Manufacturing & Service Operations Management

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Disease Bundling or Specimen Bundling? Cost- and Capacity-Efficient Strategies for Multidisease Testing with Genetic Assays

Douglas R. Bish, Ebru K. Bish, Hussein El Hajj

To cite this article:
Douglas R. Bish, Ebru K. Bish, Hussein El Hajj (2023) Disease Bundling or Specimen Bundling? Cost- and Capacity-Efficient Strategies for Multidisease Testing with Genetic Assays. Manufacturing & Service Operations Management

Published online in Articles in Advance 08 Aug 2023

. https://doi.org/10.1287/msom.2022.0296

https://pubsonline.informs.org/journal/msom

MANUFACTURING & SERVICE OPERATIONS MANAGEMENT
Articles in Advance, pp. 1–22
ISSN 1523-4614 (print), ISSN 1526-5498 (online)

Disease Bundling or Specimen Bundling? Cost- and Capacity-Efficient Strategies for Multidisease Testing with Genetic Assays

Douglas R. Bish,[a] Ebru K. Bish,[a,*] Hussein El Hajj[b]

[a] Department of Information Systems, Statistics, and Management Science, University of Alabama, Tuscaloosa, Alabama 35487; [b] Department of Information Systems and Analytics, Santa Clara University, Santa Clara, California 95053

*Corresponding author

Contact: drbish@cba.ua.edu, https://orcid.org/0000-0002-1376-8268 (DRB); ekbish@cba.ua.edu, https://orcid.org/0000-0002-2653-7135 (EKB); helhajj@scu.edu, https://orcid.org/0000-0002-8364-4586 (HEH)

Received: June 16, 2022
Revised: December 11, 2022; March 27, 2023; June 16, 2023
Accepted: July 2, 2023
Published Online in Articles in Advance: August 8, 2023

https://doi.org/10.1287/msom.2022.0296

Abstract. ***Problem definition*:** Infectious disease screening can be expensive and capacity constrained. We develop cost- and capacity-efficient testing designs for multidisease screening, considering (1) *multiplexing* (disease bundling), where one assay detects multiple diseases using the same specimen (e.g., nasal swabs, blood), and (2) *pooling* (specimen bundling), where one assay is used on specimens from multiple subjects bundled in a testing pool. A testing design specifies an *assay portfolio* (mix of single-disease/multiplex assays) and a *testing method* (pooling/individual testing per assay). ***Methodology/results*:** We develop novel models for the nonlinear, combinatorial multidisease testing design problem: a deterministic model and a distribution-free, robust variation, which both generate *Pareto* frontiers for cost- and capacity-efficient designs. We characterize structural properties of optimal designs, formulate the deterministic counterpart of the robust model, and conduct a case study of respiratory diseases (including coronavirus disease 2019) with overlapping clinical presentation. ***Managerial implications*:** Key drivers of optimal designs include the assay cost function, the tester's preference toward cost versus capacity efficiency, prevalence/coinfection rates, and for the robust model, prevalence uncertainty. When an optimal design uses multiple assays, it does so in conjunction with pooling, and it uses individual testing for at most one assay. Although prevalence uncertainty can be a design hurdle, especially for emerging or seasonal diseases, the integration of multiplexing and pooling, and the ordered partition property of optimal designs (under certain coinfection structures) serve to make the design more structurally robust to uncertainty. The robust model further increases robustness, and it is also practical as it needs only an uncertainty set around each disease prevalence. Our Pareto designs demonstrate the cost versus capacity trade-off and show that multiplexing-only or pooling-only designs need not be on the Pareto frontier. Our case study illustrates the benefits of optimally integrated designs over current practices and indicates a low price of robustness.

Funding: This work was supported by the National Science Foundation [Grant 1761842].
Supplemental Material: The online appendix is available at https://doi.org/10.1287/msom.2022.0296.

Keywords: testing design • public health screening • multiplex assay • pooled testing • assay portfolio • partition problem

1. Introduction and Motivation

Infectious disease screening improves health outcomes by identifying disease-positive subjects and their underlying pathogen. Pathogen detection is done through in vitro laboratory testing on a specimen (e.g., nasal swab, blood) collected from the subject, and it is especially important for diseases that manifest with overlapping symptoms but require different treatment. Examples include respiratory diseases, sexually transmitted diseases, transfusion-transmittable diseases, gastrointestinal diseases, blood infections, and sepsis. Consider respiratory diseases, a common cause of medical visits and hospitalizations (Templeton 2007, Chan et al. 2018). Treatment may differ depending on the virus causing the disease (e.g., influenza, respiratory syncytial virus, coronavirus disease 2019 (COVID-19)), or there may be no treatment at all (e.g., the common cold); however, only testing can distinguish one pathogen from another (Centers for Disease Control and Prevention 2022a). Further, not knowing whether the pathogen is bacterial can result in antibiotic overuse (Oved et al. 2015, Fleming-Dutra et al. 2016), contributing to the growing problem of antibiotic resistance (Llor and Bjerrum 2014). We use the terms disease and pathogen interchangeably, and we refer to a specific disease/pathogen by the most common name.

A widely used assay is the polymerase chain reaction (PCR) assay, a highly sensitive and specific genetic assay that detects a pathogen's DNA (or RNA) (Centers for Disease Control and Prevention 2021b, Food and Drug Administration 2021). Upon collection, specimens are sent to a testing laboratory, where they are prepared, and a small sample of each specimen is placed in a PCR assay (i.e., a commercial testing kit that consists of a small vial with reagents), which in turn, is placed in a PCR machine. PCR machines have capacities that range from 1 to several hundred assays per run (96 and 384 are common capacities) (e.g., Johnson 2021). Each run typically requires one to four hours (depending on the machine) to complete; the run time, along with the capacity, determines the machine's throughput. During a run, a unique section of the pathogen's DNA, if present, is amplified using reagents, such as polymerase, DNA nucleotides, and pathogen-specific primers, and pathogen-specific probes emit a signal (e.g., a fluorescence) if the pathogen is detected in the specimen. Disease screening via PCR is expensive and can be capacity restricted. Screening is often performed in state public health, hospital, blood donation agency, or commercial laboratories, which can test hundreds or more specimens per week for multiple diseases. Thus, it is essential for screening efforts to be cost and capacity efficient. In this light, we study a novel testing design problem that utilizes two *bundling* strategies to make screening more cost and capacity efficient.

1.1. Multiplexing (Disease Bundling)

One multiplex assay can detect multiple diseases using the same specimen as opposed to single-disease (*singleton*) assays. Multiplex assays are available for many disease groups, including, for example, respiratory, sexually transmitted, transfusion-transmittable, and gastrointestinal diseases (e.g., Biomérieux 2022). For instance, commercially available multiplex assays for respiratory diseases bundle 2–33 pathogens (Mahony et al. 2009, Schreckenberger and McAdam 2015, Beckmann and Hirsch 2016, Chen et al. 2017, Gonsalves et al. 2019, Kenmoe et al. 2019, Biomérieux 2022); the Alabama Department of Public Health uses a 20-plex assay to screen for respiratory pathogens, including COVID-19 (BCL-APH 2022).

1.2. Pooling (Specimen Bundling)

Under pooling, specimens from multiple subjects are bundled (i.e., pooled) and tested with one assay. Dorfman pooling (Dorfman 1943) is the most common pooling method in practice, and it is the method considered in this paper; we describe it here for a singleton assay. If the pooled test outcome is negative, then all subjects in the pool are declared negative for the disease; otherwise, each subject is retested individually via the same type of assay and is classified as negative or positive based on their individual test outcome. Because only a small sample of the specimen is needed for each test, each specimen has sufficient material for multiple tests. Examples of pooling (mostly for singleton assays) range from transfusion-transmittable diseases (e.g., West Nile virus, *Babesia microti*, Zika virus) to respiratory diseases (e.g., influenza) to sexually transmitted diseases and many others (e.g., May et al. 2010, Taylor et al. 2010, van Zyl et al. 2011, Van et al. 2012, Borges et al. 2015, Aprahamian et al. 2020, American Red Cross 2022). More recently, the Centers for Disease Control and Prevention and the Food and Drug Administration (FDA) have provided guidance for pooling for COVID-19 (Food and Drug Administration 2020, Centers for Disease Control and Prevention 2021a).

Multiplex PCR assays are widely available, and pooling with PCR is commonly used as PCR is conducive to both bundling strategies for the following reasons. (1) From a *cost perspective*, many of the reagents are not disease specific (e.g., the polymerase and the DNA nucleotides), and the required amount of these reagents is fairly insensitive to the number of diseases in the assay. The disease-specific reagents (e.g., the primers and the probes) are relatively inexpensive compared with the other reagents. Thus, multiplexing reduces the testing cost per disease over the corresponding singleton assays, leading to a concave cost structure in the number of diseases bundled. (2) From a *technology perspective*, because of amplification, PCR can detect minute quantities of the pathogen's genetic material, and thus, it is not very susceptible to dilution, preserving its high sensitivity with pooling (e.g., Nguyen et al. 2019). (3) From a *testing capacity* perspective, multiplexing leads to a more efficient use of the limited testing capacity, improving both the throughput and turnaround times. Pooling also reduces the number of tests required (hence, the testing cost and the capacity used) for low-prevalence diseases (e.g., Aprahamian et al. 2020).

When multiplexing and pooling are used separately for a given set of diseases (the disease selection problem, although of great interest, is beyond the scope of this paper), the cost and capacity efficiency objectives coincide, and the best nonintegrated design is straightforward. (i) For *multiplexing only*, bundle all the diseases into one multiplex to take advantage of the concave nature of the assay cost in the number of diseases bundled, and (ii) for *pooling only*, pool each singleton assay if the disease prevalence is below a threshold. When integrating multiplexing and pooling, however, the design problem becomes nonlinear and combinatorial because of the interdependent decisions of an *assay portfolio* (the mix and composition of singleton and multiplex assays) and a *testing method* (individual testing or pooling for each assay). Integrated multiplexing and pooling are not commonly used, although practical examples exist. We illustrate this integrated testing

design through the American Red Cross's use of a three-plex assay for hepatitis B, hepatitis C, and human immunodeficiency virus (HIV) in Dorfman pools of 16 (American Red Cross 2022). A specimen of each blood donation is sent to a testing laboratory, where samples from 16 specimens (each from a different donor) are pooled and tested with one three-plex assay for hepatitis B, hepatitis C, and HIV. If the pooled test is negative for all three diseases, then the 16 donations are declared negative for all three diseases. Otherwise (the pool tests positive for at least one of the three diseases), then each specimen (via another small sample) is individually retested using the same three-plex assay. Donations are then classified as negative or positive for each disease based on their individual test outcome. There are no guidelines on when and how these strategies should be integrated. This knowledge gap partially explains why integrated strategies are uncommon. We narrow the knowledge gap with the following methodological contributions and managerial insights.

Methodological contributions include the following: (1) development of novel multidisease testing design models for this nonlinear, partition-type combinatorial problem: a deterministic model and a distribution-free, robust variation, which both generate Pareto frontiers for cost- and capacity-efficient designs; (2) development of the deterministic counterpart of the robust model (Theorem A.1 and Corollary A.1 in Online Appendix A); and (3) characterization of structural properties of optimal testing designs as a function of key problem parameters (Theorems 1–4); and establishing the optimality of ordered partitions for certain coinfection structures (Theorem 5), leading to a polynomial-time algorithm (Corollary 1).

Managerial contributions and insights include the following. (1) Key drivers of optimal designs include the assay cost function, the tester's preference toward cost versus capacity efficiency, prevalence/coinfection rates, and for the robust model, prevalence uncertainty. Given a disease set, the most capacity-efficient design bundles all diseases into one multiplex and uses pooling only if the combined disease prevalence is below a threshold (Theorem 2), but this is not necessarily the most cost-effective design (Theorem 3). Once a combination of cost and capacity efficiency is considered, the structure of an optimal design becomes complex (Theorems 1 and 4). Now, a portfolio of multiplex and/or singleton assays can be optimal as long as pooling is used; in this case, individual testing can be used by at most one assay. Further, as coinfection rates rise, multiplexing becomes more favorable (Lemma 1). These properties highlight the interdependence between multiplexing and pooling. (2) Designs that integrate multiplexing and pooling are more structurally robust to variations in prevalence estimates than pooling alone because the pooling performance becomes a function of the assay prevalence (i.e., combined prevalence of the diseases in the assay) rather than the individual disease prevalence. Further, when ordered partitions are optimal (under certain coinfection structures) (Theorem 5), the design becomes more robust. (3) The robustness of the design can be further improved through a robust model, which is also practical, as it only needs an uncertainty set around each disease prevalence. These last two insights are especially promising for seasonal (e.g., respiratory diseases) or emerging diseases (e.g., COVID-19), for which prevalence rates are difficult to estimate. (4) The Pareto designs allow the tester to choose a design based on their cost and capacity priorities. (5) Our case study of 18 respiratory diseases illustrates that Pareto designs improve upon cost and/or capacity efficiency of multiplexing-only or pooling-only benchmarks by as much as 40%. This is despite the fact that the respiratory disease group is a relatively difficult group for testing design, compared with other disease groups, as some respiratory diseases exhibit high seasonality, high uncertainty (partially caused by COVID-19), and relatively high prevalence. Further, the price of robustness is low in our case study.

These findings underscore the value of integrated multiplexing and pooling optimization, a unique feature of our model. The use of cost- and capacity-efficient testing designs could further allow expanded screening, to cover larger populations or more diseases, providing a win-win situation for both public health practitioners (e.g., state public health laboratories, healthcare providers) and patients.

The remainder of the paper is organized as follows. Section 2 provides a review of the literature, and Section 3 provides the notation and formulates the testing design problem. Sections 4 and 5 establish properties of optimal designs and derive strategy insight. Section 6 illustrates the value of optimal designs through a case study based on publicly available data, and Section 7 concludes with research limitations and future research suggestions. Mathematical proofs and some case study results are relegated to the online appendix.

2. Literature Review

This paper is related to research on pooling and multiplexing. Although the literature in each area is large, the literature that studies their integration from a cost and capacity efficiency perspective is limited. Our review of each area is not exhaustive, but rather, it is representative of the related research so as to position our work in the literature.

We focus on PCR assays. Alternative testing technologies for pathogen identification include antigen assays and cell cultures (Ogilvie 2001, Leland and Ginocchio 2007, Gharabaghi et al. 2011, Ginocchio and McAdam 2011, Al Ghounaim et al. 2017, Das et al. 2018). Antigen

assays tend to be less expensive but also less sensitive than PCR; further, PCR can identify a larger set of pathogens because not all viral pathogens have antigen assays. On the other hand, testing via cell cultures is highly accurate (like PCR) but more expensive, with much longer turnaround times. As a result, PCR (and more recently, multiplex PCR) assays are becoming increasingly common.

Regarding pooling, Dorfman pooling was first proposed in the 1940s for syphilis screening in military inductees (Dorfman 1943). Since then, many variations of Dorfman pooling, as well as other pooling schemes, have been extensively studied, mostly for a singleton assay, to minimize the per subject expected number of tests. Although minimizing this metric improves both the cost and capacity efficiency for a singleton assay, these metrics diverge for multiple diseases because of different costs of multiplex and singleton assays; this is one of the departures of our work from this literature. It is well known that pooling reduces the expected number of tests over individual testing when disease prevalence is sufficiently low, and an optimal pool size has been analytically characterized for various pooling schemes; see Kim et al. (2007), Aprahamian et al. (2020), Bish et al. (2021), and Nguyen et al. (2021) for reviews of the pooling literature as well as the references therein.

Considering PCR multiplexing, three main research streams are of interest: (1) efficacy studies, many of which establish that multiplexing generally preserves the high sensitivity and specificity of their counterpart singleton PCR assays (e.g., Kenmoe et al. 2019); (2) design studies, which focus on compatibility from a biochemical perspective (e.g., how to design primers for different pathogens that work well together (i.e., without interference) when combined in an assay) (e.g., Rachlin et al. 2005, Yuan et al. 2021); and (3) cost-effectiveness studies, in which large multiplex assays typically offer higher operational efficiencies over their multiple singleton assay counterparts (see Section 6). Thus, much of this literature explores the benefits of multiplexing in comparison with current practices or in clinical settings, mainly in terms of whether the pathogen identification improves treatments and reduces costs. For example, a major benefit of pathogen identification is reduction in unnecessary antibiotic use (Oved et al. 2015, Fleming-Dutra et al. 2016) along with reduction in inpatient stay (Rogers et al. 2015). Multiplexing has been shown to be cost effective for many disease groups and populations (e.g., Vallières and Renaud 2013, Balakrishnan et al. 2016, Pinsky and Hayden 2019), including respiratory diseases (e.g., Subramony et al. 2016); see also the survey paper that compares assay performance and clinical (e.g., mortality rate, inpatient stay, unnecessary antibiotic use) and economic (e.g., hospitalization cost) impacts of various FDA-approved multiplex assays for different disease groups (bloodstream, respiratory, gastrointestinal, central nervous system) based on previously published studies (Ramanan et al. 2018).

In summary, much of the multiplexing literature studies existing commercial assays and does not delve into assay design; the biochemical assay design research is complementary to the operational design problem that we study. We start, as an input, with a set of pathogens with reagents compatible for multiplexing, and we construct multiplex and/or singleton assays for these pathogens so as to maximize the operational efficiencies. Importantly, we show that, although it is feasible to bundle all these diseases in one multiplex, this strategy is not necessarily optimal when considering pooling.

The operations/biostatistics literature that studies the design of multiplexing and pooling strategies from an efficiency perspective is limited. Tebbs et al. (2013) considers Dorfman pooling (with pool sizes determined via enumeration) for a two-plex for *Chlamydia trachomatis* and *Neisseria gonorrhoeae* (both sexually transmitted diseases); their numerical study shows that this two-plex with pooling reduces the number of tests over both its counterpart with individual testing and two singletons with pooling. Hou et al. (2017) and Hou et al. (2020) extend this work to other pooling schemes for the two-disease setting. Thus, all three papers (Tebbs et al. 2013, Hou et al. 2017, Hou et al. 2020) consider only the two-disease setting, where the two-plex is the only multiplex, and they do not delve into assay portfolio design. They also consider the expected number of tests; this is equivalent to the testing cost only when the two-plex and singleton assays have the same cost, which is not necessarily true (see Section 1). These papers do not provide analytical methods for optimal pool size determination. From a clinical perspective, Lindan et al. (2005) shares results from the implementation of a two-plex PCR for *C. trachomatis* and *N. gonorrhoeae*, both with Dorfman pooling (in pools of five) and with individual testing, at two clinics, and it shows that pooling reduces the number of tests by 50.3% over individual testing while achieving high efficacy. Van Hulst et al. (2009) studies the cost-effectiveness of three Dorfman pool sizes (6, 24, 48) for the hepatitis B, hepatitis C, and HIV three-plex over current blood screening practices in different countries; as a practical example, this three-plex is used by the American Red Cross in Dorfman pools of 16 (American Red Cross 2022). From a methodological perspective, our problem is also related to the partition problem discussed subsequently.

3. The Notation, Decision Problem, and Model

3.1. The Notation and the Decision Problem

We denote all vectors in boldface. Let $N = \{1, 2, \ldots, n\}$ denote the disease set for testing arranged in nonincreasing

order of disease prevalence. Coinfections are possible, and unless otherwise stated, we make no assumptions on the coinfection structure; that is, each pair of diseases in set N can be positively correlated, negatively correlated, or independent.

We study both deterministic and robust versions of the multidisease testing design problem for genetic assays considering multiplexing and Dorfman pooling (hereafter, "pooling") strategies. Although the deterministic model is the first step in the study of this novel problem, the robust model (i.e., under prevalence uncertainty) provides a realistic extension, especially for emerging or seasonal diseases, which typically exhibits high uncertainty (e.g., El Amine et al. 2017, Aprahamian et al. 2020, Nguyen et al. 2021). The objective is to minimize a *convex combination* of the two key dimensions of testing design: the expected testing cost and capacity requirement (i.e., expected number of tests) or a robust version of it, with the relative weights dictated by parameter $\lambda \in [0,1]$. The set of optimal designs ($\forall \lambda \in [0,1]$) provides the *Pareto designs*, including the cost- and test-minimizing special cases (the latter is the primary objective studied in the pooling literature) (Section 2).

A testing design corresponds to the following.

i. An assay portfolio represented by a *partition* of set N into mutually exclusive and exhaustive disease subsets, such that the diseases in each subset are bundled into one assay. Let $\boldsymbol{S} = (S^k)_{k=1,\dots,q}$, for some $q = 1,2,\dots,n$, denote a partition of set N: $S^k \cap S^l = \emptyset$, $k,l = 1,\dots,q : k \neq l$, $\cup_{k=1}^{q} S^k = N$, with $s^k = |S^k|$ denoting the size of assay S^k. Then, $s^k = 1$ corresponds to a *singleton* and $s^k \geq 2$ corresponds to a *multiplex*, or an s^k *-plex*, for the diseases in S^k.

ii. A testing method for each assay (i.e., pooling or individual testing) and if pooling, then the pool size, denoted by $t \in Z^+$. Let $\boldsymbol{t} = (t^k)_{k=1,\dots,q}$ denote the pool size vector. Then, if $t^k = 1$, assay S^k is tested individually, and if $t^k \geq 2$, assay S^k is tested via pooling, with pool size t^k.

In general, we use the superscript k for assay index and the subscript i for disease index, and we drop the indices when clear from context (e.g., we use set S when referring to an assay and a vector of sets $\boldsymbol{S}$ when referring to a partition of set N; that is, to an assay portfolio).

Each assay produces a binary outcome for each disease that it tests (e.g., an s-plex produces an s-dimensional binary test outcome vector), with a positive/negative outcome indicating the presence/absence of the corresponding disease. We assume that all assays have perfect sensitivity and specificity (reasonable for genetic assays). In pooling, if the outcome vector for the pooled test is negative for all the diseases in the assay, then all subjects in the pool are declared negative for those diseases; otherwise (the pool tests positive for at least one disease in the assay), each subject is retested individually via the *same* type of assay, and it is classified as negative or positive for each disease based on their individual test outcome vector. Assay sensitivity may reduce for large pools because of dilution. Although dilution is not pronounced for genetic assays (because of the amplification of the genetic material from the specimens), we follow the common practice and restrict pool sizes to a *pool size limit*, $\overline{M}$ (e.g., Nguyen et al. 2019, Aprahamian et al. 2020).

Assumption A. *The assay cost function for an s-plex, $c(s)$, $s \in Z^+$, is concave nondecreasing in s, with $c(1) = \gamma\,(>0)$, and $c(s) \leq \gamma \times s$, $s \in Z^+$: that is, bounded from above by a linear function.*

A concave functional form is realistic for genetic assay because many of their reagents are not disease specific and the disease-specific primers and probes are less expensive and also similar to each other (mainly consisting of simple strands of DNA); hence, they have similar cost (see Section 1). We define the *composite cost function* $\tilde{c}(s,\lambda) \equiv \lambda\, c(s) + 1 - \lambda$, $s \in Z^+, \lambda \in [0,1]$.

We express all performance metrics per testing subject. Let $T(S, t)$ and $C(S, t)$, respectively, denote the per subject expected number of tests ("expected tests") and the per subject expected testing cost for assay S and pool size t, where

$$C(S,t) = c(s) \times T(S,t). \tag{1}$$

Then, the per subject expected total testing cost ("total cost") for assay portfolio $\boldsymbol{S}$, pool size vector $\boldsymbol{t}$, and $\lambda \in [0,1]$, denoted by $TC(\boldsymbol{S},\boldsymbol{t},\lambda)$, is a convex combination of the expected testing cost and tests:

$$\begin{aligned} TC(\boldsymbol{S},\boldsymbol{t},\lambda) &= \lambda \sum_k C(S^k,t^k) + (1-\lambda)\sum_k T(S^k,t^k) \\ &= \sum_k \tilde{c}(s^k,\lambda) \times T(S^k,t^k). \end{aligned} \tag{2}$$

We use $T_D(S,t)$ to denote the per subject expected tests under pooling with pool size $t \in Z^+, t \geq 2$.

Because coinfections are possible, each subject is either free of all diseases in set N or infected by one or more diseases in this set. To represent all infection possibilities of a subject, let $N(l), l = 1,2,\dots,n$ denote the set of all l-tuples of set N (i.e., the set of all $\binom{n}{l}$ ordered elements (arranged in increasing disease index) each with l diseases). For example, $N(1) = N, N(2) = \{ij : i,j \in N, i<j\}, N(3) = \{ijr : i,j,r \in N, i<j<r\}$, and so on. Thus, indices $0, i, ij, ijr, \dots, 12\dots n$, respectively, denote the cases of no infection; infection i only; infections i and j only; infections i, j, and r only; and so on.

For the deterministic prevalence model (see Section 3.2 for the extension to stochastic prevalence), the deterministic joint prevalence vector ("joint vector"), $\boldsymbol{p} = [p_0, (p_i)_{i\in N}, (p_{ij})_{ij\in N(2)}, (p_{ijr})_{ijr\in N(3)}, \dots, p_{12\dots n}]$, denotes the probability that a random subject is in each of the 2^n infection

categories: that is, with p_0 denoting the no-infection probability; $p_i, i \in N(1)$ denoting all monoinfection probabilities; and the other elements denoting all possible coinfection probabilities, where

$$p_0 + \sum_{i\in N} p_i + \sum_{ij\in N(2)} p_{ij} + \sum_{ijr\in N(3)} p_{ijr} + \cdots + p_{12\ldots n} = 1. \tag{3}$$

We define A_i^+ as the event that a random subject is infected with disease $i \in N$ (and may be coinfected with some other diseases in set $N \setminus \{i\}$).

Definition 1. For assay $S \subseteq N$, assay prevalence, $\pi(S)$, is the probability that a random subject is infected by at least one disease in assay S (and may be coinfected with other diseases in set $N \setminus S$):

$$\pi(S) = Pr(\cup_{i\in S} A_i^+) = \sum_{i\in S} p_i + \sum_{ij\in N(2): i\in S \text{ or } j\in S} p_{ij} + \sum_{ijr\in N(3): i\in S \text{ or } j\in S \text{ or } r\in S} p_{ijr} + \cdots + p_{12\ldots n}. \tag{4}$$

The marginal prevalence of each disease, denoted $\pi_i, i \in N$, and the overall prevalence of the disease set, denoted $\pi(N)$, are given by

$$\pi_i = Pr(A_i^+) = p_i + \sum_{j:ij\in N(2)} p_{ij} + \sum_{j:ji\in N(2)} p_{ji} + \sum_{jr:ijr\in N(3)} p_{ijr} + \sum_{jr:jir\in N(3)} p_{jir} + \sum_{jr:jri\in N(3)} p_{jri} + \cdots + p_{12\cdots n}, i \in N, \tag{5}$$

$$\pi(N) = Pr(\cup_{i\in N} A_i^+) = \sum_{i\in N} p_i + \sum_{ij\in N(2)} p_{ij} + \sum_{ijr\in N(3)} p_{ijr} + \cdots + p_{12\ldots n} = 1 - p_0, \tag{6}$$

where $\boldsymbol{\pi} = (\pi_i)_{i\in N}$ is the marginal prevalence vector ("marginal vector").

The following relationship trivially holds because of coinfections and the disease indexing:

$$\pi_n \le \pi_{n-1} \le \cdots \le \pi_1 \le \pi(N) \le \sum_{i\in N} \pi_i. \tag{7}$$

Regarding the per subject expected tests function, for assay S, one test suffices for individual testing ($t = 1$), whereas for pooling with pool size $t \ge 2$, one pooled test is conducted for all t subjects, which if positive for at least one disease in assay S, is followed by an individual retest for each of the t subjects in the pool. Thus, we can write (e.g., Aprahamian et al. 2020)

$$T(S,t) = \begin{cases} T_D(S,t) = \dfrac{1}{t} + 1 - (1-\pi(S))^t, & \text{if } t \ge 2 \\ 1, & \text{if } t = 1 \end{cases}. \tag{8}$$

3.2. Model Formulations

To formulate the multidisease testing design problem, we represent a partition (assay portfolio) $\boldsymbol{S}$ by the binary decision variable vector $\boldsymbol{x} = (\boldsymbol{x}^k)_{k=1,\ldots,n}$, where each $\boldsymbol{x}^k = (x_i^k)_{i\in N}$, $\forall k$; that is, $x_i^k = 1$ if disease i is included in assay S^k and zero otherwise. Hence $\pi(S^k) = \pi(\boldsymbol{x}^k)$, $\forall k$.

The deterministic multidisease testing design problem (**TD**) is as follows:

$$\underset{\boldsymbol{x},\boldsymbol{t}}{\text{minimize}} \; TC(\boldsymbol{x},\boldsymbol{t},\lambda) = \lambda \sum_{k=1}^{n} C(\boldsymbol{x}^k, t^k) + (1-\lambda) \sum_{k=1}^{n} T(\boldsymbol{x}^k, t^k) = \sum_{k=1}^{n} \tilde{c}\left(\sum_{i\in N} x_i^k, \lambda\right) \times \min\left\{1, \frac{1}{t^k} + 1 - (1-\pi(\boldsymbol{x}^k))^{t^k}\right\} \tag{9}$$

$$\text{subject to} \quad \sum_{k=1}^{n} x_i^k = 1, \; i \in N \tag{10}$$

$$t^k \le \overline{M}, \; k = 1,\ldots,n \tag{11}$$

$$t^k \ge 0, \text{integer}, \; k = 1,\ldots,n, \tag{12}$$

$$x_i^k \text{ binary}, \; i \in N, k = 1,\ldots,n. \tag{13}$$

Objective (9) minimizes a convex combination of the expected testing cost and expected tests, with weights dictated by parameter $\lambda \in [0,1]$; thus, by varying λ, the model generates the Pareto designs, including the cost- and test-minimizing special cases. (10) ensures that each disease is tested by an assay. (11) restricts pool sizes to the pool size limit. (12) and (13) are logical constraints on integer decision variables. There is symmetry among solutions, and symmetry-breaking constraints can be added.

We now formulate the robust testing design problem under stochastic prevalence. To this end, we denote random variables in uppercase letters and their realizations in lowercase letters; use $\Omega(.)$ to denote an uncertainty set, and use ";" for probabilistic conditioning. Let $\boldsymbol{P}$ denote the continuous random joint vector, which implies, via Equation (5), a continuous random marginal vector, $\boldsymbol{\Pi} = (\Pi_i)_{i\in N}$, with respective realizations $\boldsymbol{p}$ and $\boldsymbol{\pi}$ (converging to the notation for the deterministic problem).

The robust multidisease testing design problem (**R-TD**) is as follows:

$$\underset{\boldsymbol{x},\boldsymbol{t}}{\text{minimize}} \; \max_{\boldsymbol{p}\in\Omega(\boldsymbol{P})} \left\{\lambda \sum_{k=1}^{n} C(\boldsymbol{x}^k, t^k; \boldsymbol{p}) + (1-\lambda) \sum_{k=1}^{n} T(\boldsymbol{x}^k, t^k; \boldsymbol{p})\right\} \quad \text{subject to} \quad (10),(11),(12),(13). \tag{14}$$

It is difficult to accurately estimate the distribution of the marginal vector $\boldsymbol{\Pi}$ (at disease prevalence level), let alone the distribution of the joint vector $\boldsymbol{P}$ (at mono- and coinfection probability level). Therefore, we adopt a *distribution-free* approach and construct an interval-type uncertainty set (e.g., Perakis and Roels 2008, El Amine et al. 2017, El Hajj et al. 2022a) for each random marginal prevalence, $\Pi_i, i \in N$, which then implies an equivalent uncertainty set on $\boldsymbol{P}$ (Remark A.1 in Online Appendix A). The former is relatively easier to construct (e.g., the lower and upper limits of a statistical

confidence interval (CI) on each disease prevalence can be used, as we do in the case study). (Although not practical because of the large number of coinfections, one can also start with an uncertainty set on $\boldsymbol{P}$.)

The joint optimization of multiplexing and pooling, the cost- and capacity-based objective, and the robust model are our main modeling departures from the literature (Sections 1 and 2), leading to a combinatorial problem. **TD**, with a nonlinear objective and a partition-type decision, is NP hard in general (Chakravarty et al. 1982), and **R-TD**, with a mini-max objective, is even harder.

In particular, our decision problem is related to partitioning a set of objects, each with a certain attribute, into a variable number of mutually exclusive and exhaustive groups to minimize the total cost of the partition. Within this large body of this literature, the most relevant works include Chakravarty et al. (1982), which shows the optimality of an ordered partition for a cost function that is concave in the attribute sum, and Anily and Federgruen (1991), which extends this result to a cost function that depends on the group size and the average or the maximum attribute in the group. The ordered partition property leads to an equivalent shortest path-based formulation that can be solved in polynomial time (Chakravarty et al. 1982, Anily and Federgruen 1991); see Section 5.2. Anily and Federgruen (1991) also provides the necessary conditions for the optimality of a monotone partition (i.e., an ordered partition for which objects with smaller attributes are placed in smaller groups (this property does not hold in our setting)). The aforementioned results require the concavity of the total cost function in the attribute sum: in our setting, concavity in the sum of marginal disease prevalences, which does not necessarily hold for our problem (see Equations (2), (4), and (8)), except for a special case with no coinfections (Section 5.2). Others extend the necessary conditions for the optimality of an ordered partition to different cost functions that do not apply to our setting (e.g., Hwang 1981, Hwang et al. 1985). A mini-max variation of the partition problem, when restricted to ordered partitions, is also studied (e.g., Manne and Sorevik 1995, Olstad and Manne 1995), which in our setting, corresponds to the ordered partition that minimizes the highest cost per assay for a given prevalence vector. This is in contrast with our robust model, which minimizes the highest total cost (i.e., for all assays) over all possible prevalence vectors in an uncertainty set. In the absence of the ordered partition property, the partition problem remains NP hard, and the literature develops various heuristics and exact algorithms, such as genetic algorithms (Levine 1996); linear relaxation-based methods (e.g., dual heuristics, volume algorithms) (Fisher and Kedia 1990, Chan and Yano 1992, Barahona and Anbil 2002, Boschetti et al. 2008); implicit enumeration and search trees; simplex, hybrid primal, and symmetric subgradient cutting plane methods; and column generation (Balas and Padberg 1976). We refer the interested reader to the works of Balas and Padberg (1976), Fisher and Kedia (1990), Barahona and Anbil (2002), and Lewis et al. (2008) for detailed overviews.

4. Structural Properties of Optimal Testing Designs

Section 4.1 provides preliminaries on testing method optimization, and Section 4.2 integrates assay portfolio optimization into testing design in a multidisease setting, a unique feature of our model.

4.1. Preliminaries: Optimal Testing Method—Pooling vs. Individual Testing

Given an assay portfolio $\boldsymbol{S}$, the testing design problem reduces to testing method optimization (pooling versus individual testing for each assay), and the single-disease results from the literature extend in a straightforward manner to our multidisease setting. We use t_D^* to denote the optimal Dorfman pool size (i.e., in the domain $t \in Z^+, t \geq 2$, and t^* to denote the (global) optimal pool size; i.e., in the domain $t \in Z^+, t \geq 1$ (including the individual testing option, $t = 1$)). To make the dependence of $T(.)$ and $t^*(.)$ on $\pi(S)$ explicit, in places we use $\pi(S)$ as an argument (e.g., $t^*(\pi(S))$).

Property 1 (*From Aprahamian et al. 2020*). *Consider a single disease with prevalence π. The optimal integer pool size, $t_D^*(\pi)$, that minimizes the expected tests function under Dorfman pooling (i.e., $\text{minimize}_{t\in Z^+, t\geq 2} T_D(\pi, t)$) is the solution to*

$$t_D^*(\pi) = \min\left\{ arg\ min_{t\in\{\lfloor t_{frac}\rfloor, \lceil t_{frac}\rceil, \overline{M}\}} \{T_D(\pi, t)\}, \overline{M} \right\},$$

$$where\ \ t_{frac} = \frac{2}{\ln(1-\pi)} W_0\left(-\frac{1}{2}(-\ln(1-\pi))^{1/2} \right),$$

where $W_0(\cdot)$ denotes the principle branch (i.e., the largest solution) of the Lambert function $W(x)$, defined by $x = W(x)e^{W(x)}$, $\forall x \in \mathbb{R}$ (Corless et al. 1996)).

Property 1 trivially extends to a multiplex, and a threshold for when pooling is optimal over individual testing can be derived for integer pool sizes, extending a similar result for continuous pool sizes (e.g., Aprahamian et al. 2020).

Property 2. *For any assay $S \subseteq N$, the optimal integer pool size, $t_D^*(\pi(S))$, can be derived by Property 1 by letting $\pi = \pi(S)$.*

1. *$T(\pi(S), t_D^*(\pi(S))) \leq T(\pi(S), 1) = 1$ only if $\pi(S) \leq \underline{p} = 1 - \sqrt[3]{\frac{1}{3}} \approx 0.31$, where the pooling threshold $\underline{p}$ is independent of*

assay size s: *that is*,

$$t^*(\pi(S)) = \begin{cases} t_D^*(\pi(S)), & \text{if } \pi(S) \le \underline{p} \\ 1, & \text{otherwise} \end{cases}, \quad \text{leading to}:$$

$$T(\pi(S), t^*(\pi(S))) = \begin{cases} T_D(\pi(S), t_D^*(\pi(S))), & \text{if } \pi(S) \le \underline{p} \\ 1, & \text{otherwise} \end{cases}.$$

2. (*From Aprahamian et al.* 2020, *El Hajj et al.* 2022c) $T(\pi(S), t^*(\pi(S))) = T_D(\pi(S), t_D^*(\pi(S)))$ *is strictly concave increasing in* $\pi(S) \in [0, \underline{p}]$, *and* $T(\pi(S), t^*(\pi(S))) = 1$ *for* $\pi(S) \in [\underline{p}, 1]$.

As Property 2 indicates, for a given assay portfolio $\boldsymbol{S}$, an optimal testing method can be determined *independently* for each assay in the portfolio. In the remainder of the paper, we consider that all assay portfolios use the optimal testing method delineated in Properties 1 and 2.

4.2. Optimal Testing Designs

For the robust formulation **R-TD**, under an interval-type uncertainty set for the marginal vector (Remark A.1 in Online Appendix A), we are able to characterize the worst-case solution (Theorem A.1 and Corollary A.1 in Online Appendix A). These results lead to the deterministic counterpart for **R-TD**, which corresponds to a specific instance of **TD**, as outlined in the following remark.

Remark 1.

1. By Theorem A.1 in Online Appendix A, the deterministic counterpart of **R-TD** has the following objective function:

$$\underset{\boldsymbol{x},\, \boldsymbol{t}}{\text{minimize}} \sum_{k=1}^{n} \tilde{c}\left(\sum_{i \in N} x_i^k, \lambda\right) \times \min\left\{1,\ \frac{1}{t^k} + 1 - \left(1 - \min\left\{1, \sum_{i \in N} \overline{\pi}_i x_i^k\right\}\right)^{t^k}\right\}.$$

2. Noting the equivalent objective of **TD** in (9) (i.e., a function of assay prevalences only), the deterministic counterpart of **R-TD** reduces to an instance of **TD** with specific assay prevalences: $\pi(S) = \min\{1, \sum_{i \in S} \overline{\pi}_i\}$, $\forall S \subseteq N$. This equivalence indicates that the structural properties for **TD** (Theorems 1–4) continue to hold for **R-TD**, with $\pi(S) = \min\{1, \sum_{i \in S} \overline{\pi}_i\}$, $\forall S \subseteq N$.

3. Further, for the special case where $\sum_{i \in N} \overline{\pi}_i \le 1$, the deterministic counterpart of **R-TD** also reduces to **TD** with the specific $\boldsymbol{p}$ provided in Corollary A.1 in Online Appendix A.

Importantly, the deterministic counterpart of **R-TD** requires only an upper limit on each disease prevalence, $\overline{\pi}_i, i \in N$. From an implementation point, this is highly desirable, and such upper limits can be derived, for example, from statistical CIs for disease prevalence. In general, the higher the uncertainty around a disease prevalence, the higher its upper limit will be, increasing the conservatism of the robust solution, which will be studied in Section 6.

Next, we characterize the optimal testing designs for **TD** and derive insight. (By Remark 1, Corollary A.1 in Online Appendix A, and Theorem A.1 in Online Appendix A, the structural results extend to **R-TD**.) To this end, we provide a series of definitions, which allow us to decompose the set of all possible testing designs into a number of *mutually exclusive and exhaustive design classes* based on how the disease set is partitioned and the testing methods used (see Table 1 for a mapping of all possible design classes and strategies).

Definition 2. A partition $\boldsymbol{S} = (S^k)_{k=1,\ldots,q}$ is a q-partition, $q = 1, 2, \ldots, n$, if it consists of exactly q assays, and it is an ordered q-partition if $S^1 = \{1, \ldots, s^1\}, S^2 = \{s^1 + 1, \ldots, s^1 + s^2\}, \ldots, S^q = \left\{\sum_{k=1}^{q-1} s^k + 1, \ldots, n\right\}$, for some cardinality vector $\boldsymbol{s} = (s^k)_{k=1,\ldots,q} : \sum_{k=1}^{q} s^k = n$; that is, the disease set is partitioned into q assays following a nonincreasing order of disease prevalences, $\pi_i, i \in N$.

Definition 3. Consider a q-partitioned testing design $(\boldsymbol{S}, \boldsymbol{t})$ for any $q = 1, 2, \ldots, n$.

1. If all q assays utilize pooling ($\boldsymbol{t} \ge \boldsymbol{2}$), it is a Dorfman design. If all q assays utilize individual testing ($\boldsymbol{t} = \boldsymbol{1}$), it is an individual-testing design. If some assays utilize pooling, whereas others utilize individual testing ($\boldsymbol{t} : \exists k, l = 1, \ldots, q : t^k = 1, t^l \ge 2$), it is a mixed-testing design.

2. If at least one multiplex is used ($q = 1, \ldots, n-1$), it is an mx design; otherwise ($q = n$), it is an all-singleton design.

We use the notation $D^{(q)}, I^{(q)}, q = 1, \ldots, n$ and $M^{(q)}$, $q = 2, \ldots, n$ to denote the optimal testing design when constrained to be within the q-partitioned Dorfman, individual-testing and mixed-testing design classes,

Table 1. The Mapping Between All Design Classes and Strategies

Strategy	Design class
Only multiplexing (no pooling)	$I^{(q)}, q = 1, \ldots, n-1$
Only pooling (no multiplexing)	$D^{(n)}$ and $M^{(n)}$
Both multiplexing and pooling	$D^{(q)}, q = 1, \ldots, n-1$ and $M^{(q)}, q = 2, \ldots, n-1$
No multiplexing and no pooling	$I^{(n)}$

respectively. (The $M^{(1)}$ design (i.e., one assay with a mix of pooling and individual testing) is not possible.)

Definition 4. Design class A dominates design class B, denoted $A \preceq B$, if the total cost of an optimal **TD** solution within design class A is less than or equal to that for design class B.

We are ready to provide structural properties of optimal designs; in case of multiple optimal designs, the results characterize one of the optimal designs. To put our results into perspective, we first provide the optimal design when multiplexing and pooling strategies are optimized separately (which will serve as benchmarks in the case study of Section 6), representing the current modeling of these strategies in the literature (Section 2); in this case, the cost and capacity efficiency objectives coincide.

Remark 2.

1. Multiplexing only ($\boldsymbol{t} = \mathbf{1}$). For an individual-testing design, the optimal assay portfolio is to bundle all diseases into one multiplex (see Assumption A): that is, an $I^{(1)}$ design.

2. Pooling only ($\boldsymbol{S} = (\{i\})_{i \in N}$). For an all-singleton design, the optimal testing method for each assay follows the pooling threshold policy (Properties 1 and 2); that is, each singleton is pooled only if the prevalence of its disease is sufficiently low (i.e., $I^{(n)}$, $D^{(n)}$, or $M^{(n)}$ design).

When multiplexing and pooling are optimized jointly (i.e., **TD**), a main departure from the literature, the interplay between multiplexing and pooling impacts the design, and the optimal designs in Remark 2 are no longer necessarily optimal.

Theorem 1. *Consider* ***TD****. We have that* $I^{(1)} \preceq I^{(q)}$, $\forall q = 2, \ldots, n$; *an optimal* $M^{(q)}, q = 2, \ldots, n-1$, *design is such that only one assay is individually tested and* $q - 1$ *assays are pooled,* $\forall \lambda \in [0,1]$. *Further, an optimal design class can be characterized as follows for any* $\lambda \in [0,1]$.

1. *If* $\pi(N) \leq \underline{p}$, *then* $D^{(q)}$ (*i.e., Dorfman*) *for some* $q = 1, \ldots, n$.

2. *If* $\pi_1 \leq \underline{p} < \pi(N)$, *then either* $I^{(1)}$ *or* $D^{(q)}$ *for some* $q = 2, \ldots, n$ *or* $M^{(q)}$ *for some* $q = 2, \ldots, n-1$.

3. *If* $\pi_n \geq \underline{p}$, *then* $I^{(1)}$ (*i.e., n-plex individual testing*).

4. *Otherwise* (*if* $\exists i \in \{1, \ldots, n-1\} : \pi_{i+1} < \underline{p} < \pi_i$), *then either* $I^{(1)}$ *or* $M^{(q)}$ *for some* $q = 2, \ldots, n-i+1$.

Thus, the joint optimization of multiplexing and pooling implies testing designs that span the entire spectrum of design classes (see Table 1), demonstrating the richness of this decision problem. Now, the optimal design is driven by the tension between reducing the expected number of tests versus the expected testing cost, hence the capacity versus cost trade-off. In particular, the expected tests function under pooling (at optimal pool sizes) is concave increasing in assay prevalence (Property 2), favoring multiplexing. On the other hand, the expected testing cost is the product of the assay cost function and the expected tests function (Equation (1)), both of which are concave increasing as more diseases are bundled (as this also increases the assay prevalence), but the expected testing cost is not necessarily concave in the number of bundled diseases. To see the intuition, observe that the bundling of more diseases raises the assay prevalence, thus increasing the (individual) retest probability (Equation (8)), but all individual retests use the same type of assay as the original assay, the cost of which is increasing in the number of bundled diseases. Further, if disease bundling raises an assay's prevalence too much, then the efficiency provided by pooling is lost. As a result of these tensions, an assay portfolio of multiplex and/or singleton assays can now be optimal *as long as at least one assay is pooled* (i.e., $M^{(q)}$ or $D^{(q)}$). Further, a portfolio of multiple assays and a combination of testing methods can be optimal (i.e., $M^{(q)}$) *as long as only one assay uses individual testing* (all other assays must be pooled). We also note that the integration of multiplexing and pooling makes the designs structurally robust to prevalence uncertainty because it is the assay's prevalence (i.e., the combined prevalences of the diseases in the assay) that drives the pool size and not the disaggregate disease prevalences; we quantitatively study this aspect in the case study.

Theorem 1 highlights the dependencies between multiplexing and pooling, hence the need for joint optimization. The next section discusses these dependencies in a more precise manner through the main drivers of optimal designs.

5. Design Insight: Main Drivers of an Optimal Testing Design

We now analyze the two main drivers of optimal designs: the cost structure (Section 5.1) and the disease prevalence/coinfection structure (Section 5.2).

5.1. Impact of the Cost Structure

We first study how the optimal testing design changes as the assay cost function ($c(.)$) or the tester's preference toward cost versus test minimization (parameter λ) changes. Let $\tilde{C}$ denote the infinitely many composite cost functions, $\tilde{c}(s,\lambda) = \lambda\, c(s) + 1 - \lambda$, $\forall \lambda \in [0,1]$, for which the assay cost function $c(.)$ satisfies Assumption A. The marginal differences of any composite cost function in this set are nonnegative and bounded by γ (i.e., $0 \leq \tilde{c}(s+1,\lambda) - \tilde{c}(s,\lambda) \leq \gamma$, $\forall s \in Z^+$, $\forall \tilde{c}(.) \in \tilde{C}$), and they satisfy concavity. The following definition allows us to compare cost functions in terms of their marginal differences.

Definition 5. We say that function $g(.)$ has higher differences relative to function $g'(.)$, written $g(.) \geq_{\text{diff}} g'(.)$, if $g(s+1) - g(s) \geq g'(s+1) - g'(s)$, $\forall s \in Z^+$.

Property 3. *Any composite cost function* $\tilde{c}(s,\lambda) \in \tilde{C}$ *attains lower differences as* (1) *the assay cost function* $c(s)$ *attains lower differences or* (2) λ *decreases.*

Remark 3. Among the infinitely many composite cost functions $\tilde{c}(.) \in \tilde{C}$,

1. the smallest-difference function in set $\tilde{C}$ is attained when $\lambda = 0$ or $c(s) = \gamma$, $\forall s \in Z^+$ (i.e., constant assay cost). Under this function, **TD** objective reduces to test minimization.

2. the highest-difference function in set $\tilde{C}$ is attained when $\lambda = 1$ and $c(s) = \gamma \times s$, $\forall s \in Z^+$. Under this function, **TD** objective reduces to cost minimization with a linear assay cost.

In what follows, we first characterize the optimal design for the smallest- and highest-difference composite cost functions described in Remark 3, which respectively, reduce to test minimization and cost minimization (with linear assay cost), allowing us to gain insight into the tension between these objectives. The test minimization special case is also important for positioning our work within the pooling literature, which extensively studies this objective (Section 2).

Theorem 2. *Consider the smallest-difference composite cost function in set* $\tilde{C}$*: that is, with* $\lambda = 0$ *or* $c(s) = \gamma$, $\forall s \in Z^+$. *We have that* $I^{(1)} \preceq M^{(q)}$, $\forall q = 2, \ldots, n$. *Further, if* $\pi(N) \le \underline{p}$, *then* $D^{(1)} \preceq D^{(2)} \preceq \cdots \preceq D^{(n)}$. *The optimal design class can be characterized as follows.*

1. *If* $\pi(N) \le \underline{p}$, *then* $D^{(1)}$ (*i.e., n-plex Dorfman*).

2. *Otherwise (if* $\pi(N) > \underline{p}$*), then* $I^{(1)}$ (*i.e., n-plex individual testing*).

Theorem 3. *Consider the highest-difference composite cost function in set* $\tilde{C}$*: that is, with* $\lambda = 1$ *and* $c(s) = \gamma \times s$, $\forall s \in Z^+$. *We have the following. If* $\pi(N) \le \underline{p}$, *then* $D^{(n)} \preceq D^{(n-1)} \preceq \cdots \preceq D^{(1)}$. *If* $\pi_1 \le \underline{p} < \pi(N)$, *then* $D^{(n)} \preceq D^{(q)}$, $\forall q = 1, \ldots, n-1$. *The optimal design class can be characterized as follows*:

1. *if* $\pi_1 \le \underline{p}$, *then* $D^{(n)}$ (*i.e., all-singleton Dorfman*);

2. *if* $\pi_n \ge \underline{p}$, *then any individual testing design,* $I^{(q)}$, *for any* $q = 1, \ldots, n$;

3. *otherwise (if* $\exists i \in \{1, \ldots, n-1\} : \pi_{i+1} < \underline{p} < \pi_i$*), then any* $M^{(q)}$, *for* $q = n-i+1, \ldots, n$, *which uses any individual testing design for diseases in set* $\{1, 2, \ldots, i\}$ *and pooled testing and singleton assays for each disease in set* $\{i+1 \ldots, n\}$.

Table 2 summarizes the properties of optimal designs for the general cases as well as the extreme cases (the smallest- and highest-difference composite cost functions), established in Theorems 1–3. We note that all prevalence regions in the table are relevant in practice, as it is possible for the overall prevalence to be below or above the 31% threshold, as our case study indicates (Section 6). In the extreme cases, multiplexing is in the form of "all or none." That is, either all diseases are bundled into one n-plex (1-mx design), or each disease is tested separately via its own singleton assay (all-singleton design). Such all-or-none-type assay portfolios also represent the designs when multiplexing and pooling are considered separately (Remark 2). As Theorems 2 and 3 show, this all-or-none form is driven by the tension between reducing the expected tests versus the expected testing cost. In particular, when pooling is optimal (i.e., $\pi(N) \le \underline{p}$), for the smallest-difference composite cost function (i.e., test minimization objective), one additional assay in a Dorfman design increases the expected tests (i.e., $D^{(1)} \preceq D^{(2)} \preceq \cdots \preceq D^{(n)}$); for the highest-difference composite cost function (testing cost minimization objective), one fewer assay in a Dorfman design increases the expected testing cost (i.e., $D^{(n)} \preceq D^{(n-1)} \preceq \cdots \preceq D^{(1)}$). Table 2 also provides insight

Table 2. Optimal Design Class Based on the Composite Cost Function $\tilde{c}(.)$

	Prevalence range			
		$\pi(N) > \underline{p}$		
Composite cost function	$\pi(N) \le \underline{p}$	$\pi_1 \le \underline{p}$	$\pi_{i+1} < \underline{p} < \pi_i$, for some $i \in \{1, \ldots, n-1\}$	$\pi_n \ge \underline{p}$
General $\tilde{c}(.)$ (Theorem 1)	$D^{(q)}$ (some $q = 1, \ldots, n$) pooling (mx possible)	$I^{(1)}$ or $D^{(q)}$ (some $q = 2, \ldots, n$) or $M^{(q)}$ (some $q = 2, \ldots, n-1$) mx and pooling both possible	$I^{(1)}$ or $M^{(q)}$ (some $q = 2, \ldots, n-i+1$) mx (pooling possible)	$I^{(1)}$ mx only
Special cases				
The smallest difference $\tilde{c}(.)$ ($\lambda = 0$ or $c(s) = \gamma$, $\forall s \in Z^+$) (Theorem 2)	$D^{(1)}$ both mx and pooling	$I^{(1)}$ mx only		
The highest difference $\tilde{c}(.)$ ($\lambda = 1$ and $c(s) = \gamma \times s$, $\forall s \in Z^+$) (Theorem 3)	$D^{(n)}$ pooling only	$D^{(n)}$ pooling only	Any $M^{(q)}$, $q = n-i+1, \ldots, n$ pooling (mx possible)	Any $I^{(q)}$, $q = 1, \ldots, n$ no pooling (mx possible)

Note. mx, multiplex.

into the behavior of robust designs as prevalence uncertainty rises; as the upper limit of the uncertainty set for a disease prevalence, used in the deterministic counterpart of **R-TD** (Remark 1), increases, the robust design moves from left to right through the columns of Table 2, and in the most conservative case, it defaults to the *multiplexing-only* ($I^{(1)}$) benchmark.

The tension between test versus cost minimization continues to drive an optimal design in general when the composite cost function is in between the two extremes. To demonstrate this tension clearly, we now focus on the case where the overall disease prevalence does not exceed the pooling threshold ($\pi(N) \le \underline{p} \approx 0.31$) when pooling is always optimal. In this case, the optimal design is $D^{(1)}$ for the smallest-difference composite cost function and $D^{(n)}$ for the highest-difference composite cost function (Table 2 and Theorems 2 and 3). Then, the next question is whether there exists a (unique) cost threshold function $\overline{\tilde{c}}^{q}(.)$, such that $D^{(q-1)} \preceq D^{(q)}$ ($D^{(q)} \preceq D^{(q-1)}$) for all composite cost functions with smaller (higher) differences than $\overline{\tilde{c}}^{q}(.)$. As Example 1 shows, such a threshold cost function does not exist.

Example 1. Consider disease set $N = \{1,2,3,4\}$, with equal marginal prevalences and no coinfections: $\pi_i = 0.06, i \in N$, and $\pi(N) = \sum_{i\in N}\pi_i = 0.24 \le \underline{p}$; and assay cost functions $c(.)$ and $c'(.)$ (with corresponding composite cost functions $\tilde{c}(.)$ and $\tilde{c}'(.)$): $c(1) = 1, c(2) = 1.43, c(3) = 1.75, c(4) = 2.05$, and $c'(1) = c(1)$, $c'(s) = c(s) - 0.10, s = 2,3,4$. By Definition 5, $c'(.) \le_{\text{diff}} c(.)$, hence by Property 3, $\tilde{c}'(\lambda) \le_{\text{diff}} \tilde{c}(\lambda)$, $\forall \lambda \in [0,1]$. However, at $\lambda = 1$, $D^{(1)}$ is optimal for $\tilde{c}(\lambda = 1)$, whereas $D^{(2)}$ is optimal for the smaller-difference function, $\tilde{c}'(\lambda = 1)$.

As the following result indicates, such thresholds do exist on λ.

Theorem 4. *Consider that* $\pi(N) \le \underline{p}$. *Then,* $\exists \overline{\lambda}^{(q)} \le 1, q = 1,\dots,n-1$, *such that the optimal design has the following structure*:
optimal design class:

$$\begin{cases} D^{(1)}, \text{ if } \lambda \in [0, \overline{\lambda}^{(1)}] \\ D^{(r)}, \text{ for some } r = 2,\dots,q, \quad \text{if } \lambda \in (\max\{\overline{\lambda}^{(1)}, \overline{\lambda}^{(q-1)}\}, \overline{\lambda}^{(q)}], \\ \qquad\qquad\qquad\qquad\qquad q = 2,\dots,n-1 \\ D^{(n)}, \text{ if } \lambda \in (\max\{\overline{\lambda}^{(1)}, \overline{\lambda}^{(n-1)}\}, 1]. \end{cases}$$

Remark 4.

1. Among the optimality regions delineated in Theorem 4, only the $D^{(1)}$-optimal region, $[0, \overline{\lambda}^{(1)}]$, is guaranteed to be nonempty (see the $\lambda = 0$ case in Theorem 2). That is, it is possible that there is no $\lambda \in [0,1]$ for which design $D^{(q)}, q = 2,\dots,n$, is optimal. Further, the $D^{(n)}$-optimal region is guaranteed to be nonempty only for the linear assay cost function (see Remark 3 and Theorem 3); otherwise, it is possible for $\overline{\lambda}^{(n-1)} = 1$ and hence, $(\overline{\lambda}^{(n-1)}, 1]$ to be empty.
2. When $\pi(N) \le \underline{p}$, the optimal design is completely characterized for $n \le 3$, and the number of assays is nondecreasing in λ.
   For $n = 2$, $D^{(1)}$, $\forall \lambda \in [0, \overline{\lambda}^{(1)}]$ and $D^{(2)}$, $\forall \lambda \in (\overline{\lambda}^{(1)}, 1]$.
   For $n = 3$, $D^{(1)}$, $\forall \lambda \in [0, \overline{\lambda}^{(1)}]$; $D^{(2)}$, $\forall \lambda \in (\overline{\lambda}^{(1)}, \overline{\lambda}^{(2)}]$; and $D^{(3)}$, $\forall \lambda \in (\max\{\overline{\lambda}^{(1)}, \overline{\lambda}^{(2)}\}, 1]$.
   In general, however, the number of assays in an optimal Dorfman design need not be monotone in λ. For example, for $n = 4$, the optimal design is $D^{(2)}$ or $D^{(3)}$, $\forall \lambda \in (\max\{\overline{\lambda}^{(1)}, \overline{\lambda}^{(2)}\}, \overline{\lambda}^{(3)}]$; that is, it can potentially alternate between $D^{(2)}$ and $D^{(3)}$ in this region, although such a numerical example is lacking.

5.2. Impact of the Disease Prevalence and Coinfection Structure

In this section, we discuss how the disease prevalence/coinfection structure, relatedly the correlations among disease prevalences, impacts the optimal assay portfolio, and we develop an exact, efficient algorithm for certain coinfection structures. We first make the link between the coinfection and correlation structure explicit and introduce two special cases that will guide our analysis.

Remark 5. Let $\boldsymbol{\xi} = (\xi_i)_{i\in N}$ denote a multivariate binary vector, which assumes a value of one if a random subject is infected with disease $i \in N$ (i.e., event A_i^+ occurs) and zero otherwise in accordance with the joint vector $\boldsymbol{p}$. The marginal distributions, $\xi_i \sim Bernoulli(\pi_i), i \in N$ (Dai et al. 2013).

1. The pairwise correlation coefficient, $\rho(\xi_i, \xi_j) = \frac{p_{ij} - \pi_i\pi_j}{\sqrt{\pi_i(1-\pi_i)\pi_j(1-\pi_j)}}$, $\forall ij \in N(2)$, is increasing in coinfection probability (p_{ij}) as long as the marginal prevalences (π_i, π_j) remain unchanged. ($\rho(\xi_i, \xi_j)$ may have a narrower range than $[-1,1]$ unless $\pi_i = \pi_j$ (e.g., Emrich and Piedmonte 1991).)
2. Consider two special coinfection structures.
   a. No coinfections. Events A_i^+, $\forall i \in N$, are mutually exclusive. Then, assay prevalence (Equation (4)) reduces to

$$\pi(S) = \sum_{i\in S}\pi_i, \quad \forall S \subseteq N, \tag{15}$$

and $\rho(\xi_i, \xi_j) < 0$, $\forall ij \in N(2)$.
   b. Independent diseases. Events A_i^+, $\forall i \in N$, are mutually independent. Then, assay prevalence (Equation (4)) reduces to

$$\pi(S) = \sum_{i\in S}\pi_i - \sum_{ij\in N(2):i,j\in S}\pi_i\pi_j + \sum_{ijr\in N(3):i,j,r\in S}\pi_i\pi_j\pi_r + \cdots + (-1)^{s+1}\prod_{i\in S}\pi_i, \quad \forall S \subseteq N, \tag{16}$$

and $\rho(\xi_i, \xi_j) = 0$, $\forall ij \in N(2)$.

Definition 6. We say that joint vector $\boldsymbol{p}'$ is more correlated than joint vector $\boldsymbol{p}$ if $p'_{i\cdots j} \geq p_{i\cdots j}$, $\forall i \cdots j \in N(l)$, $l = 2,\ldots,n$, and their marginal vectors are equal, $\boldsymbol{\pi}' = \boldsymbol{\pi}$; that is, each coinfection probability in $\boldsymbol{p}'$ is at least as large as its counterpart in $\boldsymbol{p}$, whereas their marginal prevalences are equal.

Lemma 1. *Consider $\pi(N) \leq \underline{p}$. The threshold $\overline{\lambda}^{(n-1)}$ is nondecreasing as $\boldsymbol{p}$ becomes more correlated.*

Thus, when pooling is optimal, as the prevalence vector becomes more correlated (i.e., some coinfection probabilities rise, whereas disease prevalences remain the same) (Definition 6), the λ-region where an all-singleton design ($D^{(n)}$) is optimal shrinks; that is, multiplexing becomes more favorable.

The optimal design is completely characterized for the smallest- and highest-difference composite cost functions (Theorems 2 and 3). For the remaining cases, however, the optimal design is either some form of Dorfman or mixed-testing design, $D^{(q)}$ or $M^{(q)}$, or n-plex individual-testing design, $I^{(1)}$ (Table 2). Hence, one must consider all possible partitions of the disease set N, and this general partition-type problem, with no assumed relationship among disease prevalences, is NP hard (Chakravarty et al. 1982). The following result shows that when disease prevalences are independent or when there are no coinfections, there exists an optimal design that uses an ordered partition (Definition 2) for both Dorfman and mixed-testing designs. This result plays a key role in the development of an efficient, exact solution procedure for these special cases. (Dorfman designs with $q = 1$ and $q = n$ are excluded from the theorem, as their assay portfolio and testing method are completely fixed.)

Theorem 5. *Suppose all disease prevalences are independent or that there are no coinfections.*

1. *Within each q-partitioned Dorfman design class, $D^{(q)}$, $q = 2,\ldots,n-1$, there exists an optimal design that uses an ordered q-partition.*

2. *Within each q-partitioned mixed-testing design class, $M^{(q)}$, $q = 2,\ldots,n$, there exists an optimal design for which diseases $\{1,\ldots,n^I\}$ (i.e., the n^I diseases with the highest prevalences) are bundled into one multiplex assay, which is individually tested, and diseases $\{n^I+1,\ldots,n\}$ are tested via pooling, following some ordered $q-1$-partition, for some $n^I, n^D \in Z^+ : n^I + n^D = n$.*

From a practical perspective, the optimality of an ordered partition indicates that small forecasting errors in disease prevalences may not have a large impact on an optimal design as long as disease ordering is mostly preserved. Intuitively, the ordered partition result holds because for both independent diseases and no coinfections cases, an ordered two partition yields the highest (lowest) possible assay prevalence for one (the other) assay, among all two partitions with fixed assay sizes (hence, fixed assay costs), and the expected tests function is concave in assay prevalence (Property 2). This reasoning extends to any q-partition, because it can be split into multiple two partitions. In particular, in the no coinfections case, the total cost function ($TC(.)$) becomes concave in the sum of disease prevalences, and the results by Chakravarty et al. (1982) and Anily and Federgruen (1991) apply to our setting. In general, however, the total cost function is not necessarily concave in the sum of disease prevalences (because of coinfections), but the ordered partition result continues to hold for independent diseases (Theorem 5). The following remark and example provide some insight on when the ordered partition result may or may not hold.

Remark 6. Consider that the prevalence vector satisfies the transitivity property; that is, if $\pi_i \geq \pi_j$ for some $i,j \in N$, then $\pi(\{i,r\}) \geq \pi(\{j,r\})$, $\forall r \in N \setminus \{i,j\}$. The transitivity property is satisfied in both the independent diseases and the no coinfections cases.

The transitivity property of the prevalence vector is necessary but not sufficient for the ordered partition result in Theorem 5 to hold, as the following example demonstrates.

Example 2. Consider disease set, $N = \{1,2,3\}$, with joint vector, $\boldsymbol{p} = (p_0 = 0.79, p_1 = 0.05, p_2 = 0.05, p_3 = 0, p_{12} = 0.06, p_{13} = 0.05, p_{23} = 0, p_{123} = 0)$ (i.e., $A_3^+ \subseteq A_1^+$) and marginal prevalences, $\pi_1 = 0.16, \pi_2 = 0.11, \pi_3 = 0.05$, which satisfy the transitivity property (Remark 6). However, for $\lambda = 0.6$ and assay cost function $c(s) = s^{0.8}, s = 1,2,3$, the optimal design is a $D^{(2)}$ design with $S_1 = \{1,3\}$ and $S_2 = \{2\}$, which is not an ordered partition.

Theorem 5 leads to an efficient algorithm.

Corollary 1. *Suppose all disease prevalences are independent or that there are no coinfections.*

1. *The problem of finding an optimal partition of set N reduces to a shortest path problem on an acyclic directed graph $G(V(N), E(N))$ with*
   - *vertex set $V(N) = N \cup \{n+1\}$ (i.e., each disease in set N represents a vertex, and vertex $n+1$ represents a dummy vertex) and*
   - *edge set $E(N) = \{(i,j) : i < j, i,j \in V(N)\}$, where edge (i,j) represents an assay for diseases $i, i+1,\ldots,j-1$ (equivalently, $S = \{i, i+1,\ldots,j-1\}$, with size $s = j-i$), with edge weight $w_{i,j} = \tilde{c}(s,\lambda) \times T(S, t^*(S))$ (i.e., the assay's expected cost at the optimal pool size).*

2. *Both the construction of graph $G(V(N), E(N))$ and solving the shortest path problem (e.g., via a topological sorting algorithm) (Corman et al. 2009) have polynomial complexity, $O(n^2)$.*

Corollary 1 holds because each ordered q-partition of set $N, q = 1,\ldots,n$, is represented by a path from vertex 1

to vertex $n+1$ on $G(V(N), E(N))$. Thus, the set of all ordered partitions of set N corresponds to the set of all paths from vertex 1 to vertex $n+1$, and an optimal partition of set N corresponds to the shortest path from vertex 1 to vertex $n+1$. Thus, when disease prevalences are independent or there are no coinfections, an optimal testing design can be found in polynomial time. When there are coinfections and disease prevalences are not independent, the problem remains NP hard, and some of the existing algorithms for the partition problem (Section 3.2) can be modified to solve our problem, which is outside the scope of this paper.

Remark 7. When $\sum_{i\in N}\overline{\pi}_i \le 1$, Theorem 5 and Corollary 1 continue to hold for the deterministic counterpart of the robust problem, **R-TD**, which in this case, is based on a deterministic joint vector $\boldsymbol{p}$ with no coinfections (Corollary A.1 in Online Appendix A). Thus, there exists an optimal robust design that is ordered, which can be determined in polynomial time.

6. A Case Study of Respiratory Diseases

Large multiplex assays are used for a variety of diseases (see Section 1). In this case study, we focus on respiratory diseases, which are challenging for the testing design problem. Some respiratory diseases exhibit seasonal prevalence fluctuations, which not only lead to high uncertainty but also, alter the disease ordering dynamically throughout the year. As a group, respiratory diseases have higher prevalences than other disease groups, making pooling less beneficial; during our study period of 2018–2021, this group not only included an emerging disease (COVID-19) but also, diseases that responded to COVID-19 mitigation measures, resulting in drastic changes in their prevalence during the study period. We present the study design in Section 6.1, the data and sources in Section 6.2, and a discussion of the results in Section 6.3.

6.1. Study Design

We consider 18 respiratory diseases (14 viral and 4 bacterial) that manifest with overlapping clinical presentation over a four-year period (2018–2021) that spans both pre- and post-COVID periods. Weekly prevalence data are available during the study period.

6.1.1. Optimal Designs. Assay design is a tactical decision, as design changes require the acquisition of assays and modification of procedures. Consequently, we produce a family of *base Pareto designs* each for 2018 (without COVID-19) and 2021 (with COVID-19), which include the deterministic (mean-based) and robust designs using models **TD** and **R-TD**, respectively, based on the yearly means and the CI upper limits of disease prevalences for $\lambda \in [0,1]$ in increments of 0.05, illustrating the trade-offs between testing cost and capacity.

6.1.2. Performance Evaluation. We evaluate the base designs using weekly data from 2018 to 2021, and in our discussion, we distinguish between two settings: (1) the *perfect information* setting, where a design is evaluated using the data from which it is derived (in this setting, the design must contend with the *weekly prevalence variations* that underscore the mean (or the CI upper limit) because of seasonality and/or other natural variations); and (2) the *imperfect information* setting, where a design is evaluated using another year's data, such that the design must also contend with *forecast error* in the yearly mean. We compare the base designs with the two benchmarks from Remark 2: *multiplexing only* ($I^{(1)}$) and *pooling only* (the best of $I^{(n)}, D^{(n)}$, and $M^{(n)}$).

The total cost ($TC(.)$) (Equation (2)), which is a convex combination of the expected testing cost and number of tests, is an abstract construct. To better illustrate the trade-offs, we report the testing cost and the number of tests (evaluated based on actual data) for all $\lambda \in [0,1]$. For reference, the $\lambda = 0$ and $\lambda = 1$ designs minimize the expected number of tests and testing cost, respectively. We also report $\forall \lambda \in [0,1]$, the *price of robustness ratio* (Table B.1 in Online Appendix B), and

$$\text{Value of joint optimization ratio } (VoJ)(\lambda)\ (\%) = \frac{[TC(I^{(1)},\lambda) - TC^X(\lambda)]}{TC(I^{(1)},\lambda)} \times 100,\ X \in \{\boldsymbol{TD}, \boldsymbol{R\text{-}TD}\},$$

and in comparison with the multiplexing-only benchmark, this consistently outperforms the pooling-only benchmark in the case study (because of high prevalence rates) and represents current testing practices.

6.2. Data: Sources and Descriptive Statistics

Table 3 reports the mean prevalences for the 18 diseases for each year in the study period along with the data sources. For the 14 viral diseases, this mean is the average of the weekly prevalences; for the four bacterial diseases (*Mycoplasma pneumoniae*, *Bordetella parapertussis*, *Bordetella pertussis*, and *Chlamydophila pneumoniae*), these data were not available, and we use the literature to estimate weekly prevalences. The diseases are indexed following a nonincreasing order of their 2018 mean prevalences (thus, COVID-19 is #18). This order does not necessarily coincide with the weekly orders throughout 2018, the mean-based orders for other years, nor the CI-based orders for 2018–2021.

Coinfection rates, needed for **TD**, are not reported in the data sources. Hence, we assume that disease prevalences are independent; that is, coinfection rates are proportional to the corresponding disease prevalences. Further, for **R-TD**, based on our data, $\sum_{i\in N}\overline{\pi}_i \le 1$ in each year of the study period. Thus, we determine the optimal **TD** and **R-TD** designs in polynomial time using Corollary 1 (i.e., by determining the best-ordered partition) (Theorem 5).

Table 3. Disease Index, Name, Yearly Mean Prevalence for 2018–2021, and Data Sources

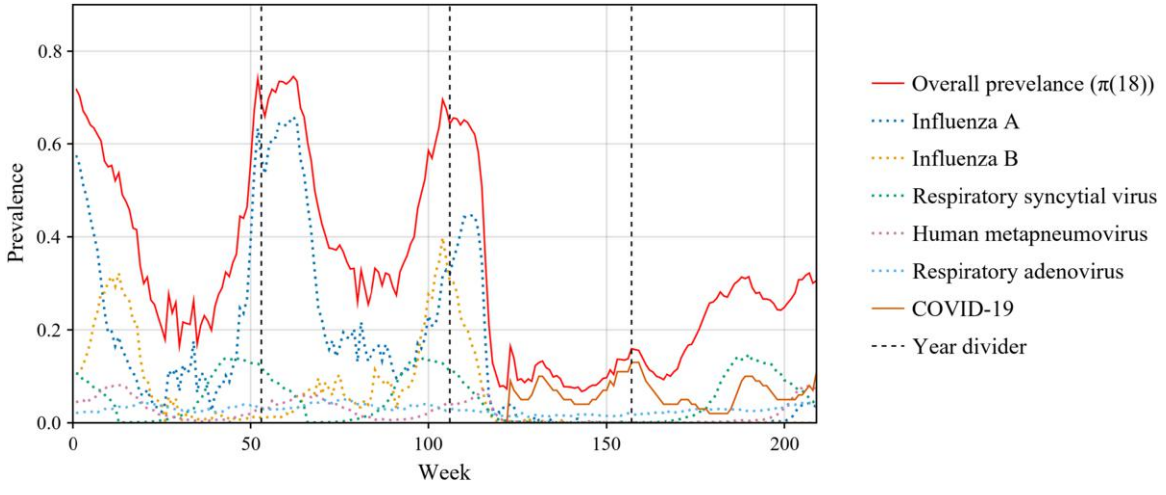
Index	Disease	Yearly mean 2018	2019	2020	2021	Data source
1	Influenza A	0.1902	0.2932	0.0930	0.0055	Centers for Disease Control and Prevention Influenza Division (2022)
2	Influenza B	0.0978	0.0887	0.0488	0.0001	Centers for Disease Control and Prevention Influenza Division (2022)
3	Respiratory syncytial virus	0.0599	0.0596	0.0207	0.0613	Centers for Disease Control and Prevention (2022b)
4	Human metapneumovirus	0.0315	0.0293	0.0149	0.0150	Centers for Disease Control and Prevention (2022b)
5	Respiratory adenovirus	0.0310	0.0368	0.0186	0.0284	Centers for Disease Control and Prevention (2022b)
6	Parainfluenza virus 3	0.0277	0.0273	0.0012	0.0340	Centers for Disease Control and Prevention (2022b)
7	CoVOC43	0.0096	0.0140	0.0023	0.0137	Centers for Disease Control and Prevention (2022b)
8	Parainfluenza virus 2	0.0093	0.0020	0.0005	0.0076	Centers for Disease Control and Prevention (2022b)
9	CoVNL63	0.0070	0.0081	0.0053	0.0069	Centers for Disease Control and Prevention (2022b)
10	CoVHKU1	0.0070	0.0062	0.0072	0.0003	Centers for Disease Control and Prevention (2022b)
11	Parainfluenza virus 4	0.0062	0.0054	0.0022	0.0048	Centers for Disease Control and Prevention (2022b)
12	*M. pneumoniae*	0.0051	0.0049	0.0050	0.0050	Schreckenberger and McAdam (2015)
13	*B. parapertussis*	0.0021	0.0021	0.0021	0.0021	Mastrantonio et al. (1998)
14	Parainfluenza virus 1	0.0019	0.0152	0.0016	0.0003	Centers for Disease Control and Prevention (2022b)
15	CoV229E	0.0015	0.0061	0.0007	0.0028	Centers for Disease Control and Prevention (2022b)
16	*B. pertussis*	0.0004	0.0004	0.0004	0.0004	Schreckenberger and McAdam (2015)
17	*C. pneumoniae*	0.0003	0.0003	0.0003	0.0004	Schreckenberger and McAdam (2015)
18	COVID-19	N/A	N/A	0.0450	0.0613	Centers for Disease Control and Prevention (2021c), Johns Hopkins University (2022)
$\pi(N)$	Overall prevalence	0.4042	0.4838	0.2421	0.2254	

Note. CoVOC43, Human coronavirus OC43; CoVNL63, Human coronavirus NL63; CoVHKU1, Human coronavirus HKU1; *M.*, *Mycoplasma*; *B. parapertussis*, *Bordetella parapertussis*; CoV229E, Human coronavirus 229E; *B. pertussis*, *Bordetella pertussis*; *C. pneumoniae*, *Chlamydia pneumoniae*; COVID-19, Coronavirus disease; N/A, non-applicable.

TD designs are based on the yearly mean prevalences in Table 3. For **R-TD**, we use the 52 weekly prevalence data each year to construct a 95% CI for each disease prevalence based on the Wald's method (Newcombe 1998), which serves as the uncertainty set for $\mathbf{\Pi}$ (Online Appendix B). Figure 1 plots the overall prevalence and select disease prevalences for each week in the study period, and it illustrates how disease prevalences, hence their ordering, fluctuate substantially throughout each year, with some diseases being highly seasonal. Therefore, even in the perfect information setting, an optimal **TD** design, which is an ordered partition based on yearly means, is not necessarily optimal for any given week in the year.

6.2.1. Cost Structure. PCR cost data from the literature (e.g., Schreckenberger and McAdam 2015) confirm the concave nature of the assay cost function in the number of diseases bundled. This is expected for genetic assays (see Section 1) because each PCR assay uses (1) common reagents and materials that are not pathogen specific (e.g., the polymerase enzyme, DNA nucleotides, vials), the required amounts of which are fairly insensitive to the number of diseases in the assay, and

Figure 1. (Color online) Overall Prevalence and Select Disease Prevalences per Week over 2018–2021



(2) pathogen-specific reagents (e.g., the primers and probes) that depend on the number of diseases, but their structure (hence, cost) is fairly similar for the different diseases. Consequently, we consider a fixed cost per assay and a variable cost per disease bundled; this function satisfies Assumption A. The cost data for medical tests are often difficult to find. Based on the limited data in Schreckenberger and McAdam (2015), we fit various assay cost functions (Online Appendix B). We demonstrate our findings for the $c(s) = 25.54 + 4.46 \times s$ function, but sensitivity analysis for other cost functions indicates similar qualitative findings. Because the number of tests (per subject) is one for the $I^{(1)}$ benchmark and the total cost is a convex combination of the testing cost and number of tests, we normalize the assay cost function so that $c(n) = 1$.

We consider a pool size limit, $\overline{M}$, of 32, which is common for PCR assays (Yelin et al. 2020, Bish et al. 2021, El Hajj et al. 2022b).

6.3. Case Study Findings

Key properties of optimal designs are established analytically (Sections 4 and 5). In Section 6.3.1, we use the perfect information setting to show additional properties; in Section 6.3.2, we compare optimal designs with benchmarks through the progression of the COVID-19 pandemic in the realistic, imperfect information setting.

6.3.1. Properties of Mean-Based and Robust Designs (Perfect Information Setting).

To motivate our discussion, Tables 4 and 5 provide the **TD** and **R-TD** Pareto designs along with the benchmarks designs and their results evaluated using 2018 (17 diseases) and 2021 (18 diseases, including COVID-19) weekly data in terms of the actual testing cost and number of tests (mean, minimum, and maximum values over the 52 weeks) under the perfect information settings (i.e., the 2018/2021 designs are evaluated using the 2018/2021 weekly data (under weekly variations but no forecast error)). Figure 2 plots the mean values for the testing cost and number of tests for each of the **TD** and **R-TD** designs to visualize the Pareto frontier. Tables 4 and 5 also report the value of joint optimization ratios (*VoJ*; in terms of its range over λ; i.e., with respect to the 17-plex (for 2018) or 18-plex (for 2021) $I^{(1)}$ benchmarks).

6.3.1.1. Mean-Based vs. Robust Designs.

Mean-based designs and their robust counterparts are quite similar, despite the differences in their inputs, including different disease orderings (e.g., for 2018, the mean overall prevalence for **TD** is $\pi(N) = 0.4042$, whereas for **R-TD**, it is 0.6079; i.e., the sum of the CI upper limits). This underscores another benefit of **TD**; by integrating multiplexing and pooling, the designs become structurally robust to variations in mean prevalences. To see this, consider that pooling is the cause of weekly variability in the metrics; individual testing always uses one test for all prevalences. Then, (1) through bundling of diseases in an assay, the assay's performance becomes a function of the assay prevalence (i.e., combined prevalences of the diseases in the assay) rather than the individual disease prevalence, which serves to increase its robustness under both weekly prevalence variations and forecast error. (2) The ordering of diseases within an assay is immaterial, and it is only when ordering

Table 4. 2018 TD and R-TD Designs, Metrics, and *VoJ* Based on 2018 Data Without COVID-19 Testing

					Perfect information setting		
Model (n = 17)	Range of λ values	Design class	Assay sizes	Pool sizes	Testing cost mean (min-max)	Number of tests mean (min-max)	$VoJ(\lambda)$ (%) range
TD	0.00–0.40	$I^{(1)}$	[17]	[1]	1.00 (1.00–1.00)	1.00 (1.00–1.00)	0.00–0.00
TD	0.45–0.55	$M^{(2)}$	[15, 2]	[1, 32]	0.93 (0.93–0.93)	1.05 (1.05–1.06)	0.11–1.35
TD	0.60–0.80	$M^{(2)}$	[12, 5]	[1, 13]	0.85 (0.84–0.89)	1.16 (1.13–1.24)	2.56–8.60
TD	0.85	$M^{(2)}$	[11, 6]	[1, 10]	0.84 (0.83–0.87)	1.21 (1.19–1.27)	10.21
TD	0.90	$M^{(3)}$	[8, 7, 2]	[1, 6, 32]	0.81 (0.76–0.97)	1.39 (1.29–1.67)	13.40
TD	0.95–1.00	$M^{(3)}$	[6, 6, 5]	[1, 5, 13]	0.79 (0.72–0.94)	1.55 (1.40–1.85)	16.70–21.50
		Disease order			1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17		
R-TD	0.00–0.40	$I^{(1)}$	[17]	[1]	1.00 (1.00–1.00)	1.00 (1.00–1.00)	0.00–0.00
R-TD	0.45–0.60	$M^{(2)}$	[15, 2]	[1, 32]	0.93 (0.93–0.93)	1.05 (1.05–1.06)	0.11–1.97
R-TD	0.65–0.80	$M^{(2)}$	[12, 5]	[1, 12]	0.85 (0.84–0.89)	1.16 (1.13–1.23)	4.03–8.57
R-TD	0.85–0.90	$M^{(2)}$	[11, 6]	[1, 9]	0.84 (0.83–0.87)	1.21 (1.19–1.26)	10.16–11.98
R-TD	0.95	$M^{(3)}$	[10, 5, 2]	[1, 8, 32]	0.83 (0.80–0.88)	1.31 (1.25–1.42)	14.75
R-TD	1.00	$M^{(3)}$	[7, 5, 5]	[1, 5, 12]	0.80 (0.74–0.96)	1.52 (1.39–1.84)	19.54
		Disease order			1, 2, 3, 4, 6, 5, 8, 7, 10, 9, 11, 12, 14, 13, 15, 16, 17		
Multiplexing only	N/A	$I^{(1)}$	[17]	[1]	1.00	1.00	N/A
Pooling only	N/A	$D^{(17)}$	17 singletons	3–32	1.24	4.20	N/A

Note. N/A, non-applicable.

Table 5. 2021 TD and R-TD Designs, Metrics, And *VoJ* Based on 2021 Data with COVID-19 Testing

					Perfect information setting		
Model (n = 18)	Range of λ values	Design class	Assay sizes	Pool sizes	Testing cost mean (min-max)	Number of tests mean (min-max)	$VoJ(\lambda)$ (%) range
TD	0.00–0.30	$D^{(1)}$	[18]	[3]	0.86 (0.59–1.02)	0.86 (0.59–1.02)	14.43–14.43
TD	0.35–0.70	$D^{(2)}$	[13, 5]	[3, 26]	0.71 (0.49–0.84)	0.93 (0.66–1.10)	14.65–22.44
TD	0.75	$D^{(2)}$	[11, 7]	[3, 13]	0.68 (0.48–0.82)	1.00 (0.71–1.21)	23.76
TD	0.80	$D^{(3)}$	[7, 6, 5]	[3, 7, 26]	0.63 (0.45–0.81)	1.21 (0.87–1.57)	25.77
TD	0.85–1.00	$D^{(3)}$	[6, 7, 5]	[3, 6, 26]	0.62 (0.45–0.80)	1.23 (0.89–1.57)	28.62–37.82
		Disease order			18, 3, 6, 4, 5, 8, 15, 9, 1, 12, 11, 10, 14, 16, 17, 13, 7, 2		
R-TD	0.00–0.30	$D^{(1)}$	[18]	[3]	0.86 (0.59–1.02)	0.86 (0.59–1.02)	14.43–14.43
R-TD	0.35–0.65	$D^{(2)}$	[13, 5]	[3, 26]	0.71 (0.49–0.84)	0.93 (0.66–1.10)	14.65–21.33
R-TD	0.70–0.75	$D^{(2)}$	[11, 7]	[3, 13]	0.68 (0.48–0.82)	1.00 (0.71–1.21)	22.14–23.76
R-TD	0.80–1.00	$D^{(3)}$	[6, 7, 5]	[3, 6, 26]	0.62 (0.45–0.80)	1.23 (0.89–1.57)	25.56–37.82
		Disease order			18, 3, 6, 5, 4, 7, 8, 9, 1, 12, 11, 15, 13, 16, 17, 10, 14, 2		
Multiplexing only	N/A	$I^{(1)}$	[18]	[1]	1.00	1.00	N/A
Pooling only	N/A	$D^{(18)}$	18 singletons	5–32	0.92	3.25	N/A

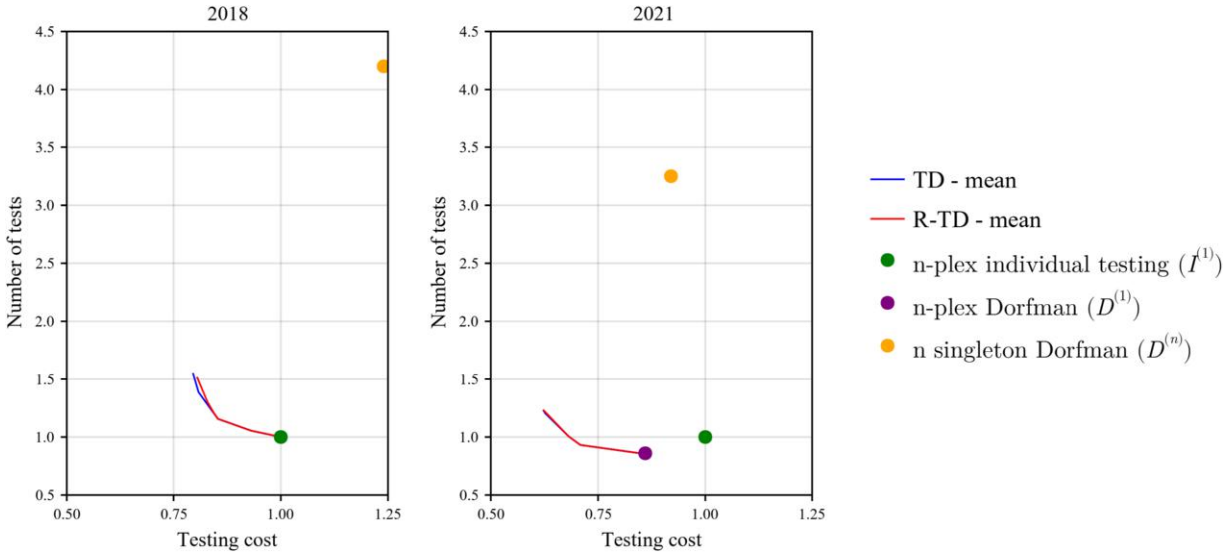
Note. N/A, non-applicable.

errors place the disease in the "wrong" assay do suboptimalities arise. (3) A particular pool size remains optimal for a range of prevalences and beneficial for an even larger range. The performances of the mean-based and robust designs are even closer in 2021 because of a lower overall prevalence and lower CI upper limits (e.g., influenza had lower prevalences and less pronounced seasonality, hence lower uncertainty). The price of robustness is also low (Table B.1 in Online Appendix B). Robust designs may even *lower* the actual cost (e.g., $\lambda = 0.70$ for 2021) (Table 5 and Table B.1 in Online Appendix B) because **TD** uses the expected costs based on yearly means, which may deviate from the actual costs because of weekly variations.

6.3.1.2. Comparison with Benchmarks.

The multiplexing-only ($I^{(1)}$) benchmark is on the Pareto frontier only for 2018 (not 2021) and only for $0 \le \lambda \le 0.4$ (Figure 2 and Theorem 2). For 2018 (Table 4), both **TD** and **R-TD** use multiple assays with mixed testing methods (pooling and individual testing) for all $\lambda \ge 0.45$; these designs have positive VoJ values and show the trade-off between testing cost and number of tests. For 2021 (Table 5), which has a lower overall prevalence than 2018, every **TD** and **R-TD** design is a Dorfman design with multiplexing, and some of these Pareto designs improve upon both the testing cost and the number of tests compared with the benchmarks. To better understand the practical impact of the trade-off between the testing cost and the number of tests, remember that

Figure 2. (Color online) 2018 and 2021 Perfect Information TD and R-TD Designs and Benchmarks



PCR machines can have large capacities (see Section 1); hence, there are many settings in which they are run below capacity, and increasing the number of tests to reduce the testing cost in these settings can be an especially good trade-off.

Pooling alone does not provide as much benefit as the integrated mean-based or robust designs; the pooling-only benchmark is not on the Pareto frontier in either year. On the other hand, pooling of the n-plex ($D^{(1)}$) is not useful in 2018; for 2021, it can improve the 18-plex benchmark by 14.43% (i.e., $VoJ(\lambda = 0)$) compared with a VoJ of 37.82% attained through integrated multiplexing and pooling.

6.3.1.3. Design Structures. Noting the mean overall prevalence of 0.4042 for 2018 and 0.2254 for 2021 (above and below the pooling threshold of 0.31, respectively), the optimal designs in Tables 4 and 5 are in agreement with Theorems 1 and 2. As λ increases, all designs move toward more assays. This is expected for the 2021 designs in light of Theorem 4 (as $\pi(18) = 0.2254 < 0.31$ for 2021), but in the absence of this condition, we still see a similar trend for the 2018 designs. The first assay almost always bundles the most diseases (which have the highest prevalence); hence, it has the largest assay prevalence and the smallest pool size. This keeps the first assay's pool size constant as λ increases, allowing the other assays to use larger pools. The robust designs tend to have smaller pools than the mean-based designs because of the use of the CI upper limits.

6.3.1.4. Effect of Seasonality. One challenge for testing design for respiratory diseases is seasonality. We explore the benefits of accounting for seasonality through the use of a separate **TD** design for the high versus low season constructed and evaluated based on the 2018 data. We define the seasons based on the pooling threshold of 31%, as this implies when pooling is beneficial for the 17-plex. The *low season* spans the period for which the weekly overall prevalence remains below 31%, which for 2018, was weeks 20–44, and the remaining weeks represent the *high season*. The VoJ range over the 17-plex $I^{(1)}$ is $[VoJ(\lambda = 0) = 0,\ VoJ(\lambda = 1) = 20.5\%]$ (Table 4) for original 2018 **TD** designs (without accounting for seasonality) and $[VoJ(\lambda = 0) = 5.3\%,\ VoJ(\lambda = 1) = 25.2\%]$ (Table B.2 in Online Appendix B) with high- and low-season designs. Thus, accounting for seasonality can further increase the benefits of optimal designs.

6.3.2. Testing Design During the Pandemic (Imperfect Information Designs).

The previous perfect information setting provides a reference point on the potential benefits of joint optimization. Next, we study testing design in the realistic, imperfect information setting, where prevalences are unknown at the time of testing design. One can integrate a sophisticated forecasting method (available in the literature for influenza and COVID-19) into our testing design models. Although this is worthy of future investigation, a main finding from the study of the imperfect information setting in this section is that *sophisticated forecasting methods may not be necessary to reap most of the benefits*. Integrating multiplexing and pooling already increases the robustness of optimal designs to forecast errors; see Section 6.3.1. As we shall see, the 2018 designs or simple modifications of them to account for COVID-19 effects on influenza rates work very well compared with the benchmarks.

6.3.2.1. 2019–2021 (Without COVID-19 Testing): Performance of 2018 Designs. We first explore the performance of the 2018 designs for the 17 diseases (from Table 4) (i.e., designs constructed using the 2018 data) on data from 2019 to 2021. Although the disease ordering in 2018 does not match that in 2019, 2020, or 2021, these **TD** and **R-TD** imperfect information designs continue to substantially outperform both benchmarks for the 2019–2021 period (Table B.3 in Online Appendix B). For example, the $VoJ(\lambda = 1)$ values for mean-based and robust designs are 20.5% and 18.4%, respectively, over the $I^{(1)}$ benchmark in this period. These benefits are similar in scale to the 2018 designs under perfect information, continuing to show that both **TD** and **R-TD** are quite robust to forecast errors.

Next, we show that we can improve upon the performance of the 2018 designs, evaluated using 2021 data, via design modifications based on minimal information on the impact of COVID-19.

6.3.2.2. 2021 (Both with and Without COVID-19 Testing): Performance of Modified 2018 Designs. Because of the pandemic, 2021 was quite an unusual year for respiratory diseases. The protective measures taken for COVID-19 also reduced the influenza rates, greatly perturbing the order of diseases from 2018. This was apparent in November to December of 2020 (the beginning of the 2020/2021 influenza season), during which the reported influenza rates were much lower than the historic averages, and such unusual influenza patterns were anticipated to continue into 2021. Thus, we now demonstrate how the testing designs for 2021 can be improved with simple forecasting. To this end, we simply set influenza A and B prevalences to their averages from November and December of 2020; for all other diseases, we continue to use the 2018 data. Using this *modified* 2018 data, we produce *modified 2018 designs* using **TD** and **R-TD**. First, we produce testing designs for the 17 diseases (without COVID-19); then, turn our attention to COVID-19. Forecasting an emerging and evolving disease is very difficult, and many COVID-19 forecasting methods in the literature turned out to be highly inaccurate. In the absence of a reliable COVID-19 forecast in 2021, we explore a simple strategy of

adding COVID-19 into the first assay in the modified 2018 designs for **TD** and **R-TD**, thus bundling it with the most prevalent diseases in each design. Retrospectively, this can be accomplished using the modified 2018 data and a mean COVID-19 prevalence of 10%. With an actual 2021 COVID-19 mean of 6.1% (unknown at the time of testing design), it is suboptimal and represents an overestimation.

Table B.4 in Online Appendix B reports the results for the modified 2018 designs and for comparison purposes, the 2021 perfect information designs both with and without COVID-19 testing, all evaluated using 2021 data. For each design with COVID-19 testing, Figure 3 also depicts the weekly performance using 2021 data in terms of the actual testing cost, the number of tests, and its mean performance over the year (in solid lines). The pooling-only benchmark is not depicted in Figure 3, as it continues to be substantially dominated by all other designs. Finally, Table 6 reports the *VoJ* for the modified 2018 and perfect information 2021 designs with respect to the $I^{(1)}$ benchmark based on the 2021 data.

6.3.2.3. Perfect Information vs. Imperfect Information Designs. The family of modified 2018 **TD** designs performs very well in 2021 both with and without COVID-19 testing (Figure 3 and Table B.4 in Online Appendix B), and their *VoJ* values are high and close to the 2021 perfect information design values, which represent the maximum possible benefits under perfect information (Table 6). These findings continue to illustrate the structurally robust nature of **TD** designs under prevalence uncertainty.

6.3.2.4. Mean-Based vs. Robust Designs. Based on the 2021 data, the price of robustness is very low for the modified 2018 and 2021 perfect information designs, with the exception of the modified 2018 **R-TD** designs with COVID-19 testing (Table 6). This is mainly because the overestimation of the COVID-19 mean, combined with higher prevalence upper limits for the other diseases (based on the modified 2018 data), leads to an overall prevalence for the robust model that exceeds the pooling threshold, making it more conservative than all other designs, which is expected based on Table 2. In particular, comparing modified 2018 **TD** and **R-TD** designs with COVID-19 testing (based on Table B.4 in Online Appendix B), for $\lambda \in [0, 0.35]$, the robust design is the 18-plex $I^{(1)}$ benchmark, whereas the mean-based design uses a pool of three; for $\lambda \in [0.40, 0.55]$, the robust design has a negative *VoJ*, unlike the mean-based design with a positive *VoJ* (in this case, the robust design reduces the testing cost but not enough to compensate for the increase in the number of tests). The robust design starts improving on the $I^{(1)}$ benchmark only at $\lambda = 0.6$, and at $\lambda = 0.95$, it performs nearly as well as the mean-based design with respect to *VoJ*. This is because as λ increases, the robust design moves away from the $I^{(1)}$ benchmark and starts using multiple pooled assays, the cost benefit of which outweighs the increase in the number of tests.

6.3.2.5. Comparison with Benchmarks. Modified 2018 **TD** designs outperform the $I^{(1)}$ benchmark on 2021 data; for the robust model, this happens mainly for the no COVID-19 testing case, as explained.

Figure 3. (Color online) Modified 2018 vs. 2021 TD Designs and $I^{(1)}$ Benchmark Based on 2021 Data

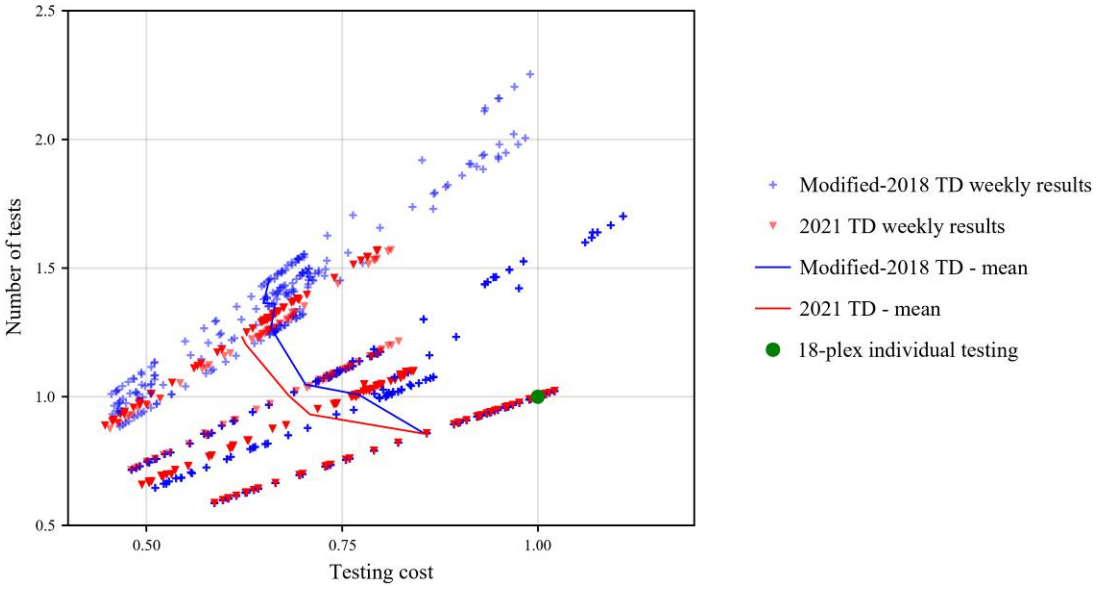


Table 6. *VoJ* for Modified 2018 and 2021 TD and R-TD Designs Based on 2021 Data with and Without COVID-19 Testing

Without COVID-19								
Modified 2018 TD	λ range	0.00–0.40	0.45–0.60	0.65–0.85	0.90–0.95	1.00		
	$VoJ(\lambda)$ (%)	24.40–24.40	19.35–22.69	26.26–32.63	34.71–37.52	39.68		
Modified 2018 R-TD	λ range	0.00–0.35	0.40–0.55	0.60–0.80	0.85	0.90	0.95	1.00
	$VoJ(\lambda)$ (%)	24.40–24.40	18.50–21.82	24.77–31.12	32.56	34.88	37.27	39.48
2021 TD	λ range	0.00–0.35	0.40–0.75	0.80	0.85–0.90	0.95–1.00		
	$VoJ(\lambda)$ (%)	24.40–24.40	25.15–32.40	33.60	35.72–38.39	41.00–43.81		
2021 R-TD	λ range	0.00–0.30	0.35–0.70	0.75	0.80	0.85	0.90–1.00	
	$VoJ(\lambda)$ (%)	24.40–24.40	24.12–31.36	32.11	33.31	35.72	38.18–43.81	
With COVID-19								
Modified 2018 TD	λ range	0.00–0.35	0.40–0.55	0.60–0.80	0.85	0.90	0.95	1.00
	$VoJ(\lambda)$ (%)	14.43–14.43	8.66–12.22	15.92–22.79	24.97	26.60	31.47	34.29
Modified 2018 R-TD	λ range	0.00–0.35	0.40–0.55	0.60–0.85	0.90	0.95	1.00	
	$VoJ(\lambda)$ (%)	0.00–0.00	(−5.23)–(−1.31)	2.91–12.67	13.71	31.54	34.09	
2021 TD	λ range	0.00–0.30	0.35–0.70	0.75	0.80	0.85–1.00		
	$VoJ(\lambda)$ (%)	14.43–14.43	14.65–22.44	23.76	25.77	28.62–37.82		
2021 R-TD	λ range	0.00–0.30	0.35–0.65	0.70–0.75	0.80–1.00			
	$VoJ(\lambda)$ (%)	14.43–14.43	14.65–21.33	22.14–23.76	25.56–37.82			

7. Conclusions, with Limitations and Future Research Directions

We develop tools and insights for multidisease testing design for public health screening practitioners. In the absence of practical guidelines on how multiplexing and pooling should be integrated, our study sheds important light on efficient testing practices, which can also benefit other stakeholders. For example, efficient testing enables timely and accurate diagnosis, leading to improved public health outcomes. Because the use of multiplexing and pooling does not require obtaining extra specimens from the patient, the patient does not notice any difference, but expanded testing improves patient outcomes and satisfaction (Schreckenberger and McAdam 2015); assay manufacturers can also benefit from expanded testing and hence, an expanded market. Quantifying such benefits for other stakeholders is an important research direction.

7.1. Limitations and Future Research Directions

To our knowledge, our model is the first mathematical model to combine multiplexing and pooling optimization in testing design for public health screening, and it is our hope that this work will spur new academic research as well as empirical explorations to further investigate the benefits of optimal testing designs. Although our model considers many realistic aspects of the testing problem, any analytical model must rely on certain assumptions. Some of these assumptions, discussed here, can be considered as limitations of this research, but they also present opportunities for future research.

To increase model realism, several assumptions can be relaxed. For example, although PCR assays are highly sensitive in general, they may miss a disease during the window period when the pathogen has extremely low concentration in the specimen. Incorporating the pathogen dynamics into testing design is an important extension. We consider that the set of diseases for screening is given; adding a disease selection component to testing design (thus, selecting from a set of *potential* diseases) would increase the practical impact of this research effort. Further, studying a stochastic formulation of the testing design problem and considering correlated disease prevalences over time (e.g., because of seasonality) are worthwhile extensions. Relatedly, although the price of robustness was low in our case study, in general a mini-max-type objective can be overly conservative, and it might be promising to explore other objectives, such as regret-based objectives (e.g., El Amine et al. 2017).

Testing design for respiratory diseases, considered in our case study, can be challenging; a primary reason for this challenge is seasonality. Our simple way of accounting for seasonality in testing design shows promise and should be explored further. Our case study also suggests that optimizing the testing design infrequently and using simple forecasting methods to deal with a rare event, such as the emergence of COVID-19, work well for testing design in the post-COVID-19 period. This is one particular strength of multiplex assays; by bundling diseases in an assay, their combined prevalence becomes more reliable. Further, small forecasting errors may not impact an optimal design as long as disease ordering is not altered much. Nevertheless, it is valuable to explore the benefits of more accurate forecasting methods for testing design or more frequent optimization of pool sizes than that for the assay portfolio because pool sizes may be easier to change on a more frequent basis. It is also important to conduct case studies of other disease groups so as to quantify the benefits of optimal testing designs in other contexts.

We study the tactical testing design decision. Complementary future research directions include operational decisions: for instance, the online optimization problem (Golrezaei et al. 2014, Elmachtoub and Levi 2016, Keyvanshokooh et al. 2021) that arises in our setting when batches of specimens arrive at the laboratory in a stochastic manner. Thus, given PCR testing machine(s) and capacity, the tester needs to decide when to run the machine (i.e., below capacity or wait for more specimens?).

Acknowledgments

The authors are grateful to the Department Editor Prof. Özlem Ergün, the associate editor, and two referees for providing suggestions and comments on this manuscript, which led to significant improvements in the analysis, the insights gained, and the presentation. The authors also thank the Centers for Disease Control and Prevention's National Respiratory and Enteric Virus Surveillance System for providing data and the Bureau of Clinical Laboratories at the Alabama Department of Public Health for insights on respiratory disease screening. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

Al Ghounaim M, Xiao Y, Caya C, Papenburg J (2017) Diagnostic yield and clinical impact of routine cell culture for respiratory viruses among children with a negative multiplex RT-PCR result. *J. Clinical Virology* 94:107–109.

American Red Cross (2022) Infectious disease, HLA and ABO donor qualification testing. Accessed June 1, 2022, https://www.redcrossblood.org/biomedical-services/blood-diagnostic-testing/blood-testing.html.

Anily S, Federgruen A (1991) Structured partitioning problems. *Oper. Res.* 39(1):130–149.

Aprahamian H, Bish DR, Bish EK (2020) Optimal group testing: Structural properties and robust solutions, with application to public health screening. *INFORMS J. Comput.* 32(4):895–911.

Balakrishnan G, Landeen C, Dy-Boarman EA, Wall GC (2016) Outcomes with multplex PCR respiratory sample testing on hospital inpatients: Outcomes and cost associated with a multiplex PCR system on nasopharyngeal swab specimens (FilmArray Respiratory Panel) in a large tertiary inpatient hospital population. *J. Acad. Hospital Medicine* 8(3):1–4.

Balas E, Padberg MW (1976) Set partitioning: A survey. *SIAM Rev.* 18(4):710–760.

Barahona F, Anbil R (2002) On some difficult linear programs coming from set partitioning. *Discrete Appl. Math.* 118(1–2):3–11.

BCL-APH (2022) Discussion with Mark Ellis of the Bureau of Clinical Laboratories—Alabama Public Health. Working paper.

Beckmann C, Hirsch HH (2016) Comparing Luminex NxTAG-Respiratory Pathogen Panel and RespiFinder-22 for multiplex detection of respiratory pathogens. *J. Medical Virology* 88(8):1319–1324.

Biomérieux (2022) FILMARRAY™ Respiratory Panel. Accessed June 1, 2022, https://www.biomerieux-nordic.com/product/filmarray-respiratory-panel.

Bish DR, Bish EK, El Hajj H, Aprahamian H (2021) A robust pooled testing approach to expand COVID-19 screening capacity. *PLoS One* 16(2):e0246285.

Borges CM, Pathela P, Pirillo R, Blank S (2015) Targeting the use of pooled HIV RNA screening to reduce cost in health department STD clinics: New York City, 2009–2011. *Public Health Rep.* 130(1):81–86.

Boschetti MA, Mingozzi A, Ricciardelli S (2008) A dual ascent procedure for the set partitioning problem. *Discrete Optim.* 5(4):735–747.

Centers for Disease Control and Prevention (2021a) Interim guidance for use of pooling procedures in sars-cov-2 diagnostic and screening testing. Accessed June 1, 2022, https://www.cdc.gov/coronavirus/2019-ncov/lab/pooling-procedures.html.

Centers for Disease Control and Prevention (2021b) Nucleic acid amplification tests (NAATs). Accessed June 1, 2022, https://www.cdc.gov/coronavirus/2019-ncov/lab/naats.html.

Centers for Disease Control and Prevention (2021c) Past weekly surveillance reports. Accessed June 1, 2022, https://www.cdc.gov/coronavirus/2019-ncov/covid-data/covidview/index.html.

Centers for Disease Control and Prevention (2022a) Similarities and differences between flu and COVID-19. Accessed June 1, 2022, https://www.cdc.gov/flu/symptoms/flu-vs-covid19.htm.

Centers for Disease Control and Prevention (2022b) The National Respiratory and Enteric Virus Surveillance System (NREVSS). Accessed June 1, 2022, https://www.cdc.gov/surveillance/nrevss/.

Centers for Disease Control and Prevention Influenza Division (2022) FluView—Weekly U.S. Influenza Surveillance Report. Accessed June 1, 2022, https://www.cdc.gov/flu/weekly/#ClinicalLaboratories.

Chakravarty AK, Orlin JB, Rothblum UG (1982) A partitioning problem with additive objective with an application to optimal inventory groupings for joint replenishment. *Oper. Res.* 30(5): 1018–1022.

Chan M, Koo SH, Jiang B, Lim PQ, Tan TY (2018) Comparison of the Biofire Film Array Respiratory Panel, Seegene AnyplexII RV16, and Argene for the detection of respiratory viruses. *J. Clinical Virology* 106:13–17.

Chan TJ, Yano CA (1992) A multiplier adjustment approach for the set partitioning problem. *Oper. Res.* 40(1 suppl 1): S40–S47.

Chen H, Weng H, Lin M, He P, Li Y, Xie Q, Ke C, Jiao X (2017) The clinical significance of FilmArray Respiratory Panel in diagnosing community-acquired pneumonia. *BioMed Res. Internat.* 2017: 7320859.

Corless RM, Gonnet GH, Hare DE, Jeffrey DJ, Knuth DE (1996) On the LambertW function. *Adv. Comput. Math.* 5(1):329–359.

Cormen TH, Leiserson CE, Rivest RL, Stein C (2009) *Introduction to Algorithms* (MIT Press, Cambridge, MA).

Dai B, Ding S, Wahba G (2013) Multivariate Bernoulli distribution. *Bernoulli* 19(4):1465–1483.

Das S, Dunbar S, Tang Y-W (2018) Laboratory diagnosis of respiratory tract infections in children—The state of the art. *Frontiers Microbiology* 9:2478.

Dorfman R (1943) The detection of defective members of large populations. *Ann. Math. Statist.* 14(4):436–440.

El Amine H, Bish EK, Bish DR (2017) Robust post-donation blood screening under prevalence rate uncertainty. *Oper. Res.* 66(1):1–17.

El Hajj H, Bish DR, Bish EK (2022a) Optimal genetic screening for cystic fibrosis. *Oper. Res.* 70(1):265–287.

El Hajj H, Bish DR, Bish EK, Aprahamian H (2022b) Screening multi-dimensional heterogeneous populations for infectious diseases under scarce testing resources, with application to COVID-19. *Naval Res. Logist.* 69(1):3–20.

El Hajj H, Bish DR, Bish EK, Kay D (2022c) Novel pooling strategies for genetic testing, with application to newborn screening. *Management Sci.* 68(11):7994–8014.

Elmachtoub AN, Levi R (2016) Supply chain management with online customer selection. *Oper. Res.* 64(2):458–473.

Emrich L, Piedmonte M (1991) A method for generating high-dimensional multivariate binary variates. *Amer. Statist.* 45(4): 302–303.

Fisher ML, Kedia P (1990) Optimal solution of set covering/partitioning problems using dual heuristics. *Management Sci.* 36(6): 674–688.

Fleming-Dutra KE, Hersh AL, Shapiro DJ, Bartoces M, Enns EA, File TM, Finkelstein JA, et al. (2016) Prevalence of inappropriate antibiotic prescriptions among US ambulatory care visits, 2010–2011. *JAMA* 315(17):1864–1873.

Food and Drug Administration (2020) Pooled sample testing and screening testing for COVID-19. Accessed June 1, 2022, https://www.fda.gov/medical-devices/coronavirus-covid-19-and-medical-devices/pooled-sample-testing-and-screening-testing-covid-19.

Food and Drug Administration (2021) Nucleic acid based tests. Accessed June 1, 2022, https://www.fda.gov/medical-devices/in-vitro-diagnostics/nucleic-acid-based-tests.

Gharabaghi F, Hawan A, Drews SJ, Richardson SE (2011) Evaluation of multiple commercial molecular and conventional diagnostic assays for the detection of respiratory viruses in children. *Clinical Microbiology Infect.* 17(12):1900–1906.

Ginocchio CC, McAdam AJ (2011) Current best practices for respiratory virus testing. *J. Clinical Microbiology* 49:S44–S48.

Golrezaei N, Nazerzadeh H, Rusmevichientong P (2014) Real-time optimization of personalized assortments. *Management Sci.* 60(6):1532–1551.

Gonsalves S, Mahony J, Rao A, Dunbar S, Juretschko S (2019) Multiplexed detection and identification of respiratory pathogens using the NxTAG® respiratory pathogen panel. *Methods* 158:61–68.

Hou P, Tebbs JM, Bilder CR, McMahan CS (2017) Hierarchical group testing for multiple infections. *Biometrics* 73(2):656–665.

Hou P, Tebbs JM, Wang D, McMahan CS, Bilder CR (2020) Array testing for multiplex assays. *Biostatistics* 21(3):417–431.

Hwang FK (1981) Optimal partitions. *J. Optim. Theory Appl.* 34(1):1–10.

Hwang FK, Sun J, Yao EY (1985) Optimal set partitioning. *SIAM J. Algebraic Discrete Methods* 6(1):163–170.

Johns Hopkins University (2022) Coronavirus Resource Center. Accessed June 1, 2022, https://coronavirus.jhu.edu/region/united-states.

Johnson M (2021) PCR machines. *Materials Methods* 3:193.

Kenmoe S, Tcharnenwa C, Monamele GC, Kengne CN, Ripa MN, Whitaker B, Alroy KA, et al. (2019) Comparison of FTD® respiratory pathogens 33 and a singleplex CDC assay for the detection of respiratory viruses: A study from Cameroon. *Diagnostic Microbiology Infectious Disease* 94(3):236–242.

Keyvanshokooh E, Shi C, Van Oyen MP (2021) Online advance scheduling with overtime: A primal-dual approach. *Manufacturing Service Oper. Management* 23(1):246–266.

Kim H-Y, Hudgens MG, Dreyfuss JM, Westreich DJ, Pilcher CD (2007) Comparison of group testing algorithms for case identification in the presence of test error. *Biometrics* 63(4):1152–1163.

Leland DS, Ginocchio CC (2007) Role of cell culture for virus detection in the age of technology. *Clinical Microbiology Rev.* 20(1):49–78.

Levine D (1996) *A Parallel Genetic Algorithm for the Set Partitioning Problem* (Springer, Boston), 23–35.

Lewis M, Kochenberger G, Alidaee B (2008) A new modeling and solution approach for the set-partitioning problem. *Comput. Oper. Res.* 35(3):807–813.

Lindan C, Mathur M, Kumta S, Jerajani H, Gogate A, Schachter J, Moncada J (2005) Utility of pooled urine specimens for detection of *Chlamydia trachomatis* and *Neisseria gonorrhoeae* in men attending public sexually transmitted infection clinics in Mumbai, India, by PCR. *J. Clinical Microbiology* 43(4):1674–1677.

Llor C, Bjerrum L (2014) Antimicrobial resistance: Risk associated with antibiotic overuse and initiatives to reduce the problem. *Therapeutic Adv. Drug Safety* 5(6):229–241.

Mahony JB, Blackhouse G, Babwah J, Smieja M, Buracond S, Chong S, Ciccotelli W, et al. (2009) Cost analysis of multiplex PCR testing for diagnosing respiratory virus infections. *J. Clinical Microbiology* 47(9):2812–2817.

Manne F, Sorevik T (1995) Optimal partitioning of sequences. *J. Algorithms* 19(2):235–249.

Mastrantonio P, Stefanelli P, Giuliano M, Herrera Rojas Y, Ciofi degli Atti M, Anemona A, Tozzi AE (1998) *Bordetella parapertussis* infection in children: Epidemiology, clinical symptoms, and molecular characteristics of isolates. *J. Clinical Microbiology* 36(4):999–1002.

May S, Gamst A, Haubrich R, Benson C, Smith DM (2010) Pooled nucleic acid testing to identify antiretroviral treatment failure during HIV infection. *J. Acquired Immune Deficiency Syndromes* 53(2):194–201.

Newcombe RG (1998) Two-sided confidence intervals for the single proportion: Comparison of seven methods. *Statist. Medicine* 17(8):857–872.

Nguyen NT, Bish EK, Bish DR (2021) Optimal pooled testing design for prevalence estimation under resource constraints. *Omega* 105:102504.

Nguyen NT, Aprahamian H, Bish EK, Bish DR (2019) A methodology for deriving the sensitivity of pooled testing, based on viral load progression and pooling dilution. *J. Translational Medicine* 17(1):1–10.

Ogilvie M (2001) Molecular techniques should not now replace cell culture in diagnostic virology laboratories. *Rev. Medical Virology* 11(6):351–354.

Olstad B, Manne F (1995) Efficient partitioning of sequences. *IEEE Trans. Comput.* 44(11):1322–1326.

Oved K, Cohen A, Boico O, Navon R, Friedman T, Etshtein L, Kriger O, et al. (2015) A novel host-proteome signature for distinguishing between acute bacterial and viral infections. *PLoS One* 10(3):e0120012.

Perakis G, Roels G (2008) Regret in the newsvendor model with partial information. *Oper. Res.* 56(1):188–203.

Pinsky B, Hayden R (2019) Cost-effective respiratory virus testing. *J. Clinical Microbiology* 57(9):e00373-19.

Rachlin J, Ding C, Cantor C, Kasif S (2005) MuPlex: Multi-objective multiplex PCR assay design. *Nucleic Acids Res.* 33(suppl 2): W544–W547.

Ramanan P, Bryson AL, Binnicker MJ, Pritt BS, Patel R (2018) Syndromic panel-based testing in clinical microbiology. *Clinical Microbiology Rev.* 31(1):e00024-17.

Rogers BB, Shankar P, Jerris RC, Kotzbauer D, Anderson EJ, Watson JR, O'Brien LA, Uwindatwa F, McNamara K, Bost JE (2015) Impact of a rapid respiratory panel test on patient outcomes. *Arch. Pathology Laboratory Medicine* 139(5):636–641.

Schreckenberger PC, McAdam AJ (2015) Point-counterpoint: Large multiplex PCR panels should be first-line tests for detection of respiratory and intestinal pathogens. *J. Clinical Microbiology* 53(10):3110–3115.

Subramony A, Zachariah P, Krones A, Whittier S, Saiman L (2016) Impact of multiplex polymerase chain reaction testing for respiratory pathogens on healthcare resource utilization for pediatric inpatients. *J. Pediatrics* 173:196–201.

Taylor SM, Juliano JJ, Trottman PA, Griffin JB, Landis SH, Kitsa P, Tshefu AK, Meshnick SR (2010) High-throughput pooling and real-time PCR-based strategy for malaria detection. *J. Clinical Microbiology* 48(2):512–519.

Tebbs JM, McMahan CS, Bilder CR (2013) Two-stage hierarchical group testing for multiple infections with application to the infertility prevention project. *Biometrics* 69(4): 1064–1073.

Templeton KE (2007) Why diagnose respiratory viral infection? *J. Clinical Virology* 40:S2–S4.

Vallières E, Renaud C (2013) Clinical and economical impact of multiplex respiratory virus assays. *Diagnostic Microbiology Infectious Disease* 76(3):255–261.

Van TT, Miller J, Warshauer DM, Reisdorf E, Jernigan D, Humes R, Shult PA (2012) Pooling nasopharyngeal/throat swab specimens

to increase testing capacity for influenza viruses by PCR. *J. Clinical Microbiology* 50(3):891–896.

Van Hulst M, Hubben GA, Sagoe KW, Promwong C, Permpikul P, Fongsatitkul L, Glynn DM, Smit Sibinga CT, Postma MJ (2009) Web interface-supported transmission risk assessment and cost-effectiveness analysis of postdonation screening: A global model applied to Ghana, Thailand, and the Netherlands. *Transfusion* 49(12):2729–2742.

van Zyl G, Preiser W, Potschka S, Lundershausen A, Haubrich R, Smith D (2011) Pooling strategies to reduce the cost of HIV-1 RNA load monitoring in a resource-limited setting. *Clinical Infectious Diseases* 52(2):264–270.

Yelin I, Aharony N, Tamar ES, Argoetti A, Messer E, Berenbaum D, Shafran E, et al. (2020) Evaluation of COVID-19 RT-qPCR test in multi sample pools. *Clinical Infectious Diseases* 71(16): 2073–2078.

Yuan J, Yi J, Zhan M, Xie Q, Zhen TT, Zhou J, Li Z, Li Z (2021) The web-based multiplex PCR primer design software Ultiplex and the associated experimental workflow: Up to 100-plex multiplicity. *BMC Genomics* 22(1):835.