScan method are the observed AE counts for each lot in the sequence of L lots. In particular, denote the per lot AE counts as the random vector $\boldsymbol{A} = (A_1, A_2, \dots, A_L)$. The number of doses per lot, which without loss of generality is assumed constant, for each lot in the sequence, is denoted by D. For simplicity of exposition, it is assumed that each dose generates either zero or one AE. The HMMScan method takes as input a specific realization of the random vector $\boldsymbol{A}$ denoted by $\boldsymbol{a} = (a_1, a_2, \dots, a_L)$. This vector is then normalized by D to form observed AE rates, i.e., the number of AEs per D doses. The method can be extended to accommodate variation in the number of doses per lot by taking as input in addition to the vector $\boldsymbol{a}$ a vector of the lot sizes also of length L. In this case, lots are given weights proportional to their number of doses during the parameter estimation process to reflect the fact that larger lots have a lower variance in the estimated AE rate.

2.2.2 Hidden Markov Models

The modeling assumption is that the distribution of the random vector $\boldsymbol{A}$ (and the corresponding observed vector $\boldsymbol{a}$) is governed by a Hidden Markov Model (HMM) [10]. The respective HMM model captures both the potential heterogeneity of the patient population receiving the drug, as well as the different states (conditions) of the respective manufacturing and supply chain processes. In particular, let $\mathcal{C} = \{1, 2, \dots, C\}$ be the set of patient subpopulations that are exposed to a given sequence of drug lots. Additionally, let $\mathcal{S} = \{1, 2, \dots, S\}$ be the set of possible states of the underlying manufacturing and supply chain conditions. For each state $s \in \mathcal{S}$ and subpopulation $c \in \mathcal{C}$, let p_{sc} be the average probability per dose of incurring an AE. Let w_{sc} be the likelihood that a lot in state s is used within a subpopulation $c \in \mathcal{C}$. The state of each lot $\ell \in \{1, 2, \dots, \mathrm{L}\}$, which is unobserved (or *hidden*), is a random variable denoted by H_ℓ. The number of observed AEs for lot ℓ given that $H_\ell = s$ is captured through a state-dependent mixture of binomials (MB) distribution. That is, for each integer $a \in \{1, 2, \dots, D\}$:

$$P(A_\ell = a | H_\ell = s) = \sum_{c=1}^{C} \left(w_{sc} \cdot \mathrm{Binomial}(a; D, p_{sc}) \right)$$

The expression $\mathrm{Binomial}(x; n, p)$ denotes the probability mass for a binomial distribution with n trials and p probability of success evaluated at x. For the remainder of the paper, the state with the lowest (highest) mean AE rate will be referred to as the "low-risk" ("high-risk") state. Additionally, the sequence of states $\{H_\ell\}_{\ell \in \{1,2,\dots,L\}}$ evolves according to a Markov transition matrix that captures the probability of moving from each state to any other state [10]. Specifically, the transition matrix captures, for each pair of states $s, s' \in \mathcal{S}$, the probability $P(H_{\ell+1} = s | H_\ell = s')$. Finally, note that the transition matrix induces a stationary

149 distribution over the states that represents the long-run frequency of each state if
150 the hidden Markov process were run on an infinitely long lot sequence.

151 ### *2.2.3 Model Selection*

152 This section describes how the newly proposed HMMScan method selects the
153 HMM model structure with the best fit to the observed sequence of per lot AE
154 rates from a set of candidate model structures. The HMMScan model selection
155 procedure in Fig. 1 takes as input the observed sequence of AE rates, $\boldsymbol{a}$, and a set
156 of HMM candidate models. The candidate models are obtained by varying the
157 assumed number of states and subpopulations (i.e., the size of S and C) over a grid
158 of potential values from 1 to S_{max} and C_{max}, respectively.
159
160 Generally, the range of plausible HMM models in the typical use-cases considered
161 in this paper can be covered by using relatively small values for S_{max} and C_{max}
162 (i.e., less than 10). The reason is that, typically, the number of clinically
163 meaningful relevant subpopulations is relatively small, and the manufacturing
164 conditions can typically be aggregated into high-level states that capture the
165 respective risk level for quality variation. Additionally, complex HMM structures
166 with many hidden states and mixture components tend to overfit. Each candidate
167 model corresponds to a hypothesis regarding the number of hidden states and
168 patient subpopulations that best describes the observed AE rate sequence.
169
170 The HMMScan model selection procedure applies two sequential steps. The first
171 step involves *Parameter Estimation* to calibrate the parameters of each candidate
172 model. In the second step, *BIC Model Fit Evaluation* is used to determine which
173 candidate models provide the best fit to the sequence of observed per lot AE rates.

During the Parameter Estimation step, maximum likelihood estimates for the respective HMM parameters are obtained via the Expectation Maximization (EM) algorithm [12]. The EM algorithm takes as input an HMM with initial parameter values and returns locally optimal parameter estimates. The HMMScan method searches for globally optimal parameter estimates by running EM with multiple random initializations. For HMMs with $S = 1$, the EM algorithm uses closed form equations to iteratively optimize the binomial mixture weights and probabilities until convergence [10]. For HMMs with $S > 1$, the Baum-Welch algorithm, a variation of EM, optimizes both the transition probabilities and the state-specific distribution parameters [13] The HMMScan implementation referenced in this paper relies on the implementations of EM and Baum-Welch in the pomegranate Python package [14]. Further details regarding parameter initialization can be found in Section S1 of Online Resource 1.

The second step of the HMMScan model selection procedure, BIC Model Fit Evaluation, compares the fitted candidate models using Bayesian Information Criteria (BIC) and selects the model with the minimum BIC value. The BIC captures a tradeoff between the explanatory power of the model with respect to the data, and the complexity of the model in terms of the number of parameters [11].

The form of the BIC is motivated by the notion that finding the best-fitting HMM structure for the observed data is equivalent to maximizing the likelihood of the data given the HMM hyperparameters S and C, $P(\boldsymbol{a}|S, C)$, over all possible combinations of S and C. The BIC approximates this likelihood, which is not

observable, using the maximum likelihood parameter estimates. The approximation consists of two terms, a negative term that depends on the likelihood of the model evaluated at the maximum likelihood parameter values, and a positive complexity term that penalizes the number of estimated parameters (i.e., hidden states and mixture components) in the model. Therefore, a lower BIC value indicates increased plausibility of the model after accounting for model complexity.

Pairwise differences in BIC values can also be translated into a more interpretable metric, the relative odds that one model fits the observed data better than the other. In [11], Raftery calculates that, for models fit on long input data sequences, a BIC difference of 10 or more indicates a greater than 99% probability that the model with the lower BIC value provides a stronger fit to the observed data.

2.2.4 Method Output

The HMMScan method outputs the best-fitting model according to the BIC, and this model can be used to detect whether there is statistical evidence in favor of serial correlation in the AE rates in the input lot sequence. If an HMM with $S > 1$ provides the best fit to the observed AE rates according to the BIC, then the HMMScan method signals that there is evidence in favor of serial correlation in AE rates for the input lot sequence. This is considered as a positive HMMScan signal for a serial correlation. On the other hand, if $S = 1$ provides the best fit, this is considered a negative HMMScan signal, i.e., no evidence of serial correlation.

In addition to indicating the potential presence of clinically relevant variation in manufacturing and supply chain conditions, the best-fitting HMM is used to

identify the most likely sequence of hidden states associated with the input lot sequence. The mostly likely state sequence is calculated using the well-known and efficient Viterbi algorithm [10], which returns the path of hidden states that maximizes the joint likelihood of the hidden state sequence and the observed AE rates given the estimated maximum likelihood parameter values. These predicted hidden states can provide important temporal information as to what lots have been produced under high-risk states, and this could be used to inform subsequent root cause analysis, as discussed in Section 4.

2.3 Validation on Simulated Data

This section describes a validation and performance assessment of the HMMScan method through simulated data that capture different conditions and data input attributes. The selected conditions for the accuracy assessment are motivated by practical scenarios for true dynamics of manufacturing and supply chain conditions. The specific instances for each respective scenario are captured through corresponding ground truth HMM models used to generate the simulated data. Specifically, the scenarios vary in the number of hidden states, the degree of similarity of the state-dependent mixtures of binomial distributions, and the structure of the underlying transition matrix of the hidden states.

HMMScan is evaluated for its ability to detect the correct model structure for sample sequences of varying length generated by each ground truth model. For each sample sequence, the Model Selection step is applied according to the description in Section 2.2.3 above. Specifically, the method fits a collection of candidate HMMs, each corresponding to a hypothesis about the structure of the ground truth HMM. This collection contains single-state models with up to six

mixture components, two-state models with up to three mixture components, and three- and four-state models with up to two mixture components. Models with additional states and components did not provide the best BIC for any of the simulated sample sequences.

The performance of the HMMScan method is evaluated according to two metrics. The first metric is *detection accuracy*, which compares the structure of the lowest BIC model to the ground truth model. The detection accuracy of HMMScan is defined for a particular ground truth model structure as the fraction of samples for which HMMScan correctly detects the model structure. The most important aspect for this metric is the ability of the method to distinguish between single and multiple state models.

The second metric is *state prediction accuracy*, which evaluates the hidden state predictions. For a given sample sequence, the HMMScan method is deemed to have correctly detected the model structure if that sample is generated by a multiple-state (single-state) model and the model with the lowest BIC also has multiple states (a single state).

The state prediction accuracy for a single sample sequence is defined as the balanced accuracy of the per lot hidden state predictions from the model with the lowest BIC. Balanced accuracy is defined as the equally weighted average of the hidden state prediction accuracies for each hidden state. This metric is used to correct for imbalance in the ground truth frequency of the hidden states in a sample sequence. In instances with a ground truth model with multiple states, and where the model with the lowest BIC, selected by the HMMScan method, has a

single-state structure, the ground truth state with the lowest mean AE rate (the low-risk state) will be predicted for all lots in the sequence. The state prediction accuracy for HMMScan for a particular ground truth model structure is defined as the mean of the state prediction accuracies across the samples generated by that model structure.

The primary accuracy assessment is performed using instances with ground truth HMMs models of one state or two states (low-risk and high-risk). The one-state ground truth HMMs have two binomial mixture components. The transition matrices associated with the two-state ground truth models are defined by three input parameters. The first parameter is the number of hidden states. The second parameter is the stationary probability of the low-risk state. Finally, the third parameter is the average number of consecutive lots in the high-risk state, often called the *mean high-risk sojourn length*. The different combinations of these inputs can be mapped to the following five practical motivating scenarios:

1. **No High-Risk Sojourns**. Sequences are generated by single-state HMMs, reflecting a process where per lot AE rates are not affected by manufacturing and supply chain variation.
2. **Short and Frequent High-Risk Sojourns.** The lots oscillate rapidly between the low-risk state and the high-risk state, simulating manufacturing and supply chain processes that lacks proper control.
3. **Short and Infrequent High-Risk Sojourns.** The process primarily operates in the low-risk state and occasionally moves into a high-risk state for a short period of time. The low sojourn time of the high-risk state indicates that the initially unobserved, or hidden, manufacturing or supply

chain issues driving the differences in AE risk are resolved promptly, but the recurrence of the high-risk state indicates that the root cause is not fully resolved or a different issue has occurred.

4. **Long and Frequent High-Risk Sojourns.** The process experiences many hidden issues that take an extended period of time to detect and resolve.
5. **Long and Infrequent High-Risk Sojourns.** The process experiences few hidden issues that take an extended period of time to detect and resolve.

Within each scenario, both the length of the input sample sequence and the similarity between the mixture components (one-state models) or state-dependent distributions (two-state models) are varied. The similarity between two distributions is controlled by setting the binomial parameters to induce a particular value of the overlapping coefficient (OVL) [15,16]. The OVL, which ranges between 0 and 1, measures the probability mass that is intersected by two probability mass functions. The length of the sample sequences is varied between 50 to 500. This range covers the sequences lengths observed in the use case data described in Section 3.2.1 (114-460 lots). Table 1 lists the specific parameter values used to define the ground truth models and sequences lengths.

2.4 Use Case Application and Validation

The HMMScan method is also applied to real field data for three sequences of lots, each consisting of a different dose form of the same drug. The three dose forms each have different manufacturing and supply chain attributes and different mean levels of AEs. The lots for each dose form were considered as a temporal sequence based on the packaging date.

The manufacturer shared the sequence of valid lot numbers, sizes, and packaging dates for three lot sequences of interest. The reported AE counts per lot were obtained from the U.S. FDA Adverse Event Reporting System (FAERS) database [17], which aggregates spontaneous AE reports from manufacturers, patients, and health care providers primarily based in the United States. Each AE report consists of one or more reactions for a single patient incident. The manufacturer data were matched to the AE information from FAERS using the lot numbers provided in both sources. Due to data privacy constraints, the manufacturer was not able to share patient covariates or additional distribution and manufacturing information related to the lots.

For each dose form, the HMMScan method returns the fitted HMM with the best BIC score. The structure of this HMM is denoted as (S_{BIC}, C_{BIC}), i.e., S_{BIC} states and C_{BIC} mixture components. A similar validation approach to Section 2.3 is then applied to the field data and the fitted HMMs to gauge the likelihood that HMMScan method has accurately identified the correct model structures. Specifically, the goal is to estimate the probability than an HMM with structure (S_{BIC}, C_{BIC}) would have been chosen if the observed lot sequence were generated by an HMM with a different structure.

To obtain this estimate, 100 sample sequences with the same length as the observed sequence are generated from each fitted candidate HMM not selected by the BIC. Consider a sample sequence generated by a specific such candidate HMM (not selected by the BIC) with structure denoted by $(S_{sampling}, C_{sampling})$. A sample sequence is considered *misidentified* if an HMM with structure (S_{BIC}, C_{BIC}) has the best BIC of all candidate models fit to that sequence. The

fraction of misidentified sample sequences gives an estimate of the probability of misidentifying a sequence generated by a $(S_{sampling}, C_{sampling})$ HMM as a sequence from a (S_{BIC}, C_{BIC}) HMM. Similarly, the *single-state false negative rate* is defined as the probability that a single-state HMM has the best BIC score when all candidate models are fit to a sample generated by the (S_{BIC}, C_{BIC}) HMM, where $S_{BIC} > 1$. The single-state false negative rate provides additional insight about how distinct the fitted (S_{BIC}, C_{BIC}) model is from candidate fitted HMMs with $S = 1$ structures. Additional false negative rates are described in Section S2 of Online Resource 1, estimating the probability that a sample from the (S_{BIC}, C_{BIC}) HMM is incorrectly identified as the wrong multiple-state model or a model with the wrong number of mixture components.

Finally, as a separate sensitivity analysis, the HMMScan method is applied to the sequence of AE counts and lot sizes and the results are compared to primary results using normalized observed AE rates.

3 Results

All analysis was conducted in R and Python, and the code is available as a GitHub repository [18]. This repository includes a tutorial for generating results for a new use case, as well as instructions for reproducing the use case and simulation results presented in this paper. The relevant data are also stored in a public repository [19]. This analysis was performed during the period between 2018 and 2023.

3.1 Validation on Simulated Data

The results of the single-state model simulations, provided in Section S3.1 of Online Resource 1, show HMMScan detection accuracies above 0.97 for all sequence lengths and all degrees of mixture component overlap. The two-state simulation results are presented in Fig. 2. The results indicate that the HMMScan method has detection accuracy greater than 0.90 for sequence lengths 100 lots or longer, where both the low-risk and high-risk states are well separated (i.e., overlapping coefficient equal to 0.05), and have similar long-term frequencies (i.e., low-risk stationary probability is between 0.5 and 0.75, high-risk sojourn length is between 4 and 25).

For models with more imbalanced long-term frequencies of the states (i.e., a high low-risk stationary probability of 0.75 or 0.90) and either very short or very long high-risk sojourn lengths (i.e., 2 and 25, respectively), input sequences of 150 lots or more are required to meet this detection threshold. Moreover, when the states are not as well-separated (overlapping coefficient of 0.25 or greater), input sequences of 200 lots or more are required. In scenarios in which the high-risk state is much less frequent (low-state stationary probability equal to 0.90), then detection accuracy of over 0.90 is achieved only if the state-dependent AE distributions are well-separated (specifically, an overlapping coefficient of 0.25 or lower) and the high-risk sojourns are of moderate length (average length between 2 and 10), even for input sequences of 300 lots or longer.

For short input sequences, the highest detection accuracy is achieved in scenarios with equal prevalence of well-separated (overlapping coefficient equal to 0.05) high and low risk states and either medium length sojourns in the high-risk state

401 (high-risk sojourn length equal 10), or rapidly oscillating states (high-risk mean
402 sojourn length equal 1.25). In contrast, very low detection accuracy (less than
403 0.10) is observed for all input sequences lengths when the high risk state is
404 infrequent (high risk mean sojourn length of 1.25 and low-risk stationary
405 probability between 0.75 and 0.90), regardless of the degree of state overlap. This
406 is somewhat expected since when the average high-risk sojourn lasts less than two
407 lots, detection of multiple states is extremely difficult unless the frequency of the
408 high and low risk states is nearly identical. Note that a generating model with two
409 states and equal transition probabilities is indistinguishable from a single state
410 model, accounting for the low accuracy in the first plot of the second row, which
411 is again expected.
412
413 Detailed state prediction accuracy results can be found in Sections S3.2 – S3.5 of
414 Online Resource 1 for multiple-state simulations, including instances with three-
415 and four-state models and multiple mixture components. These results are
416 qualitatively the same as the two-state results described above. As expected, if
417 two of the three states are very similar in a three-state generating model, it is
418 difficult for the HMMScan method to distinguish between the similar states.
419 Crucially, this does not impact HMMScan's ability to detect that these samples
420 were drawn from a multiple-state model.

421 ## 3.2 Use Case Application and Validation

422 ### *3.2.1 Data*

423 Table 2 summarizes the inclusion and exclusion criteria with respect to the AEs
424 reported from FAERS for the three dose forms of the respective biologic. First,

AE reports with a missing lot number in FAERS are excluded from the analysis, as are reports with an invalid lot number that does not appear in the manufacturer's records. AEs related solely to drug administration reactions (e.g., "wrong dose administered") or unrelated reactions (e.g., "dog bite") that are highly unlikely to reflect product quality issues are also excluded. A full list of excluded reactions can be found in Section S4 of Online Resource 1.

The primary analysis further limits the set of relevant AEs to those with at least one reaction that is either known to be associated with the drug or that involves a serious reaction. These restrictions reflect a desire to minimize the number of included AEs that are not directly related to the product without omitting very serious AEs. A list of known reactions is obtained from the drug's package label. Chest pains, pneumonia, fungal infections, malignancies, and relapse of prescribed indications are examples of known reaction categories included in the analysis. This list is augmented with the following serious reactions: loss of consciousness, arrythmia, hospitalization, and death.

A secondary robustness analysis is conducted using only AEs from expedited reports. This class of AEs contains event reports deemed both serious and unexpected by the manufacturers, and therefore manufacturers are required by regulation to report these events to the FDA. The expedited reports capture events that are most likely to be concerning to manufacturers and regulators.

After restricting the set of eligible AEs, the raw AE counts and the number of doses per lot are used to create per lot AE rates based on a normalized lot size of $D = 100{,}000$ doses. The choice of the normalization factor is due to data privacy

considerations with respect to the exact lot sizes. The final preprocessing step removes 17 lots (1.9%) with outlier AE rates from the dataset[1]. When the outlier lots are removed from a lot sequence, the lots on either side of the outliers are treated as consecutive, a method known as "gluing" [20]. Prior research indicates that applying the gluing procedure with less than 8% of lots designated as missing does not affect the likelihood or magnitude of HMM parameter estimates [20]. Retaining outlier lots and capping their AE rates at the 75th percentile plus 1.5 times the IQR was also tested as an alternative outlier preprocessing step with no meaningful changes in the results. The structures of the best-fitting HMM models and the sequences of predicted hidden states were not materially affected by the choice of outlier preprocessing method (see Section S5 of Online Resource 1 for full results using capping). Table 3 shows the distribution of the AE rates per lot for each modeled dose form.

3.2.2 Model Selection Results

For each dose form, the grid of candidate model structures is constructed by setting $S_{max} = 4$ and $C_{max} = 9$. The BIC values only degraded outside the chosen hyperparameter ranges, indicating that the complexity penalty is outweighing the likelihood gains, and likely overfitting models. Each of the candidate models is fit with 50 random initializations and the results corresponding to the parameter estimates with the highest likelihood are retained. The BIC values for the fitted candidate models are shown in Fig. 3.

[1] An outlier is defined as an AE rate greater than the 75th percentile plus 1.5 times the interquartile range (IQR) [28].

Dose Forms A and B. Multiple state HMMs have the best fit as measured by BIC for dose forms A and B ($S = 3, C = 2$ for dose form A and $S = 3, C = 3$ for dose form B). The BIC difference between the best-fitting multiple state model and the best-fitting single state model is larger than 10 for both dose forms, suggesting significantly stronger fit for a multiple state model and related serial correlation in the per lot AE rates [11]. Note for dose form B that the other multiple-state models ($S = 3, C = 2$; $S = 2, C = 3$; $S = 2, C = 4$) have similar BIC values to the best model (with $S = 3, C = 3$).

Dose Form C. A multiple-state HMM with $S = 2$ and $C = 3$ provides the best BIC for dose form C, but the BIC difference between this model structure and a single state model with $C = 3$ is lower than 10, indicating weaker evidence of serial correlation in the per lot AE rates.

The models were also estimated when the exact lot sizes were considered using lot size weights in the likelihood function, and the respective results are shown in Section S6 of Online Resource 1. The number of hidden states in the models with the lowest BIC values do not change compared to the results presented above when the inputs to the model are the normalized per lot reported AE rates.

3.2.3 Identifying States with High AE Risk

Fig. 4 and Table 4 illustrate the maximum likelihood estimated parameters for the HMM with the best BIC value for dose forms A and B. This includes the state transition matrix, the stationary distribution of the time spent in each state, and the state-dependent mixture distribution. Due to the relatively weak evidence in favor of a multiple-state state model for dose form C, the maximum likelihood

parameters are included in Section S6 of Online Resource 1. In Fig. 4, a clear separation exists for dose form A between states 3 and 1, corresponding to high and low average number of reported AEs, respectively. State 2 represents a medium-risk state. Similarly, there is clear separation between the high-risk state 3 and the low-risk state 1 for dose form B. Section S7 of Online Resource 1 provides the parameter estimates using lot size weights.

3.2.4 AE Risk Transitions

The estimated state-specific mean AE rates in Table 4 demonstrate that the ordering of the states by AE risk is robust. The estimated transition matrices both have high probabilities on the diagonal, indicating that the hidden states are all highly persistent. This suggests that high-risk and low-risk AE states tend to form long contiguous regions.

In fact, these regions are observable in Fig. 5 for both dose form A and dose form B. This figure orders the lots by packaging date for both dose forms and colors the AE rate for each lot by its most likely hidden state. Both dose forms have two clearly identifiable regions of high-risk lots as well as multiple low-risk regions at the beginning and end of the sequences (Fig. 5a and Fig. 5b). Furthermore, when the HMMScan method is performed using AE rates based solely on expedited reports, the best-fitting HMMs indicate nearly identical high-risk regions (Fig. 5d and Fig. 5e). These regions are essentially contiguous despite the presence of occasional lots with low AE rates in the high-risk regions. Analysis of multiple AE definitions (known and serious, expedited) is used to establish consistency of the state transition points. Transitions to and from the high risk state near lots 20 and 200 for dose form A, lots 25 and 140 for dose form B appear for both

definitions, signaling that these points should be prioritized for root cause investigation.

Similar persistent high-risk regions are visible for dose form C in Fig. 5c. However, the results on the expedited AE reports indicate that a single-state model has the best BIC score, further suggesting only weak evidence in favor of multiple states in the ground truth model for this lot sequence.

The most likely hidden state sequences generated by the models fit using lot size weights are available in Section S6 of Online Resource 1. The results differ only minimally compared to the Fig. 5.

3.2.5 Use Case Validation

Fig. 6 shows the estimated misidentification probability for each candidate model structure for each of the three lot sequences in the use case. Across all three lot sequences, the sample sequences generated by single-state HMMs are very rarely identified as having S_{BIC} states and C_{BIC} mixture components by HMMScan (misidentification probability ≤ 0.01). For dose form A, the misidentification probabilities for the two-state generating models are less than 0.10. The two-state generating model misidentification probabilities for dose form B are less than 0.18.

The estimated single-state false negative rates for dose forms A and B are very low, 0.0 and 0.02 respectively, and as expected much higher (0.71) for dose form C. Multiple state and mixture component false negative rates are reported in Section S2.2 of Online Resource 1.

4 Discussion

The HMMScan method is proposed as an initial signal detection tool to identify lot sequences where serial correlation in AE rates suggests the potential presence of clinically relevant variation in manufacturing and supply chain conditions. The method can be naturally extended to take variable lot sizes as input, as well as additional temporal information with respect to intermediate lots used during the manufacturing process. The method is particularly relevant for biologic drug manufacturing, where the inherent complexity compared to traditional small molecule (chemical) drugs is well-known and stems from the fact that these processes are primarily based on biological processes. In addition, often biologic drugs require special maintaining special conditions throughout the supply chain and distribution (e.g., temperature control).

In principle, a primary benefit of HMMScan is the potential to enable manufacturers and regulators to combine AE and lot-specific information to identify previously hidden signals and direct investigations in a scalable fashion across a range of pharmaceutical products. Realizing this benefit relies heavily on adverse event reporters providing lot numbers as good clinical practice. Additionally, such broad application of HMMScan would rely on a well-developed data input pipeline to gather the following information for each lot: packaging date, relevant AE counts, number of doses, and dose form. This is the minimum required data input for the method as currently constructed, though in principle the model could take additional information about the distribution patterns by lot, including more granular regional distribution information and patient characteristics. Additional information about the lot-to-lot differences in

patient populations could be used to adjust the AE counts to account for these differences. In this case, a positive signal of serial correlation in AE rates would be even more likely to correspond to variation in manufacturing and supply chain conditions. However, the use case data available from the manufacturer did not include these in-depth lot-specific data related to patient characteristics and manufacturing conditions that would enable potential root causes of the observed variation. Further collaboration between regulators, manufacturers, and academics to collect and format these data is the first step toward realizing this opportunity to augment drug safety monitoring to improve patient outcomes.

One natural question is how the proposed HMMScan compares with more naïve approaches to test for serial correlation. Indeed, simple statistical tests for serial correlation could indicate the presence or absence of serial correlation in a time series. However, hidden Markov models are also able to capture more subtle correlation structures likely to exist that differ across hidden states. Another advantage of the HMMScan method is that it also indicates which particular lots are more likely to be related to risky states of the manufacturing or supply chain condition. This could provide significant help to guiding further investigation of potential causal factors that drive the risky states.

Another natural question is whether the BIC is the appropriate model selection criterion to select the best-fitting HMM. Other model selection criteria, including the Akaike Information Criterion (AIC) and the bootstrap likelihood ratio test, have been utilized in the literature to select between latent state models like HMMs. Nylund et al. compare multiple information criteria, including BIC and AIC, to several likelihood ratio tests in the task of identifying the correct number

of hidden states in three types of latent variable models [23]. The authors identify the BIC and the bootstrap likelihood ratio test as the most accurate methods. However, the bootstrap likelihood ratio test, which takes a pair of models and provides a statistical signal about the relative fit of the pair, is difficult to apply for the HMMScan method because of the difficulty in identifying a single model from a set with the best fit to a particular sequence.

The simulated scenario validation results suggest very natural insights with respect to the expected accuracy of the HMMScan method. Specifically, they illustrate that the method's accuracy improves with longer input lot sequences, highly distinguishable low and high risk states, and balanced long-term frequency. High detection accuracy on well-separated states is important for identifying large differences between high-risk and low-risk states that are likely to correspond to high-priority investigations. On the other hand, the scenarios in which the method's accuracy is relatively lower, tend to be those in which the low and high risk states are less distinguishable, in which case, there is lower priority to detect them since the clinical impact is lower. The accuracy of the HMMScan method is also lower under scenarios with shorter input sequence and very long high-risk sojourn lengths. These results are somewhat unsurprising because in both of these scenarios there is a significantly lower number of state transitions, and thus, detecting the existence of multiple (different) states is objectively more challenging.

Additional multiple-state simulations using instances with three- and four-state models and multiple mixture components directionally support the insights described above. As expected, if two of the three states are very similar in a three-

state generating model, the HMMScan method is frequently unable to distinguish between the similar states. Most importantly, the HMMScan method exhibits high accuracy in distinguishing between a single vs. multistate scenarios. This is particularly important because the method generates a signal in cases where there seems to be evidence for the existence of low and well-separated high risk state. The simulation results also indicate that the HMMScan method has high detection accuracy for input sequence lengths that are similar to the use case data, where both the low-risk and high-risk states seem quite distinguishable and have approximately the same long-term frequency.

Applying the method to the use case data shows strong evidence in favor of multiple state HMMs as the most likely generating models for the dose form A and B AE rate sequences. The use case validation results provide support for this interpretation. The estimated use case misidentification probabilities demonstrate a low risk that the observed dose form A and B sequences were generated by a single-state stochastic process. Furthermore, the single state false negative rate estimates indicate that samples from an HMM with S_{BIC} states and C_{BIC} mixture components are very rarely identified as samples from a single state model for dose forms A and B.

Interesting, in contrast to dose forms A and B, the evidence from the detection results only weakly favor a multiple state model for dose form C. While the estimated misidentification probabilities for dose form C single state HMMs are low, the single-state false negative rate analysis indicates that samples from an HMM with S_{BIC} states and C_{BIC} mixture components are frequently mistaken for samples from single state models. Furthermore, only a single state appears in the

651 sequence of most likely states for the dose form C expedited AE sequence.
652 Overall, this analysis does not strongly support the existence of differentiated
653 manufacturing and supply chain risk states for dose form C. This type of result
654 could indicate that the mechanism of delivery may have an impact on reported AE
655 rates, suggesting that manufacturing issues could be relevant beyond the mere
656 active pharmaceutical ingredient (API).
657
658 Upon receiving a signal from the HMMScan method, a root cause investigator
659 could start by combining drug distribution data with patient population statistics
660 to check for significant changes in the patient population for the lots in the
661 vicinity of hidden state transitions. These are lots for which the HMMScan
662 method suggests clinically meaningful changes to manufacturing and supply chain
663 conditions might have occurred. Additionally, the investigator could look for
664 evidence suggesting process differences in manufacturing facilities across lots,
665 changes in raw materials suppliers and supply chain protocols, work order and
666 deviations information, numerical product quality measurements, and other data
667 available at a sub-lot level granularity to search for potential manufacturing and
668 supply chain mechanisms for changes in AE rates.
669
670 **Limitations.** One limitation of the HMMScan method is that it only raises a
671 potential signal of relevant variability but does not provide exact causes. Root
672 cause analysis utilizing additional, and likely proprietary, features of the
673 manufacturing lots would be essential to rule out patient-related factors and
674 confirm a causal relationship between manufacturing and supply chain conditions
675 and AE rate variation. The method does not account for other potential societal
676 sources of AE variability unrelated to manufacturing and patient-specific effects,

677 such as pandemics, pollution, and changes in awareness of AE reporting
678 mechanisms. Another limitation is that the method does not provide explicit
679 probabilistic guarantees about the signal.
680
681 It is important to acknowledge that underreporting of AEs to spontaneous
682 reporting systems has been a well-documented but not well-understood concern,
683 with some estimates of the underreporting rate over 90% [10]. More recent
684 research by Alatawi and Hansen continues to find wide disparities in the estimated
685 underreporting rate across products, though the authors notably do not find any
686 statistically significant underreporting for biologics [24]. In particular, if the
687 reporting rate is constant over time or known in terms of relative magnitude over
688 time, the ability for the HMMScan method to detect serial correlation is
689 unaffected by the absolute level of this rate. Moreover, while sudden, short-term
690 changes in the reporting rate could be mistaken as state transitions that affect the
691 results of the HMMScan method, long-term, moderate trends, either positive or
692 negative, should not meaningfully affect the ability of the method to detect local
693 serial correlation.
694
695 **Future directions**. A possible direction for future methodological research is to
696 increase the complexity of the candidate model structures that HMMScan
697 considers by allowing the hidden state of lot ℓ to depend on a prior history of
698 states before lot $\ell - 1$. Limited dependence on only the most recent hidden state
699 is useful because it yields the fast and well-understood Baum-Welch algorithm for
700 maximum likelihood parameter estimation. However, EM-based parameter
701 estimation algorithms for variable length HMMs, which allow state dependence
702 on history prior to the most recent state, have been proposed [25]. More recently,

a Bayesian model for variable length Markov chains was introduced [26], though this model has not been studied in a hidden Markov setting.

The HMMScan model could be further extended by considering the hidden states as partially observed, as proposed in [27]. For example, variability in the lot sizes could indicate changes in the stability of the manufacturing process. More detailed lot-specific data could also be incorporated into this framework when available.

5 Conclusion

This paper presents HMMScan, a novel pharmacovigilance method for detecting patterns in AE rates across manufacturing lots using probabilistic modeling techniques. HMMScan is a method that could be utilized by both manufacturers and regulators to automate lot variability monitoring and inform targeted root cause analysis. Specifically, HMMScan provides: (1) a reliable signal when serial correlation is detected in an observed AE rate sequence, and (2) a model to identify individual lot subsequences where variation in manufacturing and supply conditions may have contributed to higher AE rates. HMMScan's signal detection capability is validated using both simulated and field data. In a case study of three lot sequences corresponding to three dose forms of a major biologic, the strong evidence of serial correlation was detected for two of three dose forms.

Declarations

Funding. This work is supported by the U.S. Food and Drug Administration, Grant No. U01FD006483. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the financial sponsor.

Conflict of Interest Statement. Subsequent to the substantial completion of this work but prior to publication, the second, third and fourth authors (S.S., J.M.W., and R.L., respectively) received through MIT an award from the MIT-Takeda initiative to conduct research on signal detection. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Availability of Data and Materials. The datasets generated during and/or analyzed during the current study are available in the HMMScan Data Repository, doi.org/10.17632/zzd5vbj7yn.2.

Ethics Approval. Not applicable.

Consent to Participate. Not applicable.

Consent for Publication. Not applicable.

Code Availability. All code and related documentation are available at https://github.com/josh-wilde/hmmscan.

Author Contributions. Joshua Wilde, Stacy Springs, Jacqueline Wolfrum, and Retsef Levi contributed to the study conception and design. Joshua Wilde and Retsef Levi contributed to methodology development. Data collection and analysis were performed by Joshua Wilde. The first draft of the manuscript was written by Joshua Wilde and Retsef Levi and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

References

1. Harpaz R, DuMouchel W, Shah NH, Madigan D, Ryan P, Friedman C. Novel Data-Mining Methodologies for Adverse Drug Event Discovery and Analysis. Clin Pharmacol Ther [Internet]. 2012;91:1010–21. Available from: http://doi.wiley.com/10.1038/clpt.2012.50

2. Khouri C, Nguyen T, Revol B, Lepelley M, Pariente A, Roustit M, et al. Leveraging the Variability of Pharmacovigilance Disproportionality Analyses to Improve Signal Detection Performances. Front Pharmacol. 2021;12:1–7.

3. Kulldorff M, Dashevsky I, Avery TR, Chan AK, Davis RL, Graham D, et al. Drug safety data mining with a tree-based scan statistic. Pharmacoepidemiol Drug Saf [Internet]. 2013;22:517–23. Available from: https://onlinelibrary.wiley.com/doi/10.1002/pds.3423

4. Sandberg L, Taavola H, Aoki Y, Chandler R, Norén GN. Risk Factor Considerations in Statistical Signal Detection: Using Subgroup Disproportionality to Uncover Risk Groups for Adverse Drug Reactions in VigiBase. Drug Saf [Internet]. 2020;43:999–1009. Available from: https://doi.org/10.1007/s40264-020-00957-w

5. Beninger P. Opportunities for Collaboration at the Interface of Pharmacovigilance and Manufacturing. Clin Ther [Internet]. 2017;39:702–12. Available from: http://dx.doi.org/10.1016/j.clinthera.2017.03.010

6. United States FDA Center for Biologics Evaluation and Research. Best Practices in Drug and Biological Product Postmarket Safety Surveillance for FDA Staff [Internet]. 2019. Available from: https://www.federalregister.gov/documents/2019/11/07/2019-24332/best-practices-in-drug-and-biological-product-postmarket-safety-surveillance-for-food-and-drug

7. Dumouchel W, Yuen N, Payvandi N, Booth W, Rut A, Fram D. Automated method for detecting increases in frequency of spontaneous adverse event reports over time. J Biopharm Stat. 2013;23:161–77.

8. Heimann G, Belleli R, Kerman J, Fisch R, Kahn J, Behr S, et al. A nonparametric method to detect increased frequencies of adverse drug reactions over time. Stat Med. 2018;37:1491–514.

9. Mahaux O, Bauchau V, Zeinoun Z, Van Holle L. Tree-based scan statistic – Application in manufacturing-related safety signal detection. Vaccine [Internet]. 2018;37:49–55. Available from: https://doi.org/10.1016/j.vaccine.2018.11.044
10. Rabiner LR. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. Proceedings of the IEEE. 1989;77:257–86.
11. Raftery AE. Bayesian Model Selection in Social Research. Sociol Methodol [Internet]. 1995;25:111. Available from: https://www.jstor.org/stable/271063?origin=crossref
12. Dempster AP, Laird NM, Rubin DB. Maximum Likelihood from Incomplete Data Via the EM Algorithm. Journal of the Royal Statistical Society: Series B (Methodological) [Internet]. 1977;39:1–22. Available from: http://doi.wiley.com/10.1111/j.2517-6161.1977.tb01600.x
13. Baum LE, Petrie T, Soules G, Weiss N. A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains. The Annals of Mathematical Statistics [Internet]. 1970;41:164–71. Available from: http://projecteuclid.org/euclid.aoms/1177697196
14. Schreiber J, Allen PG. pomegranate: Fast and Flexible Probabilistic Modeling in Python. Journal of Machine Learning Research [Internet]. 2018 [cited 2022 Oct 13];18:1–6. Available from: https://github.com/jmschrei/pomegranate
15. Clemons TE, Bradley EL. A nonparametric measure of the overlapping coefficient. Comput Stat Data Anal [Internet]. 2000;34:51–61. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0167947399000742
16. Weitzman MS. Measures of Overlap of Income Distributions of White and Negro Families in the United States. Washington, D.C.; 1970.
17. U.S. Food and Drug Administration. FDA Adverse Event Reporting System (FAERS) [Internet]. 2022 [cited 2022 Jul 15]. Available from: https://fis.fda.gov/extensions/FPD-QDE-FAERS/FPD-QDE-FAERS.html
18. Wilde J. HMMScan: Surveillance of Adverse Event Variability across Manufacturing Lots in Biologics [Internet]. 2022 [cited 2022 Oct 30]. Available from: https://github.com/josh-wilde/hmmscan
19. Wilde J, Levi R. HMMScan Data Repository [Internet]. Mendeley Data. 2022 [cited 2022 Oct 30]. Available from: https://data.mendeley.com/datasets/zzd5vbj7yn

20. Popov AA, Gultyaeva TA, Uvarov VE. Training hidden Markov models on incomplete sequences. 2016 13th International Scientific-Technical Conference on Actual Problems of Electronics Instrument Engineering (APEIE) [Internet]. IEEE; 2016. p. 317–20. Available from: http://ieeexplore.ieee.org/document/7806478/
21. Ratcliff R, Tuerlinckx F. Estimating parameters of the diffusion model: Approaches to dealing with contaminant reaction times and parameter variability. Psychon Bull Rev [Internet]. 2002;9:438–81. Available from: http://link.springer.com/10.3758/BF03196302
22. Wagenmakers EJ, Ratcliff R, Gomez P, Iverson GJ. Assessing model mimicry using the parametric bootstrap. J Math Psychol. 2004;48:28–50.
23. Nylund KL, Asparouhov T, Muthén BO. Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. Structural Equation Modeling. 2007;14:535–69.
24. Alatawi YM, Hansen RA. Empirical estimation of under-reporting in the U.S. Food and Drug Administration Adverse Event Reporting System (FAERS). Expert Opin Drug Saf [Internet]. 2017;16:761–7. Available from: https://www.tandfonline.com/doi/full/10.1080/14740338.2017.1323867
25. Dumont T. Context Tree Estimation in Variable Length Hidden Markov Models. IEEE Trans Inf Theory [Internet]. 2014;60:3196–208. Available from: http://ieeexplore.ieee.org/document/6780620/
26. Kontoyiannis I, Mertzanis L, Panotopoulou A, Papageorgiou I, Skoularidou M. Bayesian context trees: Modelling and exact inference for discrete time series. J R Stat Soc Series B Stat Methodol [Internet]. 2022; Available from: https://onlinelibrary.wiley.com/doi/10.1111/rssb.12511
27. Monaco J V, Tappert CC. The partially observable hidden Markov model and its application to keystroke dynamics. Pattern Recognit [Internet]. 2018;76:449–62. Available from: https://www.sciencedirect.com/science/article/pii/S0031320317304752
28. Walfish S. A Review of Statistical Outlier Methods. Pharmaceutical Technology [Internet]. 2006; Available from: http://www.pharmtech.com/pharmtech/content/printContentPopup.jsp?id=384716

Fig. 1 HMMScan model selection procedure. Alternatively, lot sizes and raw AE counts, both L-dimensional vectors, can be provided as input data rather than per lot AE rates. In this case, during Parameter Estimation each lot is assigned a weight proportional to its respective number of doses.

BIC: Bayesian Information Criterion; HMM: Hidden Markov Model

Fig. 2 Method validation simulation results showing detection accuracy for HMMs with two states and one mixture component. Each point represents the HMMScan detection accuracy calculated on 100 sample sequences with the same length as denoted on the x-axis. The panels are organized in columns based on the low-risk state stationary probability and in rows by the mean high-risk state sojourn length. Each column value represents the expected fraction of lots from the low-risk state in a sample. Each row value represents the expected number of consecutive high-risk state lots observed each time the system moves to the high-risk state.

HMM: Hidden Markov Model

Fig. 3 BIC values for the candidate HMMs for each dose form. Each tile indicates the BIC value for a fitted HMM with the number of states denoted on the x-axis and the number of binomial components per state-specific mixture distribution on the y-axis. Lower BIC values indicate a better fit of the model to the data, and the candidate HMMs with the best fit are highlighted in dark red

HMM: Hidden Markov Model; BIC: Bayesian Information Criterion

Fig. 4 Fitted state-specific binomial mixture distributions for the best-fitting HMMs for dose forms A and B. Each panel shows the distribution for the state-specific distribution associated with each hidden state.

AE: Adverse Event

Fig. 5 Per lot AE rates. The top row of plots calculates per lot AE rates based on the known and serious definition, while the bottom row includes only expedited AE reports. The lots are shaded by most likely hidden state according to the HMM with the best BIC.

HMM: Hidden Markov Model; BIC: Bayesian Information Criterion

Fig. 6 Estimated misidentification probabilities for the use case method validation. Each tile indicates the misidentification probability for a given sampling model with respect to S_{BIC} and C_{BIC}. For dose form A, $S_{BIC} = 3$ and $C_{BIC} = 2$, for dose form B, $S_{BIC} = 3$ and $C_{BIC} = 3$, and for dose form C, $S_{BIC} = 2$ and $C_{BIC} = 3$

Table 1 Input parameters for two-state model validation simulated instances.

Parameter	Description	Parameter Values
All HMMs		
Sequence Length	Length of sample sequences	{50, 100, 150, ..., 500}
One-State, Two-Component HMMs		
Overlapping Coefficient	Overlap between binomial components of the mixture distribution	{0.05, 0.25, 0.50}
Two-State, One-Component HMMs		
Overlapping Coefficient	Overlap between state-specific binomial distributions	{0.05, 0.25, 0.50}
Low-Risk State Stationary Probability	Long-term frequency of lots in low-risk state	{0.50, 0.75, 0.90}
High-Risk State Mean Sojourn Length (lots)	Average number of consecutive high-risk lots in an infinitely long sample	{1.25, 2, 4, 10, 25}

HMM: Hidden Markov Model

Table 2 Count of adverse event reports by inclusion/exclusion criteria.

	Dose Form A (463 lots)	Dose Form B (271 lots)	Dose Form C (119 lots)	Missing Lot Number	Invalid Lot Number
Raw AEs from FAERS	71,890	13,582	2,789	283,888	8,653
Excluding drug administration AEs	67,402	13,184	2,562		
Relevant (known + other serious) AEs	**21,628**	**4,950**	**884**		
Expedited AEs	**7,798**	**2,051**	**437**		

AEs: Adverse Events; FAERS: FDA Adverse Event Reporting System

Table 3 Adverse Event Rates per Lot

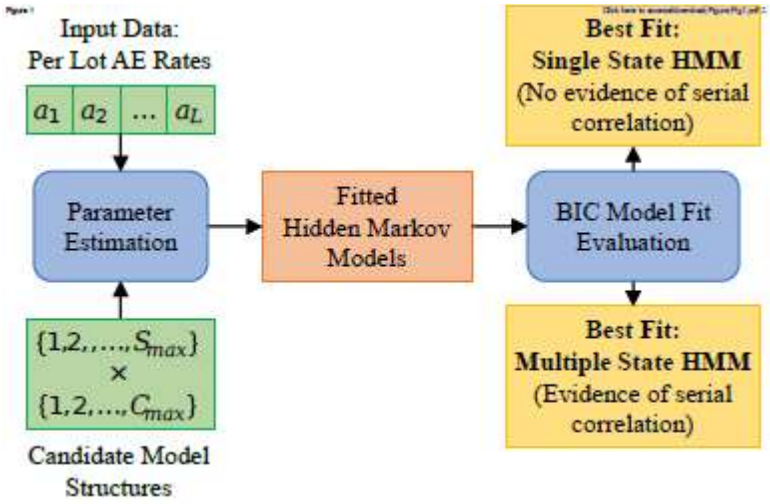
	All Lots			Outliers Removed		
	Dose Form A	Dose Form B	Dose Form C	Dose Form A	Dose Form B	Dose Form C
Minimum	0	0	0	0	0	0
25th Percentile	27	9	3	27	9	2
50th Percentile	41	18	11	41	18	10
75th Percentile	63	29	19	62	28	18
95th Percentile	85	44	47	84	39	23
Maximum	151	280	248	113	51	30
Mean	45	22	21	44	18	11
Lot Count	*463*	*271*	*119*	*459*	*264*	*113*
Lot Count, > 0 reported AEs	*453*	*234*	*93*	*449*	*227*	*87*

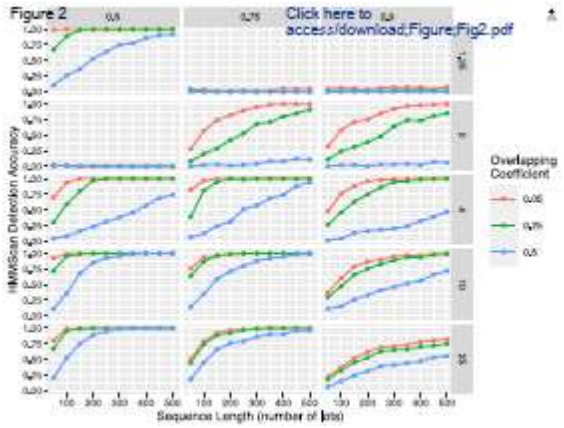
Table 4 Estimated transition matrix and state-specific mean AE rates for best-fitting HMMs, with mean AE rate 90% CIs (confidence intervals) estimated via parametric bootstrap [21,22].

	Dose Form A					Dose Form B				
	Transition Probabilities (from row state to column state)					Transition Probabilities (from row state to column state)				
Hidden State	To State: 1	To State: 2	To State: 3	Mean AE Rate (90% CI)	Stat. Prob.	To State: 1	To State: 2	To State: 3	Mean AE Rate (90% CI)	Stat. Prob.

1	0.76	0.19	0.05	9.8 (9.4 – 12.8)	0.14	0.92	0.08	0.00	6.9 (0.0 – 8.3)	0.25
2	0.06	0.90	0.04	32.9 (29.4 – 32.6)	0.43	0.06	0.85	0.09	14.5 (9.1 – 17.8)	0.30
3	0.02	0.03	0.95	66.4 (60.2 – 64.5)	0.43	0.00	0.06	0.94	26.4 (23.0 – 30.3)	0.45

HMM: Hidden Markov Model





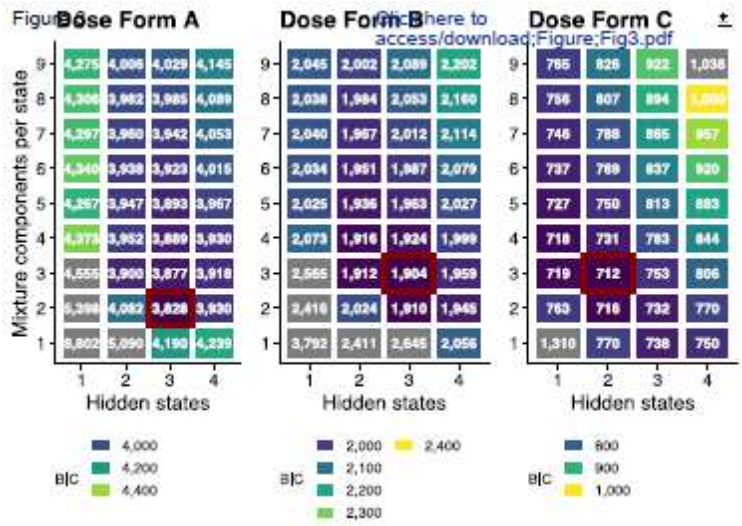
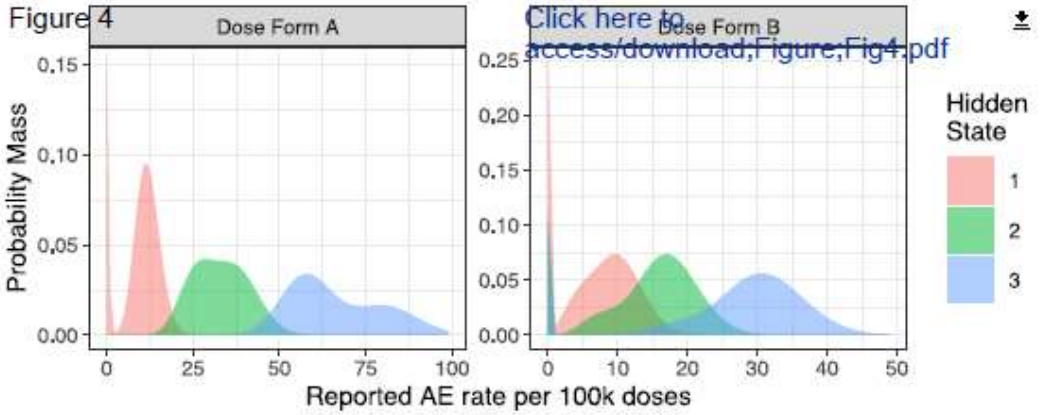
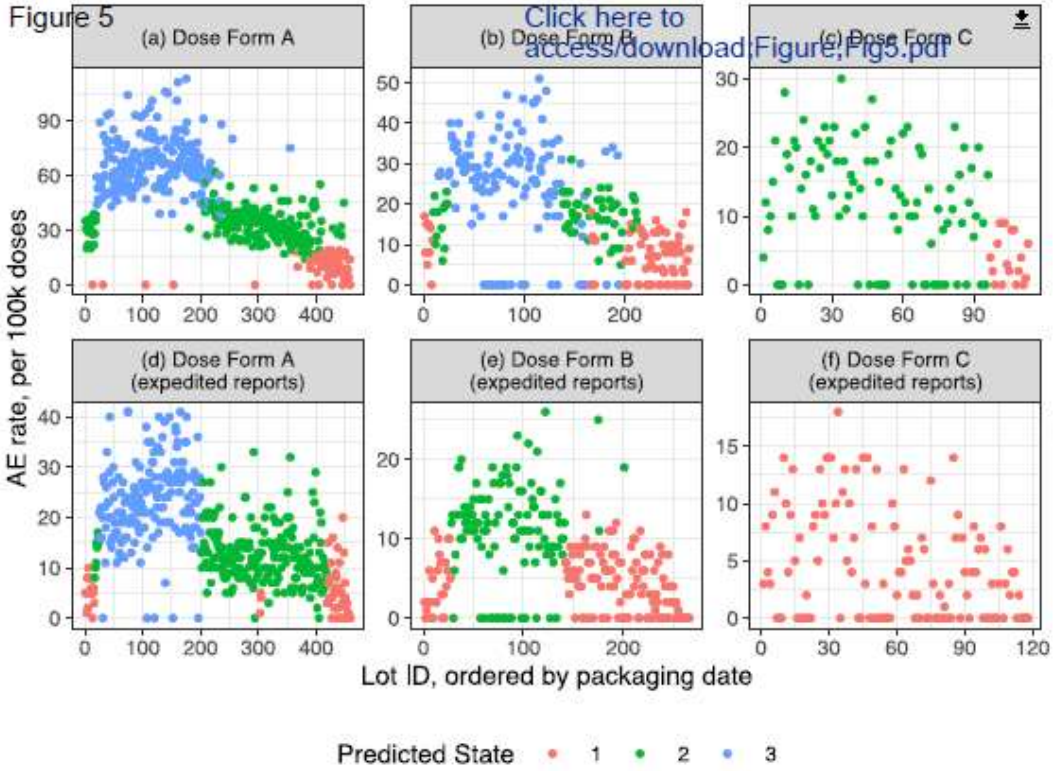


Figure 4



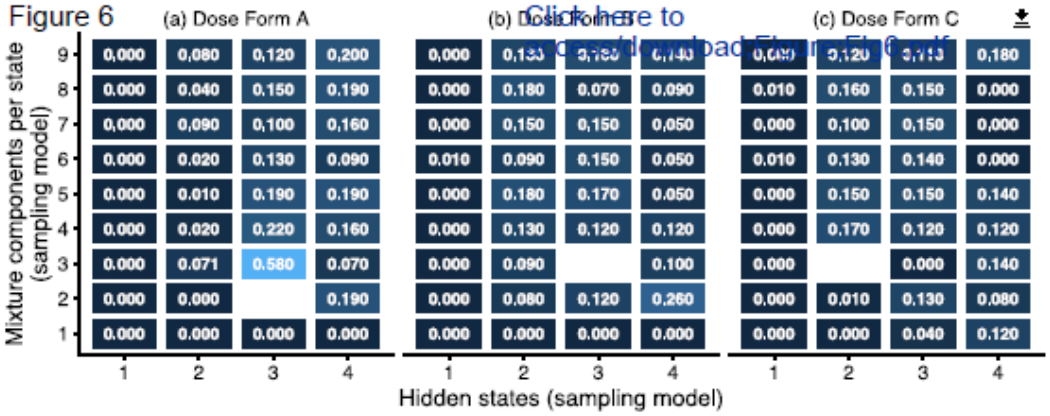
Click here to access/download;Figure;Fig4.pdf

Figure 5



Click here to access/download;Figure;Fig5.pdf

Figure 6



Click here to access/download;Figure;Fig6.pdf