Submitted to *Operations Research*
manuscript

Active Postmarketing Drug Surveillance for Multiple Adverse Events

Joel Goh
Stanford Graduate School of Business, CA 94305, joelgoh@stanford.edu

Margrét V. Bjarnadóttir
Robert H. Smith School of Business, MD 20742, margret@rhsmith.umd.edu

Mohsen Bayati
Stanford Graduate School of Business, CA 94305, bayati@stanford.edu

Stefanos A. Zenios
Stanford Graduate School of Business, CA 94305, stefzen@GSB.stanford.edu

Active postmarketing drug surveillance is important for consumer safety. However, existing methods have limitations that prevent their direct use for active drug surveillance. One important consideration that has been absent thus far is the modeling of multiple adverse events and their interactions. In this paper, we propose a method to monitor the effect of a single drug on multiple adverse events, which explicitly captures interdependence between events. Our method uses a sequential hypothesis testing paradigm, and employs an intuitive test-statistic. Stopping boundaries for the test-statistic are designed by asymptotic analysis and by reducing the design problem to a convex optimization problem. We apply our method to a dynamic version of Cox's proportional hazards model, and show both analytically and numerically how our method can be used as a test for the hazard ratio of the drug. Our numerical studies further verify that our method delivers Type I/II errors that are below pre-specified levels and is robust to distributional assumptions and parameter values.

Key words: Health Care, Stochastic Models, Drug Surveillance, Adverse Drug Events

1. Introduction

In most countries, before a drug is approved for commercial distribution, the drug has to undergo a series of clinical trials, designed to assess both its efficacy and potential side effects. However, clinical trials may fail to detect all the adverse side effects associated with the drug. Researchers have pointed to the small size and low statistical power of clinical trial populations (Wisniewski et al. 2009) and pressures for quick drug approval (Deyo 2004) as possible reasons for such failures. *Postmarketing drug surveillance* is the process of monitoring drugs that are already commercially distributed, in order to flag drugs that have potential adverse side effects. It is typically based on observational data and cannot definitively establish or refute causal relationships between drugs

and adverse events. Nevertheless, if a causal relationship exists, effective drug surveillance would provide preliminary evidence for this relationship and an early warning signal for regulators to take mitigating actions (e.g., limit the drug's distribution, issue warnings to consumers, order further studies), thereby acting as a final safeguard to protect the consumer population from lapses in the initial process of drug approval.

Drug surveillance systems may be classified as either *passive* or *active*. A passive surveillance system relies on physicians and patients to voluntarily report suspected drug-associated adverse events to the relevant health authorities. In contrast, an active surveillance system employs automated monitoring of public health databases to proactively infer associations between drugs and adverse events. In the U.S., the present system of drug surveillance is passive. The medical community, however, has long argued against passive surveillance, citing its high susceptibility to biases such as underreporting of adverse events, and has repeatedly called upon the U.S. Food and Drug Administration (FDA) to develop an active surveillance system (e.g., Brewer and Colditz 1999, Furberg et al. 2006, Brown et al. 2007, McClellan 2007). In response, the FDA, also recognizing the importance of active surveillance, launched the Sentinel Initiative in May 2008, which aims to "develop and implement a proactive system ... to track reports of adverse events linked to the use of its regulated products" (FDA 2012).

Existing methods that have been proposed for active drug surveillance still possess significant limitations. The report by the FDA (Nelson et al. 2009) presents a comprehensive account of existing methods and discusses their limitations. The report concludes that a common limitation of existing methods is the limited handling of confounding factors. Other limitations in individual methods include failure to account for the temporal sequence of adverse events and drug usage, lack of control over statistical parameters of interest (such as false detection rates), and high sensitivity to distributional parameters.

1.1. Contributions

Our central contribution in this paper is to propose a method for active drug surveillance, termed *Queueing Network Multi-Event Drug Surveillance* (QNMEDS), that overcomes some of the limitations listed above. In particular, QNMEDS allows simultaneous surveillance of the effect of a drug on multiple adverse events, and explicitly accounts the temporal dependencies between these adverse events, which is an important confounding factor that has been neglected in other methods. Specifically, QNMEDS allows an adverse event that has occurred in the past to affect the rate at which other adverse events occur in the future. Because drug surveillance uses public health data, it does not have have a well-controlled study population (unlike, e.g., clinical trials), and therefore, multiple adverse events are expected to exist in the study population. Moreover, the

epidemiological literature suggests that the magnitude of these interdependencies can be significant. For example, diabetics are known to have around two times the risk of cardiac events than non-diabetics (e.g., Abbott et al. 1987, Manson et al. 1991). One way that this interdependency complicates the problem of drug surveillance is when a drug that may be "safe" for patients who are free of adverse events could be "unsafe" for patients who had previously experienced some adverse event. For example, researchers have found that antiplatelet therapy, which is prescribed to prevent adverse cardiovascular events, actually *increased* the risk of major adverse cardiovascular events if the patient had Type II diabetes (Angiolillo et al. 2007). These considerations suggest that surveillance methods that fail to account for the interactions between multiple adverse events may make erroneous conclusions. This point is further supported by our simulation study in §7.3.

In addition, QNMEDS also possesses the following desirable features: 1) it is designed for sequentially arriving data, 2) it incorporates temporal dynamics such as the sequencing of drug treatment and adverse events, 3) it allows the user to control statistical parameters such as the false detection probability (Type I error), and 4) it is robust to distributional assumptions, unlike other tests that require strong assumptions about likelihood functions, such as the Sequential Probability Ratio Test (SPRT). With regard to the last point in particular, our numerical study on simulated data for two adverse events finds that an SPRT-based heuristic performs very poorly under mild perturbations of its parametric assumptions, even though it has excellent performance when its assumptions are met. Specifically, in simulated data where the drug increased the risk of an adverse event between 40% to 60%, if the SPRT-based heuristic assumed a slightly mis-specified distribution of random event times, it could only detect this elevated risk in 0-2% of all the simulation runs. In contrast, on the same data, QNMEDS detected this association in 100% of the runs.

Through our analysis of QNMEDS, we make two further technical contributions:

1. As part of our analysis, we had to estimate the probability of hitting a certain set for a multidimensional Brownian Motion (B.M.) process with correlated components. We develop a novel method to bound this hitting probability from above by constructing a new multidimensional B.M. process with independent components, and calculating the hitting probability for the new B.M. process, which turns out to be more tractable.
2. We introduce a *cross-hazards model* that captures the dynamic interactions between multiple adverse events. Our cross-hazards model can be viewed as a dynamic version of the classical Cox proportionate hazards model. We show how QNMEDS can be used for active drug surveillance in the context of this cross-hazards model.

1.2. Outline

We presently sketch an outline of how QNMEDS operates (details are given by Algorithm 3 in §5.4). QNMEDS receives sequentially arriving patient data on the times of adverse events and

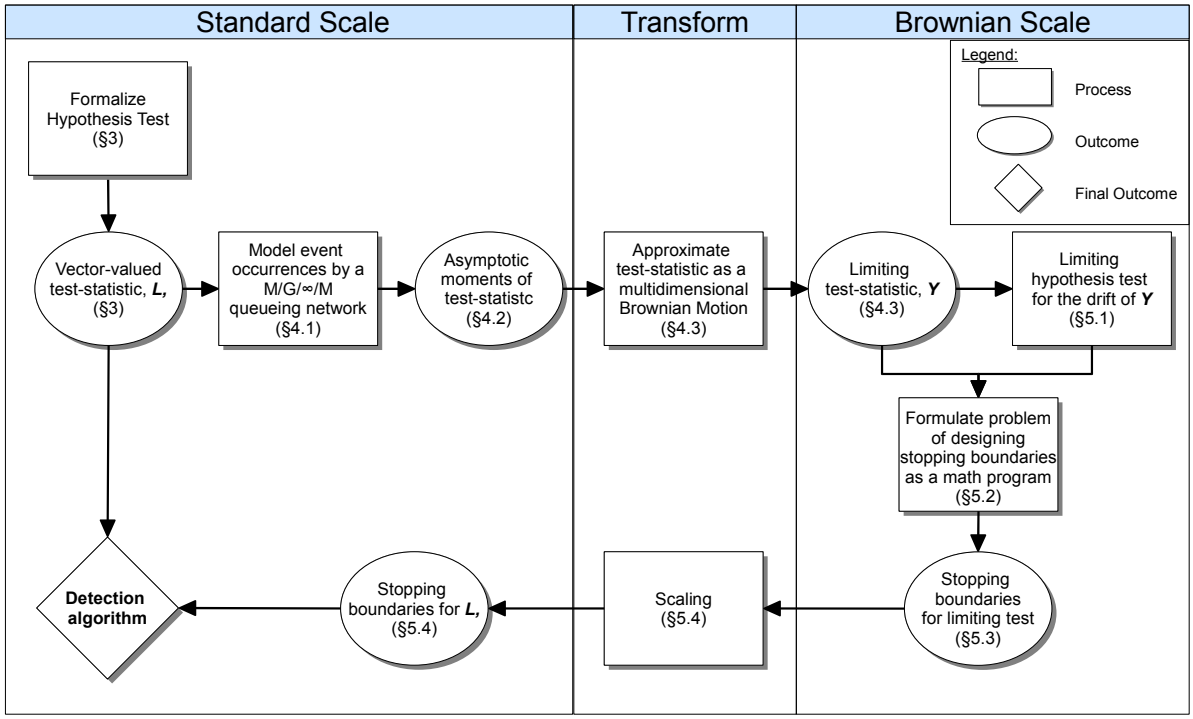

Figure 1 Illustrated overview of analytical steps in QNMEDS.

drug treatment, and uses these data to construct a certain vector-valued test-statistic, which is defined in §3. The components of this test-statistic capture the effect of the drug on each adverse event. If the test-statistic reaches a certain region of its state-space, called the *stopping region*, QNMEDS terminates and flags the drug as unsafe. The primary analysis of this paper focuses on how this stopping region is constructed in order to control the false detection rate and minimize the expected detection time.

Our analysis begins after a brief review of related methods in §2. Figure 1 illustrates the steps of our analysis, which occurs in three broad steps spanning §3 through §5. First, in §3, we describe the quantities that are to be statistically investigated, formalize a hypothesis test for these quantities, and define the test-statistic to be monitored. Second, in §4, we formulate a model of event occurrences in patients as a $M/G/\infty/M$ queueing network model with an arborescent (tree-like) structure. The queueing network models the dynamic nature of the surveillance system: patients arrive (i.e., join the surveillance system) and depart (i.e., leave the system due to death, migration, etc.) with time. Furthermore, this formulation allows us to apply results from queueing theory to characterize the first two asymptotic moments of our test-statistic (§4.2). Using this characterization, in §4.3, we proceed to apply standard convergence properties to show that our test-statistic weakly converges to a multidimensional Brownian Motion (B.M.) process as the arrival rate of

Is Feature Present ?	SPRT-based	Group sequential	Data-mining	QNMEDS
Temporal sequence of drug and adverse event	Y	Y	N	Y
Control of Type I error	Y	Y	N	Y
Robust to distributional assumptions	N	Y[a]	Y	Y
Can model multiple adverse events	N	Y[b]	Y[b]	Y
Can model multiple drugs	N	N	Y	N

[a] Depends on choice of method.

[b] Does not directly model interactions between adverse events.

Table 1 Comparison of features between our proposed method and related classes of methods.

patients into the system approaches infinity. Third, we consider a hypothesis test on the limiting test-statistic (§5.1), reformulating the problem of designing stopping boundaries for this test as a mathematical optimization problem (§5.2), and solve it (§5.3). We recover the stopping region for the original problem by rescaling and provide a summary of its implementation (§5.4, Algorithm 3). In QNMEDS, the queueing network formulation captures the interdependence between adverse events, the Brownian asymptotics confers the robustness to distributional assumptions, and the mathematical optimization allows us to introduce control parameters such as Type I errors.

In §6, we describe an example illustrating how QNMEDS can be applied to the setting of a dynamic form of Cox's proportional hazards model, in order to approximately test for the maximum hazard ratio of a drug on multiple adverse events. This setting is also used for the numerical studies described in §7. Finally, §8 concludes. Proofs of the results in the main paper are provided in Appendix A of the Electronic Companion.

2. Literature Review

We proceed to briefly review three classes of methods that can potentially be applied to the problem of active drug surveillance (for a comprehensive review, see Nelson et al. 2009). We discuss their strengths and limitations, as well as how QNMEDS integrates some of the strengths and overcomes some of the limitations of these other methods. Table 1 presents a summary of comparisons between QNMEDS and these other methods.

The first class of methods comprise the Sequential Probability Ratio Test (SPRT) by Wald (1945) and its variant, the maximized SPRT (maxSPRT) by Kulldorff et al. (2011). Both methods are general statistical inference procedures that are designed for sequential data arrival. They operate by continuously monitoring whether a test-statistic crosses certain stopping boundaries, which are designed to obtain prescribed Type I/II error rates for the test. The SPRT is a hypothesis test between two (simple) hypotheses, and has an appealing optimality property; namely, it has the smallest expected sample size of any test that has equal or less Type I/II errors (Wald and Wolfowitz 1948). The maxSPRT is an extension of the SPRT that endogenizes the parameter choice

in the alternative hypothesis, and has been applied in several studies on vaccine safety surveillance (e.g., Lieu et al. 2007, Yih et al. 2009, Klein et al. 2010, Kulldorff et al. 2011). Despite its strengths, the SPRT (and, by extension, the maxSPRT), suffers from two limitations. First, it is known to be very sensitive to distributional assumptions (e.g., Hauck and Keats 1997, Pandit and Gudaganavar 2010). This is because it is based on likelihood ratios and consequently requires the modeler to assume that the random occurrence times of adverse events follow certain distributions. A misspecification of the distribution can adversely affect its performance. Second, it is designed to test a single outcome, which in this application, is a single type of adverse event. It does not model multiple types of adverse events and their correlations, which are likely to be present in the large public health databases used for active surveillance.

The second class of methods comprise group sequential testing methods, which are reviewed in much detail by Jennison and Turnbull (1999). These methods are similar to the SPRT-based methods in that they are designed for sequential data arrival. However, they unlike the SPRT-based methods, which review the test-statistic continuously, group sequential methods review the test statistics at discrete time points. A common goal of these methods is to distribute the Type I error of the test across the review points, and different methods provide varying ways of achieving this (see, e.g., Pocock 1977, O'Brien and Fleming 1979, Lan and DeMets 1983). These methods are very flexible and can be designed to handle confounding variables. However, as in the case of SPRT-based methods, we are unaware of a group sequential test for multiple outcomes that have temporal dependencies as considered in this paper.

The third class of methods comprise data mining methods, such as the proportional reporting ratio (PRR) by Evans et al. (2001), the Bayesian confidence propagation neural network (BCPNN) by Bate et al. (1998), and the multi-item gamma Poisson shrinker (MGPS) by DuMouchel (1999). These methods are based on detecting drug-adverse event combinations that are disproportionately large compared to expected outcomes frequencies from within the database and were designed to be deployed on large datasets containing multiple drugs and multiple adverse events. While such methods do not explicitly incorporate the effect of confounding factors such as comorbidities, they are particularly amenable to subgroup analysis or stratification, which can reduce the problem of confounding factors. They typically do require some distributional assumptions, but these assumptions play a less critical role than in SPRT-based methods. However, these methods have the limitation that they were designed for hypothesis generation rather than hypothesis testing, and do not provide control over common parameters of statistical interest, such as Type I errors. A second limitation is that they were designed to find cross-sectional associations of drug-adverse event combinations, and do not incorporate considerations of temporal sequence (e.g., whether the adverse event occurred before or after the drug was taken).

QNMEDS incorporates the main strengths and overcomes the key limitations of each class of methods. Similar to the SPRT and group sequential methods, but unlike the data-mining methods, QNMEDS is based on the paradigm of sequential hypothesis testing, allows control over Type I errors, and furthermore is implicitly designed for sequentially-arriving data. Moreover, QNMEDS is designed to be robust to distributional assumptions. Finally, a unique strength of QNMEDS is that it handles not just multiple outcomes (unlike the SPRT methods), but also the temporal interdependence between these outcomes (unlike the group sequential and data-mining methods).

3. Problem Definition: Hypothesis Test Formulation

In this section, we formulate a hypothesis testing problem that determines whether a single drug increases the maximum incidence rate of a collection of adverse events beyond an exogenous threshold. We make the following assumptions in our formulation:

(A1) Patients arrive exogenously into the surveillance system (henceforth referred to as *the system*) according to a homogeneous Poisson process with a known rate.

(A2) Once in the system, patients experience the following *events*: treatment with the drug, adverse events, and departure from the system (i.e., death) at random times. Only the first occurrence time of each event is recorded for each patient. For brevity, we henceforth refer to these first occurrence times as simply the *event times*.

(A3) For each patient, the distribution of event times can depend on events that have previously occurred for the same patient.

(A4) Each patient's event times are distributed identically and independently.

Assumption (A1) is a standard assumption for arrival processes. Assumption (A2) does not limit our model because we can simply account for the second, third, etc. occurrence of the same (physcial) adverse event as separate events within our model (see §4.1 for a discussion of how this can be done). Assumption (A3) enables us to model the effect of the drug on adverse events, as well as the interdependence between adverse events. The first part of Assumption (A4), that patients are statistically identical, is reasonable if we stratify patients into relatively homogeneous socio-demographic groups with similar risk profiles, which is most feasible if the system contains many patients. The second part, that patients are statistically independent, is also reasonable since the drug-related adverse events monitored in a surveillance system are typically non-infectious conditions that do not interact between patients.

Deferring the full stochastic description of the system and these random times until §4, we proceed to define some notation and formalize our problem. Let $m \in \mathbb{N}$ represent the number of adverse events we are monitoring. Further, label the event of initiating drug treatment with index 1, and label the incidence of each adverse event with indices $j \in \mathcal{M} := \{2, \ldots, m+1\}$. For

completeness, label the event of patient arrival into the system as event 0 and departure from the system as event $m+2$.

Let λ represent the rate of (Poisson) patient arrival into the system. Also, let $S_j^{\mathbf{A}}(t)$ represent the set of patients that experienced adverse event $j \in \mathcal{M}$ *after* treatment by time t, and $S_j^{\mathbf{B}}(t)$ represent the set of patients who experienced adverse event $j \in \mathcal{M}$ *before* treatment by time t. In other words, for a given patient, if we let t_j represent the (random) occurrence time of adverse event j, and t_1 represent the (random) occurrence time of event 1 (drug treatment), then at time t, the patient belongs in set $S_j^{\mathbf{A}}(t)$ if $t_1 \le t_j \le t$, and the patient belongs in set $S_j^{\mathbf{B}}(t)$ if $t_j \le \min\{t_1, t\}$. Note that the definition of $S_j^{\mathbf{B}}(t)$ includes patients who have experienced adverse event j by time t, but have not received drug treatment by that time. For each $i \in \{A, B\}$, we refer to the monotonically increasing process $\left|S_j^i\right| := \left\{\left|S_j^i(t)\right|, t \ge 0\right\}$ as a *patient count process*, and define η_j^i as its (asymptotic) rate, normalized by the arrival rate λ. Formally,

$$\eta_j^i := \lim_{t\to\infty} \frac{1}{\lambda t} \mathbf{E}\left(\left|S_j^i(t)\right|\right) \qquad i \in \{A, B\}, j \in \mathcal{M} \tag{1}$$

The normalization by λ is not essential, but is done primarily for expositional and analytical convenience.

The parameters η_j^i, $i \in \{A, B\}, j \in \mathcal{M}$ are unknown and will be the subject of our statistical test. Our hypothesis test is

$$\begin{aligned} H_0 &: \eta_j^{\mathbf{A}} - k_j \eta_j^{\mathbf{B}} \le 0 \ \text{ for all } j \in \mathcal{M} \\ H_1 &: \eta_j^{\mathbf{A}} - k_j \eta_j^{\mathbf{B}} \ge \epsilon_j \text{ for any } j \in \mathcal{M}. \end{aligned} \tag{2}$$

where $k_j, \epsilon_j > 0$ are model parameters. Intuitively, (2) tests whether the incidence rate of any adverse event after taking the drug, $\eta_j^{\mathbf{A}}$, is higher than some fixed multiple (k_j) of the corresponding baseline rate $\eta_j^{\mathbf{B}}$ before taking the drug. For example, if $k_j = 1$ for all j, then (2) investigates whether, for any adverse event, the patients who took the drug have a higher incidence rate of that adverse event, compared to patients who did not take the drug. The parameters $\epsilon_j > 0$ represents tolerance levels for the test. In §6, we discuss a specific application that illustrates how these model parameters k_j, ϵ_j can be chosen in a principled way.

To implement this test, we monitor the $\mathbb{R}^m$-valued test-statistic process $\boldsymbol{L} := \{\boldsymbol{L}(t), t \ge 0\}$. Its jth component is the weighted difference of the number of patients who experienced adverse event j after and before treatment with the drug:

$$L_j(t) := \left|S_j^{\mathbf{A}}(t)\right| - k_j \left|S_j^{\mathbf{B}}(t)\right|. \tag{3}$$

Intuitively, if the alternative hypothesis H_1 in (2) is true, then, for some adverse event j, $L_j(t)$ should increase with t. We would then be able to decide between the two alternatives by setting an upper threshold on the values of $L_j(t)$, and choose to reject the null hypothesis once $L_j(t)$ exceeds

this threshold. Note that our choice of a unit weight on $\left|S_j^{\mathbf{A}}(t)\right|$ is without loss of generality, since it simply represents a choice of scale for the space of $\boldsymbol{L}$.

Our test-statistic was motived by several considerations. First, it is intuitive and simple to compute from data. We later show (Proposition 2) that under our model of event occurrences, the patient count processes $\left|S_j^{\mathbf{A}}\right|$ and $\left|S_j^{\mathbf{B}}\right|$ are independent for each j. Our proposed test-statistic process may be viewed as an analog of the test-statistic for the classical independent two-sample t-test, and is intuitively appealing. Second, it is not distribution-dependent, and the same test statistic applies, regardless of our distributional assumptions on various stochastic parameters. This is in contrast to the SPRT, which uses the likelihood ratio test statistic, and its performance is potentially very sensitive to the choice of the assumed distribution. This is verified in our numerical study in §7. Third, it is motivated by the structure of the optimal likelihood ratio test-statistics for a collection of simpler tests (discussed in §6.4) where event times are assumed to be exponentially distributed.

4. Modeling Event Occurrences by a M/G/∞/M Queueing Network

In this section, we describe our stochastic model of how events occur to patients. Specifically, we define an arborescent (tree-like) $M/G/\infty/M$ queueing network, which we formally construct in §4.1, and use it to model event occurrence in patients. This queueing network formulation equips us with useful analytical tools to study stochastic properties of the patient count processes, and therefore, our proposed test-statistic.

Our development in this section, illustrated in Figure 2, proceeds in three steps. First, in §4.1, we provide a formal description of this queueing network, and show how it is used to represent event occurrences in patients. Second, in §4.2, we characterize the patient count processes $\left|S_j^{\mathbf{A}}\right|$ and $\left|S_j^{\mathbf{B}}\right|$ as sums of flow processes of the network, and use this characterization to derive expressions for their first two moments. Third, in §4.3, we describe an asymptotic regime and the test-statistic's limiting distribution in this regime. Our focus throughout this section is to describe the system and derive the relationships between quantities. Hence, for now at least, it is useful to think of all the parameters of the queueing model as fully known. We return to the problem of testing for unknown model parameters in §5.

4.1. Description of the Queueing Network

To better convey the underlying intuition, we first describe the structure of the queueing network for $m = 2$ adverse events (illustrated in Figure 3). Each node represents an ordered collection of events that can occur to patients. For example, in Figure 3, if a patient is at node 21, at some time t_0, it means that he has already experienced event 2 and event 1, in that order, by time t_0. Suppose

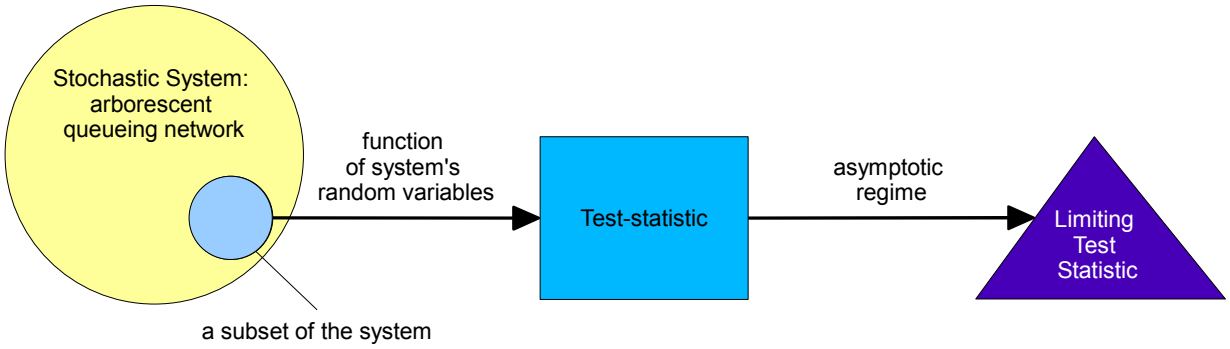

Figure 2 Illustration of steps of our development: 1) describe stochastics of queueing network, 2) express patient count processes in terms of flow processes of network, 3) describe asymptotic properties and distribution of the test statistic.

that at time $t_1 > t_0$, he experiences event 3. In the network, this is represented by him transitioning to node 213 at time t_1. Suppose at time $t_2 > t_1$, he departs the system. This is represented by him transitioning out of the system at time t_2.

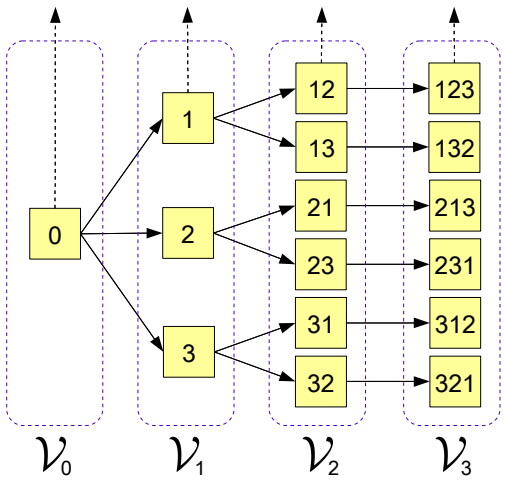

Figure 3 Illustration of queueing network for $m = 2$ adverse events, depicting the partition of the set of nodes, $\mathcal{V}$ into the sets $\mathcal{V}_k$. Upward arrows represent departures from the system. Not all departures are illustrated.

For general m, the nodes of the graph are constructed as follows. Define $\mathcal{V}_0 := \{0\}$ and for $k \in \{1, \ldots, m+1\}$, define $\mathcal{V}_k$ as the set of all k-tuples with *distinct* elements from $\{1, \ldots, m+1\}$. Then, the set of nodes of the graph are $\mathcal{V} := \biguplus_{k=0}^{m+1} \mathcal{V}_k$, where $\biguplus$ represents a disjoint union. For a node $v \in \mathcal{V}_k$, we refer to k as the *length* of node v and write $\text{len}(v) = k$. Moreover, for $v := (v_1, \ldots, v_k)$, we also use the concise representation $v = v_1 v_2 \ldots v_k$, and we let $v(j) := v_j$ denote its jth component.

Next, we describe the edges of the graph. Two nodes $u, v \in \mathcal{V}$, are joined by an edge iff $\text{len}(v) = \text{len}(u) + 1$ and $v(k) = u(k)$ for all $1 \leq k \leq \text{len}(u)$. Moreover, this edge points toward v. By construction, it is clear that for any node $v \in \mathcal{V} \setminus \{0\}$, there exists a unique directed path from 0 to v. Consequently, the graph is arborescent with node 0 as its root, and it makes sense to talk about the parent-child relationship between nodes. Specifically, for nodes u, v connected by an edge that is directed to v, we term v the child node and u the parent node.

Quantity of Interest	Queueing Network Analog
Incidence of an event	Arrival at a node
Time until the next event	Waiting time at a node
Probability that some event happens next	Routing probability to a child node

Table 2 Quantities of interest and their queueing network analogs.

Table 2 summarizes various quantities of interest and their analogs in the queueing network formulation. Since each node of the network represents an ordered collection of events that occurs to patients, the temporal interdependence between events is modeled by appropriate assignment of waiting time distributions at each node. For example, for the network illustrated in Figure 3, to capture the idea that event 3 increases the occurrence rate of event 2, the waiting time distributions for the network could be constructed such that the waiting time at node 0, conditional on being routed to node 2, is stochastically larger than the waiting time at node 3, conditional on being routed to node 32. In §6, we present an example of how waiting times for the network can be *explicitly* constructed in order to model a very natural type of interdependence between events.

As we noted previously, our model can be used to capture multiple occurrences of the same (physical) adverse event. For example, for the network illustrated in Figure 3, we can let event 2 represent the first occurrence of a certain adverse event, and event 3 represent the second occurrence of the the adverse event. We can assign routing probabilities to the network so that event 3 does not occur before event 2, (i.e., the edges from node 0 to node 3 and from node 1 to node 13 have zero probability). Moreover, any dependence between the first and second occurrences of the adverse event can be modeled through the waiting time distributions at each node.

At each node of the network we assume a general integrable waiting time distribution and independent stationary Markovian routing. For any node $v := (v_1, \ldots, v_k) \in \mathcal{V}$, we denote by p_v the *routing probability* into node v from its parent, with $p_0 := 1$ for completeness. Also, we define π_v as the *total routing probability* from node 0 to node v, which is the product of the individual routing probabilities along the unique directed path from 0 to v.

A useful stochastic process for our subsequent discussion is the process that counts the number of patients that have ever visited each node. This is defined below.

DEFINITION 1. For any node $v \in \mathcal{V}$, denote the *cumulative arrival process* into v as $A_v := \{A_v(t), t \geq 0\}$. Further, denote the *expected number of cumulative arrivals* as $\Lambda_v(t) := \mathbf{E}(A_v(t))$.

It can be shown that each A_v process is a nonhomogeneous Poisson process. This follows from the arborescent structure of the network, Poisson thinning, and Theorem 1 by Eick et al. (1993), which states that the departure processes of an $M_t/G/\infty$ queue with nonhomogeneous Poisson input is also nonhomogeneous Poisson. A useful long run asymptotic property of the expected number of cumulative arrivals, $\Lambda_v(t)$, is established in the following lemma.

LEMMA 1. *The following asymptotic relationship holds for each node $v \in \mathcal{V}$.*

$$\lim_{t\to\infty} \frac{1}{t}\Lambda_v(t) = \lambda\pi_v. \tag{4}$$

4.2. Distribution and Asymptotic Moments of Patient Count Processes

We proceed to characterize the distribution of the patient count processes $\left|S_j^i\right|$, for each $i \in \{A, B\}, j \in \mathcal{M}$, and derive expressions for its first two asymptotic moments (means, covariance matrix) in terms of the routing probabilities of the queueing network. These derivations are intermediate steps in order to characterize an approximate distribution for the test-statistic process $\boldsymbol{L}$ in §4.3. Recall from (1), the asymptotic mean is represented by the symbol η_j^i, and analogously, we define the asymptotic covariance matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{2m\times 2m}$ between the patient count processes, with components

$$\Sigma(i,j,i',j') := \lim_{t\to\infty} \frac{1}{\lambda t}\mathrm{Cov}\left(\left|S_j^i(t)\right|, \left|S_{j'}^{i'}(t)\right|\right) \qquad i,i' \in \{A,B\}, j,j' \in \mathcal{M}. \tag{5}$$

Our derivation relies on observing that $\left|S_j^i\right|$ can be decomposed into the sum of mutually independent patient arrival processes into a set of nodes, V_j^i (defined below) which furthermore implies (by Poisson superposition) that $\left|S_j^i\right|$ is a nonhomogeneous Poisson process. This is established in Proposition 1, which follows.

PROPOSITION 1. *Fixing $i \in \{A, B\}$ and $j \in \mathcal{M}$, construct the set $V_j^i \subseteq \mathcal{V}$ as $V_j^i := \biguplus_{k=1}^{m+1} V_j^i(k)$, where*

$$V_j^{\mathbf{A}}(k) := \{v \in \mathcal{V}_k : v(k) = j \text{ and } \exists k' < k, v(k') = 1\},$$
$$V_j^{\mathbf{B}}(k) := \{v \in \mathcal{V}_k : v(k) = j \text{ and } v(k') \neq 1 \ \forall k' < k\}.$$

Then, $\left|S_j^i(t)\right|$ has representation $\left|S_j^i(t)\right| = \sum_{v\in V_j^i} A_v(t)$, where the arrival processes in the set $\{A_v\}_{v\in V_j^i}$ are mutually independent.

In Proposition 1, the set $V_j^{\mathbf{A}}(k)$ is interpreted as the collection of all nodes of length k, for which the adverse event j occurs *after* drug treatment. Similarly, the set $V_j^{\mathbf{B}}(k)$ is interpreted as the collection of all nodes of length k, for which the adverse event j occurs *before* drug treatment (including the case that treatment has not occurred).

Applying Proposition 1, we can establish independence between the patient count processes $\left|S_j^{\mathbf{A}}\right|$ and $\left|S_j^{\mathbf{B}}\right|$, as well as derive expressions for their first two asymptotic moments. These are stated in the following two propositions.

PROPOSITION 2. *For each adverse event $j \in \mathcal{M}$, the processes $\left|S_j^{\mathbf{A}}\right|$ and $\left|S_j^{\mathbf{B}}\right|$ are independent.*

PROPOSITION 3. *For any $i \in \{A, B\}$ and $j \in \mathcal{M}$, we have*

$$\eta_j^i = \sum_{v \in V_j^i} \pi_v \qquad i \in \{A, B\}, j \in \mathcal{M}. \tag{6}$$

Also, for $i, i' \in \{A, B\}$ and $j, j' \in \mathcal{M}$, there exists a set $\mathcal{T}(i, j, i', j')$, with $T(i, j, i, j) = V_j^i$, such that

$$\Sigma(i, j, i', j') = \sum_{v \in \mathcal{T}(i,j,i',j')} \pi_v \qquad i, i' \in \{A, B\}, j, j' \in \mathcal{M}. \tag{7}$$

In Appendix 8, we demonstrate the explicit computation of the sets $\mathcal{T}(i, j, i', j')$ for $m = 2$ adverse events. Algorithm 1 details how the sets $\mathcal{T}(i, j, i', j')$ are constructed for general m. It uses the **commmonroot** subroutine provided in Algorithm 2.

Algorithm 1 Procedure to compute $\mathcal{T} := \mathcal{T}(i, j, i', j')$.

Require: $(i, j), (i', j') \in \{A, B\} \times \mathcal{M}$.

Compute the sets V_j^i, $V_{j'}^{i'}$ from their definitions in Proposition 1.
$\mathcal{T} \leftarrow \emptyset$.
for all Nodes $v \in V_i^j$ and $v' \in V_{i'}^{j'}$ **do**
    $\mathcal{T} \leftarrow \mathcal{T} \cup$ **commonroot**(v, v')
end for

Algorithm 2 The **commonroot** subroutine used in Algorithm 1

Require: Two nodes v and v'. **Assumes:** $\text{len}(v) \leq \text{len}(v')$.

for $k = 1$ **to** $\text{len}(v)$ **do**
    if $v(k) \neq v'(k)$ **then**
        return $\emptyset$
    end if
end for
return v'

4.3. Asymptotic Distribution of the Test Statistic Process

We have already characterized the first two moments of the patient count processes and the test-statistic process. However, exact analysis with the test statistic process remains difficult, because we were not able to explicitly characterize its distribution. Therefore, we pursue an approximate analysis in a *Large System Regime*, defined below, where the arrival rate of patients into the system goes to infinity.

The underlying intuition of this approximation is that our actual system may be viewed as a time-scaled version of a hypothetical system with a unit arrival rate. Suppose that we can evaluate the asymptotic distribution of the test-statistic for the hypothetical system, in the limit as time and space are both appropriately scaled. Then we can use this asymptotic distribution as an approximate distribution for the test-statistic in the actual system, after a suitable re-scaling. In the remainder of this section, we proceed to show that in this asymptotic regime, the test-statistic process for the normalized system weakly converges to a multidimensional B.M.. In §5, we will use this limiting test-statistic process to derive stopping boundaries, and also show how to map them into stopping boundaries for the original test-statistic.

Let us first establish some notation. We use a "hat", i.e. the symbol $\widehat{\cdot}$, to represent normalized processes under a *unit arrival rate*. Formally, we have $\widehat{L}_j(t) := L_j(t/\lambda), \left|\widehat{S}^i_j(t)\right| := \left|S^i_j(t/\lambda)\right|$ for each adverse event $j \in \mathcal{M}$ and $\widehat{\Lambda}_v(t) := \Lambda_v(t/\lambda)$ for each node $v \in \mathcal{V}$ of the queueing network. Note that all the results thus far also extend to the "hatted" symbols as well.

Consider a sequence of systems indexed by λ, such that the λth system has arrival rate λ. We associate with the λth system a *scaled test-statistic process* $\widehat{\boldsymbol{L}}^{(\lambda)}$ and *asymptotic rate parameters* $\eta_j^{\mathbf{A},\lambda}$ and $\eta_j^{\mathbf{B},\lambda}$ (these play the role of $\eta_j^{\mathbf{A}}, \eta_j^{\mathbf{B}}$ in our original system), and where $\widehat{\boldsymbol{L}}^{(\lambda)}$ is a time and space-scaled version of $\widehat{\boldsymbol{L}}$, which is explicitly given by

$$\widehat{\boldsymbol{L}}^{(\lambda)} := \left\{\widehat{\boldsymbol{L}}^{(\lambda)}(t) : t \geq 0\right\} \qquad \text{with} \qquad \widehat{\boldsymbol{L}}^{(\lambda)}(t) := \lambda^{-1/2}\widehat{\boldsymbol{L}}(\lambda t),$$

and we assume that $\eta_j^{\mathbf{A},\lambda}$, $\eta_j^{\mathbf{B},\lambda}$ vary with λ in the following *Large System Regime*:

$$\lambda \to \infty, \text{ and} \tag{8a}$$

$$\lim_{\lambda\to\infty} \sqrt{\lambda}\left(\eta_j^{\mathbf{A},\lambda} - k_j\eta_j^{\mathbf{B},\lambda}\right) = c_j, \tag{8b}$$

for some $c_j \in \mathbb{R}$, for all $j \in \mathcal{M}$.

The first part (8a) is an assumption about the true parameter value of λ, and is most appropriate when the true value of λ is large relative to the service rates of the queueing network. This is likely to be true in practice, since a "service" in our context represents an experience of an adverse event, which should occur much more infrequently than arrivals of patients into the system. The

second part (8b) is an assumption about how the parameters $\eta_j^{\mathbf{A},\lambda}$, $\eta_j^{\mathbf{B},\lambda}$ vary with λ. Since all we require is that c_j exists, and do not make any restrictions on its sign or magnitude, (8b) will hold without any further data requirement. Instead, (8b) will play a role in the design of our limiting test, a discussion that we defer until §5. For now, we apply these constructs to establish that $\widehat{\boldsymbol{L}}$ is distributed asymptotically as a multidimensional B.M..

PROPOSITION 4. *Let* $\boldsymbol{c} := (c_j)_{j\in\mathcal{M}}$ *be defined through assumption (8b), and further define* $\boldsymbol{Q} := \boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{V}^T$, *where:*

1. *The covariance matrix* $\boldsymbol{\Sigma}$ *is defined as in (5),*
2. *The* $m \times 2m$ *matrix* $\boldsymbol{V}$ *comprises two diagonal blocks,* $\boldsymbol{V} := [\boldsymbol{I}, -\boldsymbol{K}]$, *and*
3. $\boldsymbol{K}$ *is a diagonal matrix with diagonal entries* $K_{jj} := k_j$ *for each* $j \in \mathcal{M}$.

Let $\boldsymbol{Y} := \{\boldsymbol{Y}(t), t \geq 0\}$ *represent a* ($\boldsymbol{c}$, $\boldsymbol{Q}$) *multidimensional B.M.. Then,* $\widehat{\boldsymbol{L}}^{(\lambda)}$ *converges weakly to* $\boldsymbol{Y}$ *as* $\lambda \to \infty$.

The multi-dimensional B.M., $\boldsymbol{Y}$, defined in Proposition 4, will be termed the *limiting test-statistic process*. From the same proposition, we observe that $\boldsymbol{Y}$ admits the decomposition $\boldsymbol{Y}(t) = \boldsymbol{c}t + \boldsymbol{Q}^{1/2}\boldsymbol{W}(t)$, where $\boldsymbol{W} := \{\boldsymbol{W}(t), t \geq 0\}$ is a standard m-dimensional B.M., and $\boldsymbol{Q}^{1/2}$ is a decomposition of $\boldsymbol{Q}$ such that $\boldsymbol{Q}^{1/2}(\boldsymbol{Q}^{1/2})^T = \boldsymbol{Q}$.

5. Optimal Boundary Design for the Limit Problem

We now propose a sequential hypothesis test (called the *limiting test*) for the parameters of the limiting test-statistic process $\boldsymbol{Y}$, and show that this limiting test approximates the hypothesis test (2). After defining the limiting test in §5.1, we show in §5.2 how the problem of designing stopping boundaries for this limiting test can be posed as a mathematical optimization problem, and proceed to solve this problem in §5.3. The section concludes with §5.4, which summarizes the results of our analysis and concisely describes how our drug surveillance method is implemented.

5.1. Limiting Hypothesis Test on the Drift of $\boldsymbol{Y}$

We propose the limiting test

$$\begin{array}{l} H_0 : c_j \leq 0 \ \text{ for } \textit{all } j \in \mathcal{M}, \\ H_1 : c_j \geq \bar{c}_j \text{ for } \textit{any } j \in \mathcal{M}, \end{array} \tag{9}$$

where the constants $\bar{c}_j$ in the alternate hypothesis are defined as

$$\bar{c}_j := \sqrt{\lambda}\epsilon_j. \tag{10}$$

This test is an appropriate approximation of (2) as a result of the following lemma, which implies that in the limit as $\lambda \to \infty$, a correct choice between the null and the alternative on the limiting test (9) leads to a correct choice between the null and the alternative in the hypothesis test (2).

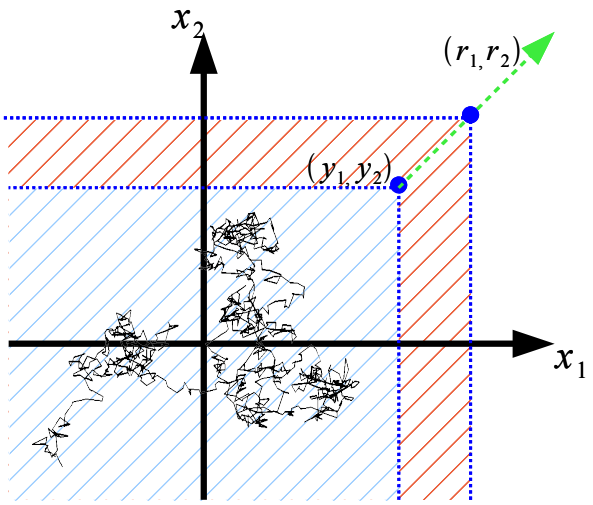

Figure 4 Sample path of Brownian motion for $m = 2$ and moving continuation region $\widetilde{\mathcal{C}}(t)$, which starts at point (y_1, y_2) and drifts at rate (r_1, r_2) per unit time.

LEMMA 2. *For each $j \in \mathcal{M}$, let c_j satisfy (8b). Then,*

1. *If $c_j < 0$ for all $j \in \mathcal{M}$, then $\eta_j^{\mathbf{A},\lambda} - k_j \eta_j^{\mathbf{B},\lambda} < 0$ for all large enough λ, for all $j \in \mathcal{M}$.*
2. *If for some $j^* \in \mathcal{M}$, $c_{j^*} > \bar{c}_{j^*}$, then $\eta_j^{\mathbf{A},\lambda} - k_j \eta_j^{\mathbf{B},\lambda} > \epsilon_j$ for all large enough λ.*

5.2. Boundary Design as an Optimization Problem

For some $\boldsymbol{r}, \boldsymbol{y} \in \mathbb{R}_+^m$, we consider a moving continuation region $\widetilde{\mathcal{C}}(t) := \{\boldsymbol{x} \in \mathbb{R}^m : \boldsymbol{x} \le \boldsymbol{r}t + \boldsymbol{y}\}$ (the vector inequality denotes a component-wise inequality), where we stop sampling and reject H_0 once $\boldsymbol{Y}(t) \notin \widetilde{\mathcal{C}}(t)$. This continuation region is illustrated in Figure 4 for $m = 2$ dimensions.

Our goal is to choose parameters $\boldsymbol{r}$ and $\boldsymbol{y}$ that are "optimal" in a sense that will be formally defined. We will do this in 3 steps: First, we apply a change of axes to transform the moving continuation region to a fixed continuation region. Second, we relate the decision parameters $\boldsymbol{r}, \boldsymbol{y}$ to testing parameters such as the Type I/II errors and expected detection time. Third, we formulate a mathematical optimization problem to choose $\boldsymbol{r}$ and $\boldsymbol{y}$.

Step 1: Change of axes. We use a *fixed continuation region* $\mathcal{C} := \{\boldsymbol{x} \in \mathbb{R}^m : \boldsymbol{x} \le \boldsymbol{y}\}$ instead of the moving continuation region $\widetilde{\mathcal{C}}(t)$, and simultaneously subtract $\boldsymbol{r}t$ from our limiting test-statistic process $\boldsymbol{Y}$. Our problem is therefore to choose optimal boundary parameters $\boldsymbol{r}$ and $\boldsymbol{y}$ that define the fixed continuation region $\mathcal{C}$, and test-statistic, $\boldsymbol{Z}(t) := \boldsymbol{Y}(t) - \boldsymbol{r}t = (\boldsymbol{c} - \boldsymbol{r})t + \boldsymbol{Q}^{1/2}\boldsymbol{W}(t)$. We stop sampling and reject H_0 once $\boldsymbol{Z}(t) \notin \mathcal{C}$. This step is not strictly necessary, but will simplify our subsequent analysis.

Step 2: Relate decisions to testing parameters. We seek a stopping rule that has a controlled false detection probability and a minimized true detection time. Specializing this definition to our

current context, we will design the test to have a Type I error below an exogenous parameter $\alpha \in (0,1)$, a Type II error of zero, and a minimized expected detection time under the alternative hypothesis, H_1, of (9). Lemma 3, which follows, relates the choice of boundary parameters $\boldsymbol{r}, \boldsymbol{y}$ to the Type I/II errors and expected time of rejection under H_1 for the test.

LEMMA 3. *For the sequential test (9), suppose that $r_j < \overline{c}_j$ for all j. Further, let $\boldsymbol{S}$ be any diagonal matrix such that $\boldsymbol{S} \succeq \boldsymbol{Q}$, and let its corresponding diagonal entries be represented by $\sigma_j^2 := S_{jj}$. Then, the Type II error is zero, the worst-case expected time of rejection under H_1 is given by $\max_{j\in\mathcal{M}} \frac{y_j}{\overline{c}_j - r_j}$, and the Type I error is bounded above by $1 - \prod_{j\in\mathcal{M}} \left(1 - \exp\left(-\frac{2r_j y_j}{\sigma_j^2}\right)\right)$.*

One way to choose $\boldsymbol{S}$ is to solve the following semidefinite programming (SDP) problem:

$$\begin{aligned} \min_{\boldsymbol{S}} \ & \sum_{j\in\mathcal{M}} S_{jj} \\ \text{s.t. } & \boldsymbol{S} \succeq \boldsymbol{Q} \\ & \boldsymbol{S} \text{ is diagonal.} \end{aligned} \tag{11}$$

Since we can interpret the positive semidefinite matrix $\boldsymbol{Q}$ as representing an ellipsoid, intuitively, problem (11) finds an ellipsoid (represented by $\boldsymbol{S}$) that is aligned to the principal coordinate axes and that covers the ellipsoid represented by $\boldsymbol{Q}$. The objective minimizes the sum of eigenvalues of $\boldsymbol{S}$, and intuitively finds a "small" covering ellipsoid. Other objectives can be used as well (e.g., other matrix norms, product of eigenvalues).

Step 3: Formulate Mathematical Optimization Problem. Lemma 3 implies that to compute the optimal boundary parameters for the test (9), we need to solve the following optimization problem.

$$\begin{aligned} \min_{\boldsymbol{r},\boldsymbol{y}} \ & \max_{j\in\mathcal{M}} \frac{y_j}{\overline{c}_j - r_j} \\ \text{s.t. } & \prod_{j\in\mathcal{M}} \left(1 - \exp\left(-\frac{2r_j y_j}{\sigma_j^2}\right)\right) \geq 1 - \alpha, \\ & \boldsymbol{y} \geq \boldsymbol{0} \\ & 0 \leq r_j < \overline{c}_j \quad j \in \mathcal{M}. \end{aligned} \tag{12}$$

In problem (12), we note that the objective is exactly the worst-case expected detection time under H_1 and the first constraint is a statement that the Type I error should not exceed α. Also, in the LHS of the same constraint, the σ_j parameters are completely determined by problem data, and obtained by the solution of the SDP (11).

5.3. Solution

We now proceed to solve (12).

PROPOSITION 5. *The optimal solution to (12) is*

$$(r_j^*, y_j^*) = \left(\frac{\overline{c}_j}{2}, \frac{\overline{c}_j}{2} s^*\right) \qquad j \in \mathcal{M}.$$

where s^* solves the transcendental equation

$$\prod_{j\in\mathcal{M}}\left(1-\exp\left(-\frac{\bar{c}_j^2 s^*}{2\sigma_j^2}\right)\right)=1-\alpha, \tag{13}$$

and is also the optimal value of (12), which represents the expected detection time under H_1.

Proposition 5 allows us to analyze certain comparative statics, summarized in the following proposition.

PROPOSITION 6. *The expected detection time, s^*, increases as σ_j increases for any j. The expected detection time also increases as $m=|\mathcal{M}|$, the number of effects being monitored, increases.*

The optimal value of $\boldsymbol{r}^*$ from Proposition 5 is quite intuitive. Recall that $\boldsymbol{r}$ represents the additional negative drift imparted to the B.M. test statistic. Since we are distinguishing between the case of zero drift and a maximum drift of $\bar{c}_j$ for the jth component, it makes sense that we would impart the B.M. with an additional negative drift that is between the two extremes. The conclusions of Proposition 6, however, might seem rather unintuitive at first glance. As the system experiences more volatility (higher σ_j) or if there are more adverse events monitored (higher m), one might expect the test-statistic process to fluctuate more wildly, and consequently exit the continuation region sooner. This intuition would be true if the continuation region remained unchanged. However, the continuation region does change as σ_j and m changes. In particular, it changes to maintain the Type I error at α. As the proposition shows, the latter effect is in fact dominant, and the expected time to detection increases as either σ_j or m increases.

5.4. Implementation

In practice, the sequential test can be implemented by a simple algorithm, which updates the test statistic at prespecified discrete time points, $\{t^{(n)}\}$, with $t^{(0)}:=0$ and $t^{(n)}\le t^{(n+1)}$. This algorithm is described in Algorithm 3.

6. Application: Test for Hazards Ratio

In this section, we apply our surveillance method to a concrete example of practical interest. In epidemiological studies, the *hazard ratio* is a common measure of the incremental risk associated with a particular drug or treatment. We will demonstrate how QNMEDS can be used to approximately test if the hazard ratio of a drug on any of a collection of adverse events is higher than a pre-specified clinical threshold. We do this in four steps: First, in §6.1, we introduce a specific statistical model (termed the *cross-hazards model*) of event occurrences in patients. Second, in §6.2, we show that the cross-hazards model is a special case of our general queueing network formulation. Third, in §6.3, we formally state the test for the hazard ratios and describe how, by an

Algorithm 3 Sequential Detection Algorithm for Multiple Adverse Events

Require: Sequentially-arriving occurrence times of adverse events and times of drug treatment.

1: For each $j \in \mathcal{M}$, set $\overline{c}_j$ by (10), as

$$\overline{c}_j \leftarrow \sqrt{\lambda}\epsilon_j.$$

2: Compute the covariance matrix $\boldsymbol{Q}$ from Proposition 4 and by Algorithm 1.

3: Obtain a diagonal matrix $\boldsymbol{S}$ such that $\boldsymbol{S} \succeq \boldsymbol{Q}$ by solving the SDP (11). Set $\sigma_j^2 \leftarrow S_{jj}$.

4: $r_j \leftarrow \overline{c}_j/2$, $y_j \leftarrow \overline{c}_j s^*/2$ where s^* solves the transcendental equation

$$\prod_{j\in\mathcal{M}} \left(1 - \exp\left(-\frac{\overline{c}_j^2 s^*}{2\sigma_j^2}\right)\right) = 1 - \alpha.$$

5: Initialize $n \leftarrow 0$,

6: **repeat**

7: $n \leftarrow n + 1$.

8: Compute the test statistic $\boldsymbol{L} := (L_j)_{j\in\mathcal{M}}$ as

$$L_j = \left|S_j^{\mathbf{A}}(t^{(n)})\right| - k_j \left|S_j^{\mathbf{B}}(t^{(n)})\right| \quad \forall j \in \mathcal{M}.$$

9: **until** $L_j > r_j t^{(n)}\sqrt{\lambda} + y_j\sqrt{\lambda}$ for some $j \in \mathcal{M}$

10: Stop the test and reject H_0.

appropriate choice of parameters k_j and ϵ_j, the hypothesis test (2) can be used as an approximate test for the hazard ratio. Fourth, in §6.4, we provide additional motivation for the structure of the test-statistic process.

6.1. Description of Cross-hazards Model of Event Occurrences

We proceed to describe our cross-hazards model, which captures the interdependence between occurrence times of events. Our model assumes that past events have a multiplicative effect on the hazard rate of future events, through a *cross-hazard matrix* $\boldsymbol{\Phi} \in \mathbb{R}_+^{(m+1)\times(m+2)}$. The cross-hazards matrix has components $\phi_\ell(j) := \phi(\ell, j)$, which represents the fractional increase in hazard rate of event j after event ℓ has occurred, or more concisely, the *hazard ratio* (HR) of event j due to event ℓ.

Consider a generic patient in the system at some time $t_0 \geq 0$ just after the occurrence of some event in $\{0, \ldots, m+1\}$. For this patient, let $\mathcal{A} \subseteq \{0, \ldots, m+1\}$ represent the set of events that has *already* occurred, and $\mathcal{Y} \subseteq \{1, \ldots, m+2\}$ the set of events that has *yet* to occur. For any non-negative r.v. T with a well-defined density, its *hazard rate function* is also well-defined, and we

denote it by h_T. We model the time to the next event, τ, as the r.v. $\tau := \min_{j\in\mathcal{Y}} T_j$, where $\{T_j\}_{j\in\mathcal{Y}}$ are mutually independent r.v.s, and each T_j is distributed such that $h_{T_j}(t)$, is given by

$$h_{T_j}(t) = h_{T_j^{\mathbf{base}}}(t) \prod_{\ell\in\mathcal{A}} \phi_\ell(j) \quad \forall t \geq 0 \tag{14}$$

where $T_j^{\mathbf{base}}$ is a non-negative r.v. that represents the random occurrence time of each event, in the absence of previously-occurring events.

Our model makes some technical assumptions about the collection of r.v.s, $\left\{T_j^{\mathbf{base}}\right\}_{j=1}^{m+2}$, which requires the following definition.

Definition 2. For a set $\mathcal{J}$ of non-negative r.v.s $\{T_j\}_{j\in\mathcal{J}}$, we say that they satisfy a *proportional hazards condition* if there exists positive constants $\{\alpha_j\}_{j\in\mathcal{J}}$, not all zero, and a function $g : [0,\infty) \to \mathbb{R}_+$ such that

$$h_{T_j}(t) = \alpha_j g(t) \qquad t \geq 0, j \in \mathcal{J}. \tag{15}$$

We assume that the r.v.s, $\left\{T_j^{\mathbf{base}}\right\}_{j=1}^{m+2}$, (A) are independent and (B) satisfy a proportional hazards condition with constants $\left\{\alpha_j^{\mathbf{base}}\right\}_{j=1}^{m+2}$. Note that (B) is not overly restrictive and encompasses standard distributions such as the exponential, Weibull, and Pareto distributions. Finally, we also assume that the T_j r.v.s constructed via (14) are all integrable. This is a regularity condition for event times, and holds without further qualification for a variety of distributions of $T_j^{\mathbf{base}}$ (e.g., Weibull).

Our cross-hazard model (14) of interdependence between adverse effects was chosen for several reasons. First, from a practical standpoint, the hazard ratio is a well-established reporting metric in the medical literature, especially for empirical studies that analyze risk factors for diseases (e.g. Frasure-Smith et al. 1993, Haffner et al. 1998, Luchsinger et al. 2001). Consequently, such data can be estimated with reasonable accuracy from existing medical studies. For two events ℓ and j for which no association has been conclusively established, a hazard ratio of unity can be used to model their independence. Second, from a theoretical perspective, our cross-hazards model is a form of Cox's proportional hazards model (Cox 1972). In our model, the $\phi_\ell(j)$ parameters for past events ℓ are exactly the multiplicative factors in Cox's model for the hazard rate of event j. Consequently, our model may be viewed as a dynamic form of Cox's model, with sequentially-updated multiplicative factors as events occur to patients (we refer readers to Appendix D for a more detailed description of Cox's model and how our model compares with it). It is precisely the dynamic updating of multiplicative factors in our model that makes the standard Cox regression unsuitable for our model. We verify this point in our numerical study. Finally, as will be shown in the next subsection, our cross-hazards model is a special case of the arborescent $M/G/\infty/M$ queueing network model discussed earlier, will allow us to employ the analytic tools that we have developed to perform hypothesis testing.

6.2. Cross-Hazards Model as a Queueing Network

The cross-hazards model of event occurrences is a special case of the $M/G/\infty/M$ network that we described in §4. The following proposition establishes the main analytical step for this result.

PROPOSITION 7. *Let $t_0 \geq 0$ be the occurrence time of any event in $\{0,\ldots,m+1\}$, and $\mathcal{Y} \subseteq \{1,\ldots,m+2\}$ be the set of events that have yet to occur by t_0. Then, the collection $\{T_j\}_{j\in\mathcal{Y}}$, defined through (14), satisfies a proportional hazards condition for some constants $\{\alpha_j\}_{j\in\mathcal{Y}}$, defined by $\alpha_j := \alpha_j^{\mathbf{base}} \prod_{\ell\notin\mathcal{Y}} \phi_\ell(j)$ for each $j \in \mathcal{Y}$. Moreover, the time until the next event, $\tau := \min_{j\in\mathcal{Y}} T_j$, satisfies*

$$\mathbf{P}\left(\tau = T_j, \tau > t\right) = q_j \mathbf{P}\left(\tau > t\right), \tag{16}$$

where $q_j := \alpha_j / \sum_{\ell\in\mathcal{Y}} \alpha_\ell$.

Suppose that an event occurs to a given patient at time $t_0 \geq 0$. In the equivalent network formulation, this is represented by the patient arriving at some node at time t_0. Proposition 7 shows that the probability that an event $j \in \mathcal{Y}$ is the next to occur is constant in time and independent of the distribution of τ, the time until the next event. This exactly fits the probabilistic description of a node in our $M/G/\infty/M$ queueing network, where τ represents the service time at the node and $\{q_j\}_{j\in\mathcal{Y}}$ the routing probabilities after service.

In particular, Proposition 7 shows that for a node $v = (v_1,\ldots,v_k) \in \mathcal{V}$, we can explicitly represent the node routing probabilities into that node, p_v (defined in §4.1), in terms of the cross-hazard matrix $\mathbf{\Phi}$ and proportionality constants $\alpha_j^{\mathbf{base}}$ for the base event times $T_j^{\mathbf{base}}$ (defined through Definition 2), as

$$p_v := \frac{\alpha_{v_k}^{\mathbf{base}} \prod_{\ell=v_1,\ldots,v_{k-1}} \phi_\ell(v_k)}{\sum_{j\neq v_1,\ldots,v_{k-1}} \left(\alpha_j^{\mathbf{base}} \prod_{\ell=v_1,\ldots,v_{k-1}} \phi_\ell(j)\right)}. \tag{17}$$

6.3. Using (2) to Test for Hazard Ratio

Since $\phi_1(j)$ represents the increase in hazard of event $j \in \mathcal{M}$ due to treatment with the drug (i.e., event 1), the hypothesis test for hazard ratios can be stated as

$$\begin{aligned} H_0 &: \max_{j\in\mathcal{M}} \phi_1(j) \leq 1, \\ H_1 &: \max_{j\in\mathcal{M}} \phi_1(j) \geq \Theta, \end{aligned} \tag{18}$$

where $\Theta > 1$ represents a clinically meaningful threshold parameter. The null hypothesis states that the maximum hazard ratio of the drug is less than unity, while the alternate hypothesis states the maximum hazard ratio of the drug is greater than the threshold Θ. Using (17), we can relate the hazard ratios $\phi_1(j)$ of present interest to the rates $\eta_j^{\mathbf{A}}$ and $\eta_j^{\mathbf{B}}$ from (2), and furthermore choose parameters k_j, ϵ_j such that the test (2) approximates (18).

Proposition 8, which follows, collects two rather intuitive properties capturing the relationship between $\phi_1(j)$ and η_j^i. For $i \in \{A, B\}$, since η_j^i represents the asymptotic rate at which subjects are being accumulated into the set $S_j^i(t)$, it is not surprising that $\eta_j^{\mathbf{B}}$, the rate of adverse event j occurring in patients who have not been previously treated with the drug, is unaffected by $\phi_1(j')$ for any $j' \in \mathcal{M}$. Similarly, it is not surprising that $\eta_j^{\mathbf{A}}$, the rate of adverse event j occurring in patients that have been treated with the drug, is increasing in $\phi_1(j)$.

PROPOSITION 8. *For any $i \in \{A, B\}$ and $j \in \mathcal{M}$, write $\eta_j^i(\phi_1(2), \ldots, \phi_1(m+1))$ as the value of η_j^i as a function of the hazard ratios $(\phi_1(2), \ldots, \phi_1(m+1))$. Then,*

1. *$\eta_j^{\mathbf{B}}(\phi_1(2), \ldots, \phi_1(m+1))$ is constant with respect to $\phi_1(j')$ for any $j' \in \mathcal{M}$, and*
2. *$\eta_j^{\mathbf{A}}(\phi_1(2), \ldots, \phi_1(m+1))$ is increasing in $\phi_1(j)$.*

REMARK 1. A quick perusal of the proof reveals that in non-degenerate cases, $\eta_j^{\mathbf{A}}(\phi_1(2), \ldots, \phi_1(m+1))$ in fact *strictly* increases in $\phi_1(j)$.

We write $\eta_j^{\mathbf{A}}(\theta) := \eta_j^{\mathbf{A}}(1, \ldots, \theta, \ldots, 1)$, with θ in the j^{th} coordinate on the RHS, and note that Proposition 8 implies that $\eta_j^{\mathbf{A}}(\theta)$ increases in θ. Using this notation, we choose

$$k_j = \frac{\eta_j^{\mathbf{A}}(1)}{\eta_j^{\mathbf{B}}} \quad \text{and} \quad \epsilon_j = \eta_j^{\mathbf{A}}(\Theta) - \eta_j^{\mathbf{A}}(1), \tag{19}$$

We note that for any fixed value of θ, $\eta_j^{\mathbf{A}}(\theta)$ is a well-defined constant by equations (6) and (17). Hence, the parameters k_j and ϵ_j as defined through (19), are also well-defined constants.

Under these choices of k_j and ϵ_j, the test (2) approximates test (18) in the following sense. We want the null hypothesis of (18) to be rejected as long as the hazard ratio of the drug for *any* adverse adverse event exceeds Θ, even if the drug has no effect on the other adverse events (i.e., has a unit hazard ratio for the other events). From Proposition 8 and the definitions of k_j, ϵ_j, we see that the test (2) will indeed reject the null hypothesis in this circumstance. Specifically, if it is true that for some $j \in \mathcal{M}$, $\phi_1(j) \geq \Theta$, and $\phi_1(j') = 1$ for all $j' \neq j$, then it follows that $\eta_j^{\mathbf{A}} - k_j \eta_j^{\mathbf{B}} \geq \epsilon_j$.

6.4. Motivation for Test-Statistic

Using the cross-hazards model of event occurrences, we obtain a further motivation for the form of our test-statistic $\boldsymbol{L}$. Specifically, our test-statistic has a similar structure to the optimal test-statistics for a collection of $|\mathcal{M}|$ simpler tests, which we refer to as *marginal tests*. In marginal test $j \in \mathcal{M}$, we test for the increase in hazard rate of event j due to treatment with the drug, assuming that the base event times $\left\{T_j^{\mathbf{base}}\right\}_{j=1}^{m+2}$ are exponentially distributed, but with unknown rate, and that there are no cross-hazard effects between adverse events. We note that these specialized assumptions are only made to develop these marginal tests (i.e., within this section) in order to motivate the structure of the test-statistic, but are not used for the rest of our analysis.

In the rest of this section, we will formally define these marginal tests and derive their optimal test-statistics in four steps. First, we define the marginal tests. Second, we derive the maximum likelihood estimator (MLE) for the unknown base event rates. Third, we use the MLE to derive the optimal test statistic for the marginal tests. Finally, we use the structural insight from the marginal test statistics to motivate the structure of our proposed test-statistic.

Step 1: Description of Marginal Test

Fix some $j \in \mathcal{M}$ and let $\zeta_j > 1$ be some fixed parameter. Consider the case that the base event time $T_j^{\mathbf{base}}$ is exponentially-distributed with *apriori* unknown rate parameter $\mu_j^{\mathbf{base}} := \frac{1}{\mathbf{E}\left(T_j^{\mathbf{base}}\right)}$ and assume that there are no cross-hazard effects between adverse events. To investigate the effect of the drug on the hazard rate of j, we have the simple hypothesis test

$$\begin{aligned} H_0 &: \phi_1(j) = 1, \\ H_1 &: \phi_1(j) = \zeta_j. \end{aligned} \tag{20}$$

The optimal sequential test (Wald 1945, Wald and Wolfowitz 1948) for such problems is Wald's SPRT. In the SPRT, the relevant test-statistic is the sequentially-updated likelihood ratio between the two alternatives, or equivalently, the log-likelihood ratio (LLR). When the LLR exceeds some fixed interval, computed from the exogenous Type I/II errors for the test, the test terminates and we conclude in favor of either the null or the alternative, depending on whether the LLR exits the interval through its upper or lower boundary.

Step 2: MLE for $\mu_j^{\mathbf{base}}$

Suppose patients are indexed in increasing order of their arrival into the system. For any event $j \in \{0, \ldots, m+2\}$, let t_j^k represent the time of event j for patient k. Also, recall from Definition 1 that $A_0(t)$ counts represent the total number of patients that have ever visited node 0 by time t, which is equivalent to the number of patients that have ever entered the system by time t.

PROPOSITION 9. *The MLE, $\widehat{\mu}_j^{\mathbf{base}}$, for $\mu_j^{\mathbf{base}}$, is given by*

$$\widehat{\mu}_j^{\mathbf{base}} = \frac{\left|S_j^{\mathbf{B}}(t)\right|}{\sum_{k=1}^{A_0(t)} (t_1^k \wedge t_j^k \wedge t_{m+2}^k \wedge t - t_0^k)}, \tag{21}$$

Step 3: LLR for Marginal Tests

PROPOSITION 10. *The LLR, $L_j(t)$, for the simple hypothesis test (20) is given by*

$$L_j(t) = \left|S_j^{\mathbf{A}}(t)\right| \log \zeta_j - \left|S_j^{\mathbf{B}}(t)\right| (\zeta_j - 1) \frac{\sum_{k=1}^{A_0(t)} (t_1^k \wedge t_j^k \wedge t_{m+2}^k \wedge t - t_1^k)}{\sum_{k=1}^{A_0(t)} (t_1^k \wedge t_j^k \wedge t_{m+2}^k \wedge t - t_0^k)} \tag{22}$$

Step 4: Motivation of Test Statistic

The Wald-statistic for the marginal tests appear complicated at first glance. Nevertheless, from (22), we observe that the statistic is in fact quite intuitive. At each time t, it is a (weighted) difference of the counts of two different groups of patients: those who have experienced the side effect *after* taking the drug and those who have experienced the side effect *before* taking the drug. This further motivates the structural form of our proposed test-statistic in (3).

7. Numerical Study

We describe three simulation studies which investigate the performance of QNMEDS in the context of the cross-hazards model described in §6. These studies demonstrate four important features of QNMEDS:

1. QNMEDS has controlled false detection rates (Type I errors below an exogenously-specified level of 10%) and 100% true detection rates (zero Type II errors),
2. The approximations made in applying QNMEDS to the cross-hazards model of §6 do not detract from its performance,
3. QNMEDS is robust to assumptions about distributional shapes, and
4. QNMEDS is robust to values of its input parameters.

In particular, Study 1 demonstrates features 1 through 3, whereas Studies 2 and 3 demonstrates features 1, 2 and 4.

We conduct our numerical study using the cross-hazards model of §6 with $m = 2$ adverse effects. Let adverse event 2 represent the incidence of diabetes, and adverse event 3 represent the incidence of a cardiac event. The medical literature consistently shows that diabetics have about 2 times the risk of cardiac events than non-diabetics (e.g., Kannel and McGee 1979, Abbott et al. 1987, Barrett-Connor and Khaw 1988, Manson et al. 1991). We assume that the hazard ratio of mortality (departure from the system) due to cardiac events is about 5, while all other hazard ratios are unity. This leads to a cross-hazards matrix, $\mathbf{\Phi}$, given by

$$\mathbf{\Phi} = \begin{bmatrix} 1 & \psi_2 & \psi_3 & 1 \\ 1 & 1 & 2 & 1 \\ 1 & 1 & 1 & 5 \end{bmatrix}, \tag{23}$$

where ψ_2, ψ_3 represent the *aprioi* unknown hazard ratios of adverse events 2 and 3 respectively due to treatment with the drug. We will test whether either quantity exceeds the hazard ratio threshold of $\Theta = 1.4$. Specifically, for these numerical studies, we test

$$\begin{aligned} H_0 &: \max\{\psi_2, \psi_3\} \leq 1, \\ H_1 &: \max\{\psi_2, \psi_3\} \geq 1.4, \end{aligned}$$

which is exactly the test (18) using the present parameters (i.e., $m = 2$ and $\Theta = 1.4$).

In §7.1, we describe our data generation procedure, and in §7.2 through §7.4, we will proceed to describe the three studies in detail. All simulations and numerical analyses were done in MATLAB, and we used `CVX` (Grant and Boyd 2011) running the SDPT3 optimizer (Toh et al. 1999, Tütüncü et al. 2003) to solve the SDP (11).

7.1. Data Generation

We proceed to detail the general setup of our data generation procedure. Our data are divided into individual *datasets*. Each dataset comprises 50 independent and statistically identical simulation runs and each simulation run comprises 200,000 simulated patients, arriving at a rate of $\lambda = 100$. We used a fixed total patient size since we could not simulate indefinitely.

The random base event times $\left\{T_j^{\mathbf{base}}\right\}_{j=1}^{4}$ were modeled as Weibull random variables with rates (reciprocal of their means) of $(\mu_1^{\mathbf{base}}, \mu_2^{\mathbf{base}}, \mu_3^{\mathbf{base}}, \mu_4^{\mathbf{base}}) = (4, 0.5, 0.2, 0.1)$ that were held fixed across all datasets. The rates on the event times are much smaller in magnitude than the patient arrival rate (λ), to model the fact that diabetes, cardiac events, and mortality are relatively rare events. The relative magnitudes of the rates were chosen so that the more serious adverse events would occur at a slower rate.

Across datasets, we varied the shape parameter of the Weibull distribution, κ, as well as the hazard ratio of the drug on each adverse event, ψ_2 and ψ_3. The parameter values used in each dataset are reported together in their respective results tables.

7.2. Study 1: Sensitivity Analysis on Distributional Assumptions

The first study investigated the robustness of QNMEDS against distributional assumptions. For this study, we simulated 18 separate datasets, varying the shape parameter of the Weibull distribution, κ, as well as the hazard ratio of the drug on adverse event 2, ψ_2.

We used QNMEDS as described in Algorithm 3, controlling for a Type I error of $\alpha = 0.10$. As a benchmark, we used the following SPRT-based method. For each adverse event $j \in \{2, 3\}$, we computed the likelihood ratio (LLR) for the marginal hypothesis test in (20), assuming exponentially-distributed base event times. The method terminates and rejects H_0 when the LLR for either adverse events 2 or 3 exceeds the boundary $\log(2/\alpha)$, which approximately controls the Type I error below α (Siegmund 1985, Chapter II). The additional factor of 2 represents a Bonferroni correction (see, e.g., Shaffer 1995). Exponential base event times were assumed (even if the data were not exponentially-generated) in order to investigate the sensitivity of distributional assumptions for the SPRT-based method. In these cases, the assumed exponential distributions were fitted by equating means.

Data Parameters				QNMEDS		SPRT-based method	
κ	$H_?$	ψ_2	ψ_3	Rate (95% CI)	Time (95% CI)	Rate (95% CI)	Time (95% CI)
0.5	H_0	0.8	1.0	0.0 (0.0, 0.0)	100.0 (100.0, 100.0)	**100.0 (100.0, 100.0)**	0.1 (0.1, 0.1)
		0.9	1.0	2.0 (0.0, 5.9)	98.1 (94.4, 100.0)	**100.0 (100.0, 100.0)**	0.1 (0.1, 0.1)
		1.0	1.0	8.0 (0.4, 15.6)	92.7 (85.8, 99.6)	**100.0 (100.0, 100.0)**	0.1 (0.1, 0.1)
	H_1	1.4	1.0	100.0 (100.0, 100.0)	12.2 (10.9, 13.6)	100.0 (100.0, 100.0)	0.1 (0.1, 0.1)
		1.5	1.0	100.0 (100.0, 100.0)	8.9 (7.9, 9.8)	100.0 (100.0, 100.0)	0.1 (0.1, 0.1)
		1.6	1.0	100.0 (100.0, 100.0)	7.1 (6.4, 7.7)	100.0 (100.0, 100.0)	0.1 (0.1, 0.1)
1.0	H_0	0.8	1.0	4.0 (0.0, 9.5)	96.4 (91.5, 100.0)	2.0 (0.0, 5.9)	98.0 (94.1, 100.0)
		0.9	1.0	0.0 (0.0, 0.0)	100.0 (100.0, 100.0)	2.0 (0.0, 5.9)	98.0 (94.1, 100.0)
		1.0	1.0	8.0 (0.4, 15.6)	92.8 (86.0, 99.6)	4.0 (0.0, 9.5)	96.0 (90.5, 100.0)
	H_1	1.4	1.0	100.0 (100.0, 100.0)	13.0 (11.7, 14.4)	100.0 (100.0, 100.0)	0.1 (0.1, 0.1)
		1.5	1.0	100.0 (100.0, 100.0)	8.9 (8.0, 9.7)	100.0 (100.0, 100.0)	0.1 (0.1, 0.1)
		1.6	1.0	100.0 (100.0, 100.0)	7.7 (7.0, 8.4)	100.0 (100.0, 100.0)	0.1 (0.1, 0.1)
1.5	H_0	0.8	1.0	0.0 (0.0, 0.0)	100.0 (100.0, 100.0)	0.0 (0.0, 0.0)	100.0 (100.0, 100.0)
		0.9	1.0	6.0 (0.0, 12.6)	95.0 (89.4, 100.0)	0.0 (0.0, 0.0)	100.0 (100.0, 100.0)
		1.0	1.0	2.0 (0.0, 5.9)	98.7 (96.0, 100.0)	0.0 (0.0, 0.0)	100.0 (100.0, 100.0)
	H_1	1.4	1.0	100.0 (100.0, 100.0)	14.9 (13.3, 16.5)	**0.0 (0.0, 0.0)**	100.0 (100.0, 100.0)
		1.5	1.0	100.0 (100.0, 100.0)	10.2 (9.2, 11.1)	**0.0 (0.0, 0.0)**	100.0 (100.0, 100.0)
		1.6	1.0	100.0 (100.0, 100.0)	8.7 (7.9, 9.4)	**2.0 (0.0, 5.9)**	98.0 (94.1, 100.0)

Table 3 Detection rates (%) and times (%) for QNMEDS and SPRT-based method for Weibull-distributed base event times. Detection rates in green font and regular typeface are rates that are within the test specifications, whereas those in red font and bold typeface are rates that are out of the test specifications. The shape parameter of the Weibull distribution, κ, and the hazard ratio of the drug on adverse event 2, ψ_2, are varied across datasets. The hazard ratio of the drug on adverse event 3, ψ_3, was fixed at unity.

Results of both methods are reported in Table 3. Each row of the table represents a dataset and the first four columns represent the parameters used to generate the data. In particular, column 2 with heading $H_?$, denotes which hypothesis (H_0 or H_1) is true for the given dataset. For each dataset, we report two performance metrics: the rate and time of detection (rejection of H_0). Both quantities are reported as percentages: The detection rate is reported as a percentage of the total number of simulation runs and the detection time is reported as a percentage of the length of the total observation window. Specifically, for datasets where H_0 is true (i.e., $\max\{\psi_2, \psi_3\} \leq 1$), the detection rate represents the Type I error, while for datasets where H_1 is true (i.e., $\max\{\psi_2, \psi_3\} \geq 1.4$), the detection rate represents the power. We report 95% confidence intervals for all computed metrics.

From Table 3, we observe that for the datasets where the shape parameter $\kappa = 1.0$ (rows 7–12), both methods perform as expected (and as designed) in terms of Type I/II errors, with false detections rates below 10% and true detection rates that are 100%. For QNMEDS, the true detection time, corresponding to datasets with $\psi_2 \geq 1.4$, ranges from about 7% to 15% of the total observation window. As expected, the performance of QNMEDS is robust as κ varies from 0.5 to 1.5, performing similarly across these datasets. The benchmark SPRT-based method has

an extremely good detection time when its distributional assumptions are fulfillled ($\kappa = 1.0$, rows 10–12). However, it is highly sensitive to distributional assumptions, and performed very poorly on non-exponential datasets. On datasets with $\kappa = 0.5$ (rows 4–6), it registered a 100% false detection rate, while the converse held for datasets with $\kappa = 1.5$, it essentially failed to render any true detections (rows 16–18).

These results show that despite the strong theoretical merits of SPRT-based methods, they exhibit extreme sensitivity to distributional assumptions, which may be difficult to empirically justify. This limits the direct applicability of SPRT-based methods to the problem of drug surveillance. In contrast, QNMEDS, which is designed for robustness, strikes a balance between robustness and efficiency, and performs well across varied datasets.

We also conducted a non-sequential retrospective regression analysis on these datasets. Specifically, for each simulation run of each dataset, we ran two separate right-censored Cox regressions on the full data for each simulation run. For the first regression, the dependent variable was the length of time from a patient's entry into the system until occurrence of adverse event 2. In the case that the patient departed from the system before adverse event 2 occurred, then, the observation is flagged as right-censored and the dependent variable was the length of time that the patient spent in the system. Two binary predictor variables were used in the regression: (1) Whether drug treatment preceded adverse event 2, and (2) whether adverse event 3 preceded adverse event 2. The second regression was identical except that the roles of adverse events 2 and 3 were reversed. This analysis was defined to have made a detection if the regression coefficient for the first predictor in either regression (i.e., the log hazard ratio of the drug on the target adverse event for that regression) was significantly positive.

We found that across all datasets, even those where $\max\{\psi_2, \psi_3\}$ was as high as 1.6, this retrospective analysis yielded zero detections (results not tabulated). These results are not surprising: Cox regression is not designed for dynamically updating event times, which is a feature that is present in our stochastic model of event occurrences. We note that this regression analysis was fails even though (unlike the sequential methods) it had the unfair advantage of using all the data over the entire simulation time horizon. Therefore, this analysis would fail even if it was modified to operate as a group sequential test. These negative results underscore the importance of using a detection method that is tailored to the stochastics of the system.

7.3. Study 2: Mis-specification of Cross-Hazards

SPRT-based methods (e.g., Wald and Wolfowitz 1948, Kulldorff et al. 2011) are designed detecting single adverse events, and have very good theoretical properties. One way to extend them to handle multiple adverse events is to ignore any possible correlations between the adverse events and

perform multiple single-hypothesis tests on each adverse event. This is equivalent to (incorrectly) assuming that adverse events are independent. In the present numerical study, we investigate the performance degradation incurred by QNMEDS and the SPRT-based method when both methods make this fallacious assumption.

Specifically, our present study corresponds to assuming a cross-hazards matrix of

$$\widehat{\mathbf{\Phi}} = \begin{bmatrix} 1 & \psi_2 & \psi_3 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

for both methods instead of the true $\mathbf{\Phi}$ that generates the random events (from (23)). We emphasize that, if the actual cross-hazards matrix were indeed given by $\widehat{\mathbf{\Phi}}$, the SPRT-based method corresponds exactly to the classic SPRT with a Bonferroni correction for multiple hypotheses. We repeat the data generation procedure in the previous subsection, with the following modifications. The distributional family is fixed to be exponential (i.e., $\kappa = 1.0$), and we vary both the hazard ratio of the drug on adverse events 2 and 3 (ψ_2, ψ_3) across datasets.

Results are presented in Table 4. The SPRT-based method performs very poorly, exhibiting a 100% false detection rate throughout (rows 1–3, 7–9, 13–15). QNMEDS also saw an elevated false detection rate, especially in datasets at the "boundary" of the testing parameters, where $\max\{\psi_2, \psi_3\} = 1.0$.

These results speak to the importance of accounting for the correlations between adverse events: A surveillance method that speciously ignores correlations between adverse events can perform very poorly. The results also motivate the necessity of incorporating such correlations in designing QNMEDS. The results also suggest that the SPRT-based method is not only sensitive to distributional assumptions (as shown in §7.2), but also to its input parameters.

To further investigate the sensitivity of the SPRT-based method, we re-ran it on the same datasets using a small (20%) perturbation of the elements of $\mathbf{\Phi}$ (i.e., $\Phi(2,3) = 1.6, \Phi(3,4) = 4.0$). Even with this small perturbation, the SPRT-based method still registered a 100% (false) detection rate (results not tabulated) for all datasets with $\max\{\psi_2, \psi_3\} = 1.0$. These results further underscore the sensitivity of the SPRT-based method, and further limits its practical applicability.

7.4. Study 3: Sensitivity Analysis on Parameter Values

In the third study, we investigate the robustness of QNMEDS to perturbations in its input parameter values. We repeat the data generation procedure in the previous subsection, and additionally vary the arrival rate λ across datasets. Data parameters and results are reported in Table 5.

The results show consistently that QNMEDS is robust to perturbations in its input parameters, and has good performance over a range of parameter values. As before, QNMEDS features Type I errors below 10% as designed, and has a 100% true detection rate, with true detection occurring within 7% to 15% of the entire observation window.

Data Parameters			QNMEDS		SPRT-based method	
$H_?$	ψ_2	ψ_3	Rate (95% CI)	Time (95% CI)	Rate (95% CI)	Time (95% CI)
H_0	0.8	1.0	4.0 (0.0, 9.5)	96.3 (91.3, 100.0)	**100.0 (100.0, 100.0)**	0.2 (0.2, 0.2)
	0.9	1.0	0.0 (0.0, 0.0)	100.0 (100.0, 100.0)	**100.0 (100.0, 100.0)**	0.2 (0.2, 0.2)
	1.0	1.0	10.0 (1.6, 18.4)	91.1 (83.7, 98.6)	**100.0 (100.0, 100.0)**	0.2 (0.2, 0.2)
H_1	1.4	1.0	100.0 (100.0, 100.0)	12.1 (10.7, 13.5)	100.0 (100.0, 100.0)	0.1 (0.1, 0.1)
	1.5	1.0	100.0 (100.0, 100.0)	7.7 (6.9, 8.5)	100.0 (100.0, 100.0)	0.1 (0.1, 0.1)
	1.6	1.0	100.0 (100.0, 100.0)	6.7 (6.1, 7.3)	100.0 (100.0, 100.0)	0.1 (0.1, 0.1)
H_0	0.8	1.0	0.0 (0.0, 0.0)	100.0 (100.0, 100.0)	**100.0 (100.0, 100.0)**	0.6 (0.5, 0.6)
	0.9	1.0	2.0 (0.0, 5.9)	98.1 (94.3, 100.0)	**100.0 (100.0, 100.0)**	0.2 (0.2, 0.3)
	1.0	1.0	10.0 (1.6, 18.4)	91.1 (83.7, 98.6)	**100.0 (100.0, 100.0)**	0.2 (0.2, 0.2)
H_1	1.4	1.0	100.0 (100.0, 100.0)	13.4 (10.4, 16.3)	100.0 (100.0, 100.0)	0.1 (0.1, 0.1)
	1.5	1.0	100.0 (100.0, 100.0)	10.0 (8.3, 11.7)	100.0 (100.0, 100.0)	0.1 (0.1, 0.1)
	1.6	1.0	100.0 (100.0, 100.0)	9.0 (7.5, 10.4)	100.0 (100.0, 100.0)	0.1 (0.1, 0.1)
H_0	0.8	0.8	0.0 (0.0, 0.0)	100.0 (100.0, 100.0)	**100.0 (100.0, 100.0)**	0.9 (0.7, 1.1)
	0.9	0.9	0.0 (0.0, 0.0)	100.0 (100.0, 100.0)	**100.0 (100.0, 100.0)**	0.3 (0.2, 0.3)
	1.0	1.0	10.0 (1.6, 18.4)	91.1 (83.7, 98.6)	**100.0 (100.0, 100.0)**	0.2 (0.2, 0.2)
H_0	1.4	1.4	100.0 (100.0, 100.0)	9.9 (8.2, 11.7)	100.0 (100.0, 100.0)	0.1 (0.1, 0.1)
	1.5	1.5	100.0 (100.0, 100.0)	7.2 (6.1, 8.3)	100.0 (100.0, 100.0)	0.1 (0.1, 0.1)
	1.6	1.6	100.0 (100.0, 100.0)	6.7 (5.8, 7.6)	100.0 (100.0, 100.0)	0.1 (0.1, 0.1)

Table 4 Detection rates (%) and times (%) for QNMEDS and SPRT-based method, assuming that adverse events are independent, with exponential base event times and varying input parameter values. Detection rates in green font and regular typeface are rates that are within the test specifications, whereas those in red font and bold typeface are rates that are out of the test specifications.

8. Conclusion

In this paper we have presented a method, QNMEDS, for surveillance of a drug's effect on multiple adverse events. QNMEDS uses an intuitive vector-valued test-statistic: its components are weighted differences between the number of patients who have experienced the adverse event after and before taking. By analyzing the properties of this test-statistic in a limiting regime, we design stopping boundaries for the test-statistic that has a minimizes its expected detection time, subject to constraints its Type I error. Finally, we introduced our cross-hazards model, an adaptation of the Cox proportional hazards model to a dynamic setting, and showed both analytically and numerically how QNMEDS can be used as a test of the drug's hazard ratio on each adverse event.

We verified QNMEDS's functionality and performance on simulated data. The simulations highlighted several desirable features of QNMEDS for practical applications. Namely, QNMEDS is robust to both distributional assumptions and to its parameter values. In comparison, an SPRT-based heuristic is very sensitive to distributional and parameter assumptions. Practically, it is unlikely that the modeler can get the distribution shapes or input parameters exactly right, or (as is required by SPRT-based methods) that the likelihood function has an analytically tractable form. These considerations suggest that QNMEDS is better suited for practical drug surveillance than SPRT-based methods.

Data Parameters				QNMEDS		Data Parameters				QNMEDS	
λ	$H_?$	ψ_2	ψ_3	Rate (95% CI)	Time(95% CI)	λ	$H_?$	ψ_2	ψ_3	Rate (95% CI)	Time(95% CI)
100	H_0	0.8	1.0	4.0 (0.0, 9.5)	96.4 (91.5, 100.0)	200	H_0	0.8	1.0	0.0 (0.0, 0.0)	100.0(100.0, 100.0)
		0.9	1.0	0.0 (0.0, 0.0)	100.0(100.0, 100.0)			0.9	1.0	2.0 (0.0, 5.9)	98.7 (96.1, 100.0)
		1.0	1.0	8.0 (0.4, 15.6)	92.8 (86.0, 99.6)			1.0	1.0	2.0 (0.0, 5.9)	98.2 (94.7, 100.0)
	H_1	1.4	1.0	100.0(100.0, 100.0)	13.0 (11.7, 14.4)		H_1	1.4	1.0	100.0(100.0, 100.0)	14.5 (13.0, 15.9)
		1.5	1.0	100.0(100.0, 100.0)	8.9 (8.0, 9.7)			1.5	1.0	100.0(100.0, 100.0)	11.2 (10.2, 12.2)
		1.6	1.0	100.0(100.0, 100.0)	7.7 (7.0, 8.4)			1.6	1.0	100.0(100.0, 100.0)	8.2 (7.5, 9.0)
100	H_0	1.0	0.8	0.0 (0.0, 0.0)	100.0(100.0, 100.0)	200	H_0	1.0	0.8	0.0 (0.0, 0.0)	100.0(100.0, 100.0)
		1.0	0.9	0.0 (0.0, 0.0)	100.0(100.0, 100.0)			1.0	0.9	2.0 (0.0, 5.9)	98.3 (94.9, 100.0)
		1.0	1.0	8.0 (0.4, 15.6)	92.8 (86.0, 99.6)			1.0	1.0	2.0 (0.0, 5.9)	98.2 (94.7, 100.0)
	H_1	1.0	1.4	100.0(100.0, 100.0)	15.3 (12.0, 18.6)		H_1	1.0	1.4	100.0(100.0, 100.0)	19.2 (15.6, 22.8)
		1.0	1.5	100.0(100.0, 100.0)	11.6 (9.7, 13.4)			1.0	1.5	100.0(100.0, 100.0)	12.9 (11.0, 14.9)
		1.0	1.6	100.0(100.0, 100.0)	9.9 (8.3, 11.4)			1.0	1.6	100.0(100.0, 100.0)	7.9 (7.0, 8.8)
100	H_0	0.8	0.8	0.0 (0.0, 0.0)	100.0(100.0, 100.0)	200	H_0	0.8	0.8	0.0 (0.0, 0.0)	100.0(100.0, 100.0)
		0.9	0.9	0.0 (0.0, 0.0)	100.0(100.0, 100.0)			0.9	0.9	0.0 (0.0, 0.0)	100.0(100.0, 100.0)
		1.0	1.0	8.0 (0.4, 15.6)	92.8 (86.0, 99.6)			1.0	1.0	2.0 (0.0, 5.9)	98.2 (94.7, 100.0)
	H_1	1.4	1.4	100.0(100.0, 100.0)	10.8 (9.0, 12.6)		H_1	1.4	1.4	100.0(100.0, 100.0)	11.4 (9.8, 13.0)
		1.5	1.5	100.0(100.0, 100.0)	7.9 (6.8, 9.0)			1.5	1.5	100.0(100.0, 100.0)	8.7 (7.6, 9.8)
		1.6	1.6	100.0(100.0, 100.0)	7.4 (6.5, 8.3)			1.6	1.6	100.0(100.0, 100.0)	7.2 (6.4, 8.1)

Table 5 Detection rates (%) and detection times (%) using QNMEDS on datasets with exponentially-distributed base times and varying input parameter values. Detection rates in green font and regular typeface are rates that are within the test specifications, whereas those in red font and bold typeface are rates that are out of the test specifications.

A limitation of QNMEDS is that it is designed to monitor a single drug, and does not explicitly capture any interactive effects of multiple drugs that can either accentuate or attenuate adverse events. Incorporating such interactions into our model is a subject of future research. QNMEDS was also designed to have a Type II error of zero, and this was motivated because postmarketing surveillance represents the last line of protection for consumers and because the health and economic consequences of having a false negative (i.e., an "unsafe" drug that escapes detection) are potentially enormous. However, it may be practically infeasible to continue monitoring a drug indefinitely, especially if there is overwhelming evidence that it is safe. An interesting question for future research is how QNMEDS can be extended to incorporate non-zero Type II errors. Finally, we are presently also working on empirically validating QNMEDS on a large U.S. health insurance claims database.

References

Abbott, R., R. Donahue, S. MacMahon, D. Reed, K. Yano. 1987. Diabetes and the risk of stroke. *J Amer Med Assoc* **257**(7) 949.

Angiolillo, D. J., E. Bernardo, M. Sabaté, P. Jimenez-Quevedo, M. A. Costa, J. Palazuelos, R. Hernández-Antolin, R. Moreno, J. Escaned, F. Alfonso, et al. 2007. Impact of platelet reactivity on cardiovascular

outcomes in patients with type 2 diabetes mellitus and coronary artery disease. *Journal of the American College of Cardiology* **50**(16) 1541–1547.

Barrett-Connor, E., K.-T. Khaw. 1988. Diabetes mellitus: an independent risk factor for stroke? *Am J Epidemiol* **128**(1) 116.

Bate, A., M. Lindquist, I. Edwards, S. Olsson, R. Orre, A. Lansner, R. De Freitas. 1998. A bayesian neural network method for adverse drug reaction signal generation. *European Journal of Clinical Pharmacology* **54**(4) 315–321.

Billingsley, P. 2008. *Probability and measure*. John Wiley and Sons.

Brewer, T., G. Colditz. 1999. Postmarketing surveillance and adverse drug reactions. *J Amer Med Assoc* **281**(9) 824–829.

Brown, J., M. Kulldorff, K. Chan, R. Davis, D. Graham, P. Pettus, S. Andrade, M. Raebel, L. Herrinton, D. Roblin, et al. 2007. Early detection of adverse drug events within population-based health networks: application of sequential testing methods. *Pharmacoepidem Dr S* **16**(12) 1275–1284.

Cox, D. R. 1972. Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)* 187–220.

Deyo, R. 2004. Gaps, tensions, and conflicts in the fda approval process: implications for clinical practice. *J Am Board Fam Med* **17**(2) 142.

DuMouchel, W. 1999. Bayesian data mining in large frequency tables, with an application to the FDA spontaneous reporting system. *The American Statistician* 177–190.

Efron, B. 1977. The efficiency of Cox's likelihood function for censored data. *Journal of the American Statistical Association* **72**(359) 557–565.

Eick, S., W. Massey, W. Whitt. 1993. The physics of the $M_t/\mathrm{G}/\infty$ queue. *Oper Res* **41** 731–742.

Evans, S., P. Waller, S. Davis. 2001. Use of proportional reporting ratios (PRRs) for signal generation from spontaneous adverse drug reaction reports. *Pharmacoepidemiology and drug safety* **10**(6) 483–486.

FDA. 2012. FDA's Sentinel Initiative. URL `http://www.fda.gov/Safety/FDAsSentinelInitiative/`. Last Accessed Apr 10, 2012.

Frasure-Smith, N., F. Lespérance, M. Talajic. 1993. Depression following myocardial infarction. *J Amer Med Assoc* **270**(15) 1819.

Furberg, C., A. Levin, P. Gross, R. Shapiro, B. Strom. 2006. The FDA and drug safety: A proposal for sweeping changes. *Arch Intern Med* **166** 1938–1942.

Grant, M., S. Boyd. 2011. CVX: Matlab software for disciplined convex programming, version 1.21. URL `http://cvxr.com/cvx`.

Haffner, S., S. Lehto, T. Rönnemaa, K. Pyörälä, M. Laakso. 1998. Mortality from coronary heart disease in subjects with type 2 diabetes and in nondiabetic subjects with and without prior myocardial infarction. *N Engl J Med* **339** 229–234.

Harrison, J. 1985. *Brownian motion and stochastic flow systems*. John Wiley and Sons.

Hauck, D. J., J. B. Keats. 1997. Robustness of the exponential sequential probability ratio test (sprt) when weibull distributed failures are transformed using a "known" shape parameter. *Microelectronics Reliability* **37**(12) 1835 – 1840. doi:http://dx.doi.org/10.1016/S0026-2714(96)00287-9. URL `http://www.sciencedirect.com/science/article/pii/S0026271496002879`.

Jennison, C., B. W. Turnbull. 1999. *Group Sequential Methods with Applications to Clinical Trials*. Chapman and Hall/CRC.

Kannel, W., D. McGee. 1979. Diabetes and cardiovascular disease. *J Amer Med Assoc* **241**(19) 2035.

Klein, N., B. Fireman, W. Yih, E. Lewis, M. Kulldorff, P. Ray, R. Baxter, S. Hambidge, J. Nordin, A. Naleway, et al. 2010. Measles-mumps-rubella-varicella combination vaccine and the risk of febrile seizures. *Pediatrics* **126**(1) e1–e8.

Kulldorff, M., R. Davis, M. Kolczak, E. Lewis, T. Lieu, R. Platt. 2011. A maximized sequential probability ratio test for drug and vaccine safety surveillance. *Sequential Analysis* **30**(1) 58–78.

Lan, K. G., D. L. DeMets. 1983. Discrete sequential boundaries for clinical trials. *Biometrika* **70**(3) 659–663.

Lieu, T., M. Kulldorff, R. Davis, E. Lewis, E. Weintraub, K. Yih, R. Yin, J. Brown, R. Platt, et al. 2007. Real-time vaccine safety surveillance for the early detection of adverse events. *Med Care* **45**(10) S89–S95.

Luchsinger, J., M. Tang, Y. Stern, S. Shea, R. Mayeux. 2001. Diabetes mellitus and risk of alzheimer's disease and dementia with stroke in a multiethnic cohort. *Am J Epidemiol* **154**(7) 635.

Manson, J., G. Colditz, M. Stampfer, W. Willett, A. Krolewski, B. Rosner, R. Arky, F. Speizer, C. Hennekens. 1991. A prospective study of maturity-onset diabetes mellitus and risk of coronary heart disease and stroke in women. *Arch Intern Med* **151**(6) 1141.

McClellan, M. 2007. Drug safety reform at the fdapendulum swing or systematic improvement? *New England Journal of Medicine* **356**(17) 1700–1702.

Nelson, J., A. Cook, O. Yu. 2009. Evaluation of signal detection methods for use in prospective post licensure medical product safety surveillance. *FDA Sentinel Initiative Safety Signal Identification Contract* 1–40.

O'Brien, P. C., T. R. Fleming. 1979. A multiple testing procedure for clinical trials. *Biometrics* 549–556.

Pandit, P. V., N. V. Gudaganavar. 2010. On robustness of a sequential test for scale parameter of gamma and exponential distributions. *Applied Mathematics* **1**(4) 274–278.

Pang, G., R. Talreja, W. Whitt. 2007. Martingale proofs of many-server heavy-traffic limits for markovian queues. *Probability Surveys* **4** 193–267.

Pocock, S. J. 1977. Group sequential methods in the design and analysis of clinical trials. *Biometrika* **64**(2) 191–199.

Shaffer, J. 1995. Multiple hypothesis testing. *Annu Rev Psychol* **46**(1) 561–584.

Siegmund, D. 1985. *Sequential analysis: tests and confidence intervals*. Springer.

Toh, K., M. Todd, R. Tütüncü. 1999. SDPT3 – a MATLAB software package for semidefinite programming, version 1.3. *Optim Method Softw* **11**(1) 545–581.

Tütüncü, R., K. Toh, M. Todd. 2003. Solving semidefinite-quadratic-linear programs using SDPT3. *Math Program* **95**(2) 189–217.

Wald, A. 1945. Sequential tests of statistical hypotheses. *Ann Math Stat* **16**(2) pp. 117–186.

Wald, A., J. Wolfowitz. 1948. Optimum character of the sequential probability ratio test. *Ann Math Stat* **19**(3) pp. 326–339.

Whitt, W. 1980. Some useful functions for functional limit theorems. *Math Oper Res* 67–85.

Wisniewski, S., A. Rush, A. Nierenberg, B. Gaynes, D. Warden, J. Luther, P. McGrath, P. Lavori, M. Thase, M. Fava, et al. 2009. Can phase III trial results of antidepressant medications be generalized to clinical practice? A STAR* D report. *Am J Psychiat* **166**(5) 599.

Yih, W., J. Nordin, M. Kulldorff, E. Lewis, T. Lieu, P. Shi, E. Weintraub. 2009. An assessment of the safety of adolescent and adult tetanus-diphtheria-acellular pertussis (tdap) vaccine, using active surveillance for adverse events in the vaccine safety datalink. *Vaccine* **27**(32) 4257–4262.

This page is intentionally blank. Proper e-companion title page, with INFORMS branding and exact metadata of the main paper, will be produced by the INFORMS office when the issue is being assembled.

Appendix A: Proofs of Results

Proof of Lemma 1

Proof. Let the service distribution at any node $v \in \mathcal{V}$ be represented as F_v and let $H(t) = \mathbb{1}_{\{t \geq 0\}}$ represent the Heaviside (unit step) function. Let $\star$ represent the convolution operator, which has well-known algebraic properties such as associativity and distributivity (see, e.g. Billingsley 2008, §20).

Fix a node $v \in \mathcal{V}$, with representation $v = (v_1, \ldots, v_k) = v_1 v_2 \ldots v_k$. The arrival process into node v_1 is a thinned departure process (with probability p_{v_1}) out of node 0, and hence is nonhomogeneous Poisson with rate $\lambda p_{v_1}[H \star F_{v_1}(t)]$. Similarly, the arrival process into node $v_1 v_2$ is a thinned departure process (with probability $p_{v_1 v_2}$) out of node v_1, and is nonhomogeneous Poisson with rate $\lambda p_{v_1} p_{v_1 v_2}[H \star F_{v_1} \star F_{v_1 v_2}(t)]$. Repeating this, the instantaneous arrival rate at node v may be represented as

$$\lambda_v(t) = \lambda p_{v_1} p_{v_1 v_2} \ldots p_{v_1 v_2 \ldots v_k}[H \star F_{v_1} \star F_{v_1 v_2} \star \ldots \star F_{v_1 v_2 \ldots v_k}(t)] = \lambda \pi_v[H \star G_v(t)],$$

where $G_v(t)$ is a convolution of a finite number of service time distributions (i.e. it is the distribution of the sum of a finite number of service time distributions) at each node.

Note that $\lim_{t \to \infty}[H \star G_v(t)] = \int_0^\infty dG_v(t) = 1$, where the final equality is because $G_v(\cdot)$ is a distribution function. Observe that $\lambda_v(t)$ is uniformly bounded. The required result then follows from consistency of Césaro averages. $\square$

Proof of Proposition 1

Proof. Let V_j^i be as defined in the proposition statement. V_j^i is constructed as a disjoint union because $\{\mathcal{V}_k\}$ partitions $\mathcal{V}$. Moreover, the representation $\left|S_j^i(t)\right| = \sum_{v \in V_j^i} A_v(t)$ follows immediately from the definitions of $S_j^i(t)$ and V_j^i. We will proceed to prove the independence assertion by recursively showing that the subtrees rooted at any two distinct nodes in $\biguplus_{k=1}^{m'+1} V_j^i(k)$ are disjoint for any $m' \in \{0, \ldots, m\}$. Independence of the arrival processes then follows from Lemma EC.1.1.

Consider the case that $m' = 0$. For $i = A$, we have $V_j^{\mathbf{A}} = V_j^{\mathbf{A}}(1) = \emptyset$, while for $i = B$, we have $V_j^{\mathbf{B}} = V_j^{\mathbf{B}}(1) = \{j\}$, and both are trivially disjoint. Next, fix some $m' \in \{0, \ldots, m-1\}$, and suppose that the subtrees rooted at any node in $\biguplus_{k=1}^{m'+1} V_j^i(k)$ are disjoint. Further suppose for a contradiction that for some $v \in V_j^i(m'+2)$, we also had $v \in \text{tree}\left(\biguplus_{k=1}^{m'+1} V_j^i(k)\right)$. The first inclusion implies that $v(m'+2) = j$. The second inclusion implies that $v(k) = j$ for some $k \in \{0, \ldots, m'+1\}$, a contradiction. Thus, it is necessary that $V_j^i(m'+2)$ is not contained in any subtree with root in $\biguplus_{k=2}^{m'+1} V_i^j(k)$. $\square$

Proof of Proposition 2

Proof. Fix $j \in \mathcal{M}$. By Proposition 1, there exists sets $V_j^{\mathbf{A}}, V_j^{\mathbf{B}} \subseteq \mathcal{V}$ such that $\left|S_j^i(t)\right| = \sum_{n \in V_j^i} A_v(t)$ for each $i \in \{A, B\}$. We claim that $\Delta(V_j^{\mathbf{A}}, V_j^{\mathbf{B}}) = \emptyset$. Suppose for a contradiction that $\exists n \in \Delta(V_j^{\mathbf{A}}, V_j^{\mathbf{B}}) = \text{tree}\left(V_j^{\mathbf{A}}\right) \cap \text{tree}\left(V_j^{\mathbf{B}}\right)$. Let k_1, k_j be defined such that $n(k_1) = 1$ and $n(k_j) = j$. Since $n \in \text{tree}\left(V_j^{\mathbf{A}}\right)$, we must have $k_1 < k_j$, while since $n \in \text{tree}\left(V_j^{\mathbf{B}}\right)$, we have $k_1 > k_j$, a contradiction. Hence, $\Delta(V_j^{\mathbf{A}}, V_j^{\mathbf{B}}) = \emptyset$. The proposition follows by Corollary EC.1.1. $\square$

Proof of Proposition 3

Proof. The first part of Proposition 3 follows immediately from Proposition 1, the linearity of expectations, and Lemma 1. For the second part, fix $i, i' \in \{A, B\}$, and $j, j' \in \mathcal{M}$. By Proposition 1 and Corollary EC.1.2, $\exists$ a set of nodes $\mathcal{T}(i, j, i', j') \subseteq \mathcal{V}$ such that

$$\begin{aligned}
\Sigma(i,j,i',j') &= \lim_{t\to\infty} \frac{1}{\lambda t} \text{Cov}\left(\left|S_j^i(t)\right|, \left|S_{j'}^{i'}(t)\right|\right) && \text{[By definition (7)]} \\
&= \lim_{t\to\infty} \frac{1}{\lambda t} \sum_{v \in \mathcal{T}(i,j,i',j')} \Lambda_v(t) && \text{[Proposition 1 and Corollary EC.1.2]} \\
&= \sum_{v \in \mathcal{T}(i,j,i',j')} \pi_v && \text{[Lemma 1].}
\end{aligned}$$

Finally, the result that $T(i, j, i, j) = V_j^i$ follows directly from Corollary EC.1.2. $\square$

Proof of Proposition 4

Proof. From Proposition 1, and the linearity of expectations, we have $\mathbf{E}\left(\left|\widehat{S}_j^i(t)\right|\right) = \sum_{v \in V_j^i} \widehat{\Lambda}_v(t)$. Consider the vector-valued (martingale) process with components

$$\frac{\left|\widehat{S}_j^i(\lambda t)\right| - \sum_{v \in V_j^i} \widehat{\Lambda}_v(\lambda t)}{\sqrt{\lambda}} \qquad \forall (i,j) \in \{A, B\} \times \mathcal{M}.$$

This is a martingale because $\widehat{S}_j^i$ is a nonhomogeneous Poisson process and $\sum_{v \in V_j^i} \widehat{\Lambda}_v$ is the compensator for the process. By applying the FCLT for multidimensional martingales (Pang et al. 2007, Theorem 8.1(ii)) to this process, we can show that this converges weakly to a $(\mathbf{0}, \mathbf{\Sigma})$ Brownian Motion, as $\lambda \to \infty$.

By adding and subtracting terms, and recalling that $\eta_j^i = \sum_{v \in V_j^i} \pi_v$ from (6), we have

$$\begin{aligned}
\widehat{L}_j^{(\lambda)}(t) &= \frac{\left|\widehat{S}_j^{\mathbf{A}}(\lambda t)\right|}{\sqrt{\lambda}} - k_j \frac{\left|\widehat{S}_j^{\mathbf{B}}(\lambda t)\right|}{\sqrt{\lambda}} \\
&= \frac{\left|\widehat{S}_j^{\mathbf{A}}(\lambda t)\right| - \sum_{v \in V_j^{\mathbf{A}}} \widehat{\Lambda}_v(\lambda t)}{\sqrt{\lambda}} - k_j \frac{\left|\widehat{S}_j^{\mathbf{B}}(\lambda t)\right| - \sum_{v \in V_j^{\mathbf{B}}} \widehat{\Lambda}_v(\lambda t)}{\sqrt{\lambda}} \\
&\quad + \frac{\sum_{v \in V_j^{\mathbf{A}}} \left[\widehat{\Lambda}_v(\lambda t) - \pi_v \lambda t\right]}{\sqrt{\lambda}} + k_j \frac{\sum_{v \in V_j^{\mathbf{B}}} \left[\widehat{\Lambda}_v(\lambda t) - \pi_v \lambda t\right]}{\sqrt{\lambda}} + t\sqrt{\lambda}\left[\eta_j^{\mathbf{A},\lambda} - k_j \eta_j^{\mathbf{B},\lambda}\right].
\end{aligned}$$

Hence, the desired convergence follows from the continuous mapping theorem (see Whitt 1980) and by assumption (8b) if we can prove that

$$\lim_{\lambda\to\infty} \frac{\widehat{\Lambda}_v(\lambda t) - \pi_v \lambda t}{\sqrt{\lambda}} = 0 \quad \forall v \in \mathcal{V}. \tag{EC.1}$$

As in Lemma 1, let H represent the Heaviside function, $\star$ the convolution operator. Also, let G_v be the distribution function defined in Lemma 1, and $\overline{G}_v = 1 - G_v$ its complement. Then,

$$\begin{aligned}
\lim_{\lambda\to\infty} \left|\widehat{\Lambda}_v(\lambda t) - \pi_v \lambda t\right| &\le \lim_{\lambda\to\infty} \int_0^{\lambda t} \left|\widehat{\lambda}_v(s) - \pi_v\right| ds \\
&= \int_0^{\lambda t} |\pi_v [H \star G_v(s)] - \pi_v| \, ds \qquad \text{[See proof of Lemma 1]} \\
&\le \int_0^{\infty} |\pi_v [H \star G_v(s)] - \pi_v| \, ds \qquad \text{[Integrand is positive]} \\
&= \pi_v \int_0^{\infty} H \star (1 - G_v)(s) ds \\
&= \pi_v \int_0^{\infty} \int_0^{s} d\overline{G}_v(\tau) d\tau ds \\
&= \pi_v \int_0^{\infty} \overline{G}_v(s) ds \\
&< \infty,
\end{aligned}$$

where the final inequality is due to the assumption of integrable event times. Hence, (EC.1) holds and the proof is complete. □

Proof of Lemma 2

Proof. Suppose that $\boldsymbol{c} = (c_j)_{j\in\mathcal{M}} < \mathbf{0}$. Then (8b) implies that given any $0 < \delta < -\min_j \{c_j\}$, for all sufficiently large λ, we have

$$\left|\sqrt{\lambda}\eta_j^{\mathbf{A},\lambda} - \sqrt{\lambda} k_j \eta_j^{\mathbf{B},\lambda} - c_j\right| < \delta \implies \eta_j^{\mathbf{A},\lambda} \le k_j \eta_j^{\mathbf{B},\lambda} + \frac{c_j + \delta}{\sqrt{\lambda}} < k_j \eta_j^{\mathbf{B},\lambda} \quad \forall j,$$

where the last inequality is by the condition on δ.

Similarly, if $c_{j^*} > \overline{c}_{j^*}$, then for any $0 < \delta < c_{j^*} - \overline{c}_{j^*}$ and sufficiently large λ,

$$\left|\sqrt{\lambda}\eta_j^{\mathbf{A},\lambda} - \sqrt{\lambda} k_j \eta_j^{\mathbf{B},\lambda} - c_j\right| < \delta \implies \eta_{j^*}^{\mathbf{A},\lambda} \ge k_{j^*} \eta_{j^*}^{\mathbf{B},\lambda} + \frac{c_{j^*} - \delta}{\sqrt{\lambda}}$$

The result follows from noting that

$$\begin{aligned}
k_{j^*}\eta_{j^*}^{\mathbf{B},\lambda} + \frac{c_{j^*} - \delta}{\sqrt{\lambda}} &= k_{j^*}\eta_{j^*}^{\mathbf{B},\lambda} + \frac{\overline{c}_{j^*}}{\sqrt{\lambda}} + \frac{c_{j^*} - \overline{c}_{j^*} - \delta}{\sqrt{\lambda}} \\
&> k_{j^*}\eta_{j^*}^{\mathbf{B},\lambda} + \frac{\overline{c}_{j^*}}{\sqrt{\lambda}} \qquad \text{[By } \delta < c_{j^*} - \overline{c}_{j^*}\text{]} \\
&= k_{j^*}\eta_{j^*}^{\mathbf{B},\lambda} + \epsilon_j \qquad \text{[Definition of } \overline{c}_{j^*}\text{]}.
\end{aligned}$$

□

Proof of Lemma 3

Proof. First, assume H_0 is true. Fix some drift vector $\boldsymbol{c} \leq \boldsymbol{0}$ of $\boldsymbol{Y}$

Let $\boldsymbol{Q} = \boldsymbol{U}\boldsymbol{D}^2\boldsymbol{U}^T$ represent the decomposition of $\boldsymbol{Q}$ into a unitary matrix $\boldsymbol{U}$ and a diagonal matrix $\boldsymbol{D}$. Also let $\widetilde{\boldsymbol{D}}$ be a diagonal matrix such that $\widetilde{\boldsymbol{D}}^2 = \boldsymbol{S}$. Noting that by definition, $\boldsymbol{Z}(t) = (\boldsymbol{c}-\boldsymbol{r})\,t + \boldsymbol{U}\boldsymbol{D}\boldsymbol{W}(t)$, we define the B.M. process $\widetilde{\boldsymbol{Z}}$ analogously, as $\widetilde{\boldsymbol{Z}}(t) := (\boldsymbol{c}-\boldsymbol{r})\,t + \widetilde{\boldsymbol{D}}\boldsymbol{W}(t)$. Also define the stopping times $\tau := \inf\{t \geq 0 : \boldsymbol{Z}(t) \in \partial\mathcal{C}\}$, and $\theta := \inf\left\{t \geq 0 : \widetilde{\boldsymbol{Z}}(t) \in \partial\mathcal{C}\right\}$.

The probability of (incorrectly) rejecting H_0 is exactly $\mathbf{P}(\tau < \infty)$. We will show that this quantity can be bounded above by $\mathbf{P}(\theta < \infty)$, which in turn can be computed as

$$\mathbf{P}(\theta < \infty) = \begin{cases} 1 - \prod_{j \in \mathcal{M}} \left(1 - \exp\left(-\frac{2(r_j - c_j)y_j}{\sigma_j^2}\right)\right) & \text{if } c_j \leq r_j, \\ 1 & \text{otherwise,} \end{cases}$$

by applying a well-known result (e.g., Harrison 1985, Chapter 3) for hitting probabilities of a one-dimensional B.M., and by the independence of the components of $\boldsymbol{W}(t)$. The statement of the lemma would then follow by observing that the expression for $\mathbf{P}(\theta < \infty)$ is monotonically increasing in each c_j, which would give us

$$\sup_{\boldsymbol{c} \leq \boldsymbol{0}} \mathbf{P}(\tau < \infty) \overset{(*)}{\leq} \sup_{\boldsymbol{c} \leq \boldsymbol{0}} \mathbf{P}(\theta < \infty) = 1 - \prod_{j \in \mathcal{M}} \left(1 - \exp\left(-\frac{2r_j y_j}{\sigma_j^2}\right)\right),$$

as long as we can prove the inequality $(*)$ in the display above.

It remains to show that the inequality $\mathbf{P}(\tau < \infty) \leq \mathbf{P}(\theta < \infty)$ holds. Since $\boldsymbol{S} \succeq \boldsymbol{Q}$, there exists some matrix $\boldsymbol{A} := [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m] \in \mathbb{R}^{m \times m}$ such that $\boldsymbol{S} - \boldsymbol{Q} = \boldsymbol{A}\boldsymbol{A}^T$, i.e., $\widetilde{\boldsymbol{D}}^2 = \boldsymbol{U}\boldsymbol{D}^2\boldsymbol{U}^T + \boldsymbol{A}\boldsymbol{A}^T$. Let $\widetilde{\boldsymbol{W}} := \left\{\widetilde{\boldsymbol{W}}(t), t \geq 0\right\}$ be a multidimensional B.M., independent of $\boldsymbol{W}$, and let $\mathcal{F}_t^W, t \geq 0$ be the filtration generated by $\boldsymbol{W}(t)$, and $\mathcal{F}_\infty^W := \sigma\left(\bigcup_{t \geq 0} \mathcal{F}_t\right)$. Then, by conditioning on the path of the original B.M., we have the following inequality:

$$\begin{aligned}
\mathbf{P}(\theta < \infty) &= \mathbf{P}\left(\exists t \geq 0 : (\boldsymbol{c}-\boldsymbol{r})\,t + \widetilde{\boldsymbol{D}}\boldsymbol{W}(t) \in \mathcal{C}^c\right) && \text{[By definition of } \theta\text{]} \\
&= \mathbf{P}\left(\exists t \geq 0 : (\boldsymbol{c}-\boldsymbol{r})\,t + \boldsymbol{U}\boldsymbol{D}\boldsymbol{W}(t) + \boldsymbol{A}\widetilde{\boldsymbol{W}}(t) \in \mathcal{C}^c\right) && \text{[By } \widetilde{\boldsymbol{D}}^2 = \boldsymbol{U}\boldsymbol{D}^2\boldsymbol{U}^T + \boldsymbol{A}\boldsymbol{A}^T\text{]} \\
&= \mathbf{P}\left(\exists t \geq 0 : \boldsymbol{Z}(t) + \boldsymbol{A}\widetilde{\boldsymbol{W}}(t) \in \mathcal{C}^c\right) && \text{[By definition of } \boldsymbol{Z}\text{]} \\
&= \mathbf{E}\left(\mathbf{P}\left(\exists t \geq 0 : \boldsymbol{Z}(t) + \boldsymbol{A}\widetilde{\boldsymbol{W}}(t) \in \mathcal{C}^c | \mathcal{F}_\infty^W\right)\right) && \text{[Tower Property]} \\
&\geq \mathbf{E}\left(\mathbf{P}\left(\exists t \geq 0 : \boldsymbol{Z}(t) + \boldsymbol{A}\widetilde{\boldsymbol{W}}(t) \in \mathcal{C}^c, \tau < \infty | \mathcal{F}_\infty^W\right)\right) \\
&= \mathbf{E}\left(\mathbb{1}_{\{\tau < \infty\}}\mathbf{P}\left(\exists t \geq 0 : \boldsymbol{Z}(t) + \boldsymbol{A}\widetilde{\boldsymbol{W}}(t) \in \mathcal{C}^c | \mathcal{F}_\infty^W\right)\right)
\end{aligned}$$

Next, define the event $H := \left\{\exists t : \boldsymbol{Z}(t) + \sum_{j \in \mathcal{M}} \boldsymbol{a}_j \widetilde{W}_j(t) \in \mathcal{C}^c\right\}$. Substituting the definition of H into both sides of the inequality above, we observe that the key inequality $(*)$ would follow if we can show that

$$\mathbb{1}_{\{\tau < \infty\}}\mathbf{P}\left(H | \mathcal{F}_\infty^W\right) \overset{a.s.}{=} \mathbb{1}_{\{\tau < \infty\}}.$$

Fix a path $\{\boldsymbol{Z}(t), t \geq 0\} \in \{\tau < \infty\} \in \mathcal{F}_\infty^W$. This fixes the stopping time $\tau < \infty$, and the hitting point $\boldsymbol{Z}(\tau) \in \partial\mathcal{C}$. Let the unit vector $\boldsymbol{h} \in \mathbb{R}^m$ represent a supporting hyperplane of $\mathcal{C}^c$ at $\boldsymbol{Z}(\tau)$. Since $\mathcal{C}^c$ is the finite union of halfspaces, we may choose the components of $\boldsymbol{h}$ to be signed such that an arbitrary $\boldsymbol{x} \in \mathcal{C}^c$ if and only if $\langle \boldsymbol{h}, \boldsymbol{x} - \boldsymbol{Z}(\tau)\rangle \leq 0$, where $\langle \cdot, \cdot \rangle$ represents the inner product.

Path continuity of $\boldsymbol{Z}$ implies that for any $\epsilon > 0$, there exists some $\delta > 0$ such that $\|\boldsymbol{Z}(t) - \boldsymbol{Z}(\tau)\| \leq \epsilon$ for all $|t - \tau| \leq \delta$. Hence, for $t \in [\tau - \delta, \tau + \delta]$,

$$\begin{aligned}\left\langle \boldsymbol{h}, \boldsymbol{Z}(t) + \sum_{j\in\mathcal{M}} \boldsymbol{a}_j \widetilde{W}_j(t) - \boldsymbol{Z}(\tau) \right\rangle &= \sum_{j\in\mathcal{M}} \widetilde{W}_j(t) \langle \boldsymbol{h}, \boldsymbol{a}_j\rangle + \langle \boldsymbol{h}, \boldsymbol{Z}(t) - \boldsymbol{Z}(\tau)\rangle \\ &\leq \sum_{j\in\mathcal{M}} \widetilde{W}_j(t) \langle \boldsymbol{h}, \boldsymbol{a}_j\rangle + \|\boldsymbol{Z}(t) - \boldsymbol{Z}(\tau)\| \\ &\leq \Omega \sum_{j\in\mathcal{M}} \widetilde{W}_j(t) + \epsilon,\end{aligned} \tag{EC.2}$$

where $\Omega := \max_{j\in\mathcal{M}} |\langle \boldsymbol{h}, \boldsymbol{a}_j\rangle|$. For notational brevity, define $\widehat{W} := \frac{1}{m}\sum_{j\in\mathcal{M}} \widetilde{W}_j(t)$ and note that $\widehat{W}$ is a standard (single-dimensional) B.M.. Moreover, we observe that $H \supseteq F_\epsilon$, where F_ϵ is defined by

$$F_\epsilon := \left\{ \widehat{W} \leq -\frac{\epsilon}{m\Omega} \text{ for some } t \in [\tau - \delta, \tau + \delta] \right\}.$$

This is because if the event F_ϵ occurs, then the RHS of (EC.2) is negative, which would imply that $\boldsymbol{Z}(t) + \sum_{j\in\mathcal{M}} \boldsymbol{a}_j \widetilde{W}_j(t) \in \mathcal{C}^c$.

Next, consider the following event

$$G_\epsilon := \left\{ \inf_{0\leq t\leq \tau-\delta} \widehat{W}(t) \leq -\frac{\epsilon}{m\Omega} \right\}.$$

A B.M. has level sets that contain no isolated point and is unbounded (Billingsley 2008, Theorem 37.4), and thus, as long as the B.M. $\widehat{W}$ hits level $-\frac{\epsilon}{m\Omega}$ before time $\tau - \delta$, it will a.s. hit this same level again within the time interval $[\tau - \delta, \tau + \delta]$. Thus, $\mathbf{P}\left(G_\epsilon \setminus F_\epsilon | \mathcal{F}_\infty^W\right) = 0$ for all ϵ. As $\epsilon \downarrow 0$, $\mathbf{P}\left(G_\epsilon | \mathcal{F}_\infty^W\right) \uparrow 1$ and thus $\mathbf{P}\left(F_\epsilon | \mathcal{F}_\infty^W\right) \uparrow 1$, which implies that $\mathbf{P}\left(H | \mathcal{F}_\infty^W\right) = 1$ on $\{\tau < \infty\}$. Thus, we have $\mathbb{1}_{\tau<\infty}\mathbf{P}\left(H | \mathcal{F}_\infty^W\right) \overset{a.s.}{=} \mathbb{1}_{\tau<\infty}$ as required.

Now assume H_1 is true, which means that there exists some $j \in \mathcal{M}$ such that $c_j \geq \overline{c}_j$. Since $\boldsymbol{r} < \overline{\boldsymbol{c}}$, $\boldsymbol{Z}(t)$ is a multi-dimensional B.M., with a drift vector possessing at least one strictly positive component. Thus, w.p.1., the test statistic will exit $\mathcal{C}$ eventually, i.e. the Type II error is zero.

Define the hitting time of boundary j, as a function of the drift c_j as $T_j(c_j) := \inf\{t \geq 0 : Z_j(t) = y_j\}$. If j is such that $c_j > r_j$, by optional sampling and monotone convergence, we can easily see that $T_j(c_j)$ is integrable with $\mathbf{E}\left(T_j(c_j)\right) = \frac{y_j}{c_j - r_j}$. Conversely, if j is such that $c_j \leq r_j$, then there exists some non-trivial probability that $T_j(c_j) = +\infty$ and therefore $\mathbf{E}\left(T_j(c_j)\right) = +\infty$. The overall stopping time for the test is the earliest of these hitting times, i.e. $T(\boldsymbol{c}) := \min_{j\in\mathcal{M}} T_j(c_j)$.

Since the alternate hypothesis is a composite hypothesis, the objective of the optimization is to minimize the *worst-case* expected time among all $\boldsymbol{c}$ that satisfies H_1. Defining the set $\mathcal{R} := \{\boldsymbol{c} \in \mathbb{R}^m : \max_{j\in\mathcal{M}}(c_j - \overline{c}_j) \geq 0\}$, the optimization objective is exactly the quantity,

$$\sup_{\boldsymbol{c}\in\mathcal{R}} \mathbf{E}\left(\min_{j\in\mathcal{M}}\{T_j(c_j)\}\right).$$

Define $\partial\mathcal{R} := \{\boldsymbol{c} \in \mathbb{R}^m : \max_{j\in\mathcal{M}}(c_j - \overline{c}_j) = 0\}$. It is immediate from the geometry of $\mathcal{R}$ that

$$\sup_{\boldsymbol{c}\in\mathcal{R}} \mathbf{E}\left(\min_{j\in\mathcal{M}}\{T_j(c_j)\}\right) = \sup_{\boldsymbol{c}\in\partial\mathcal{R}} \mathbf{E}\left(\min_{j\in\mathcal{M}}\{T_j(c_j)\}\right),$$

because each T_j is monotonically decreasing in its argument.

Hence, the result would follow if we can show that

$$\sup_{\boldsymbol{c}\in\partial\mathcal{R}} \mathbf{E}\left(\min_{j\in\mathcal{M}}\{T_j(c_j)\}\right) = \max_{j\in\mathcal{M}} \mathbf{E}\left(T_j(\overline{c}_j)\right). \tag{EC.3}$$

For any $j^* \in \arg\max_j \mathbf{E}\left(T_j(\overline{c}_j)\right)$, consider the feasible point $\boldsymbol{c} \in \mathcal{R}$, constructed as follows: $c_{j^*}^{(n)} = \overline{c}_{j^*}$ and $c_j^{(n)} = r_j$ for all $j \neq j^*$. Hence, $\min_j\{T_j(c_j)\} = T_{j^*}(\overline{c}_{j^*})$. Taking expectations on both sides yields $\mathbf{E}\left(\min_j\{T_j(c_j)\}\right) = \mathbf{E}\left(T_{j^*}(\overline{c}_{j^*})\right) = \max_{j\in\mathcal{M}} \mathbf{E}\left(T_j(\overline{c}_j)\right)$. Since $\boldsymbol{c}$ is an arbitrary feasible point in $\mathcal{R}$, we have shown that

$$\sup_{\boldsymbol{c}\in\mathcal{R}} \mathbf{E}\left(\min_{j\in\mathcal{M}}\{T_j(c_j)\}\right) \geq \max_{j\in\mathcal{M}} \mathbf{E}\left(T_j(\overline{c}_j)\right).$$

To show the converse, observe that for any $\boldsymbol{c} \in \partial\mathcal{R}$, there exists some $\ell \in \mathcal{M}$, such that $c_\ell = \overline{c}_\ell$, which implies that $T_\ell(c_\ell) = T_\ell(\overline{c}_\ell)$. Thus, for each $\boldsymbol{c} \in \mathcal{R}$, we may write $\min_{j\in\mathcal{M}}\{T_j(c_j)\} = T_\ell(\overline{c}_\ell) \wedge \min_{j\neq\ell}\{T_j(c_j)\}$, for some $\ell \in \mathcal{M}$. Now, define for each $\ell \in \mathcal{M}$, the sets $U_\ell := \{\boldsymbol{c} \in \mathbb{R}^m : c_j \leq \overline{c}_j, \forall j \neq \ell\}$, and note that each $U_\ell \supseteq \partial\mathcal{R}$. Then, we have

$$\begin{aligned}
\sup_{\boldsymbol{c}\in\partial\mathcal{R}} \mathbf{E}\left(\min_{j\in\mathcal{M}}\{T_j(c_j)\}\right) &\leq \max_{\ell\in\mathcal{M}} \sup_{\boldsymbol{c}\in U_\ell} \mathbf{E}\left(T_\ell(\overline{c}_\ell) \wedge \min_{j\neq\ell}\{T_j(c_j)\}\right) && \text{[Since } U_\ell \supseteq \partial\mathcal{R} \text{ for each } \ell\text{]} \\
&\leq \max_{\ell\in\mathcal{M}} \sup_{\boldsymbol{c}\in U_\ell} \mathbf{E}\left(T_\ell(\overline{c}_\ell)\right) && \text{[Monotonicity of expectation]} \\
&= \max_{\ell\in\mathcal{M}} \mathbf{E}\left(T_\ell(\overline{c}_\ell)\right) && \text{[}\mathbf{E}\left(T_\ell(\overline{c}_\ell)\right) \text{ is constant in } \boldsymbol{c}\text{]}
\end{aligned}$$

thus establishing (EC.3). □

Proof of Proposition 5

Proof. Our proof proceeds by first simplifying problem (12). First, we claim that there is no loss of generality in restricting to the linear subspace where $r_j = \overline{c}_j/2$ for all j. To see this, apply a

change-of-variable, $r_j = \frac{\overline{c}_j}{2}(1+\delta_j)$, and $w_j = y_j(1+\delta_j)$. In terms of variables (δ_j, w_j) problem (12) becomes

$$\begin{aligned}
&\min_{\boldsymbol{\delta},\boldsymbol{w}} \max_{j\in\mathcal{M}} \frac{2w_j}{(1-\delta_j^2)\overline{c}_j} \\
&\text{s.t.} \prod_{j\in\mathcal{M}} \left(1-\exp\left(-\frac{\overline{c}_j w_j}{\sigma_j^2}\right)\right) \geq 1-\alpha, \\
&\boldsymbol{w} \geq \mathbf{0} \\
&-1 \leq \delta_j < 1 \quad \forall j \in \mathcal{M}.
\end{aligned}$$

From the above, it is clear that any non-zero value of δ_j can only increase the objective.

Hence, under the restriction of $r_j = \overline{c}_j/2$, problem (12) becomes

$$\begin{aligned}
&\min_{\boldsymbol{y}} \max_{j\in\mathcal{M}} \left[\frac{2y_j}{\overline{c}_j}\right] \\
&\text{s.t.} \sum_{j\in\mathcal{M}} \log\left(1-\exp\left(-\frac{\overline{c}_j y_j}{\sigma_j^2}\right)\right) \geq \log(1-\alpha), \\
&\boldsymbol{y} \geq \mathbf{0}.
\end{aligned}$$

We linearize the problem above by introducing an auxiliary variable s, and apply the change-of-variables $x_j = y_j/\sigma_j^2$ to transform it into a more amenable form for optimization,

$$\begin{aligned}
&\min_{s,\boldsymbol{x}} s \\
&\text{s.t.} \frac{2\sigma_j^2}{\overline{c}_j} x_j \leq s \quad \forall j \in \mathcal{M} \\
&\sum_{j\in\mathcal{M}} \log\left(1-\exp\left(-\overline{c}_j x_j\right)\right) \geq \log(1-\alpha), \\
&\boldsymbol{x} \geq \mathbf{0}.
\end{aligned} \tag{EC.4}$$

The KKT conditions are necessary for optimality of (EC.4) due to Slater-type constraint qualifications and for sufficiently large $\boldsymbol{x}, s$. Moreover, (EC.4) is a convex program, and hence, the KKT conditions are also sufficient for optimality. We will proceed to derive the KKT optimality conditions. We associate dual variables $\nu_j, j \in \mathcal{M}$ with the first set of constraints and ξ with the second constraint. The KKT conditions read

$$\begin{aligned}
s: &\quad \sum_{j\in\mathcal{M}} \nu_j = 1 \\
x_j: &-\nu_j \frac{2\sigma_j^2}{\overline{c}_j} + \xi \frac{\overline{c}_j e^{-\overline{c}_j x_j}}{1-e^{-\overline{c}_j x_j}} = 0 \ \forall j
\end{aligned} \tag{EC.5}$$

together with the complementary slackness conditions. Suppose for a contradiction that $\xi = 0$. Then we must have that $v_j = 0$ for all $j \in \mathcal{M}$, which violates the first condition. Hence, we must have $\xi > 0$, and by complementary slackness, the second constraint of (EC.4) must bind, i.e., we have $\sum_{j\in\mathcal{M}} \log\left(1-\exp\left(-\overline{c}_j x_j\right)\right) = \log(1-\alpha)$.

Furthermore, this implies that $v_j > 0$ for all $j \in \mathcal{M}$, otherwise, the second set of constraints of

(EC.5) will be violated. By complementary slackness of the first set of constraints of (EC.4), we require

$$\frac{2\sigma_j^2 x_j}{\overline{c}_j} = s \implies x_j = \frac{\overline{c}_j}{2\sigma_j^2} s \quad \forall j \in \mathcal{M},$$

where s solves the equation

$$\sum_{j\in\mathcal{M}} \log\left(1-\exp\left(-\frac{\overline{c}_j^2 s}{2\sigma_j^2}\right)\right) = \log(1-\alpha).$$

By the transformation $y_j = \sigma_j^2 x_j$ we recover the solution to (12). $\square$

Proof of Proposition 6

Proof. Consider the function $f(\sigma) := \log\left(1-\exp\left(-\frac{\overline{c}_j^2 s}{2\sigma^2}\right)\right)$ for any fixed $\overline{c} > 0$ and $s > 0$. Direct computation shows that $f'(\sigma) \le 0$. Hence, the function $\sum_{j\in\mathcal{M}} \log\left(1-\exp\left(-\frac{\overline{c}_j^2 s}{2\sigma_j^2}\right)\right)$, which increases with s, decreases with *each* σ_j. Recall that s^* denotes a solution of (13). Hence, for each j, *ceteris paribus*, as $\sigma_j \uparrow$, we also have $s^* \uparrow$.

Consider a fixed number of adverse events $|\mathcal{M}|$, and fixed σ_j, $j \in \mathcal{M}$. Let s^* solve (13) for these parameters. For a new pair $(\overline{c}, \sigma) > 0$, consider

$$\left(1-\exp\left(\frac{-\overline{c}^2 s^*}{2\sigma^2}\right)\right)\prod_{j\in\mathcal{M}}\left(1-\exp\left(\frac{-\overline{c}_j^2 s^*}{2\sigma_j^2}\right)\right) = \left(1-\exp\left(\frac{-\overline{c}^2 s^*}{2\sigma^2}\right)\right)(1-\alpha) < (1-\alpha).$$

Hence, for s to solve (13) with $|\mathcal{M}|+1$ adverse events, it must necessarily be larger than s^*. $\square$

Proof of Proposition 7

Proof. Equations (15) and (14) for $\left\{T_j^{\mathbf{base}}\right\}_{j=1}^{m+2}$ directly imply that $\{T_j\}_{j\in\mathcal{Y}}$ satisfies the proportional hazards condition with constants α_j as defined in the proposition statement.

Fix $t \ge 0$ and $j \in \mathcal{Y}$. For each $\ell \in \mathcal{Y}$, define $\overline{F}_\ell(t) := \mathbf{P}(T_\ell > t)$, and let f_ℓ be the density of T_ℓ. We observe that by independence, $\mathbf{P}(\tau > t) = \prod_{\ell\in\mathcal{Y}} \overline{F}_\ell(t)$, and

$$\begin{aligned}
\mathbf{P}(\tau = T_j, \tau > t) &= \mathbf{P}\left(\{T_j > t\} \cap \bigcap_{\ell\neq j}\{T_\ell > T_j\}\right) \\
&= \int_t^\infty \prod_{\ell\neq j} \mathbf{P}(T_\ell > s)\, f_j(s) ds \qquad \text{[Independence]} \\
&= \int_t^\infty \prod_{\ell\neq j} \overline{F}_\ell(s) f_j(s) ds \qquad \text{[Definition of } \overline{F}_\ell\text{]} \\
&= \alpha_j \int_t^\infty \prod_{\ell\in\mathcal{Y}} \overline{F}_\ell(s) g(s) ds \qquad \text{[By (15) for } \{T_j\}_{j\in\mathcal{Y}}\text{]}
\end{aligned}$$

Sum across j on both sides. The LHS simplifies since τ must take the value of some T_j, and the RHS simplifies directly. This yields

$$\mathbf{P}(\tau > t) = \left(\int_t^\infty \prod_{\ell\in\mathcal{Y}} \overline{F}_\ell(s) g(s) ds\right)\left(\sum_{j\in\mathcal{Y}} \alpha_j\right).$$

The result follows by direct substitution. □

Proof of Proposition 8

Proof. Recall from (6), that $\eta_j^{\mathbf{B}} = \sum_{v \in V_j^{\mathbf{B}}} \pi_v$. To show part 1, it suffices to note that from the definition of $V_j^{\mathbf{B}}$ from Proposition 1, for any node $v \in V_j^{\mathbf{B}}$, there does not exist a k such that $v(k) = 1$. Consequently, for any node $v \in V_j^{\mathbf{B}}$, the routing probability π_v is constant with respect to $\phi_1(j')$ for any $j' \in \mathcal{M}$.

To prove part 2, fix some $j \in \mathcal{M}$. Define for any node $v \in \mathcal{V}$,

$$P(v) := V_j^{\mathbf{A}} \cap \{u \in \mathcal{V} : \text{len}\,(u) \geq \text{len}\,(v)\,, u(k) = v(k), 1 \leq k \leq \text{len}\,(v)\}\,.$$

In words, $P(v)$ is the subset of nodes in $V_j^{\mathbf{A}}$ that have v as a *prefix*. Also define

$$\kappa(v) := \begin{cases} \dfrac{1}{\pi_v} \displaystyle\sum_{u \in P(v)} \pi_u & \text{if } \pi_v > 0, \\ 0 & \text{otherwise.} \end{cases}$$

In words, $\kappa(v)$ is the sum of routing probabilities over the set $P(v)$, and normalized by π_v. Alternatively, $\kappa(v)$ represents the sum of routing probabilities from node v to its child nodes that have j after 1. We note that by construction, $P(0) := V_j^{\mathbf{A}}$ and $\kappa(0) := \eta_j^{\mathbf{A}}$. Finally, define $\mathcal{Z}$ as the set of nodes that do not contain either 1 or j.

For the rest of the proof, to keep notation manageable, we use the following two shorthands. First, for any node $v \in \mathcal{V}$, we write the condition $\forall j' \notin v$ as a shorthand to mean that j' is the subset of elements of $\{1, \ldots, m+1\}$ such that v has no component equal to j'. Second, for any node $v \in \mathcal{V}$, we write the shorthand $\partial_j p_v$ to mean the partial derivative $\partial p_v / \partial \phi_1(j)$.

The following three preliminaries hold by straightforward manipulations and their proofs will only be sketched.

$$\kappa(v1j') = \kappa(vj'1) \qquad \forall v \in \mathcal{Z}, \forall j' \notin v1j, \tag{EC.6}$$

$$\partial_j p_{v1j'} \leq 0 \qquad \forall v \in \mathcal{Z}, \forall j' \notin v1j, \tag{EC.7}$$

$$\partial_j p_v = 0 \qquad \text{if } \nexists k < \text{len}\,(v) \text{ such that } v(k) = 1. \tag{EC.8}$$

Firstly, (EC.6) follows from fact that $p_{vj} = p_{\sigma(v)j}$ for any permutation $\sigma(\cdot)$, which in turn follows immediately from the definition of p_v in (17) and the fact that products commute. Secondly, both (EC.7) and (EC.8) follows immediately from the definition of p_v in (17).

Finally, to prove the proposition statement, we claim that

$$\partial_j \kappa(v) \geq 0 \quad \text{and} \quad \partial_j \kappa(v1) \geq 0 \tag{EC.9}$$

for all nodes $v \in \mathcal{V}$ that do not contain 1 or j. Indeed, this implies the proposition statement because $\eta_j^{\mathbf{A}} = \kappa(0)$. We shall prove our claim by backward induction on the length of v.

First, consider any node $v \in \mathcal{V}$, with length $m-1$, that does not contain 1 or j. The node $v1$ only has a single child, namely $v1j$, and hence, $\kappa(v1) = p_{v1j}$. Further, by (17), the routing probability to that child node is

$$p_{v1j} = \frac{\phi_1(j)\left[\prod_{j' \neq j} \phi_{j'}(j)\alpha_j\right]}{\phi_1(j)\left[\prod_{j' \neq j} \phi_{j'}(j)\alpha_j\right] + \left[\prod_{j' \neq j} \phi_{j'}(m+2)\alpha_{m+2}\right]}.$$

Note that $p_{v1j} \neq 1$ in general because there is some probability of departing the system. Nevertheless, we still have $\partial_j p_{v1j} \geq 0$ from the expression above, since the terms in square brackets are constants with respect to $\phi_1(j)$. Similarly, the node v only has two children, $v1$ and vj. However, since the node $vj \notin V_j^{\mathbf{A}}$, hence $vj \notin P(n)$. Therefore $\kappa(v) = \kappa(v1)$, and hence $\partial_j \kappa(v) = \partial_j \kappa(v1) \geq 0$.

To prove the result for nodes v of general length, suppose that (EC.9) holds for all nodes that do not contain 1 or j and have length strictly greater than some ℓ. Let v be a node of length ℓ, that does not contain 1 or j. We note that the following recursions hold.

$$\kappa(v) = p_{v1}\kappa(v1) + \sum_{j' \notin vj} p_{vj'}\kappa(vj'), \tag{EC.10}$$

and

$$\kappa(v1) = p_{v1j} + \sum_{j' \notin v1j} p_{v1j'}\kappa(v1j'). \tag{EC.11}$$

Then, by the chain rule, we get

$$\begin{aligned}
\partial_j \kappa(v1) &= \partial_j p_{v1j} + \sum_{j' \notin v1j} \kappa(v1j')\partial_j p_{v1j'} + \sum_{j' \notin v1j} p_{v1j'}\partial_j \kappa(vij') && \text{[Chain rule]} \\
&= \partial_j p_{vij} + \sum_{j' \notin vij} \kappa(vij')\partial_j p_{vij'} + \sum_{j' \notin vij} p_{vij'}\partial_j \kappa(vj'1) && \text{[By (EC.6)]} \\
&\geq \partial_j p_{vij} + \sum_{j' \notin vij} \kappa(vij')\partial_j p_{vij'} && \text{[By inductive hypothesis]} \\
&= \sum_{j' \notin vij} (\kappa(vij') - 1)\partial_j p_{vij'} && \text{[By } p_{vij} = 1 - \textstyle\sum_{j' \notin vij} p_{vij'}\text{]} \\
&\geq 0 && \text{[By (EC.7) and } \kappa(\cdot) \in [0,1]\text{]}.
\end{aligned}$$

Noting that by (EC.8) that p_{v1} and $p_{vj'}$ are constant with respect to $\phi_1(j)$, we have

$$\partial_j \kappa(v) = p_{v1}\partial_j \kappa(v1) + \sum_{j' \notin vj} p_{vj'}\partial_j \kappa(vj') \geq 0,$$

by the inductive hypothesis and since we just showed that $\partial_j \kappa(v1) \geq 0$. $\square$

Proof of Proposition 9

Proof. Throughout this proof, let C represent some generic constant free of μ_j. Fix a patient $k \in \mathbb{N}$. Clearly, $k \leq A_0(t)$ to contribute anything to the log-likelihood function (LLF) of μ_j.

First, consider $k \in S_j^{\mathbf{B}}(t)$. The (additive) contribution of this patient to the LLF is $\log \mu_j - \mu_j(t_j^k - t_0^k) + C$. Second, consider $k \notin S_j^{\mathbf{B}}(t)$. The contribution of this patient to the LLF is $-\mu_j(t_1^k \wedge t_{m+2}^k \wedge t - t_0^k) + C$, since if either treatment or departure occurs, there is no further contribution to the LLF of μ_j.

Thus, the LLF for μ_j is

$$LLF_j(t) := \left|S_j^{\mathbf{B}}(t)\right| \log \mu_j - \mu_j \sum_{k=1}^{A_0(t)} (t_1^k \wedge t_j^k \wedge t_{m+2}^k \wedge t - t_0^k) + C,$$

and the expression for the MLE is obtained via straightforward differentiation. □

Proof of Proposition 10

Proof. Fix a patient $k \in \mathbb{N}$ and suppose μ_j is known. Clearly patient k only contributes to the LLR if $t_1^k < t$. First, if $k \in S_j^{\mathbf{A}}(t)$, the additive contribution to the LLR is $\log \zeta_j - \mu_j(\zeta_j - 1)(t_j^k - t_1^k)$. For $k \notin S_j^{\mathbf{A}}(t)$ and $t_1^k < t$, the additive contribution to the LLR is $-\mu_j(\zeta_j - 1)(t_{m+2}^k \wedge t - t_1^k)$.

To summarize, the overall LLR is

$$LLR_j(t) = \left|S_j^{\mathbf{A}}(t)\right| \log \zeta_j - \mu_j(\zeta_j - 1) \sum_{k=1}^{A_0(t)} (t_1^k \wedge t_j^k \wedge t_{m+2}^k \wedge t - t_1^k).$$

Finally, (22) is obtained by substituting the MLE (21) for μ_j. □

Appendix B: Technical Lemmas

In this Appendix, we establish several technical results that are used as components of proofs in the main paper. These results show how the position of nodes within the arborescent queueing network affect the correlation of their arrival processes. We begin with several definitions.

Definitions

For nodes $u, v \in \mathcal{V}$, such that u is the (unique) parent of v, we define $\text{parent}(\cdot), \text{child}(\cdot)$ as $u := \text{parent}(v)$ and $v \in \text{child}(u)$. Also, for any node $v \in \mathcal{V}$, define $\text{tree}(v)$ the set of nodes that are descendants of v, i.e.

$$\text{tree}(v) := \{u \in \mathcal{V} : \text{len}(u) \geq \text{len}(v), u(k) = v(k), 1 \leq k \leq \text{len}(v)\}.$$

Similarly, for any finite collection of nodes $\mathcal{A} \subseteq \mathcal{V}$, define $\text{tree}(\mathcal{A}) := \bigcup_{v \in \mathcal{A}} \text{tree}(v)$. Moreover, for $T = \text{tree}(v)$ for some $v \in \mathcal{V}$, we define $\text{root}(T) := v$.

Also, we define the *tree intersection* for two nodes $u, v \in \mathcal{V}$ as

$$\Delta(u,v) := \text{tree}\,(u) \cap \text{tree}\,(v)\,.$$

We apply the same notation to sets, i.e. for $U, V \subseteq \mathcal{V}$, $\Delta(U,V) := \text{tree}\,(U) \cap \text{tree}\,(V)$. For $\Delta(u,v) \neq \emptyset$, define $r(u,v) := \text{root}\,(\Delta(u,v))$. Finally, note that if $u \in \text{tree}\,(v)$, then $r(u,v) = u$, and vice-versa.

Three Technical Results

LEMMA EC.1. *Given two nodes $u, v \subseteq \mathcal{V}$, let $T := \Delta(u,v)$ and define processes $X := \{X(t), t \geq 0\}$, $Y := \{Y(t), t \geq 0\}$, with $X(t) := A_u(t)$, $Y(t) := A_v(t)$. Then,*

1. *If $T = \emptyset$, then X and Y are independent.*
2. *If $T \neq \emptyset$, then $\text{Cov}\,(X(t), Y(t)) = \Lambda_{r(u,v)}(t)$ for all $t \geq 0$.*

Proof. First, consider the case that u and v have the same lengths. The arrivals to u and v are then thinned processes from some common (and generically nonhomogeneous) Poisson process. Thus, A_u and A_v are independent, unless $u = v$, in which case $A_u = A_v$ and $\text{Cov}\,(A_u(t), A_v(t)) = \text{Var}\,(A_u(t)) = \Lambda_u(t)$ for all $t \geq 0$.

Next, suppose u and v have different lengths. WLOG, let $\text{len}\,(u) < \text{len}\,(v)$. By going up the tree, we can find the ancestor v' such that $v \in \text{tree}\,(v')$, and v' is in the same generation as u, i.e. $\text{len}\,(v') = \text{len}\,(u)$. First, suppose $T = \emptyset$. Then, it is necessary that $v' \neq u$. Otherwise, $v \in \text{tree}\,(u)$. Thus, from above, the processes $A_{v'}$ and A_u are independent by Poisson thinning. Since A_v is derived from $A_{v'}$ by a sequence of repeated queueing and thinning, it is also independent of A_u.

Second, suppose that $T \neq \emptyset$. Then, it is necessary that $v \in \text{tree}\,(u)$, and thus $T = \text{tree}\,(v)$. Letting $K := \text{len}\,(v) - \text{len}\,(u)$, define $u^K := v$, and recursively $u^{k-1} := \text{parent}\,(u^k)$ for each $k \in \{1, \ldots, K\}$. Then we have $u^0 = u$, and u^k are ancestors of v for each k.

Fix some $t \geq 0$. We claim that for each $k \in \{0, \ldots, K\}$, we may write

$$A_{u^k}(t) = B_{u^k}(t) + A_{u^{k+1}}(t). \tag{EC.12}$$

where $B_{u^k}(t)$ and $A_{u^{k+1}}(t)$ are independent. Indeed, letting $Q_{u^k} := \{Q_{u^k}(t), t \geq 0\}$ represent the queue length process at node u^k, and $D_{u^k} := \{D_{u^k}(t), t \geq 0\}$ the departure process from that node, we have

$$A_{u^k}(t) = Q_{u^k}(t) + D_{u^k}(t) + A_{u^{k+1}}(t) + \sum_{\substack{w \in \text{child}\left(u^k\right) \\ w \neq u^{k+1}}} A_w(t).$$

By a Poisson thinning argument, $A_{u^{k+1}}(t)$ is independent of $A_w(t)$ for all $w \in \text{child}\,(u^k) \setminus \left\{u^{k+1}\right\}$ and also independent of $D_{u^k}(t)$. Moreover, by Eick et al. (1993, Theorem 1), $A_{u^{k+1}}(t)$ is independent of

$Q_{u^k}(t)$. Hence, if we define $B_{u^k}(t) := A_{u^k}(t) - A_{u^{k+1}}(t)$, then $B_{u^k}(t)$ and $A_{u^{k+1}}(t)$ are independent. By recursively expanding (EC.12), we get

$$A_u(t) = \sum_{k=0}^{K-1} B_{u^k}(t) + A_v(t),$$

with $A_v(t)$ independent of $\sum_{k=0}^{K-1} B_{u^k}(t)$. Thus, $\mathrm{Cov}\left(A_u(t), A_v(t)\right) = \mathrm{Var}\left(A_v(t)\right) = \Lambda_v(t) = \Lambda_{r(u,v)}(t)$.

□

This Lemma is extended by the following corollary.

COROLLARY EC.1. *Given two sets of nodes $U, V \subseteq \mathcal{V}$, let $T := \Delta(U,V)$ and define processes $X := \{X(t), t \geq 0\}$, $Y := \{Y(t), t \geq 0\}$, with $X(t) := \sum_{u\in U} A_u(t)$, $Y(t) := \sum_{v\in V} A_v(t)$. Then,*

1. *If $T = \emptyset$, then X and Y are independent.*
2. *If $T \neq \emptyset$, then $\mathrm{Cov}\left(X(t), Y(t)\right) = \sum_{(u,v)\in\mathcal{T}} \Lambda_{r(u,v)}(t)$ for all $t \geq 0$, where*

$$\mathcal{T} := \left\{(u,v) \in U \times V : \Delta(u,v) \neq \emptyset\right\}.$$

Proof. Note that U, V are finite sets, and admit the representation $U = \biguplus_{u\in U} \{u\}$ and $V = \biguplus_{v\in V} \{v\}$. By definition,

$$\begin{aligned} T &= \mathrm{tree}\,(U) \cap \mathrm{tree}\,(V) \\ &= \left(\biguplus_{u\in U} \mathrm{tree}\,(u)\right) \cap \left(\biguplus_{v\in V} \mathrm{tree}\,(v)\right) \\ &= \biguplus_{u\in U} \biguplus_{v\in V} \left(\mathrm{tree}\,(u) \cap \mathrm{tree}\,(v)\right) \\ &= \biguplus_{(u,v)\in\mathcal{T}} \Delta(u,v) \end{aligned}$$

Hence, if $T = \emptyset$, we have $\mathrm{tree}\,(u) \cap \mathrm{tree}\,(v) = \emptyset$ for all $u \in U$, $v \in V$. By Lemma EC.1, the vector processes $(A_u)_{u\in U}$ and $(A_v)_{v\in V}$ are independent. Since Borel functions applied separately to independent vectors preserves their independence, and X and Y are sums over the components of $(A_u)_{u\in U}$ and $(A_v)_{v\in V}$ respectively, X and Y are also independent.

Conversely, if $T \neq \emptyset$, then $\mathcal{T} \neq \emptyset$ either. Thus, for any $t \geq 0$,

$$\begin{aligned} \mathrm{Cov}\left(X(t), Y(t)\right) &= \sum_{u\in U}\sum_{v\in V} \mathrm{Cov}\left(A_u(t), A_v(t)\right) \text{ [Bilinearity of Cov}\,(\cdot)] \\ &= \sum_{(u,v)\in\mathcal{T}} \mathrm{Cov}\left(A_u(t), A_v(t)\right) \quad \text{[Lemma EC.1.1]} \\ &= \sum_{(u,v)\in\mathcal{T}} \Lambda_{r(u,v)}(t). \qquad \text{[Lemma EC.1.2]} \end{aligned}$$

□

Appendix C: Illustration with m = 2 Adverse Events

In this Appendix, we consider a small example with $m = 2$ adverse events, labelled with indices $\{2, 3\}$. Index 0 represents arrival to the system, index 1 represents treatment with the drug, and index 4 represents departure from the system. The $M/G/\infty/M$ queueing network for this small example is illustrated in Figure 3 of the main text. Through this example, we demonstrate how to compute the means and covariances of the the patient count processes $\left|S_j^i(t)\right|$ in Proposition 3. This small and concrete example serves to illustrate the key ideas invoked in the derivations for the case of general m.

Decomposition of Patient Counts into Cumulative Arrivals

We begin by exhibiting the result of Proposition 1, by expressing the patient count processes $\left|S_j^i(t)\right|$ as sums of cumulative arrival processes A_v (defined in Definition 1).

First, $|S_2^{\mathbf{B}}(t)|$, the number of patients who experienced adverse event 2 *before* treatment, can be obtained as the sum

$$|S_2^{\mathbf{B}}(t)| = A_2(t) + A_{32}(t), \tag{EC.13}$$

Note that each of the arrival processes are all independent, because of Poisson thinning. Similarly, $|S_2^{\mathbf{A}}(t)|$, the number of patients who experienced side effect 2 *after* treatment, can be obtained as the sum

$$|S_2^{\mathbf{A}}(t)| = A_{12}(t) + A_{132}(t) + A_{312}(t). \tag{EC.14}$$

Analogous expressions can be derived for adverse event 3.

$$\left|S_B^3(t)\right| = A_3(t) + A_{23}(t), \tag{EC.15}$$

and

$$\left|S_A^3(t)\right| = A_{13}(t) + A_{123}(t) + A_{213}(t). \tag{EC.16}$$

Covariance Calculations

First, consider the covariance between $|S_2^{\mathbf{B}}(t)|$ and $|S_2^{\mathbf{A}}(t)|$ (Eqs. (EC.13) and (EC.14)). The arcs used are depicted in Figure EC.1. From the figure, it is clear that $|S_2^{\mathbf{A}}(t)|$ is independent of $|S_2^{\mathbf{B}}(t)|$, illustrating the result of Lemma 2. A completely analogous argument shows that the same holds for adverse event 3: $|S_3^{\mathbf{B}}(t)|$ and $|S_3^{\mathbf{A}}(t)|$ are independent.

Second, consider the covariance between $|S_2^{\mathbf{B}}(t)|$ and $|S_3^{\mathbf{B}}(t)|$ (Eqs. (EC.13) and (EC.15)). The relevant processes are depicted in Figure EC.2. While there are no overlapping arcs used, there is still dependence. For example, the processes $A_2(t)$ and $A_{23}(t)$ are generally not independent. Indeed

$$A_2(t) = A_{23}(t) + A_{21}(t) + A_{24}(t) + Q_2(t).$$

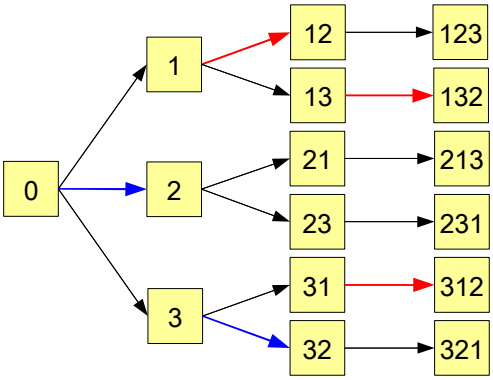

Figure EC.1 Blue: Arcs used in constructing $\left|S_2^{\mathbf{B}}(t)\right|$. Red: Arcs used in constructing $\left|S_2^{\mathbf{A}}(t)\right|$. Departures not illustrated.

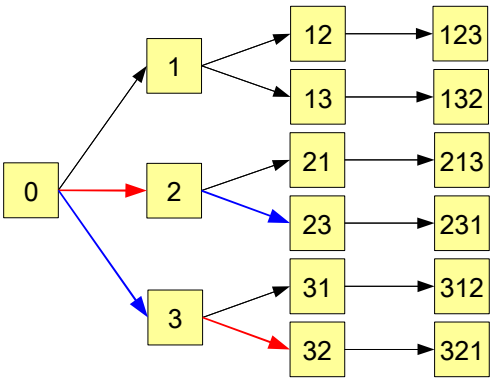

Figure EC.2 Red: Arcs used in constructing $\left|S_2^{\mathbf{B}}(t)\right|$. Blue: Arcs used in constructing $\left|S_3^{\mathbf{B}}(t)\right|$. Departures not illustrated

Note that $A_{24}(t)$ represents the "arrival" to node 24, which is the external departure process from Node 2. Node 2 behaves like a $M_t/G/\infty$ queue. Eick et al. (1993, Theorem 1) show that for such queues, the queue length process is independent of the departure process from the queue. Thus, A_{23} and Q_2 are independent. Moreover, A_{21}, A_{23}, and A_{24} are mutually independent by Poisson thinning. Hence, $\mathrm{Cov}(A_2(t), A_{23}(t)) = \mathrm{Var}(A_{23}(t))$. A similar relationship holds for A_3 and A_{32}. Thus, the covariance can be computed as

$$\begin{aligned}
&\mathrm{Cov}(|S_2^{\mathbf{B}}(t)|, |S_3^{\mathbf{B}}(t)|) \\
&\quad = \mathrm{Cov}(A_2(t) + A_{23}(t), A_3(t) + A_{32}(t)) && \text{[(EC.13) and (EC.15))]} \\
&\quad = \mathrm{Cov}(A_{23}(t), A_{23}(t)) + \mathrm{Cov}(A_{32}(t), A_{32}(t)) && \text{[Independence]} \\
&\quad = \mathrm{Var}(A_{23}(t)) + \mathrm{Var}(A_{32}(t)) \\
&\quad = \Lambda_{23}(t) + \Lambda_{32}(t).
\end{aligned}$$

Third, consider the covariance between $|S_2^{\mathbf{A}}(t)|$ and $|S_3^{\mathbf{A}}(t)|$ (Eqs. (EC.14) and (EC.16)). The relevant processes are depicted in Figure EC.3. Again, the processes $A_{12}(t)$ and $A_{123}(t)$ are generally not independent. Indeed, we express

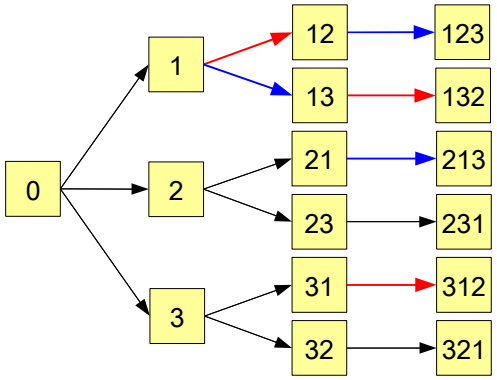

Figure EC.3 Red: Arcs used in constructing $\left|S_2^{\mathbf{A}}(t)\right|$. Blue: Arcs used in constructing $\left|S_3^{\mathbf{A}}(t)\right|$. Departures not illustrated

$$A_{12}(t) = A_{123}(t) + A_{124}(t) + Q_{12}(t),$$

and note that A_{123} and Q_{12} are independent by Eick et al. (1993, Theorem 1), while A_{123} and A_{124} are independent by Poisson thinning. Hence, $\mathrm{Cov}(A_{12}(t), A_{123}(t)) = \mathrm{Var}(A_{123}(t))$. A similar relationship holds for A_{13} and A_{132}. Thus, the covariance can be computed as

$$\begin{aligned}
&\mathrm{Cov}(\left|S_2^{\mathbf{A}}(t)\right|, \left|S_3^{\mathbf{A}}(t)\right|) \\
&\quad = \mathrm{Cov}(A_{12}(t) + A_{132}(t) + A_{312}(t), A_{13}(t) + A_{123}(t) + A_{213}(t)) && \text{[(EC.14) and (EC.16))]} \\
&\quad = \mathrm{Var}(A_{123}(t)) + \mathrm{Var}(A_{132}(t)) && \text{[Independence]} \\
&\quad = \Lambda_{123}(t) + \Lambda_{132}(t).
\end{aligned}$$

Fourth, consider the correlation between $S_2^{\mathbf{B}}(t)$ and $S_3^{\mathbf{A}}(t)$ (Eqs. (EC.13) and (EC.16)). The relevant processes are depicted in Figure EC.4. From the Figure, it is clear that by the same

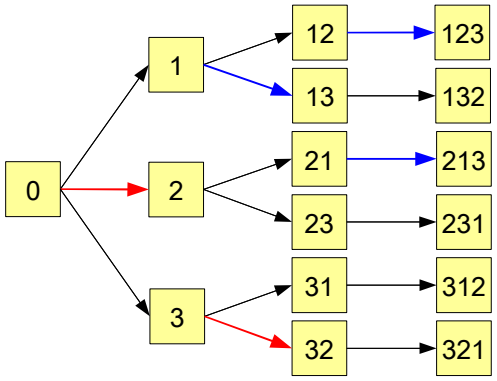

Figure EC.4 Red: Arcs used in constructing $\left|S_2^{\mathbf{B}}(t)\right|$. Blue: Arcs used in constructing $\left|S_3^{\mathbf{A}}(t)\right|$.

argument, the only correlation that can arise comes from the processes A_2 and A_{213}, and this gives us the covariance

$$\begin{aligned}
\mathrm{Cov}(\left|S_2^{\mathbf{B}}(t)\right|, \left|S_3^{\mathbf{A}}(t)\right|) &= \mathrm{Var}(A_{213}(t)) \\
&= \Lambda_{213}(t).
\end{aligned}$$

Finally, consider the covariance between $|S_2^{\mathbf{A}}(t)|$ and $|S_3^{\mathbf{B}}(t)|$ (Eqs. (EC.14) and (EC.15)). The relevant processes are depicted in Figure EC.5. Their covariance is

$$\begin{aligned}\mathrm{Cov}(|S_2^{\mathbf{A}}(t)|, |S_3^{\mathbf{B}}(t)|) &= \mathrm{Var}(A_{312}(t)) \\ &= \Lambda_{312}(t).\end{aligned}$$

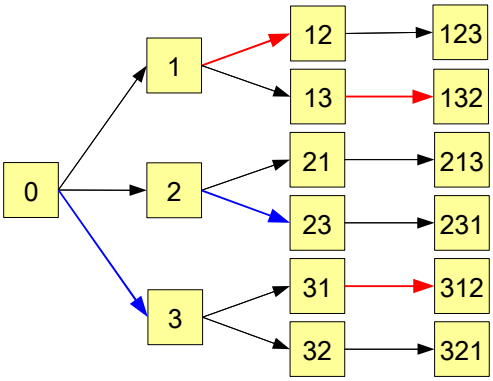

Figure EC.5 Red: Arcs used in constructing $|S_2^{\mathbf{A}}(t)|$. Blue: Arcs used in constructing $|S_3^{\mathbf{B}}(t)|$.

Appendix D: Hazard Rates and the Cox Proportional Hazards Model

For a nonnegative random variable T that has a density f_T and cumulative distribution function F_T, the *hazard rate function* of T is a non-negative valued function $h_T : \mathbb{R}_+ \to \mathbb{R}_+$ that is defined as

$$h_T(s) := \frac{f_T(s)}{1 - F(s)}.$$

The Cox proportional hazards model is a model of the relationship between a nonnegative random variable, T (which usually has the interpretation as the random time to an event) and a vector of predictor variables, which we will denote as $\boldsymbol{x} := (x_1, \ldots, x_N)$. The model assumes that these predictors have a multiplicative effect on the hazard rate of T, h_T. More precisely, it assumes that there is a baseline hazard rate function, $h_{T\mathbf{base}}$, such that h_T is given by

$$h_T(s) = h_{T\mathbf{base}}(s) e^{\boldsymbol{\beta}' \boldsymbol{x}}, \tag{EC.17}$$

where the vector of coefficients $\boldsymbol{\beta} := (\beta_1, \ldots, \beta_N)$, is estimated from data. The usual method of estimating these coefficients is through maximizing a quantity known as the partial likelihood. This model has several other attractive properties: It can handle right-censored data, and the estimation procedure for $\boldsymbol{\beta}$ does not require specification of the baseline hazard rate function $h_{T\mathbf{base}}$. We refer interested readers to Cox (1972) or Efron (1977) for details.

It is useful to compare the classical version of the proportional hazards model in (EC.17) from our dynamic version of the proportional hazards model (14). In the classical version, predictor variables can be real-valued or binary-valued, and are assumed to be known at time 0. In our model (14), the "predictor variables" are binary-valued and dynamically updated as the stochastic system evolves.