
EFFECTO DE LA CORRELACIÓN Y LOS FALSOS NEGATIVOS EN LA ESTRATEGIA DE TESTEO GRUPAL PARA DETECCIÓN DE COVID-19

LEONARDO BASSO * †

VICENTE SALINAS * †

DENIS SAURÉ * §

CHARLES THRAVES * §

NATALIA YANKOVIC ¶

Resumen

La pandemia de COVID-19 ha puesto a prueba los sistemas de salud en todo el mundo. En situaciones como la actual, donde el testeo activo forma parte de las estrategias de respuesta y la disponibilidad de tests de detección es limitada, el método de testeo grupal surge como una alternativa para aumentar la capacidad de detección. La ganancia en eficiencia obtenida mediante el testeo grupal, derivada de testear múltiples muestras combinadas simultáneamente, depende del estado de la epidemia en la población objetivo. La progresión de la infección en poblaciones cerradas, tales como Establecimientos de Larga Estadía para Adultos Mayores (ELEAM), anticipa una correlación significativa en el contagio. Frente a ello, se desarrolló un modelo probabilístico que incorpora dicha correlación en el cálculo del tamaño de grupo óptimo para esquemas de testeo grupal de dos etapas. A su vez, en base a una experiencia previa implementando pilotos de testeo grupal en el SENAMA (Servicio Nacional del Adulto Mayor), se desarrolló un modelo de simulación de la dinámica de contagios en un ELEAM y se estimó la correlación del contagio entre residentes y funcionarios. Los resultados muestran que es posible aplicar el testeo grupal de forma efectiva en el contexto local, considerando distintos escenarios prevalencia y correlación, y manteniendo bajo control el riesgo de falsos negativos, y que la correlación en contagio puede ser alta, dependiendo de la política de cuarentena implementada por el ELEAM, validando la utilización de modelos que incorporan correlación en el cálculo de tamaños de grupo.

Palabras Clave: Testeo grupal, COVID-19, Simulación, Beta-Binomial

*Instituto Sistemas Complejos de Ingeniería (ISCI), Santiago, Chile.

†Departamento de Ingeniería Civil - Transporte, FCFM, Universidad de Chile, Chile.

‡Departamento de Física, FCFM, Universidad de Chile, Chile.

§Departamento de Ingeniería Industrial, FCFM, Universidad de Chile, Chile.

¶ESE Business School, Universidad de los Andes, Chile.

1. Motivación y antecedentes

El virus causante de la pandemia de COVID-19, SARS-CoV-2, ha puesto bajo estrés a prácticamente la totalidad de los sistemas de salud del mundo. Los recursos de todo tipo han mostrado ser escasos y la detección y trazabilidad de los posibles contagiados se ha presentado como una estrategia efectiva para la contención de la pandemia.

En este contexto, las políticas de contención propuestas a la fecha consideran típicamente el testeo de todos los contactos estrechos de pacientes confirmados o sospechosos de COVID-19 [2]. En Chile, los contactos estrechos se definen como todas las personas que han estado en contacto con un caso confirmado con COVID-19, entre dos días antes del inicio de síntomas y 14 días después del inicio de los síntomas del enfermo, o 14 días después de confirmar el diagnóstico¹ en caso de ser un paciente asintomático [4]. El contacto estrecho supone haber mantenido: (i) más de 15 minutos de contacto cara a cara, a menos de un metro, sin mascarilla; (ii) haber compartido un espacio cerrado por dos horas o más, en lugares como oficinas, trabajos, reuniones, colegios, entre otros, sin mascarilla; (iii) haberse trasladado en cualquier medio de transporte cerrado a una proximidad menor de un metro con otro ocupante del medio de transporte que esté contagiado, sin mascarilla; o (iv) vivir o pernoctar en el mismo hogar o lugares similares a un hogar, tales como, hostales, internados, instituciones cerradas, ELEAM, hoteles, residencias, entre otros.

Bajo esta definición de contacto estrecho, es necesario testear poblaciones completas viviendo juntas, poniendo bajo presión la capacidad de realizar las pruebas diagnósticas de manera oportuna. En este contexto, es razonable esperar un alto grado de **correlación** en el contagio de ciertos grupos de pacientes, como por ejemplo aquellos habitando Establecimientos de Larga Estadía para Adultos Mayores (ELEAM) gestionados por el Servicio Nacional del Adulto Mayor (SENAMA).

En situaciones donde la disponibilidad de tests de detección para el cribado de pacientes infectados de una población es limitado, una posible estrategia que permite ahorrar, potencialmente, un número significativo de tests es el **testeo grupal**. En esta estrategia, las muestras individuales se combinan y se analizan de manera agrupada en búsqueda de un patógeno: si el resultado del test es negativo, se considera que todos los pacientes del grupo son negativos;

¹La confirmación de los casos de COVID-19 está basado en la detección de secuencias únicas del ARN del virus mediante pruebas de amplificación del ácido nucleico (PCR).

por otro lado, si el resultado del test es positivo, se concluye que al menos uno de los pacientes es positivo. Cuando el objetivo del test es diagnosticar a cada paciente, un resultado positivo requiere pruebas adicionales para determinar a los pacientes infectados.

A modo de ejemplo, considere un grupo de 10 pacientes cuyas muestras se analizan en un único grupo: un resultado negativo implica un ahorro de un 90 % de tests (relativo al testeo individual), incluyendo ahorros en reactivos asociados y tiempos de proceso. En la medida que un resultado negativo es dominante (lo cual depende de los supuestos subyacentes, ver Sección 1.2), esta estrategia aparece como un método efectivo para ampliar la capacidad de testeo de un sistema de salud, lo que posibilitaría la implementación de estrategias de testeo masivas, como las que se han llevado a cabo en otros países [16], o las que están siendo impulsadas por el gobierno [10].

El problema de minimizar el número de tests requeridos para diagnosticar un conjunto de pacientes mediante la estrategia de testeo grupal ha sido estudiado extensamente en los últimos 80 años (ver la Sección 1.2). Sin embargo, la estrategia de testeo óptima es muy difícil de encontrar, su cálculo es sumamente caro, en término de recursos computacionales y sumamente compleja de implementar, debido a su carácter adaptativo.

En el caso de los pacientes con COVID-19 se presentan una serie de dificultades para la implementación del testeo grupal. Por una parte, los falsos negativos han sido una preocupación en la lucha contra la pandemia [15] y en la práctica, las estrategias basadas en testeo grupal pueden ser más arriesgadas en relación a los potenciales falsos negativos. Por otra parte, la existencia de correlación entre pacientes ha sido evidente (por ejemplo, los ELEAM han sido foco de gran cantidad de contagios y muertes por COVID-19 en particular en España e Italia), pero no ha sido incorporada en los modelos. Finalmente, la toma de muestra nasofaríngea necesaria para el PCR es un procedimiento molesto y desagradable, que requiere de un personal sanitario cualificado, por lo que repetir la toma de muestras debería evitarse, limitando el diseño de la estrategia de testeo grupal a la clásica de dos etapas.

1.1. Objetivo

Este trabajo busca presentar una estrategia de testeo grupal que se haga cargo de la realidad de los pacientes COVID-19 en entornos donde se realiza testeo activo por contagios². Para esto, En primer lugar, se plantea un modelo probabilístico de contagio en una población cerrada, que incluye como

²Para realizar el análisis, no se necesita la revisión de una Junta de Repaso Institucional (IRB, por sus siglas en inglés) al tratarse de un trabajo de modelamiento y optimización teórico y predictivo, donde no se han usado datos individuales de pacientes.

input los parámetros de operación de la técnica de testeo, así como también la correlación en la infección en la población objetivo, extendiendo así modelos tradicionales para el cálculo de tamaños de grupo óptimo que no incluyen correlación. El modelo toma como base el modelo tradicional utilizado en la literatura de testeo grupal, el que asume independencia en el contagio entre los individuos en la población y lleva a una distribución binomial del número total de contagios, e introduce correlación en el contagio incorporando aleatoriedad en la probabilidad de contagio de los individuos. Dicho modelo permite estudiar, entre otras cosas, el efecto de la correlación en los tamaños de grupo óptimos y en el riesgo de falsos negativos.

En segundo lugar, se desarrolla un modelo de simulación de la dinámica de contagios en una población cerrada. Dicho modelo es capaz de incorporar distintas dinámicas de interacción entre subgrupos de la población, por ejemplo, funcionarios y residentes de un ELEM, así como replicar distintas políticas de cuidado, cuarentena y testeo dentro de dicha población. Utilizando este modelo es posible, entre otras cosas, estimar el grado de correlación en el contagio en la población, posibilitando la validación de la intuición al respecto, y estudiar el efecto de las distintas estrategias de contención del virus en dicha correlación, y la progresión del contagio en general.

1.2. Contribución

Como resultado del trabajo realizado, se espera extender los cálculos de tamaños de grupo óptimos a escenarios con correlación en el contagio en la población objetivo, contribuyendo a la literatura existente en el tema. Así mismo, el análisis incluye aristas de la estrategia no cubierta por la literatura en testeo grupal. Este es el caso del riesgo de falsos negativos, cuyo análisis es de interés independiente de la extensión a escenarios con correlación positiva. En relación a este punto, una segunda contribución consiste en mostrar la existencia de un trade-off entre el número esperados de tests a utilizar y la varianza del número de falsos negativos, que se acentúa en escenarios con correlación positiva. Al respecto, el análisis indica que focalizar la elección de tamaño de grupo solamente en minimizar el número esperado de tests a utilizar puede resultar en decisiones arriesgadas, que apuestan fuertemente en la generación de falsos negativos para efectuar dicha minimización. Este es sin duda un factor a considerar al momento de decidir una estrategia de testeo, y sin duda el presente análisis es novedoso y de interés para los tomadores de decisiones.

Por último, el trabajo contribuye en mostrar que es posible enfrentar escenarios con alta correlación en contagio cuando la población objetivo es relati-

vamente cerrada, como lo es el caso de los ELEAM operados por SENAMA. En particular, se muestra cómo la prevalencia y correlación observada al momento de realizar campañas de testeo dependen de la política general de cuidado, cuarentena y testeo en poblaciones cerradas. De forma prescriptiva, el análisis puede ser utilizado para recomendar técnicas y tamaños de grupo para realizar dichas campañas, de forma de minimizar la utilización de recursos escasos.

El trabajo está organizado de la siguiente forma. En la Sección 2 se revisa la literatura relevante. Luego, en la Sección 3, se presenta el modelo probabilístico utilizado para calcular medidas de desempeño asociadas a un tamaño de grupo bajo la estrategia de testeo grupal en dos etapas, esto es, el número esperado de tests, y el número esperado y varianza de falsos negativos. A continuación, en la Sección 4, se analizan los resultados de utilizar la estrategia propuesta, calculando los potenciales ahorros en el número de tests y el riesgo en el número de falsos negativos. En la Sección 5 se incluye una aplicación de la estrategia propuesta aplicada en el contexto de los ELEAM gestionados por el SENAMA. Finalmente, en la Sección 6, se presentan las reflexiones finales y las conclusiones de este trabajo.

2. Revisión bibliográfica

Las estrategias de testeo grupal no son nuevas. En [5] se presenta la estrategia original de dos etapas para pacientes infectados de sífilis en el contexto del ejército de los Estados Unidos. El modelo considera sensibilidad perfecta del test, por lo que los falsos negativos no son un problema, y se muestra que el número de tests necesarias para testear una población usando la estrategia de muestreo grupal es mucho menor con respecto a la estrategia tradicional de testeo individual, especialmente en situaciones de baja prevalencia.

A partir de este estudio seminal, numerosos trabajos han abordado el problema de testeo grupal, realizando modificaciones relacionadas con qué hacer en caso de que el test agrupado resulte positivo. Por ejemplo, [7] y [13] proponen continuar testeando de manera agrupada, ya sea en un único test adicional o recursivamente en dos subgrupos, toda vez que el resultado de una muestra agrupada sea positivo.

Las políticas de testeo grupal pueden ser clasificadas en adaptativas y no adaptativas. El primer caso se caracteriza por permitir más de una secuencia de testeo, donde los agrupamientos sucesivos son dependientes de los resultados de los tests previamente realizados. En este sentido, el modelo clásico de testeo grupal propuesto en [5] corresponde a una política de testeo adaptativa con

dos etapas. Otros trabajos como [14] han aplicado programación dinámica para optimizar la selección de subgrupos con el objetivo de minimizar los costos de falsos-negativos, falsos-positivos, y número de tests.

En el segundo caso, el de políticas no adaptativas, los grupos a testear son definidos previo a conocerse cualquier resultado intermedio de testeo. En [1] se presenta una revisión bibliográfica sobre políticas no adaptativas.

Los problemas de testeo grupal también pueden clasificarse según los supuestos sobre la población contagiada, la que puede ser tratada de manera probabilística o combinatorial. En el caso probabilístico, se define una distribución de probabilidad para modelar la condición de salud de los pacientes (infectados o no), la que por lo general considera que las probabilidades de infección son independientes. En el caso combinatorial, se considera que existe un conjunto (no conocido) de individuos que está infectado y es ese conjunto el que se desea identificar.

El testeo grupal adaptativo ha sido estudiado para pacientes COVID-19. Casos con múltiples etapas de testeo grupal han sido analizadas teóricamente por [9]. La estrategia clásica de testeo grupal en dos etapas para casos de COVID-19 ha sido validada internacionalmente en estudios realizados en Israel [17]. En Chile, se ha validado el PCR para muestras agrupadas, donde las muestras individuales se dividen en dos: una para el análisis agrupado y la otra para realizar un PCR individual en caso que el resultado del test agrupado sea positivo [6]. En [3] se utiliza simulación para evaluar el impacto de las políticas de testeo agrupado versus el testeo individual. En este trabajo no se incluye la posible correlación de la población ni tampoco se realiza una estimación del riesgo de falsos negativos de manera explícita.

3. Modelo de prevalencia y correlación en infección

Considerando el problema de testear a una población de N pacientes mediante la técnica de testeo grupal, y siguiendo la literatura, se supondrá que cada paciente en la población presenta el patógeno (a niveles detectables por la prueba PCR) con probabilidad $p \in (0, 1)$. Esto es, p representa la prevalencia de SARS-CoV-2 en la población. Sin embargo, se tiene como supuesto que existe correlación en la presencia del patógeno en dos individuos cualesquiera en la población. Se denota $\rho \in [0, 1)$ a esta correlación, y se restringe su rango a valores no-negativos, bajo la lógica que, en las aplicaciones de interés en donde hay algún grado de contacto entre los integrantes de una población, la

presencia del patógeno en un individuo aumenta la posibilidad de su presencia en el resto.

Formalmente, se introduce la correlación ρ mediante el siguiente modelo. Para $i \leq N$, se define

$$X_i := \begin{cases} 1 & \text{si paciente } i \text{ presenta el patógeno,} \\ 0 & \sim, \end{cases} \quad i \leq N,$$

y se asume que, dado un valor $q \in (0, 1)$, X_i distribuye *Bernoulli*(q) (es decir, $\mathbb{P}\{X_i = 1\} = q$) para $i \leq N$ y la secuencia $\{X_i, i \leq N\}$ es independiente e idénticamente distribuida (notar que esta independencia está condicionada en el valor de q).

Cuando $\rho = 0$, se tiene que $q = p$, y el número de pacientes con el patógeno en un grupo de tamaño n distribuye *Binomial*(n, p). El análisis de este caso es el presentado en [5]. Por otro lado, cuando $\rho > 0$, se asume que q es una variable aleatoria distribuida *Beta*(α, β), donde

$$\alpha = p(\rho^{-1} - 1), \quad \beta = (1 - p)(\rho^{-1} - 1), \tag{1}$$

(En rigor, tanto α como β son funciones de la prevalencia p y la correlación ρ , sin embargo se omitirá esta dependencia, para mantener la claridad de la exposición.) Esto tiene dos importantes consecuencias: i) la aleatoriedad en q introduce correlación en la presencia del patógeno en la población; y ii) la distribución del número de pacientes con el patógeno en un conjunto de tamaño k sigue una distribución *BetaBinomial*(k, α, β). El siguiente lema formaliza estas propiedades.

Lema 1. *Considerando un conjunto de pacientes $S \subseteq \{1, \dots, N\}$, y $X(S) := \sum_{i \in S} X_i$, la cantidad de pacientes en S con el patógeno. Se tiene que $X(S) \sim \text{BetaBinomial}(|S|, \alpha, \beta)$, esto es*

$$\mathbb{P}\{X(S) = k\} = \binom{|S|}{k} \frac{B(k + \alpha, |S| - k + \beta)}{B(\alpha, \beta)}, \quad k \leq |S|,$$

donde $|S|$ denota la cardinalidad del conjunto S y $B(\cdot, \cdot)$ es la función *Beta*. Adicionalmente, para $i \neq j$ se tiene que

$$\text{Corr}(X_i, X_j) := \rho, \quad \mathbb{E}\{X_i\} = p, \quad \text{y} \quad \text{Var}(X_i) = p(1 - p).$$

Prueba. Antes de comenzar, notar que si se toma el caso $\rho > 0$ y, utilizando la Ecuación (1), se puede mostrar que

$$p = \frac{\alpha}{(\alpha + \beta)}, \quad \rho = \frac{1}{(\alpha + \beta + 1)}.$$

Estas relaciones se utilizarán en reiteradas ocasiones en el resto del documento.

La primera parte del Lema es consecuencia directa de la definición de una distribución BetaBinomial. Se tiene que

$$\begin{aligned}
 \mathbb{P}\{X(S) = k\} &= \int_0^1 \mathbb{P}\{X(S) = k|q = x\} \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)} dx \\
 &= \int_0^1 \binom{|S|}{k} x^k (1-x)^{|S|-k} \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)} dx \\
 &= \binom{|S|}{k} \frac{B(\alpha+k, \beta+|S|-k)}{B(\alpha, \beta)} \int_0^1 \frac{x^{\alpha+k-1}(1-x)^{\beta+|S|-k-1}}{B(\alpha+k, \beta+|S|-k)} dx \\
 &= \binom{|S|}{k} \frac{B(\alpha+k, \beta+|S|-k)}{B(\alpha, \beta)}.
 \end{aligned}$$

En este desarrollo, primero se condiciona sobre el valor de q (utilizando la densidad de una variable aleatoria $Beta(\alpha, \beta)$), y después se usa el hecho que, condicional en el valor de q , $X(S)$ distribuye binomial. Notar que, la última igualdad de la ecuación anterior, sigue de reconocer la integral de la densidad de una variable aleatoria $Beta(\alpha+k, \beta+|S|-k)$ sobre su dominio.

Respecto a la segunda parte del Lema 1, se tiene que

$$\begin{aligned}
 \mathbb{E}\{X_i\} &= \int_0^1 \mathbb{E}\{X_i|q = x\} \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)} dx \\
 &= \int_0^1 x \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)} dx = \frac{\alpha}{\alpha+\beta} = p,
 \end{aligned}$$

donde en la última igualdad se encuentra la esperanza de una variable aleatoria de distribución $Beta(\alpha, \beta)$ y se toma la definición de α y β en términos de p y ρ . Por otro lado, se tiene que

$$\begin{aligned}
 \mathbb{E}\{X_j X_i\} &= \int_0^1 x^2 x^{\alpha-1}(1-x)^{\beta-1} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} dx \\
 &= \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)} \int_0^1 x^{\alpha+1}(1-x)^{\beta-1} \frac{\Gamma(\alpha+\beta+2)}{\Gamma(\alpha+2)\Gamma(\beta)} dx \\
 &= \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)},
 \end{aligned}$$

donde se establece $Beta(\cdot, \cdot)$ en términos de la función $\Gamma(\cdot)$ (gamma). Con esto, se tiene que

$$\begin{aligned}
 \text{Cov}\{X_i, X_j\} &= \mathbb{E}\{X_i X_j\} - \mathbb{E}\{X_i\} \mathbb{E}\{X_j\} \\
 &= \frac{\alpha}{\alpha+\beta} \left(\frac{\alpha+1}{\alpha+\beta+1} - \frac{\alpha}{\alpha+\beta} \right) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} = p(1-p)\rho.
 \end{aligned}$$

Debido al carácter binario de X_i , se tiene que $\mathbb{E}\{X_i^2\} = \mathbb{E}\{X_i\} = \alpha/(\alpha + \beta)$. Esto implica que

$$\text{Var}(X_i) = \frac{\alpha}{\alpha + \beta} \left(1 - \frac{\alpha}{\alpha + \beta}\right) = \frac{\alpha\beta}{(\alpha + \beta)^2} = p(1 - p).$$

Finalmente, se llega a que, para $i \neq j$,

$$\text{Corr}(X_i, X_j) = \frac{\text{Cov}(X_i, X_j)}{\sqrt{\text{Var}(X_i) \text{Var}(X_j)}} = \frac{\text{Cov}(X_i, X_j)}{\text{Var}(X_j)} = \rho.$$

Lo que concluye la prueba del Lema 1. □

Considerando el caso $\rho > 0$ y n el tamaño de grupo utilizado en la estrategia de testeo grupal de dos etapas. Para calcular el número esperados de tests a utilizar bajo esta estrategia, se considera la *especificidad* y la *sensibilidad* de la técnica de testeo. Esto es, se denota por $S_e \in [0, 1]$ a la probabilidad que un test entregue un resultado positivo cuando la muestra efectivamente contiene el patógeno (la sensibilidad del test), y por $S_p \in [0, 1]$ a la probabilidad que un test entregue un resultado negativo cuando la muestra efectivamente no contiene el patógeno (la especificidad del test). Al respecto, se toma como supuesto que, estos parámetros de operación, no son afectados por el tamaño del agrupamiento, y que cada test falla en el diagnóstico de forma independiente, incluso si se utiliza la misma muestra (en un testeo sucesivo).

Prueba. *Estos supuestos son razonables en el marco de la especificidad del test PCR, puesto que los falsos positivos ocurren mayoritariamente debido a problemas en el manejo de las muestras. Los supuestos son más fuertes en el caso de la sensibilidad, puesto que los falsos negativos ocurren cuando, por ejemplo, una de las muestras incluidas en el agrupamiento se encuentra muy cerca pero bajo del umbral de detección³ (por lo que la muestra, testeada de forma individual, entrega un resultado positivo); en este caso es posible que se produzca dilución en la muestra, lo que ubique a la muestra combinada ligeramente por sobre el umbral de detección, con lo que la muestra es incorrectamente etiquetada como libre del patógeno. En la práctica, sin embargo, es muy difícil encontrar muestras cercanas al umbral de detección [6].* □

Sea T el número de tests utilizados para diagnosticar a toda la población. El Lema 2 entrega un expresión para la esperanza de T .

³La técnica de testeo PCR mide cuantos ciclos (de calentamiento/enfriamiento de la muestra) se requieren para amplificar la presencia del patógeno de forma de hacerla detectable. Por esto, si dicho tiempo (en ciclos) es menor que cierto umbral, se concluye que el resultado es positivo.

Lema 2. *Suponiendo que n es múltiplo de N , entonces,*

$$\mathbb{E}\{T\} = N \left(\frac{1}{n} + S_e + (1 - S_e - S_p) \frac{B(\alpha, n + \beta)}{B(\alpha, \beta)} \right).$$

Prueba. Sea S_k el conjunto de pacientes incluidos en el grupo k a testear, y se denota por T_k al número de tests necesarios para diagnosticar a los pacientes en el grupo k . (Notar que $\{S_k, k = 1 \dots N/n\}$ forma una partición de la población). Se tiene que

$$\begin{aligned} \mathbb{E}\{T\} &= \sum_{k=1}^{N/n} \mathbb{E}\{T_k\} \\ &= \sum_{k=1}^{N/n} ((1 + n(1 - S_p)) \mathbb{P}\{X(S_k) = 0\} + (1 + nS_e) (1 - \mathbb{P}\{X(S_k) = 0\})) \\ &= \sum_{k=1}^{N/n} 1 + nS_e + n(1 - S_e - S_p) \mathbb{P}\{X(S_k) = 0\} \\ &= N \left(\frac{1}{n} + S_e + (1 - S_p - S_e) \frac{B(\alpha, n + \beta)}{B(\alpha, \beta)} \right). \end{aligned}$$

Donde la primera igualdad viene de la linealidad de la esperanza, la segunda de condicionar sobre el número de pacientes con el patógeno en el grupo k y la última del Lema 1, y hecho que el número de infecciones en un grupo se distribuye de igual forma en cada grupo. \square

Cuando no existe correlación, se obtiene el resultado presentado en [3], que extiende el desarrollo en [5] considerando la especificidad y sensibilidad del test.

Notar que la expresión arriba es muy fácil de evaluar (la función *Beta* viene incorporada en la mayoría de los software estadísticos). Considerando que existen limites naturales al tamaño de grupo a utilizar en la práctica (por ejemplo, a la fecha, el método de testeo grupal ha sido válido en laboratorio para grupos de hasta 32 muestras sin modificar el protocolo del test: adicionalmente, en ausencia de tecnologías de procesamiento automatizado, el personal de laboratorio puede manejar tamaños de grupo relativamente bajos), esta expresión puede ser usada directamente para el cálculo del tamaño óptimo, vía enumeración.

Podemos obtener del Lema 2 que, para un tamaño de grupo dado, los parámetros de operación de la técnica de testeo afectan directamente el número esperado de tests requeridos. En particular, mientras mayor sea la especificidad del test, menor es el número esperado de tests requeridos (en la práctica, la especificidad del test PCR es cercana al 100%). Por otro lado, el efecto de la

sensibilidad es el inverso, y a mayor sensibilidad, mayor es el número esperado de tests a utilizar (debido al aumento del riesgo de obtener falsos negativos). Ciertamente, el riesgo de falsos negativos es algo que debiese considerarse al momento de determinar tamaños de grupo. A continuación se realiza un análisis de este riesgo. Se denota por F el número de falsos negativos asociados al diagnóstico de la población. El Lema 3 caracteriza el valor esperado y la varianza de F , en función de los parámetros de operación, y el tamaño de grupo.

Lema 3. *Suponiendo que n es múltiplo de N , entonces se tiene que $\mathbb{E}\{F\} = N(1 - S_e^2)$ y*

$$\text{Var}(F) = N(1 - S_e^2)p - N(1 - S_e)(1 + S_e - S_e^2 - nS_e^3)(p^2 + p(1 - p)\rho) + N^2(1 - S_e^2)^2p(1 - p)\rho.$$

Prueba. Siguiendo la prueba del Lema 2, sea F_k el número de falsos negativos obtenidos al testear al grupo k . Se tiene que

$$\begin{aligned} \mathbb{E}\{F\} &= \sum_{k=1}^{N/n} \mathbb{E}\{F_k\} \\ &= \sum_{k=1}^{N/n} \sum_{i=1}^n \mathbb{E}\{F_k | X(S_k) = i\} \mathbb{P}\{X(S_k) = i\} \\ &\stackrel{(a)}{=} \sum_{k=1}^{N/n} \sum_{i=1}^n (i(1 - S_e) + i(1 - S_e)S_e) \mathbb{P}\{X(S_k) = i\} \\ &= \sum_{k=1}^{N/n} (1 - S_e^2) \mathbb{E}\{X(S_k)\} = \frac{N}{n} (1 - S_e^2) n \frac{\alpha}{\alpha + \beta} = N(1 - S_e^2)p. \end{aligned}$$

Se observa que en (a) se utiliza el hecho que, cuando $X(S_k) = i$, si el test grupal resulta en un falso negativo (lo que ocurre con probabilidad $(1 - S_e)$), esto resulta en i falsos negativos, y cuando el test grupal entrega el resultado correcto (lo que ocurre con probabilidad S_e), esto resulta, en promedio, en $(1 - S_e)i$ falsos negativos, provenientes de los tests individuales. La última igualdad utiliza el hecho que la esperanza de una variable aleatoria distribuida *BetaBinomial*(k, α, β) es $k(\alpha/(\alpha + \beta))$.

Considerar ahora el cálculo de la varianza de F . Primero, se debe notar que condicional en q , los falsos negativos en cada grupo son variables aleatorias independientes, por lo que se tiene que

$$\mathbb{E}\{F^2 | q\} = \sum_{k=1}^{N/n} E\{F_k^2 | q\} + \frac{N}{n} \left(\frac{N}{n} - 1 \right) (nq(1 - S_e^2))^2.$$

Para desarrollar el término asociado a cada grupo, recordar que si $X \sim \text{Binomial}(n, q)$, entonces

$$E \{X^2\} = \text{Var}(X) + \mathbb{E} \{X\}^2 = nq(1 - q) + (nq)^2.$$

Luego, tomando que, condicional en $X(S_k) = i$ y el test grupal no falló, el número de falsos negativos obtenidos en el grupo k distribuye $\text{Binomial}(i, (1 - S_e))$. Dejando a G denotar al evento en el que test grupal no falla, se tiene que

$$\begin{aligned} E \{F_k^2|q\} &= \sum_{i=1}^n ((1 - S_e)\mathbb{E} \{F_k^2|X(S_k) = i, \bar{G}\} + S_e\mathbb{E} \{F_k^2|X(S_k) = i, G\}) \\ &\quad \mathbb{P} \{X(S_k) = i|q\} \\ &= \sum_{i=1}^n ((1 - S_e)i^2 + S_e(i S_e(1 - S_e) + i^2(1 - S_e^2))) \mathbb{P} \{X(S_k) = i|q\} \\ &= S_e^2(1 - S_e)nq + (1 - S_e)(1 + S_e(1 - S_e))(q(1 - q)n + n^2q^2) \\ &= (S_e^2(1 - S_e) + (1 - S_e)(1 + S_e(1 - S_e)))nq \\ &\quad - ((1 - S_e)(1 + S_e(1 - S_e)))q^2n(1 - n) \\ &= (1 - S_e^2)nq - ((1 - S_e)(1 + (1 - S_e)S_e))q^2n(1 - n). \end{aligned}$$

Con esto, tenemos que

$$\begin{aligned} \mathbb{E} \{F^2|q\} &= N(1 - S_e^2)q - ((1 - S_e)(1 + (1 - S_e)S_e))q^2N(1 - n) \\ &\quad + N^2q^2((1 - S_e^2))^2 - Nnq^2((1 - S_e^2))^2. \end{aligned}$$

Ahora, notamos que

$$\begin{aligned} \mathbb{E} \{q^2\} &= \text{Var}(q) + \mathbb{E} \{q\}^2 = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)} \\ &\quad + \frac{\alpha^2}{(\alpha + \beta)^2} = p(1 - p)\rho + p^2. \end{aligned}$$

Finalmente, tomando la esperanza (respecto a q) sobre $\mathbb{E} \{F^2|q\}$, y sustrayendo $\mathbb{E} \{F\}^2$, se tiene

$$\begin{aligned} \text{Var} \{F\} &= N(1 - S_e^2)p - \\ &\quad \left(((1 - S_e)(1 + (1 - S_e)S_e)(1 - n) + N^2((1 - S_e^2))^2 - Nn((1 - S_e^2))^2) \right. \\ &\quad \left. (p(1 - p)\rho + p^2) - N^2(1 - S_e^2)^2p^2 \right) \\ &= N(1 - S_e^2)p - N(1 - S_e)(1 + S_e - S_e^2 - nS_e^3)(p^2 + p(1 - p)\rho) \\ &\quad + N^2(1 - S_e^2)^2p(1 - p)\rho, \end{aligned}$$

donde la última igualdad viene de agrupar términos de acuerdo a sus dependencias, después de algo de álgebra. \square

Notar que el número esperado de falsos negativos es independiente del tamaño de grupo utilizado. Erróneamente se podría concluir que el tamaño de grupo no afecta a F . La segunda parte del Lema 3, sin embargo, nos dice que a medida que el tamaño de grupo aumenta, también lo hace la varianza de F . La consecuencia para la selección del tamaño del agrupamiento es directa: a mayor tamaño mayor es el riesgo asociado de tener falsos negativos (tomar como referencia el caso extremo donde $n = N$). Estos aspectos se revisarán numéricamente en la Sección 4.

4. Resultados

En la Tabla 1 se presentan los resultados del modelo obtenidos con una población de 100 pacientes ($N = 100$) para prevalencias que varían desde 0,01 % a 40 %, considerando 4 niveles de correlación. Se incluye el agrupamiento óptimo, el número esperado de tests y los ahorros en el número de tests comparado con la estrategia de testeo individual y la estrategia de testeo grupal, pero utilizando el agrupamiento que se obtiene si se ignora la correlación. Adicionalmente, en el caso que el test no sea perfecto y pueda arrojar falsos negativos, se incluye el valor esperado de los falsos negativos y su desviación estándar. Para evaluar el impacto del testeo grupal en el riesgo de los falsos negativos se consideró $S_e = 0,7$. En la práctica, la sensibilidad del test PCR varía dependiendo del método de recolección de la muestra, e.g. raspado nasofaríngeo, saliva, etc. En este modelo se consideró el escenario $S_e = 0,7$ como un extremo y $S_e = 1,0$ el otro.

Los ahorros en el número de tests pueden ser tan grandes como del 99 % en comparación a realizar pruebas individuales y de hasta el 60 % en comparación con realizar testeo grupal pero utilizando el agrupamiento calculado a partir de un modelo que no considere la correlación.

Adicionalmente, se observa que para prevalencias bajas los grupos de testeo son grandes, iguales al tamaño de la población. El tamaño del agrupamiento óptimo disminuye al aumentar la prevalencia para $S_e = 1$. En el caso de tener tests imperfectos ($S_e < 1$) se produce una discontinuidad y vuelve a ser óptimo testear a toda la población en un único grupo al aumentar la prevalencia.

La Figura 1 muestra los ahorros en el número de tests al utilizar el agrupamiento óptimo del modelo (que incluye explícitamente la correlación) versus el caso donde la correlación no se considera, en función de la correlación para distintos niveles de prevalencia ($S_e = 0,7$, $N = 100$).

Si bien el número esperado de falsos negativos no depende del tamaño del grupo, su desviación estándar es creciente en n . En la Figura 2 se presenta el agrupamiento óptimo en función de la prevalencia, para distintos niveles de correlación. Se ve claramente que el testeo grupal en escenarios de alta prevalencia y de baja correlación se produce una apuesta por tener un resultado negativo en un único test agrupado (i.e., $n = 100$), el que con alta probabilidad será un falso negativo producto de la sensibilidad del test ($S_e = 0,7$, $N = 100$).

La Figura 3 muestra como la desviación estándar de falsos negativos disminuye al reducir el tamaño del grupo, a la vez que el número esperado de tests crece. Podemos ver que usando $n = 6$ en vez del agrupamiento óptimo ($n^* = 100$) aumentamos en 8,5% el número esperado de tests (de 71 a 77) pero disminuimos la desviación estándar de los falsos negativos en un 39% (de 15 a 9,2).

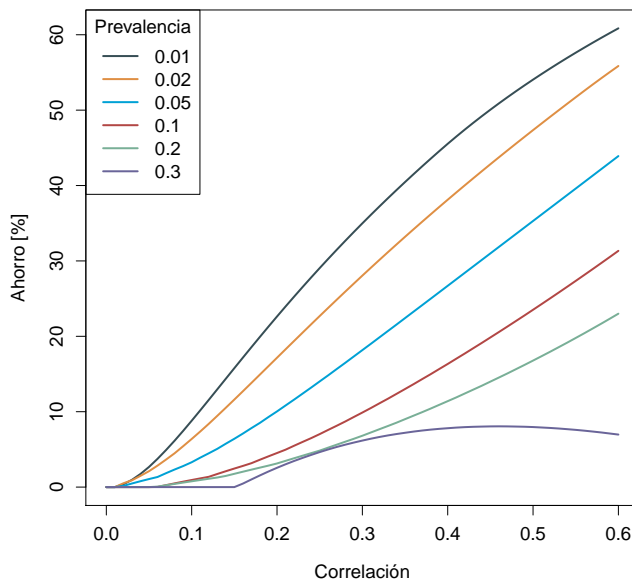


Figura 1: Ahorros en el número esperado de tests considerando la correlación de manera explícita vs testeo grupal sin incluir correlación ($S_e = 0,7$, $N = 100$)

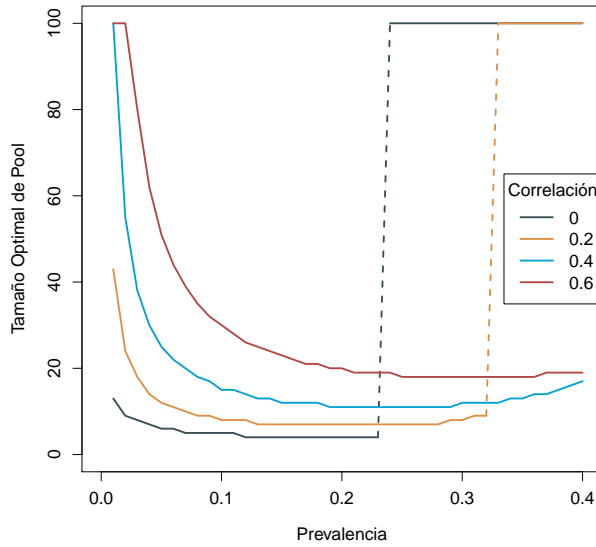


Figura 2: Tamaño del agrupamiento óptimo en función de la prevalencia ($S_e = 0,7, N = 100$)

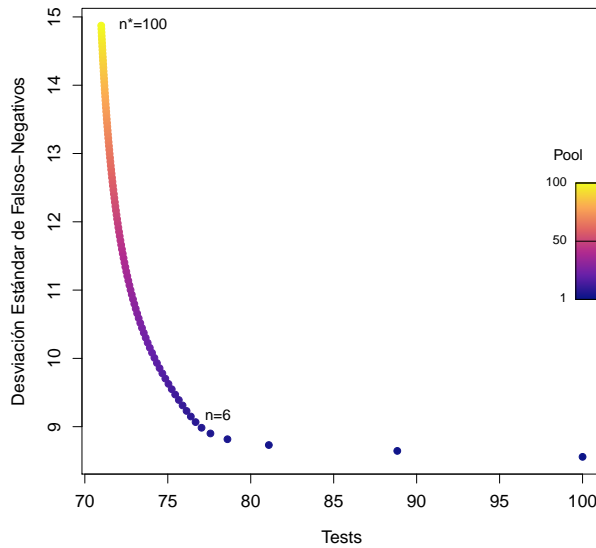


Figura 3: Desviación estándar y número esperado de tests para distintos tamaños de grupos ($S_e = 0,7, N = 100$)

Tabla 1: Agrupamiento óptimo, número esperado de tests y porcentaje de ahorro en relación a estrategia de testeo individual y de la estrategia de testeo grupal ignorando la correlación. A la izquierda $S_e = 1$ y a la derecha $S_e = 0,7$ donde se reporta además el número esperado de falsos negativos y su desviación estándar ($N = 100$).

Prevalencia	n*	Ahorros [%]			n*	Ahorros [%]			Falsos Negativos Exp. (Std.)
		Exp. tests	Testeo ind.	$\rho = 0$		Exp. tests	Testeo ind.	$\rho = 0$	
Correlación = 0									
0,0001	100	2	98	0	100	2	98	0	0 (0)
0,001	32	6	94	0	39	5	95	0	0 (0)
0,01	11	20	80	0	13	16	84	0	1 (1)
0,02	8	27	73	0	9	23	77	0	1 (1)
0,05	5	43	57	0	6	35	65	0	3 (2)
0,1	4	59	41	0	5	49	51	0	5 (2)
0,2	3	82	18	0	4	66	34	0	10 (3)
0,3	3	99	1	0	100	71	29	0	15 (10)
0,4	1	100	0	0	100	71	29	0	20 (13)
Correlación = 0.2									
0,0001	100	1	99	0	100	1	99	0	0 (0)
0,001	100	2	98	42	100	2	98	40	0 (1)
0,01	31	12	88	18	43	9	91	23	1 (3)
0,02	17	19	81	12	24	15	85	17	1 (3)
0,05	9	34	66	7	12	27	73	10	3 (5)
0,1	6	51	49	3	8	40	60	5	5 (7)
0,2	4	73	27	2	7	57	43	3	10 (10)
0,3	4	90	10	1	8	68	32	3	15 (11)
0,4	1	100	0	0	100	71	29	0	20 (19)
Correlación = 0.4									
0,0001	100	1	99	0	100	1	99	0	0 (0)
0,001	100	2	98	54	100	1	99	50	0 (1)
0,01	72	8	92	40	100	6	94	46	1 (4)
0,02	38	13	87	31	55	10	90	38	1 (5)
0,05	18	25	75	22	25	19	81	27	3 (8)
0,1	11	41	59	13	15	31	69	16	5 (10)
0,2	7	62	38	10	11	47	53	11	10 (14)
0,3	7	78	22	6	12	58	42	8	15 (16)
0,4	8	90	10	10	17	66	34	3	20 (18)
Correlación = 0.6									
0,0001	100	1	99	0	100	1	99	0	0 (0)
0,001	100	1	99	60	100	1	99	54	0 (1)
0,01	100	5	95	58	100	4	96	61	1 (5)
0,02	81	9	91	49	100	6	94	56	1 (7)
0,05	36	18	82	39	51	13	87	44	3 (10)
0,1	20	31	69	28	30	23	77	31	5 (13)
0,2	13	50	50	21	20	37	63	23	10 (17)
0,3	12	65	35	15	18	48	52	7	15 (19)
0,4	12	77	23	23	19	56	44	5	20 (21)

5. Caso de estudio: Correlación en un ELEAM

El modelo de prevalencia y correlación en los contagios presentado en la Sección 3 está motivado por la realidad de los ELEAM gestionados por el SENAMA. En estos centros, un grupo de adultos mayores vive bajo los cuidados de un equipo de profesionales de la salud.

Para utilizar la estrategia de testeo grupal analizado, es necesario conocer la prevalencia y la correlación de la población para poder escoger adecuadamente el tamaño del grupo. Con este propósito, se desarrolló una herramienta que permite simular la evolución de infectados y del número de pacientes sintomáticos entre la población que compone un ELEAM. A continuación se detalla dicha herramienta, y los resultados obtenidos.

5.1. Detalles de la rutina de simulación

En la simulación se consideraron dos grupos de individuos: residentes y funcionarios. En ella, se generó una matriz de interacciones entre la población, donde se establece la probabilidad que dos individuos (funcionarios o residentes) entren en contacto durante un turno: mientras mayor sea la probabilidad de las interacciones, se espera una difusión de los contagios más rápida. Con ello, se supuso que la probabilidad de interacción diaria entre dos miembros cualesquiera de la población es fija (se realizan simulaciones con diferentes valores para este valor).

En términos del modelo epidemiológico, el tiempo de incubación se modeló con una distribución Lognormal [8], mientras que la capacidad de contagio sigue la evolución de la infección del paciente de manera escalada [11]. A su vez, todos los pacientes contagiados tienen la capacidad de contagiar, aunque existe una proporción que nunca presenta síntomas. La simulación consideró que un 30% de los pacientes asintomáticos, independiente del grupo al que pertenezcan (en línea con la evidencia internacional [12]). Luego, los principales supuestos y parámetros utilizados fueron⁴:

Supuestos.

- Las personas se infectan sólo una vez. Al recuperarse, quedan inmunizados.

⁴El detalle de esta simulación se encuentra disponible en <https://covidanalytics.isci.cl/simulacion>.

- Las tareas de los funcionarios contagiados y detectados son cubiertas por el resto de los funcionarios.
- La infecciosidad, largo del período de incubación y presencia/ausencia de síntomas son independientes entre individuos.
- El tiempo de incubación sigue una distribución Lognormal [8] y el tiempo de recuperación distribuye uniforme entre 2 y 4 semanas.
- La probabilidad de testear positivo está basada en la curva de positividad presentada en [11], considerando la evolución de la enfermedad en un paciente.
- La curva de infecciosidad es proporcional a la curva de positividad.

Parámetros.

- Matriz de turnos: indica las horas de interacción dentro de la población cerrada. Por ejemplo, para residentes, es una matriz de unos, y para funcionarios contiene unos (ceros) por las noches y ceros (unos) para el resto del día, para el turno nocturno (diurno).
- Matriz de interacción: indica la probabilidad de interacción entre individuos de distintos subgrupos de la población.
- Grupos: grupos considerados, en este caso, residentes (28) y funcionarios (22).
- Pacientes asintomáticos: Cada paciente es asintomático de forma independiente, con probabilidad 30 % [12].
- Probabilidad de contagio externa funcionarios: corresponde a la probabilidad diaria de contagio externa igual a 0.1 %.
- Cuarentena preventiva de pacientes sintomáticos: corresponde a todos los pacientes dos días tras desarrollar síntomas comienzan una cuarentena preventiva durante 14 días.

El proceso de simulación, se realizó mediante una simulación de Montecarlo para estudiar la evolución diaria de la infección. A su vez, se registró el número de personas infectadas (independiente de la presencia de síntomas), y de su estado de cuarentena. Cada ejecución de la simulación se comienza con toda la población no infectada.

- En $t = 0$ se asignó al azar la condición de sintomático/asintomático a cada individuo, que se utiliza en caso el individuo se contagie en algún momento durante la simulación.

- Para cada día, se calculó la probabilidad que cada paciente se contagie durante el transcurso de la jornada.
- Los contagios se generaron de acuerdo a las probabilidades calculadas arriba. Para cada persona infectada, se generaron tiempos de incubación, infecciosidad y recuperación.
- Tras esto, la simulación se mueve al siguiente día.

5.2. Caso de estudio

Como caso de estudio, se consideró un ELEAM con 28 residentes y 22 funcionarios en dos turnos de 11 personas cada uno. La simulación se inicia con toda la población del centro libre del virus, donde los funcionarios son quienes potencialmente introducen la infección con una probabilidad que depende de la prevalencia global.

Los residentes sólo pueden contagiarse por interacciones con funcionarios o residentes del centro, mientras que los funcionarios pueden contagiarse en sus interacciones en el centro o de manera exógena fuera del trabajo. En este escenario, las interacciones entre dos residentes tienen un mayor riesgo en comparación con las existentes entre un funcionario y un residente o entre dos funcionarios con igual turno, pues los contactos entre residentes suelen ser más largos. En ese sentido, un individuo sano puede contagiarse si entra en contacto con un paciente con potencial infeccioso, donde la probabilidad de contagio dependerá de la capacidad de contagio y del riesgo de la interacción.

Se toma como supuesto que todo paciente sintomático al segundo día de presentar síntomas se aísla en cuarentena preventiva y no tiene la posibilidad de contagiar. Esto último, busca replicar la estrategia de testeo reactivo que se ha aplicado en general por SENAMA, donde se testea de manera selectiva a quienes presentan síntomas de la enfermedad.

La Figura 4 presenta la evolución del número de casos activos y de los casos sintomáticos para tres niveles de probabilidad de interacción, desde el primer día que un paciente presenta síntomas. Es posible observar que, en caso de un alto nivel de interacciones, si no se realiza ninguna intervención al cabo de 22 días prácticamente la totalidad de la población estaría contagiada.

La Figura 5 presenta la evolución de la prevalencia y de la correlación para la población total del centro ($N = 28 + 11 + 11 = 50$) en función de los días de simulación, mientras que la Figura 6 considera como primer día de referencia aquel cuando se presenta el primer caso sintomático (el que es sospechoso de COVID-19, pero no ha sido confirmado con una prueba PCR). Ambas gráficas fueron construidas para tres niveles en la probabilidad de interacción

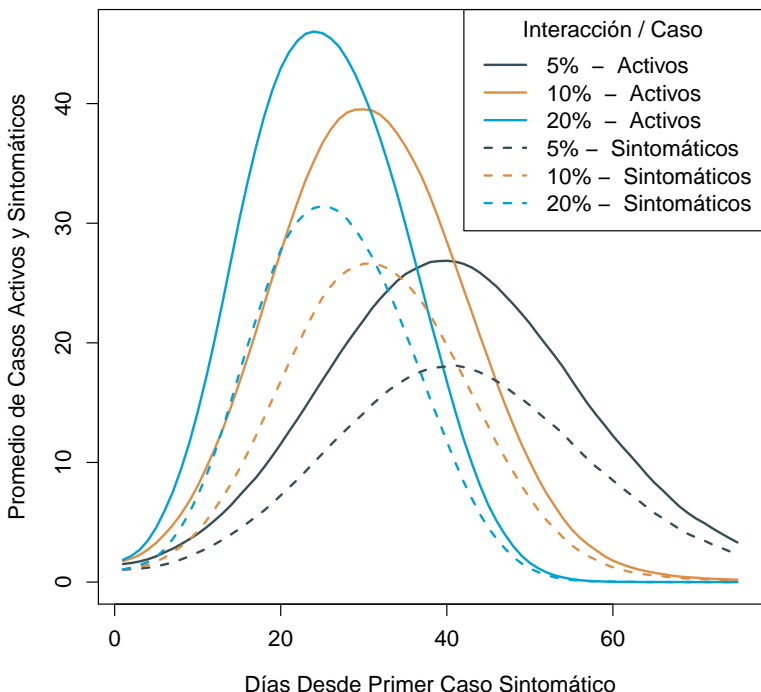


Figura 4: Número esperado de casos activos y de casos sintomáticos, para probabilidades de interacciones diarias del 5, 10 y 20% ($N = 50$).

entre individuos, 5%, 10% y 20%, entre cualquier par de miembros de la comunidad.

Es posible ver que, tanto la prevalencia como la correlación, aumentan a medida que avanzan los días desde el primer paciente con síntomas, hasta llegar a un peak y luego decaer. Cabe destacar además que, dicho peak, tanto para la prevalencia como para correlación, ocurren casi en el mismo día (para un mismo grado de interacción). Además, se aprecia que el valor máximo de la prevalencia está estrechamente relacionado de manera creciente con el grado de interacción (ver panel izquierdo de las Figuras 5 y 6). A su vez, se aprecia que el peak de la correlación (con respecto del día de inicio de la simulación), aumenta con el grado de interacción de los individuos, ver panel derecho de la Figura 5, pero no ocurre lo mismo al considerar el día con respecto al primer caso sintomático. Otro elemento a destacar, es la tasa a la que crece la prevalencia de la enfermedad y la correlación en ambas figuras, donde se observa que la prevalencia crece más rápido en el caso de considerar los días con res-

pecto al primer caso sintomático, mientras que, en el caso de la correlación, se tiene un crecimiento más rápido en el caso contrario. Es importante destacar que considerar el primer día de simulación o el primer día en que existe un caso sintomático sólo cambia el horizonte temporal de las Figuras 5 y 6, por lo que se esperaría similar comportamiento si se dejara evolucionar la simulación más allá del día 50 desde su inicio.

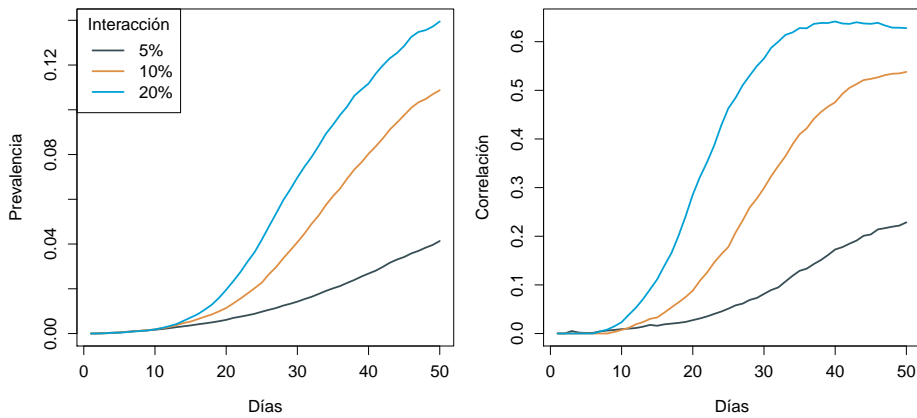


Figura 5: Prevalencia y correlación en la población, para probabilidades de interacciones diarias del 5, 10 y 20% considerando el primer día de simulación ($N = 50$).

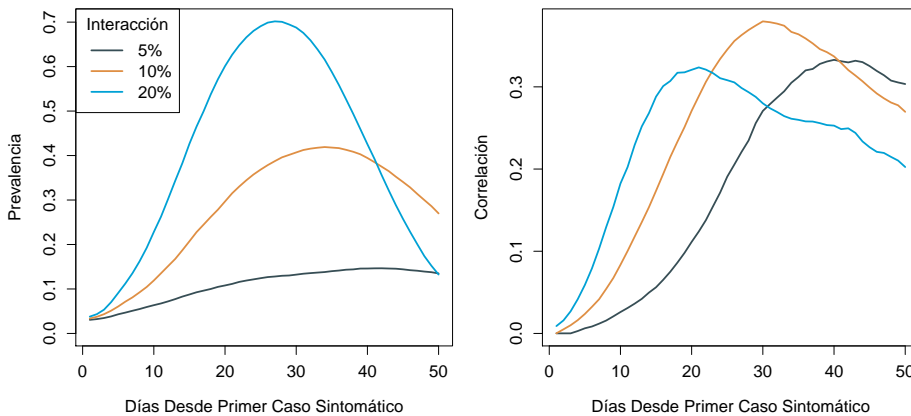


Figura 6: Prevalencia y correlación en la población, para probabilidades de interacciones diarias del 5, 10 y 20% considerando el primer día en que existen pacientes sintomáticos ($N = 50$).

Dada la evolución temporal de la prevalencia y de la correlación, la situa-

ción epidemiológica de la población bajo estudio puede ser muy diferente entre un día y otro. Estos resultados confirman la intuición detrás de la definición de contacto estrecho y las recomendaciones de testeo establecidas por el Ministerio de Salud [10], puesto que se observa una correlación significativa en contagios. Esto implica que, los agrupamientos recomendados, si se utiliza una estrategia de testeo grupal, serán diferentes dependiendo de la etapa en la que se encuentre la población.

Las Tablas 2 y 3 presentan la prevalencia, correlación y los agrupamientos recomendados por el modelo presentado en la Sección 3 y el modelo de agrupamiento que no incorpora explícitamente la correlación. En la última columna se incluye el ahorro en el número esperado de tests de seguir el agrupamiento óptimo que considera la correlación versus el caso de agrupamiento que omite dicha correlación (es decir, considera $\rho = 0$). Este análisis supone que $S_e = 1$, para prevenir el efecto indeseable de “apostar” a un resultado negativo de todo el grupo producto de la sensibilidad de la prueba, como se discutió en la Sección 4. En el caso de la Tabla 2, los días hacen referencia al inicio de la simulación, mientras que, en la Tabla 3, se considera el primer día como aquel en que se tiene un paciente sintomático en la población.

Tabla 2: Para cada día de simulación: prevalencia, correlación, número óptimo de grupo (n^*), número óptimo de grupo omitiendo correlación (n^\diamond), y ahorro en número de tests a usar. Probabilidades de interacciones diarias del 5, 10 y 20% de izquierda a derecha.

Día	Prev.	ρ	n^*	n^\diamond	Ahorro [%]	Prev.	ρ	n^*	n^\diamond	Ahorro [%]	Prev.	ρ	n^*	n^\diamond	Ahorro [%]
5	0,00	0,00	46	46	0,0	0,00	0,00	47	47	0,0	0,00	0,00	50	50	0
10	0,00	0,01	25	25	0,0	0,00	0,01	24	24	0,0	0,00	0,02	31	24	2,3
15	0,00	0,02	20	17	0,8	0,01	0,03	18	14	2,1	0,01	0,11	25	12	11,8
20	0,01	0,03	16	13	1,5	0,01	0,09	15	10	5,1	0,02	0,29	25	8	20,5
25	0,01	0,05	14	11	1,9	0,02	0,18	14	7	10,8	0,04	0,46	25	5	32,3
30	0,01	0,08	13	9	3,7	0,04	0,30	15	6	13,0	0,07	0,57	24	4	34,6
35	0,02	0,13	13	8	5,2	0,06	0,41	15	5	17,7	0,09	0,63	24	4	31,9
40	0,03	0,17	12	7	7,2	0,08	0,48	16	4	23,5	0,11	0,64	22	4	28,3
45	0,03	0,20	12	6	9,3	0,10	0,52	16	4	22,4	0,13	0,64	19	3	34,1
50	0,04	0,23	11	5	12,9	0,11	0,54	15	4	20,9	0,14	0,63	18	3	31,3

Se puede apreciar, de la Tabla 2, cómo el tamaño de grupo recomendado, tanto considerando como no considerando correlación, baja a medida que avanzan los días. Esto se debe al aumento sostenido de prevalencia y correlación. Además, se aprecia que los casos con mayor correlación ocurren al asumir un mayor grado de interacción (probabilidad de interacción del 20%). Esto trae como consecuencia que omitir tales niveles de correlación implican una pérdida de eficiencia en el número esperado de tests a utilizar. Por ejemplo, se aprecia que luego de un mes de la simulación, usar una estrategia de tes-

Tabla 3: Para cada día de simulación con respecto al primer caso sintomático: prevalencia, correlación, número óptimo de grupo (n^*), número óptimo de grupo omitiendo correlación (n^\diamond), y ahorro en número de tests a usar. Probabilidades de interacciones diarias del 5, 10 y 20 % de izquierda a derecha.

Día	5%					10%					20%				
	Prev.	ρ	n^*	n^\diamond	Ahorro [%]	Prev.	ρ	n^*	n^\diamond	Ahorro [%]	Prev.	ρ	n^*	n^\diamond	Ahorro [%]
5	0,04	0,01	5	5	0,0	0,06	0,02	5	5	0,0	0,09	0,06	4	4	0
10	0,06	0,03	5	5	0,0	0,12	0,08	4	4	0,0	0,23	0,18	4	3	1,4
15	0,09	0,06	4	4	0,0	0,21	0,17	4	3	1,6	0,43	0,29	9	1	1,3
20	0,11	0,11	5	4	0,2	0,30	0,27	5	3	2,5	0,60	0,32	1	1	0
25	0,12	0,19	5	3	6,2	0,37	0,35	7	1	10,1	0,69	0,31	1	1	0
30	0,13	0,27	6	3	9,1	0,41	0,38	8	1	8,2	0,69	0,28	1	1	0
35	0,14	0,31	7	3	10,3	0,42	0,36	8	1	6,3	0,59	0,26	1	1	0
40	0,15	0,33	7	3	10,8	0,39	0,34	7	1	6,9	0,43	0,25	1	1	0
45	0,14	0,33	7	3	10,6	0,34	0,30	6	1	10,5	0,26	0,23	4	3	1,8
50	0,14	0,30	7	3	10,5	0,27	0,27	5	3	2,7	0,13	0,20	5	3	5,9

teo que considere la correlación puede traer un ahorro en el número de tests de 34,6 % (versus testeo grupal miope a la correlación). Al realizar el mismo ejercicio sobre los resultados de la Tabla 3, se observa que el ahorro en número esperado de tests es de menor magnitud que en el caso anterior. Esto ocurre porque los niveles de correlación son menores al mirar los escenarios desde el día del primer caso sintomático versus desde el día de inicio de la simulación.

6. Discusión y conclusiones

Frente a la preocupante progresión de la infección por COVID-19 en poblaciones cerradas, tales como establecimientos de larga estadía para adultos mayores (ELEAM), este trabajo presenta un modelo de testeo grupal de dos etapas que incorpora de manera explícita la correlación en la probabilidad de estar infectado, y que puede ser usado para minimizar el número esperado de tests. En caso de tests con sensibilidades menores a uno se presenta una fórmula explícita para evaluar el riesgo en los falsos negativos, que es creciente con el tamaño del agrupamiento. Lo anterior hace evidente el trade-off que puede existir entre minimizar el número esperado de tests versus el número de falsos negativos.

El contexto de aplicación de este modelo son las poblaciones (parcialmente) cerradas, que en nuestro caso de estudio representa la población de un ELEAM. Para estimar la prevalencia y la correlación presente en un ELEAM se construyó un modelo de simulación que permite seguir la evolución de los pacientes contagiados, utilizando parámetros de la literatura y de las políticas

implementadas en la realidad. Ajustando una distribución Beta-Binomial, se estimó la prevalencia y correlación para cada día y se obtuvo como resultado el tamaño del agrupamiento óptimo en el contexto de testeo grupal de dos etapas.

El análisis desarrollado caracteriza el ahorro en el número de tests a utilizar al usar la política óptima de testeo grupal versus su versión tradicional que omite correlación. Lo anterior es significativamente más pronunciado al mirar los datos de las simulaciones desde el inicio de ésta. Además, los resultados resaltan la importancia de acudir prontamente a los ELEAM una vez que se han detectado casos sintomáticos, debido al rápido crecimiento de la prevalencia en los días inmediatamente posteriores como se ilustra en la Figura 6.

En este mismo sentido, los resultados oportunos son los que realmente permiten gestionar los aislamientos preventivos, puesto que de nada sirve testear a una población si no se es capaz de minimizar los riesgos de contagio mientras se está a la espera de los resultados.

Estos resultados resaltan la importancia de disponer de un mecanismo que permita anticiparse a un elevado número de contagios, por ejemplo, realizando con frecuencia tests a todos los miembros de la población o recomendando testeos grupales periódicos en grupos cerrados, cada dos o tres semanas en nuestro caso de estudio. De esta manera, sería posible identificar a tiempo cualquier potencial brote utilizando un número muy limitado de tests, puesto que esto “reiniciaría” la dinámica de la evolución de la infección (tal como se ilustra en la Figura 5).

En términos de direcciones futuras de investigación, el siguiente paso natural consiste en la validación de la dinámica de los contagios en el modelo de simulación presentado en este trabajo. Una vez validado, el modelo puede servir de base para la evaluación de estrategias de testeo y de cuarentenas preventivas, que podrían incluir el testeo agrupado de toda la población bajo estudio. Por otro lado, el modelo toma un número de supuestos respecto a la evolución temporal de la infección, y a la dinámica de contagio, basado en evidencia parcial, recolectada hasta el momento sobre la pandemia. En la medida que el conocimiento acerca del virus se perfeccionen, nuevos y mejores modelos podrán ser planteados y utilizados.

Agradecimientos: Esta investigación ha sido financiada en parte por el Instituto Sistemas Complejos de Ingeniería ISCI (ICM-FIC: P05-004-F, CONICYT: FB0816). Los autores agradecen el apoyo de ANID PIA/BASAL AFB180003.

Referencias

- [1] D. Balding, W. Bruno, D. Torney, y E. Knill. A comparative survey of non-adaptive pooling designs. En *Genetic mapping and DNA sequencing*, páginas 133–154. Springer, 1996.
- [2] CDC. Contract tracing plan, covid-19. Accessed July 22, 2020, 2020.
- [3] A. Cherif, N. Grobe, X. Wang, y P. Kotanko. Simulation of pool testing to identify patients with coronavirus disease 2019 under conditions of limited test availability. *JAMA Network Open*, 6(3), 2020.
- [4] Diario Oficial. Resolución 424 del 9 de junio de 2020, ministerio de salud, gobierno de chile, 2020.
- [5] R. Dorfman. The detection of defective numbers of large populations. *Annals of Mathematical Statistics*, 1(14):436–440, 1943.
- [6] M. Farfan, J. Torres, M. O’Ryan, M. Olivares, P. Gallardo, y S. C. Optimizing rt-pcr detection of sars-cov-2 for developing countries using pool testing. medRxiv 2020.04.15.20067199, 2020.
- [7] A. Gill y D. Gottlieb. The identification of a set by successive intersections. *Information and Control*, 24(1):20–35, 1974.
- [8] X. He, E. H. Lau, P. Wu, X. Deng, J. Wang, X. Hao, Y. C. Lau, J. Y. Wong, Y. Guan, X. Tan, et al. Temporal dynamics in viral shedding and transmissibility of covid-19. *Nature medicine*, 26(5):672–675, 2020.
- [9] C. Mentus, M. Romeo, y C. DiPaola. Analysis and applications of adaptive group testing methods for covid-19. *medRxiv*, 2020.
- [10] Ministerio de Salud. Protocolo de coordinación para acciones de vigilancia epidemiológico durante la pandemia covid-10 en chile: Estrategia nacional de testeo, trazabilidad y aislamiento, 2020.
- [11] N. Sethuraman, S. S. Jeremiah, y A. Ryo. Interpreting Diagnostic Tests for SARS-CoV-2. *JAMA*, 323(22):2249–2251, 2020.
- [12] Q. Stephen A. Lauer, Kyra H. Grantz, Bi, et al. The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application. *Annals of Internal Medicine*, 172(9):577–582, 2020.
- [13] A. Sterrett. On the detection of defective members of large populations. *The Annals of Mathematical Statistics*, 28(4):1033–1036, 1957.

- [14] L. M. Wein y S. A. Zenios. Pooled testing for hiv screening: capturing the dilution effect. *Operations Research*, 44(4):543–569, 1996.
- [15] S. Woloshin, N. Patel, y A. Kesselheim. False negative tests for sars-cov-2 infection - challenges and implications. *New England Journal of Medicine*, 38(383), 2020.
- [16] WSJ. Wuhan tests nine million people for coronavirus in 10 days. Accessed September 11, 2020, 2020.
- [17] I. Yelin, N. Aharony, E. Shaer-Tamar, A. Argoetti, E. Messer, D. Berenbaum, E. Shafran, A. Kuzli, N. Gandali, T. Hashimshony, et al. Evaluation of covid-19 rt-qpcr test in multi-sample pools. *MedRxiv*, 2020.