A Mathematical Model for Projecting the Replenishment of Compounds in a Sample Bank

Gilles Gnacadja
Research and Development Information Systems, Amgen, Inc.
One Amgen Center Drive, Thousand Oaks, California 91320-1799, USA
gilles.gnacadja@gmail.com

Mark Gulbronson
Discovery Technologies, Amgen, Inc.
One Amgen Center Drive, Thousand Oaks, California 91320-1799, USA

Published in *Chemometrics and Intelligent Laboratory Systems*
DOI 10.1016/j.chemolab.2013.03.004
Preprint revision F.07 2 May 2013

Abstract

Sample banks are repositories of compounds that are screened for therapeutic potential in the drug discovery process. Effectively managing a sample bank requires good intelligence to plan the storage of compounds and the workload of personnel and equipment. We help address this concern with a mathematical model to project compound depletion and replenishment as a function of screening activity and compound copy count. The model is deterministic and formulated in terms of probabilities. Its results show good agreement with those of randomization-based simulations of sample bank operations. This suggests that the model produces reliable projections.

Keywords. Sample Bank; Compound Management; High Throughput Screening; Inventory Replenishment; Probability.
Mathematics Subject Classification (2010): 97M60, 60C05, 05A15.

1 Introduction

A sample bank is a repository of compounds used in the drug discovery process. The number of compounds in the collection ranges from thousands to millions and the sample age ranges from days to decades. The drug discovery process begins with the selection of an unmet medical need. Scientists identify molecular targets or proteins believed to be critical to initiation or progression of a disease upon which a potential drug may have a therapeutic effect. Researchers then test hundreds of thousands of chemical compounds with the goal of finding one or more molecules that have desired biological activity. This is done in a series of incrementally stringent experiments known as a screening campaign. When a promising candidate is selected, chemists routinely synthesize a series of structurally similar compounds in an effort to increase potency or reduce unwanted side effects.

Screening campaigns nowadays are conducted in high throughput, using a set of methods, equipment, and techniques that emerged a few decades ago and has since undergone considerable evolution. It is the subject of several historical perspectives, experience reports and reviews, e.g. Pereira and Williams [6], Houston et al. [2], Mayr and Fuerst [4], and Mayr and Bojanic [3]. High throughput screening has become a central function in research laboratories and requires a materials management organization that is responsible for supplying test compounds. This organization must balance several constraints in response to changing scientific demand. These constraints include storage capacity, compound preservation, and staff

workload. In our case, the annual screening rate has changed significantly and the compound collection has grown tenfold in the last ten years. We store compounds in three forms, listed here in the order they are created. (SF stands for Storage Form.)

SF1: Solid powder

SF2: Frozen DMSO solution in large volume archive tubes for multiple use

SF3: Frozen DMSO solution in small volume operational tubes for single use

By default, sample requests are fulfilled from storage form SF3. The necessary operational tubes are thawed to prepare aliquots and any remaining solution is discarded. If operational tubes are not available, they are replenished from storage form SF2. The necessary archive tubes are thawed and any remaining solution is frozen (provided the acceptable limit of freeze-thaw cycles has not been reached). When archive tubes are depleted, they and operational tubes are replenished from storage form SF1. This second kind of replenishment is particularly labor intensive and time consuming because it requires precise weighing. We implemented the REMP 384 Tube Technology™ to reduce freeze-thaw cycles in storage form SF2 and to take advantage of the increased throughput of the automation. Some organizations do not use storage form SF3 and instead prepare aliquots from storage form SF2.

Prior to the development of the model presented in this paper, the number of operational copies stored was largely based on guesswork and available capacity. Reassessment was triggered when the operational store was approaching capacity more quickly than anticipated. Analysis of historical order data gave some indication of tube usage, but the data were incomplete. The goal was to create a model that predicts the number of compounds requiring replenishment as a function of variable collection size, screening rate, and operational tube copy count. Armed with this mathematical model, we can develop reliable projections on several figures of concern, such as:

- The impact of screening activity patterns on compound usage;
- The cost and workload associated with replenishment over time; and
- The operational cost of alternate configurations of the sample bank within and across storage sites.

We describe the model and illustrate its projections in Section 2, and discuss its implications in Section 3. We present and prove the mathematical results the model is based upon in Sections 4 and 5, respectively. In addition, we provide in a supplementary article the mathematical material required to prove the last of the five mathematical results.

2 Method and Illustrations

The model presented here is used for the replenishment of archive tubes from solid, i.e. of storage form SF2 from storage form SF1. For an organization, such as ours, that uses storage form SF3, *tube capacity* in our discussion refers to the number of operational tubes (SF3) that are made from each archive tube (SF2). For an organization that does not use storage form SF3, *tube capacity* will stand for the number of aliquots that can be extracted from each archive

tube (SF2). We assume that all tubes are full before the first screening campaign, and that archive tubes are replenished to capacity as they are depleted. Because the first two stages of a screening campaign are responsible for the bulk of compound consumption, we base our model on the simplifying consideration that a campaign consists of a *primary stage* followed by a *secondary stage*. These stages go by various names in drug discovery, e.g. hit confirmation for the primary stage and potency determination for the secondary stage. We list model parameters in Table 1.

Notation	Description
n	Number of compounds in library
m_1	Number of compounds selected at the primary stage of a campaign
m_2	Number of compounds selected at the secondary stage of a campaign
p_1	Probability that a compound is selected at the primary stage of a campaign
p_2	Probability that a compound already selected at the primary stage of a campaign is selected at the secondary stage of the same campaign
q_1	$q_1 = 1 - p_1$
q_2	$q_2 = 1 - p_2$
t	Tube capacity
r	Number of campaigns
k	Number of times a compound is selected

Table 1: Model parameters

We introduce some terminology in order to facilitate our exposition. An r-scenario will be a sequence of r campaigns. More specifically:

- A *type-1 r-scenario* is a succession of $r-1$ screening campaigns followed by the primary stage of an r-th campaign.
- A *type-2 r-scenario* is a succession of r screening campaigns.

Thus, a type-2 r-scenario consists of r primary screening stages and r secondary screening stages in alternating succession, and a type-1 r-scenario consists of a type-2 $(r-1)$-scenario followed by one primary screening stage.

The parameters listed in Table 1 satisfy the following conditions.

- $n, m_1, m_2, t, r, k \in \mathbb{Z}$ and $p_1, p_2, q_1, q_2 \in \mathbb{R}$
- $1 \leqslant m_1 \leqslant m_2 \leqslant n$, $t \geqslant 1$, and $r \geqslant 1$
- $0 \leqslant k \leqslant 2r-1$ in a type-1 r-scenario and $0 \leqslant k \leqslant 2r$ in a type-2 r-scenario
- $0 \leqslant p_1, p_2, q_1, q_2 \leqslant 1$

Generically, but not necessarily, we have $p_1 = m_1/n$ and $p_2 = m_2/m_1$. We work mainly with the probabilities p_1 and p_2; the numbers m_1 and m_2 are not explicitly involved in the development of the model.

Notation	Description
$f_1(p_1, p_2, r, t)$	Probability that, at the completion of a type-1 r-scenario, an initially full tube of capacity t requires replenishment
$f_2(p_1, p_2, r, t)$	Probability that, at the completion of a type-2 r-scenario, an initially full tube of capacity t requires replenishment
$g(p_1, p_2, r, k)$	Probability that, in a type-2 r-scenario, a compound is selected precisely k times
$h(p_1, p_2, r, t, s)$	Probability that, in a type-2 r-scenario, the number of times a compound is selected is congruent to s modulo t ($s = 0, \ldots, t-1$)

Table 2: Probability functions in the model

We define several probability functions in Table 2. With the probability function f_i, where $i = 1$ or $i = 2$, we get that $f_i(p_1, p_2, r, t) \cdot n$ is the projected number of compounds that need replenishment upon the completion of a type-i r-scenario. Therefore the number of compounds that need replenishment because of an r-th campaign is projected to be

$$F(n, p_1, p_2, r, t) \quad := \quad \big(f_1(p_1, p_2, r, t) + f_2(p_1, p_2, r, t)\big) \cdot n \; . \tag{1}$$

Figure 1 shows the evolution of $F(n, p_1, p_2, r, t)$ with respect to the number r of campaigns for particular values of tube capacity t and of the numbers of compounds in the library and selected at the primary and secondary stages. In order to evaluate the pertinence of this deterministic model to sample bank operations, we simulated screening campaigns by random calculations. The outcome, shown in Figure 2, suggests that the numbers the deterministic model produces are reasonable projections of what to expect in reality. The deterministic model provides reproducibility of what one would obtain by averaging all random simulations. It also gives the ability to obtain derived projections by analytical means (e.g. Results 4 and 5), which one cannot obtain systematically from random simulations alone.

Following is a description of how screening campaigns are simulated by random calculations:

1. We number all compounds consecutively from 1 to n. Each is assigned a selection count, which initially is zero.

2. To simulate the primary stage of a campaign, we randomly select a list of m_1 numbers in the range $1, \ldots, n$. The list is selected by a pseudo-random selection routine based on a uniform distribution; it is unsorted. The selection count of each selected compound is incremented by one.

3. To simulate the secondary stage, we select the first m_2 entries of the above list. Again, the selection count of each selected compound is incremented by one.

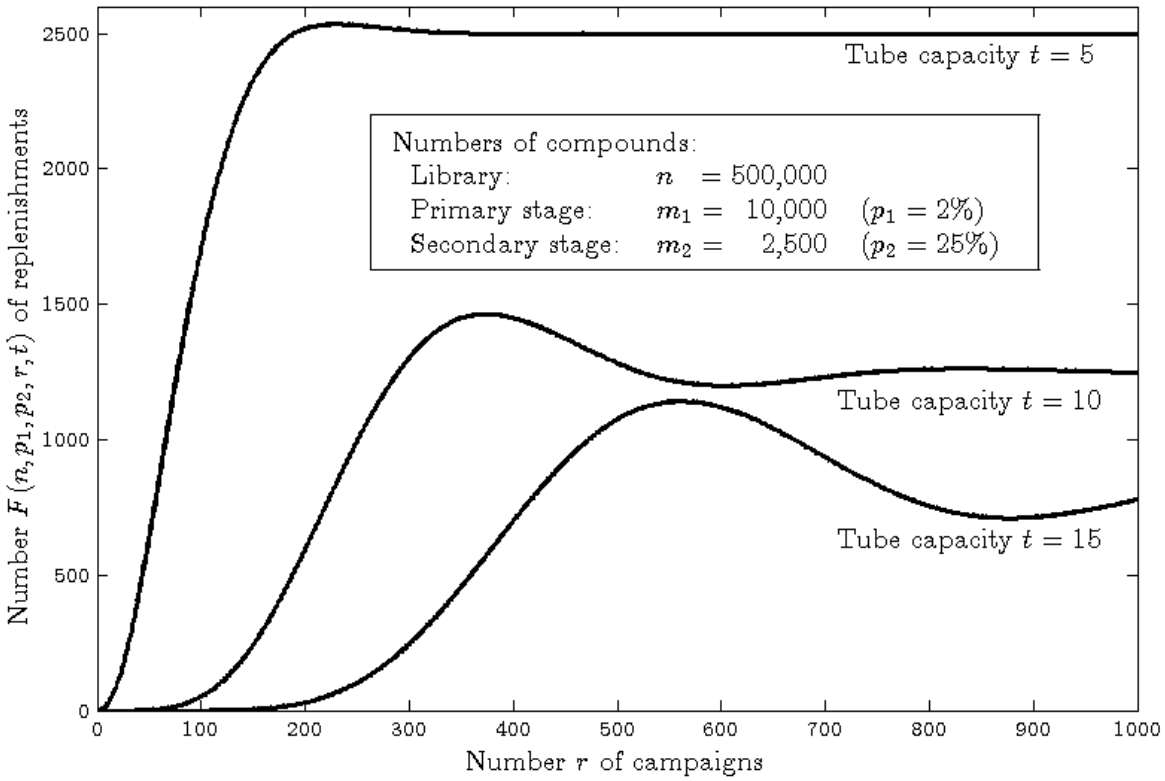

Figure 1: This figure shows the number of compounds that require replenishment as a result of running a campaign (both primary and secondary stages), as projected by the deterministic model of Equation (1). All parameters are specified on the figure and axes labels. The curve for a tube capacity of one has the distinct feature of being a horizontal line at the level $F(n, p_1, p_2, r, 1) = np_1(1 + p_2) = m_1 + m_2 = 12{,}500$. It is not shown here because the significant discrepancy in vertical scale would lessen the features of the other curves.

4. The number of required replenishments is found as the number of compounds whose selection count has become a multiple of the tube capacity at the primary stage or at the secondary stage. A compound is counted twice if the multiplicity condition is satisfied at both stages, a possibility, and in fact a certainty, if and only the tube capacity is $t = 1$.

Calculating F from Equation (1) requires f_1 and f_2. The functions f_1 and f_2 have simple expressions in terms of the function h, which are recorded in Results 1 and 2, respectively. We present in Result 3 an iterative method to calculate the function h, and thereafter in Equation (2) an alternate expression of the function h in terms of the function g.

We provide in Result 4 a very simple expression for the long-term value of F, i.e. the steady-state number of compounds that require replenishment. If the probabilities p_1 and p_2 have their generic values $p_1 = m_1/n$ and $p_2 = m_2/m_1$, then this steady-state number is $(m_1 + m_2)/t$. It is noteworthy that this number is independent of n. Thus, the model projects that one should not expect to affect the long-term pace of replenishment in an existing sample bank solely by increasing the size of its library.

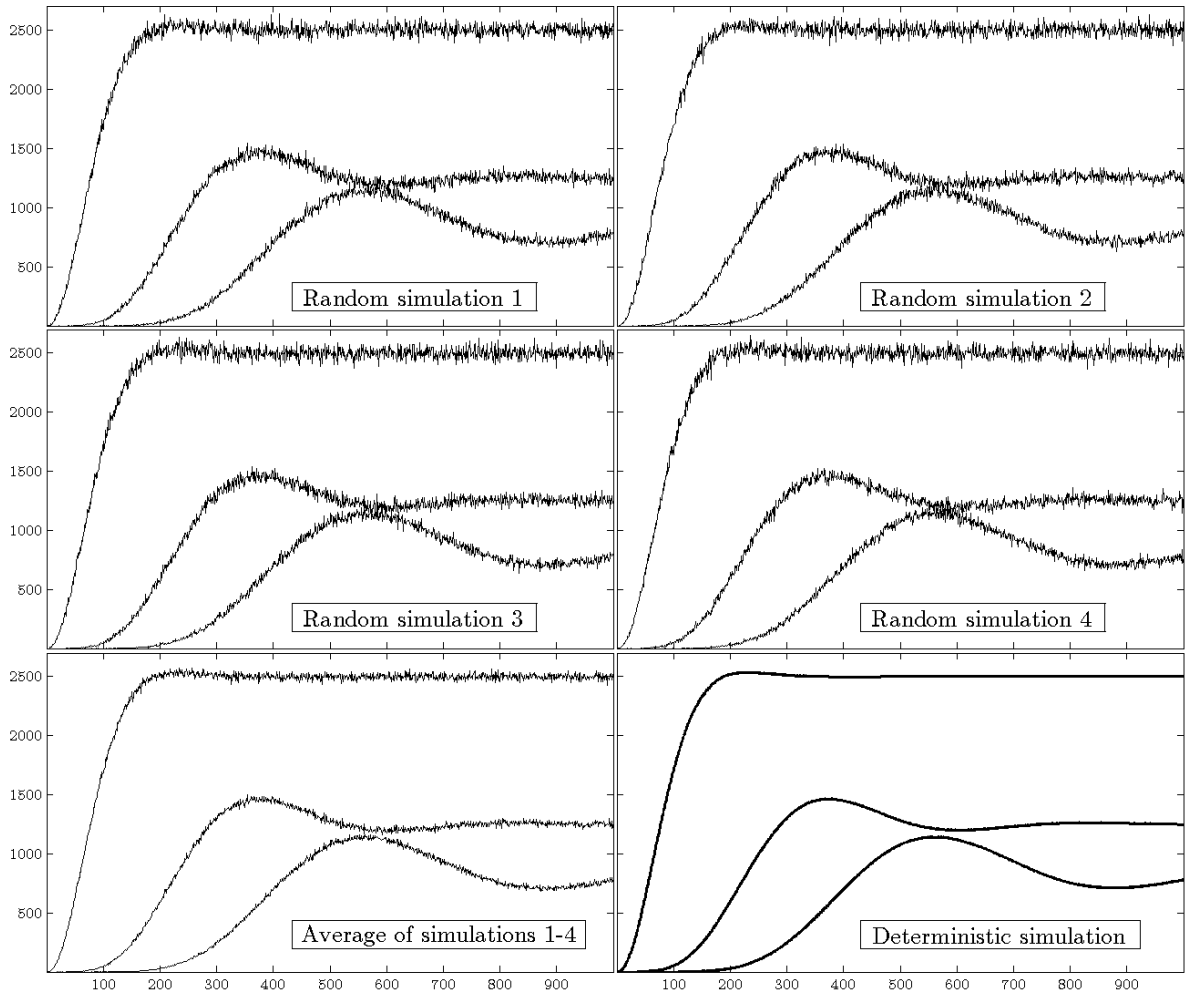

Figure 2: The six panels on this figure show projections of the number of compounds that require replenishment as a result of running a campaign. The parameters are not noted on the figure in the interest of clearness, but the axes and the fixed parameters are exactly as on Figure 1. In particular, the upper, middle, lower curves in each panel are for the tube capacities $t = 5, 10, 15$, respectively. The four upper panels show projections based on random calculations. The lower left panel shows the average of the four random calculations. The lower right panel is the same as Figure 1 and is included here for convenient visual comparison. The apparent agreement between the six panels and the even closer agreement between the two lower panels suggest that the deterministic model generates reasonable projections.

Result 5 concerns another practical aspect of long-term behavior. It addresses the question of when the projected number of replenishments becomes confined within a prescribed interval around steady-state, i.e. what reaching steady state means in practice. These are valuable projections, for instance when planning for a new sample bank. Figure 3 illustrates such projections.

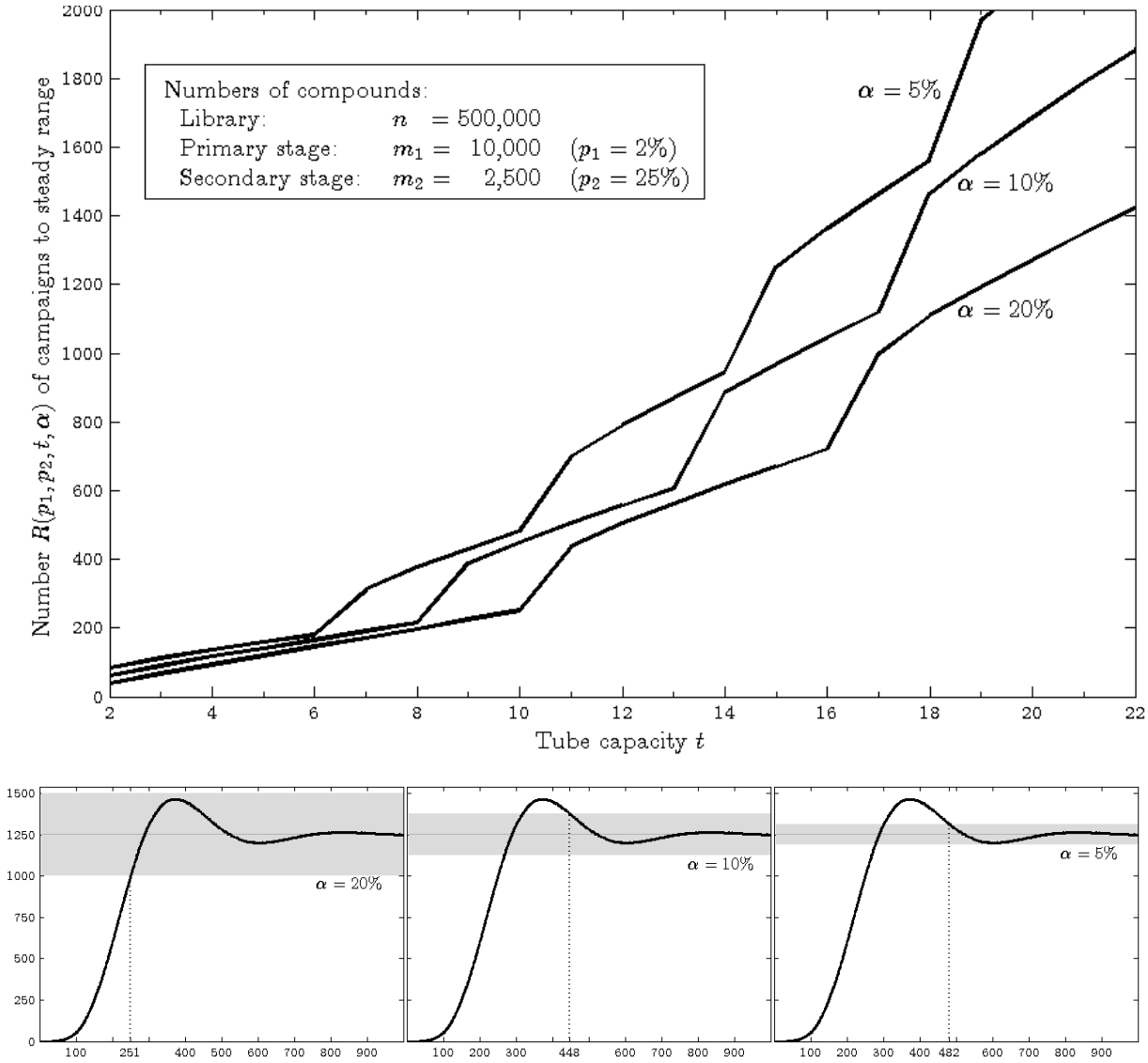

Figure 3: The panels on this figure illustrate the concept of a steady range. A steady range is an interval centered on the steady state. In practice, approaching the steady state means entering and staying in a steady range of prescribed width. The upper panel shows $R(p_1, p_2, t, \alpha)$, the number of the campaign projected to be the entry point into a steady range of replenishments, as a function of the tube capacity t and for selected values of the steady range width factor α. This factor is the ratio of the half-width of the steady range to the steady state. The function $R(p_1, p_2, t, \alpha)$ is calculated using Result 5. For instance, with a tube capacity of $t = 10$, the projection is that starting with campaigns 251, 448, 482, the number of replenishments should stay within 20%, 10%, 5%, respectively, of steady state. These three cases are illustrated on the three lower panels with the projected replenishment curve for $t = 10$ (already shown in Figure 1) and the respective steady ranges and entry points.

3 Discussion

Several factors affect the workload and cost of compound replenishment, including the number of compound copy counts in the operational store and the relative locations of the archival

and operational stores. The model presented here provides the raw data needed to reliably make workload and cost projections. For instance, we see on Figure 1 that with a high copy count, the need for replenishment is low or even nonexistent for a transcient period of time, the duration of which depends on the intensity of screening activity. But a high copy count requires more storage capacity in the operational store. The model can help balance the ongoing and upfront monetary and physical-space costs associated with these competing considerations.

Figure 1 shows that the number of replenishments can feature damped oscillations as it settles to steady state. This may be explained heuristically by the fact that we start the simulations from a state where all tubes are full. Indeed, because of this, there is an overshoot transition from no-to-few compounds needing replenishments, to large numbers of compounds needing replenishment at about the same time. This causes the first bump, and the damped oscillations follow to stabilize the demand to steady state. A more mathematically rigorous explanation is possible. One such explanation involves the fact that probability distributions constructed by iterative convolution can be multimodal. This is beyond the scope of this paper. For the interested reader, we suggest Odlyzko and Richmond [5] and the cited and citing literature, plus Gnacadja [1] as a bridge between that body of work and the present paper. The specifics for the model herein could be the subject of further investigations.

4 Mathematical Results

This section consists of Results 1-5. One can calculate F with Equation (1) if f_1 and f_2 are known. Results 1 and 2 show how to calculate f_1 and f_2, respectively, provided h is known. Result 3 shows how to calculate h. Results 4 and 5 are concerned with the long-term behavior of F. All results are proved in Section 5.

Result 1. *The probability $f_1(p_1, p_2, r, t)$ has the following expressions.*

$$\begin{aligned}
\textit{For } r = 1 \textit{ and } t \geqslant 1 : \quad f_1(p_1, p_2, 1, t) &= \begin{cases} p_1 & \textit{for } t = 1 \\ 0 & \textit{for } t \geqslant 2 \end{cases} \\
\textit{For } r \geqslant 1 \textit{ and } t = 1 : \quad f_1(p_1, p_2, r, 1) &= p_1 \\
\textit{For } r \geqslant 2 \textit{ and } t \geqslant 2 : \quad f_1(p_1, p_2, r, t) &= p_1\, h(p_1, p_2, r-1, t, t-1)
\end{aligned}$$

Result 2. *The probability $f_2(p_1, p_2, r, t)$ has the following expressions.*

$$\begin{aligned}
\textit{For } r = 1 \textit{ and } t \geqslant 1 : \quad f_2(p_1, p_2, 1, t) &= \begin{cases} p_1\, p_2 & \textit{for } t = 1 \textit{ and } t = 2 \\ 0 & \textit{for } t \geqslant 3 \end{cases} \\
\textit{For } r \geqslant 1 \textit{ and } t = 1 : \quad f_2(p_1, p_2, r, 1) &= p_1\, p_2 \\
\textit{For } r \geqslant 2 \textit{ and } t \geqslant 2 : \quad f_2(p_1, p_2, r, t) &= p_1\, p_2\, h(p_1, p_2, r-1, t, t-2)
\end{aligned}$$

We need some preparation in order to state Result 3. Let

$$a_0 := q_1\,, \quad a_1 := p_1\, q_2\,, \quad a_2 := p_1\, p_2$$

and

$$h(p_1, p_2, r, t) := \big(h(p_1, p_2, r, t, s)\big)_{s=0,\ldots,t-1}\,.$$

Observe that $a_0 + a_1 + a_2 = 1$ and that $h(p_1, p_2, r, t)$ is a probability t-vector.
Let the $t \times t$ matrix $H(p_1, p_2, t)$ be given as follows.

$$
\begin{aligned}
H(p_1, p_2, 1) &= \begin{pmatrix} 1 \end{pmatrix} \\
H(p_1, p_2, 2) &= \begin{pmatrix} a_0 + a_2 & a_1 \\ a_1 & a_0 + a_2 \end{pmatrix} \\
H(p_1, p_2, t) &= \begin{pmatrix}
a_0 & a_1 & a_2 & & & \\
 & a_0 & a_1 & a_2 & & \\
 & & \ddots & \ddots & \ddots & \\
 & & & a_0 & a_1 & a_2 \\
a_2 & & & & a_0 & a_1 \\
a_1 & a_2 & & & & a_0
\end{pmatrix} \quad \text{if } t \geqslant 3
\end{aligned}
$$

Result 3. *For any $t \geqslant 1$, the probability vectors $h(p_1, p_2, r, t)$ are given iteratively as follows.*

$$
\begin{aligned}
h(p_1, p_2, 1, 1) &= (1) \\
h(p_1, p_2, 1, 2) &= (a_0 + a_2, a_1) \\
h(p_1, p_2, 1, t) &= (a_0, a_1, a_2, \underbrace{0, \ldots, 0}_{t-3}) \quad \textit{for } t \geqslant 3 \\
h(p_1, p_2, r, t) &= h(p_1, p_2, r-1, t) \cdot H(p_1, p_2, t) \quad \textit{for } r \geqslant 2
\end{aligned}
$$

Observe that $H(p_1, p_2, t)$ is the circulant matrix associated with the vector $h(p_1, p_2, 1, t)$. This is to say that in $H(p_1, p_2, t)$, the top row is $h(p_1, p_2, 1, t)$ and each subsequent row is obtained from the preceding one by circularly shifting the entries rightward.

We have in Equation (2) another expression for h. It is in terms of g and follows from the very definition of g and h in Table 2.

$$
h(p_1, p_2, r, t, s) = \sum_{\substack{0 \leqslant k \leqslant 2r \\ k \in s + t\mathbb{Z}}} g(p_1, p_2, r, k) = \sum_{\mu=0}^{\texttt{floor}((2r-s)/t)} g(p_1, p_2, r, \mu t + s) \; . \tag{2}
$$

The recursive equation for h in Result 3 is more suited than Equation (2) for calculations. But as we see next, Equation (2) is useful to justify that the recursion starts as asserted. Let

$$
g(p_1, p_2, r) := \big(g(p_1, p_2, r, k)\big)_{k=0,\ldots,2r} \; .
$$

By interpreting the definition of g from Table 2, we obtain

$$
g(p_1, p_2, 1) = (a_0, a_1, a_2) \; .
$$

Then, by applying Equation (2) with $r = 1$, we obtain the expressions of the vector $h(p_1, p_2, 1, t)$ asserted in Result 3.

We now have all the results to lay out the following calculation roadmap. It assumes that the the parameters p_1, p_2, r, t, and n are given.

1. Use Result 3 to calculate the probability h.
2. With the probability h, use Results 1 and 2 to calculate the probabilities f_1 and f_2, respectively.
3. With the probabilities f_1 and f_2, use Equation (1) to calculate F, the projected number of compounds that require replenishment.

We implemented this roadmap in a simulation software. The execution time is barely noticeable.

In preparation for Results 4 and 5, we set

$$F_{\mathrm{ss}}(n, p_1, p_2, t) \quad := \quad n\, p_1\, (1 + p_2)\, /\, t\ .$$

Result 4. $\lim_{r\to\infty} F(n, p_1, p_2, r, t) \ = \ F_{\mathrm{ss}}(n, p_1, p_2, t)\ .$

Thus, $F_{\mathrm{ss}}(n, p_1, p_2, t)$ is the projected steady-state number of compounds that require replenishment. If the probabilities p_1 and p_2 have their generic values $p_1 = m_1/n$ and $p_2 = m_2/m_1$, then

$$F_{\mathrm{ss}}(n, p_1, p_2, t) \ = \ F_{\mathrm{ss}}(n, m_1/n, m_2/m_1, t) \ = \ (m_1 + m_2)\, /\, t\ ;$$

the steady-state number depends only on the tube capacity t and the numbers m_1 and m_2 of compounds selected at the primary and secondary stages. It does not depend on the size n of the library (as long as $n \geqslant m_1$).

In practice, just knowing the steady state is not enough. We also need to know when we reach steady state, or more accurately, when we reach and stay within a prescribed interval around the steady state.

For $\alpha \in \mathbb{R}_{>0}$, the positive integer $R(p_1, p_2, t, \alpha)$ be uniquely defined by the following conditions.

$$\exists\, r < R(p_1, p_2, t, \alpha) : \ \left|F(n, p_1, p_2, r, t) - F_{\mathrm{ss}}(n, p_1, p_2, t)\right| > \alpha\, F_{\mathrm{ss}}(n, p_1, p_2, t) \tag{3a}$$

$$\forall\, r \geqslant R(p_1, p_2, t, \alpha)\ , \ \left|F(n, p_1, p_2, r, t) - F_{\mathrm{ss}}(n, p_1, p_2, t)\right| \leqslant \alpha\, F_{\mathrm{ss}}(n, p_1, p_2, t) \tag{3b}$$

This definition says that, starting with campaign number $R(p_1, p_2, t, \alpha)$, and not before, all replenishment counts are projected to be within a factor α of the steady state $F_{\mathrm{ss}}(n, p_1, p_2, t)$.

Now let $\kappa(p_1, p_2, t)$ be defined as follows.

$$\begin{aligned} \kappa(p_1, p_2, 2) \quad &:= \quad |1 - 2a_1| \\ \kappa(p_1, p_2, t) \quad &:= \quad \max_{1\leqslant s\leqslant t-1} \left| a_0 + a_1 \exp\left(\frac{2s\pi\mathrm{i}}{t}\right) + a_2 \exp\left(\frac{4s\pi\mathrm{i}}{t}\right)\right| \quad \text{for } t \geqslant 3 \end{aligned}$$

Then, for $t \geqslant 2$, let

$$\hat{R}(p_1, p_2, t, \alpha) \quad := \quad 1 + \frac{\ln\left(\dfrac{\alpha}{\sqrt{(t-1)t}}\right)}{\ln\left(\kappa(p_1, p_2, t)\right)}\ .$$

Result 5. *Let* $\alpha \in \mathbb{R}_{>0}$.

If $\alpha \geqslant \sqrt{(t-1)t}$, *and in particular if* $t = 1$, *then* $R(p_1, p_2, t, \alpha) = 1$.

Suppose $t \geqslant 2$ *and* $\alpha < \sqrt{(t-1)t}$. *Then* $\hat{R}(p_1, p_2, t, \alpha) > 1$ *and*

$$\forall\, r \geqslant \hat{R}(p_1, p_2, t, \alpha)\,,\ \left|F(n, p_1, p_2, r, t) - F_{\text{ss}}(n, p_1, p_2, t)\right| \ \leqslant\ \alpha\, F_{\text{ss}}(n, p_1, p_2, t)\ .$$

Equivalently,

$$R(p_1, p_2, t, \alpha) \quad \leqslant \quad \hat{R}(p_1, p_2, t, \alpha)\ .$$

Result 5 gives rise to an algorithm for finding $R(p_1, p_2, t, \alpha)$ in the nontrivial case. Starting with $r = \texttt{ceiling}\left(\hat{R}(p_1, p_2, t, \alpha)\right)$, decrement r while the condition

$$r \geqslant 1 \quad \text{and} \quad \left|F(n, p_1, p_2, r, t) - F_{\text{ss}}(n, p_1, p_2, t)\right| \leqslant \alpha\, F_{\text{ss}}(n, p_1, p_2, t)$$

is satisfied. Then $R(p_1, p_2, t, \alpha)$ equals the value that r had just before the condition failed.

It is easy to satisfy Condition (3a). However, without a finite upper bound for $R(p_1, p_2, t, \alpha)$, e.g. as provided by Result 5, one cannot be certain to satisfy Condition (3b).

5 Proofs of Mathematical Results

Proving Results 1 and 2 amounts to articulating what they say. We use separately published work of the first author to prove Results 3 and 4. The proof of Result 5 requires a stronger form of this published work, which we provide in the supplementary article.

Proof of Result 1. We provide justification for each of the three equations in the result.

$f_1(p_1, p_2, 1, 1) = p_1$ and $f_1(p_1, p_2, 1, t) = 0$ for $t \geqslant 2$:

A compound requires replenishment at the completion of the primary stage of the first campaign if and only if the following two conditions are met:

- The compound is selected at that stage, an event probability of p_1.
- The tube capacity is one.

$f_1(p_1, p_2, r, 1) = p_1$ for $t = 1$ and $r \geqslant 1$:

With a tube capacity of one, a compound requires replenishment at the completion of the primary stage of any campaign if and only if it is selected at that stage, an event of probability p_1.

Equation for $f_1(p_1, p_2, r, t)$ for $t \geqslant 2$ and $r \geqslant 2$:

A type-1 r-scenario consists of a type-2 $(r-1)$-scenario followed by the primary stage of an r-th campaign. A compound requires replenishment at the completion of a type-1 r-scenario if and only if the number of times the compound has been selected in the type-1 r-scenario has just become a positive multiple of t. This means that the following two conditions are met:

- The number of times the compound was selected in the type-2 $(r-1)$-scenario is congruent to $t-1$ modulo t, an event of probability $h(p_1, p_2, r-1, t, t-1)$.
- The compound was selected at the primary stage of the r-th campaign, an event of probability p_1.

This completes the proof of Result 1.

Proof of Result 2. We provide justification for each of the three equations in the result.

$f_2(p_1, p_2, 1, t) = p_1\, p_2$ for $t = 1, 2$ and $f_2(p_1, p_2, 1, t) = 0$ for $t \geqslant 3$:

A compound requires replenishment at the completion of the first campaign if and only if the following two conditions are met:

- The compound is selected at both the primary and the secondary stages of the campaign, an event of probability $p_1\, p_2$.
- The tube capacity is either one or two.

$f_2(p_1, p_2, r, 1) = p_1\, p_2$ for $t = 1$ and $r \geqslant 1$:

With a tube capacity of one, a compound requires replenishment at the completion of any campaign if and only if it is selected at both the primary and the secondary stages of that campaign, an event of probability $p_1\, p_2$.

Equation for $f_2(p_1, p_2, r, t)$ for $t \geqslant 2$ and $r \geqslant 2$:

A type-2 r-scenario consists of a type-2 $(r-1)$-scenario followed by an r-th campaign. A compound requires replenishment at the completion of a type-2 r-scenario if and only if the number of times the compound has been selected in the type-2 r-scenario has just become a positive multiple of t. This means that the following two conditions are met:

- The number of times the compound was selected in the type-2 $(r-1)$-scenario is congruent to $t-2$ modulo t, an event of probability $h(p_1, p_2, r-1, t, t-2)$.
- The compound was selected at both the primary and the secondary stages of the r-th campaign, an event of probability $p_1\, p_2$.

This completes the proof of Result 2.

Proof of Results 3 and 4. The work of the first author in Gnacadja [1] is directly applicable. With φ and Φ as defined in this reference, and by using the very definitions of h and H herein, we have

$$h(p_1, p_2, r, t) \;=\; \varphi\big((a_0, a_1, a_2), r, t\big) \quad \text{and} \quad H(p_1, p_2, t) \;=\; \Phi\big((a_0, a_1, a_2), 1, t\big)\ .$$

The iterative equation in Result 3 is an instance of Proposition 2 from the reference. With Theorem 1 from the same reference, we have

$$\lim_{r\to\infty} h(p_1, p_2, r, t, s) \quad = \quad 1/t$$

for any $s = 0, \ldots, t-1$. Applying this for $s = t-1$ and $s = t-2$ yields Result 4.

Proof of Result 5. We will use the supplementary article. Consider the polynomial

$$Q_{p_1,p_2,t}(X) \;=\; \begin{cases} a_0 + a_2 + a_1 X & \text{if } t = 2 \\ a_0 + a_1 X + a_2 X^2 & \text{if } t \geqslant 3\,. \end{cases}$$

With $f := h(p_1, p_2, 1, t)$, which is known explicitly from Result 3, we match the relevant notations of the supplementary article as follows.

$$\begin{aligned} h(p_1, p_2, r, t) &= \varphi(f, r, t) \\ Q_{p_1,p_2,t}(X) &= P_{f,t}(X) \\ \kappa(p_1, p_2, t) &= \gamma(f, t) \end{aligned}$$

We then apply Theorem 1 from the supplementary article. We have $\kappa(p_1, p_2, t) < 1$ and

$$\forall\, r \geqslant 1\,,\ \big\|h(p_1, p_2, r, t) - (1/t)(\underbrace{1, \ldots, 1}_{t})\big\|_2 \;\leqslant\; \big(\kappa(p_1, p_2, t)\big)^r \sqrt{(t-1)/t}\,.$$

On another hand, for $r \geqslant 2$ and $t \geqslant 2$, we have

$$\begin{aligned} F(n, p_1, p_2, r, t) &= n\left(f_1(p_1, p_2, r, t) + f_2(p_1, p_2, r, t)\right) \\ &= n\, p_1 \left(h(p_1, p_2, r-1, t, t-1) + p_2\, h(p_1, p_2, r-1, t, t-2)\right)\,, \end{aligned}$$

whence

$$\begin{aligned} &F(n, p_1, p_2, r, t) - F_{\text{ss}}(n, p_1, p_2, t) \\ &\quad= n\, p_1 \left(h(p_1, p_2, r-1, t, t-1) + p_2\, h(p_1, p_2, r-1, t, t-2) - (1 + p_2)\,/\,t\right) \\ &\quad= n\, p_1 \left(\big(h(p_1, p_2, r-1, t, t-1) - 1/t\big) + p_2 \left(h(p_1, p_2, r-1, t, t-2) - 1/t\right)\right)\,. \end{aligned}$$

Then,

$$\begin{aligned} &\big|F(n, p_1, p_2, r, t) - F_{\text{ss}}(n, p_1, p_2, t)\big| \\ &\quad\leqslant\; n\, p_1 \left(\big|h(p_1, p_2, r-1, t, t-1) - 1/t\big| + p_2 \left|h(p_1, p_2, r-1, t, t-2) - 1/t\right|\right) \\ &\quad\leqslant\; n\, p_1\, (1 + p_2)\, \big\|h(p_1, p_2, r-1, t) - (1/t)(1, \ldots, 1)\big\|_2 \\ &\quad\leqslant\; n\, p_1\, (1 + p_2)\, \big(\kappa(p_1, p_2, t)\big)^{r-1} \sqrt{(t-1)/t} \\ &\quad=\; F_{\text{ss}}(n, p_1, p_2, t)\, \big(\kappa(p_1, p_2, t)\big)^{r-1} \sqrt{(t-1)t}\,. \end{aligned}$$

Therefore,

$$\begin{aligned} &\big|F(n, p_1, p_2, r, t) - F_{\text{ss}}(n, p_1, p_2, t)\big| \;\leqslant\; \alpha\, F_{\text{ss}}(n, p_1, p_2, t) \\ &\qquad\Leftarrow\; \big(\kappa(p_1, p_2, t)\big)^{r-1} \sqrt{(t-1)t} \;\leqslant\; \alpha \\ &\qquad\Leftrightarrow\; \big(\kappa(p_1, p_2, t)\big)^{r-1} \;\leqslant\; \alpha\,/\,\sqrt{(t-1)t} \\ &\qquad\Leftrightarrow\; (r-1)\, \ln\big(\kappa(p_1, p_2, t)\big) \;\leqslant\; \ln\big(\alpha\,/\,\sqrt{(t-1)t}\big) \\ &\qquad\Leftrightarrow\; r - 1 \;\geqslant\; \frac{\ln\big(\alpha\,/\,\sqrt{(t-1)t}\big)}{\ln\big(\kappa(p_1, p_2, t)\big)} \\ &\qquad\Leftrightarrow\; r \;\geqslant\; \hat{R}(p_1, p_2, t, \alpha)\,. \end{aligned}$$

The proof of Result 5 is thus complete.

Acknowledgments

The authors gratefully acknowledge their colleague Steve Young for helping articulate the context around screening in the introduction, and Editor Clifford H. Spiegelman, Associate Editor Tom Burr, and the two anonymous reviewers for helping improve the paper overall.

References

[1] G. Gnacadja, *Asymptotic Equidistribution of Congruence Classes with respect to the Convolution Iterates of a Probability Vector*, Statistics and Probability Letters **82** (2012), no. 10, 1849–1852, DOI 10.1016/j.spl.2012.05.025.

[2] J. G. Houston et al., *Case Study: Impact of Technology Investment on Lead Discovery at Bristol-Myers Squibb, 1998-2006*, Drug Discovery Today **13** (2008), no. 1-2, 44–51, DOI 10.1016/j.drudis.2007.11.004.

[3] L. M. Mayr and D. Bojanic, *Novel Trends in High-Throughput Screening*, Current Opinion in Pharmacology **9** (2009), no. 5, 580–588, DOI 10.1016/j.coph.2009.08.004.

[4] L. M. Mayr and P. Fuerst, *The Future of High-Throughput Screening*, Journal of Biomolecular Screening **13** (2008), no. 6, 443–448, DOI 10.1177/1087057108319644.

[5] A. M. Odlyzko and L. B. Richmond, *On the Unimodality of High Convolutions of Discrete Distributions*, The Annals of Probability **13** (1985), no. 1, 299–306, DOI 10.1214/aop/1176993082.

[6] D. A. Pereira and J. A. Williams, *Origin and Evolution of High Throughput Screening*, British Journal of Pharmacology **152** (2007), no. 1, 53–61, DOI 10.1038/sj.bjp.0707373.