

# Técnicas matemáticas para diagnosis médica

TRABAJO FIN DE GRADO

Curso 2020/2021



**UNIVERSIDAD COMPLUTENSE  
MADRID**

FACULTAD DE CIENCIAS MATEMÁTICAS

GRADO EN INGENIERÍA MATEMÁTICA

Alumna: Celia Sierra García

Tutora: Ana María Carpio Rodríguez

Madrid, 26 de junio de 2021

## **Resumen:**

Un diagnóstico correcto es de vital importancia para la vida del paciente, pues de lo contrario se le puede negar una terapia oportuna o administrar medicamentos potencialmente tóxicos. No obstante, en la mayoría de los casos, lograr un diagnóstico certero puede ser de gran complejidad, pues el médico debe tener en cuenta numerosos y complejos factores. A lo largo de este trabajo se verá cómo, mediante la aplicación de técnicas matemáticas, se puede disminuir la incertidumbre que acompaña a los médicos en la toma de decisiones. Se estudiarán los pasos a seguir en el caso del diagnóstico de la tuberculosis, para después extrapolar los procedimientos seguidos a otras enfermedades infecciosas. En concreto, se indicarán los factores a tener en cuenta para un diagnóstico correcto del COVID-19.

**PALABRAS CLAVE:** diagnosis, factores de riesgo, regresión logística, tuberculosis, COVID-19.

## **Abstract:**

A correct diagnosis is extremely important to the patient's life; otherwise a convenient therapy could be denied to him or he could receive toxic medication. However, most of the time it is incredibly difficult to get an accurate diagnosis, as physicians need to take into account numerous factors. During this project it will be seen how we can apply mathematics to reduce the uncertainty around physicians' decisions. It will be shown what steps must be followed to diagnose active tuberculosis with the goal of being able to follow the same procedure with other infectious diseases such as COVID-19.

**KEYWORDS:** diagnosis, risk factors, logistic regression, tuberculosis, COVID-19.

# Índice

Índice de tablas	4
Índice de figuras	5
<b>1. Antecedentes</b>	<b>6</b>
<b>2. Objetivo y plan del trabajo</b>	<b>6</b>
<b>3. Tipos de estudio</b>	<b>6</b>
<b>4. Pruebas diagnósticas</b>	<b>8</b>
4.1. Pruebas con resultados dicotómicos . . . . .	8
4.1.1. Cálculo de los falsos positivos . . . . .	10
4.2. Pruebas con más de dos resultados . . . . .	10
4.3. Pruebas múltiples . . . . .	11
<b>5. Factores de riesgo y riesgo relativo</b>	<b>11</b>
<b>6. Diagnóstico de la tuberculosis</b>	<b>13</b>
6.1. Pequeña introducción a la tuberculosis . . . . .	13
6.2. Cálculo del riesgo acumulativo de desarrollar tuberculosis . . . . .	14
6.3. Cálculo de la incidencia anual acumulada de la tuberculosis . . . . .	14
6.4. Cálculo del valor predictivo positivo de la TST . . . . .	14
6.4.1. Cálculo de la prevalencia de tuberculosis latente . . . . .	14
6.4.1.1. Cálculo del riesgo anual de infección . . . . .	15
6.4.1.2. Cálculo de la prevalencia de infección entre contactos . . . . .	15
6.4.2. Cálculo de los falsos positivos . . . . .	16
6.4.2.1. Falsos positivos por vacunación . . . . .	16
6.4.2.2. Falsos positivos por exposición a NTM . . . . .	17
6.5. Cálculo del valor predictivo positivo . . . . .	17
6.6. Cálculo del nivel endémico . . . . .	18
6.7. Factores de riesgo y riesgo relativo . . . . .	18
6.8. RR debido a la etnia . . . . .	19
<b>7. Diagnóstico del COVID-19</b>	<b>19</b>
7.1. Pequeña introducción al COVID-19 . . . . .	19
7.2. Datos empleados . . . . .	20
7.3. Cálculo del valor predictivo positivo . . . . .	20
7.3.1. Cálculo de la seroprevalencia . . . . .	20
7.3.2. Cálculo de los falsos positivos . . . . .	20
7.4. Tasa de ataque secundario . . . . .	21
7.5. Análisis de conglomerados . . . . .	23
7.6. Factores de riesgo y riesgo relativo . . . . .	25
7.7. Análisis de la participación en el estudio serológico . . . . .	28
7.8. Cálculo del riesgo de infección y proporción de asintomáticos . . . . .	29

<b>Siglas</b>	<b>38</b>
<b>Referencias</b>	<b>39</b>

## Índice de figuras

1.	Tipos de estudios [3] . . . . .	7
2.	Relación del ARI con la prevalencia e incidencia de TBC (Styblo) [8] . . . . .	15
3.	Sensibilidad y especificidad pruebas COVID [11] . . . . .	20
4.	Árbol de probabilidades [15] . . . . .	31
5.	Trace plot para $\pi$ [15] . . . . .	35
6.	Trace plot para $s$ [15] . . . . .	35
7.	Trace plot para $\tau$ [15] . . . . .	35
8.	Trace plot para $\sigma$ [15] . . . . .	35
9.	Trace plot para $p$ [15] . . . . .	36
10.	Trace plot para $\nu$ [15] . . . . .	36
11.	Trace plot para $y$ [15] . . . . .	36

## Índice de tablas

1.	Resultados de la prueba en función de la existencia de la enfermedad . . . . .	8
2.	Cálculo del riesgo relativo . . . . .	11
3.	Cálculo del Odds Ratio . . . . .	12
4.	Obtención falsos positivos por BCG . . . . .	16
5.	Falsos positivos por vacunación [9] . . . . .	17
6.	Factores de riesgo para la TB [9] . . . . .	18
7.	Factores de riesgo para la TAS [15] . . . . .	22
8.	Factores de riesgo para la TAS (interacciones) [15] . . . . .	22
9.	Odds de TAS [15] . . . . .	23
10.	Factores de riesgo entre convivientes [15] . . . . .	26
11.	Factores de riesgo entre no convivientes [15] . . . . .	27
12.	Clasificación de los contactos en grupos [15] . . . . .	30
13.	Probabilidades de pertenecer a cada subgrupo [15] . . . . .	32
14.	Valor de los parámetros a posteriori [15] . . . . .	34
15.	Valores derivados de los parámetros a posteriori [15] . . . . .	34

## 1. Antecedentes

En este trabajo se seguirán dos investigaciones previamente realizadas. Una primera, publicada en 2008, en la que se desarrolló un algoritmo capaz de interpretar los resultados de la prueba de la tuberculina (TST) a partir del tamaño de la reacción en la piel, el valor predictivo positivo de dicha prueba y el riesgo de desarrollar la enfermedad [9]. Una segunda, publicada en 2021, en la que se investigan los factores de riesgo en la transmisión del COVID-19 entre contactos cercanos; así como la seroprevalencia y la tasa de ataque secundario del mismo [15].

## 2. Objetivo y plan del trabajo

La RAE define *diagnóstico* como la determinación de la naturaleza de una enfermedad mediante la observación de sus síntomas; o bien, como la calificación que da el médico a la enfermedad según los signos que advierte.

Es evidente la importancia de obtener un diagnóstico correcto, pues un diagnóstico tardío o un error en este pueden ocasionar que se niegue al paciente una terapia oportuna o que se le administren medicamentos incorrectos y potencialmente tóxicos [1].

Jerome P. Kassirer afirma que la obtención de un diagnóstico absolutamente certero es inalcanzable, pues se trata de una hipótesis acerca de la naturaleza de la enfermedad de un paciente que se deriva de observaciones a través del uso de la inferencia, siendo el objetivo de los médicos reducir el nivel de incertidumbre lo suficiente como para poder tomar una decisión terapéutica óptima [2].

Un caso que ejemplifica lo indicado previamente es el de la tuberculosis, donde la dificultad para analizar los resultados de sus pruebas diagnósticas han llevado al desarrollo de un algoritmo capaz de interpretar dichos resultados y de predecir el riesgo que tiene un individuo de desarrollar tuberculosis activa (TB) a lo largo de su vida.

Este trabajo constará de tres partes:

1. Introducción de los conceptos claves para poder llevar a cabo el estudio de una enfermedad infecciosa y su correspondiente diagnóstico.
2. Ejemplificación de la primera parte mediante el análisis del algoritmo mencionado previamente.
3. Extrapolación de los cálculos estudiados en la segunda parte a otras enfermedades. Concretamente, al COVID-19.  
Nótese que, a pesar de que se seguirá el esquema determinado por el análisis de la TB, se introducirán algunos conceptos diferentes debido a la caracterización del COVID-19 como pandemia.

## 3. Tipos de estudio

Los estudios epidemiológicos se clasifican en descriptivos o analíticos.

- **Los estudios descriptivos** valoran la frecuencia y la distribución de las enfermedades en las poblaciones. Estos pueden aportar indicios que contribuyan a generar hipótesis sobre asociaciones entre factores de exposición y estados de salud o enfermedad.
- **Los estudios analíticos** comprueban o rechazan las hipótesis generadas por los estudios descriptivos, con el objetivo de identificar factores de riesgo o de protección de una enfermedad.

Para el desarrollo de este trabajo se emplearán dos tipos de estudios analíticos:

- **Estudio de casos y controles:** estudio en el cual los sujetos se seleccionan en función de que tengan (casos) o no tengan (controles) una determinada enfermedad. Una vez seleccionados los individuos en cada grupo, se investiga si estuvieron expuestos o no a una característica de interés y se compara la proporción de expuestos o no en el grupo de casos frente a la del grupo de controles.
- **Estudio de cohortes:** estudio en el que se hace una comparación de la frecuencia de aparición de un evento entre dos grupos, uno de los cuales está expuesto a un factor que no está presente en el otro grupo. Los individuos que componen los grupos de estudio se seleccionan en función de la presencia o ausencia de la exposición evaluada. El estudio de cohortes puede ser prospectivo o retrospectivo:

- En el estudio prospectivo se elige una muestra de individuos en los cuales se mide la exposición a estudio y se les sigue durante un periodo de tiempo para ver si se produce un determinado efecto.
- En el estudio retrospectivo se selecciona una muestra de individuos y partiendo de información ya existente (registros, historias clínicas, etc.), se reconstruye lo que ocurrió con esa cohorte al ser sometida a un factor de exposición, observando si hubo diferencias con el grupo no expuesto en la aparición del efecto.

La diferencia se puede ver claramente en la siguiente imagen:

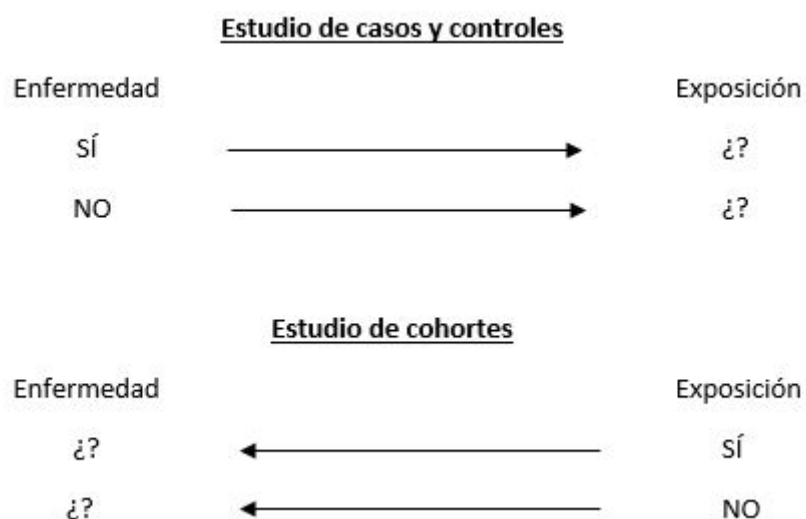


Figura 1: Tipos de estudios [3]



## 4. Pruebas diagnósticas

### 4.1. Pruebas con resultados dicotómicos

Se llama prueba diagnóstica (PD) a cualquier proceso que pretenda determinar en un paciente la presencia de cierta condición, supuestamente patológica, no susceptible de ser observada directamente (es decir, con alguno de los cinco sentidos elementales) [4].

A lo largo del trabajo se empleará la siguiente notación:  $T^+$  y  $T^-$  para un resultado positivo y negativo en el test respectivamente;  $E^+$  para una persona con la enfermedad y  $E^-$  para una persona que no sufre la enfermedad.

	Enfermos	No enfermos	Total
Positivos	a (VP)	b (FP)	a+b
Negativos	c (FN)	d (VN)	c+d
Total	a+c	b+d	a+b+c+d

Tabla 1: Resultados de la prueba en función de la existencia de la enfermedad

siendo VP los verdaderos positivos, FP los falsos positivos, FN los falsos negativos y VN los verdaderos negativos.

Para una correcta evaluación de una prueba diagnóstica se deben conocer los siguientes conceptos:

- **Sensibilidad:** Es la capacidad de una prueba para identificar correctamente aquellos individuos que tienen la enfermedad. Es decir, se trata de la probabilidad de que para un sujeto enfermo se obtenga en la prueba un resultado positivo.

$$P(T^+|E^+)$$

Su estimación es:

$$S = \frac{\text{Verdaderos positivos}}{\text{Total de enfermos}} = \frac{VP}{VP + FN} = \frac{a}{a + c} \quad (1)$$

A continuación, se calcula su intervalo de confianza.

Sea

$$X_i = \begin{cases} 1 & \text{si el paciente } i - \text{ésimo enfermo tiene un resultado positivo en el test} \\ 0 & \text{si un paciente } i - \text{ésimo enfermo tiene un resultado negativo en el test} \end{cases} \quad (2)$$

un suceso que se repite  $N$  veces, con probabilidad  $S$  de ocurrencia constante. Suponiendo que las experiencias sucesivas son independientes, se asume un modelo binomial.

Sea

$$\hat{S} = \frac{\sum_1^N X_i}{N} \quad (3)$$

un estimador sin sesgo de  $S$  cuya varianza se estima por

$$\frac{\hat{S} \times (1 - \hat{S})}{N} \quad (4)$$

Por el Teorema Central del Límite (siendo  $N \geq 30$ ) se obtiene:

$$IC_{(1-\alpha) \times 100\%} = \left( \hat{S} - z_{\frac{\alpha}{2}} \times \sqrt{\frac{\hat{S} \times (1 - \hat{S})}{N}}, \hat{S} + z_{\frac{\alpha}{2}} \times \sqrt{\frac{\hat{S} \times (1 - \hat{S})}{N}} \right) \quad (5)$$

- **Especificidad:** Es la capacidad de una prueba para identificar correctamente aquellos individuos que no tienen la enfermedad. Es decir, se trata de la probabilidad de que para un sujeto sano se obtenga en la prueba un resultado negativo.

$$P(T^- | E^-)$$

Su estimación es:

$$S = \frac{\text{Verdaderos negativos}}{\text{Total de no enfermos}} = \frac{VN}{VN + FP} = \frac{d}{b + d} \quad (6)$$

El cálculo de los intervalos de confianza es análogo al caso de la sensibilidad.

- **Prevalencia de la enfermedad:** Es la proporción de la población de individuos que en un intervalo dado y en una región determinada se consideran enfermos. Por tanto, se identifica con la probabilidad de que un individuo padezca la enfermedad antes de realizar la prueba (probabilidad a priori).

$$P(E^+)$$

Su estimación es:

$$P = \frac{\text{Total de enfermos}}{\text{Total de individuos}} = \frac{a + c}{a + b + c + d} \quad (7)$$

El cálculo de los intervalos de confianza es análogo al caso de la sensibilidad.

- **Incidencia:** Es el número de casos nuevos de una enfermedad que se desarrollan en una población durante un período de tiempo determinado. Se puede medir con dos índices: incidencia acumulada (IA) o densidad (o tasa) de incidencia (DI).

- *Indidencia acumulada:* proporción de individuos que desarrollan la enfermedad durante el periodo de seguimiento.

Su estimación es:

$$IA = \frac{\text{Nuevos casos}}{\text{Total de individuos susceptibles}} \quad (8)$$

Se emplea cuando la cohorte es fija, es decir, cuando no se permiten entradas de nuevos individuos durante el seguimiento.

- *Densidad de incidencia*: cociente entre el número de casos nuevos ocurridos durante el periodo de seguimiento y la suma de todos los tiempos de observación. Su estimación es:

$$DI_{(t_0, t_1)} = \frac{\text{Nuevos casos}}{\sum_{i=1}^N \Delta t_i} \quad (9)$$

Se emplea cuando la cohorte no es fija.

- **Valor predictivo positivo**: Es la probabilidad que tiene un individuo de estar enfermo cuando el resultado de la prueba es positivo.

$$PPV = P(E^+|T^+)$$

Se obtendrá la fórmula para calcular el PPV a partir del teorema de Bayes.

$$\begin{aligned} PPV = P(E^+|T^+) &= \frac{P(T^+|E^+) \times P(E^+)}{P(T^+)} = \frac{P(T^+|E^+) \times P(E^+)}{P(T^+|E^+) \times P(E^+) + P(T^+|E^-) \times P(E^-)} = \\ &= \frac{P(T^+|E^+) \times P(E^+)}{P(T^+|E^+) \times P(E^+) + P(T^+ \cap E^-)} = \frac{P(T^+|E^+) \times P(E^+)}{P(T^+|E^+) \times P(E^+) + FP} = \\ &= \frac{\text{sensibilidad} \times \text{prevalencia}}{\text{sensibilidad} \times \text{prevalencia} + FP} \end{aligned} \quad (10)$$

El cálculo de los intervalos de confianza es análogo al caso de la sensibilidad.

#### 4.1.1. Cálculo de los falsos positivos

Para poder calcular los falsos positivos se necesita, previamente, identificar las distintas causas de que un individuo que no tenga la enfermedad obtenga un resultado positivo en la prueba.

$$FP = \sum_i FP_i \quad (11)$$

siendo  $FP_i$  los falsos positivos debidos a la causa  $i$ -ésima.

## 4.2. Pruebas con más de dos resultados

En el caso de que los resultados de una prueba diagnóstica sean variables continuas es necesario determinar un nivel umbral o punto de corte (PC) encima del cual el resultado de la prueba es considerado positivo y por debajo de este es considerado negativo.

La relación entre la sensibilidad y la especificidad de dichas pruebas dependerá del punto de corte seleccionado, lo que plantea la necesidad de establecer puntos de corte óptimos. Para ello se emplean las curvas ROC.

Las curvas ROC se construyen mediante la representación de los puntos  $(x, y)$  dados por  $(1 - \text{Especificidad}, \text{Sensibilidad})$  para cada punto de corte en el rango de resultados observados. El gráfico expresará mayor poder discriminatorio cuanto mayor sea el área bajo la curva y el mejor punto de corte será aquel que se aleje más de la diagonal.

### 4.3. Pruebas múltiples

En ocasiones, se recurre al uso de pruebas múltiples. Existen dos posibilidades:

- **Pruebas en paralelo:** Se hacen simultáneamente dos o más pruebas en el mismo individuo a fin de aumentar la sensibilidad diagnóstica.  
Se considera negativa la prueba en aquellos sujetos que tengan resultados negativos en todas las pruebas y positiva en aquellos que presenten resultados positivos en al menos una de ellas.
- **Pruebas en serie:** Se aplican secuencialmente con la finalidad de aumentar la especificidad (es decir, para confirmar un primer resultado positivo).  
Se considera positiva la prueba en aquellos sujetos que tengan resultados positivos en todas las pruebas y negativa en aquellos que presenten resultados negativos en las pruebas confirmatorias.

## 5. Factores de riesgo y riesgo relativo

Un factor de riesgo es cualquier elemento que aumenta la posibilidad de desarrollar una enfermedad; mientras que el riesgo relativo (RR) es la probabilidad de que se desarrolle la enfermedad en individuos expuestos a algún factor de riesgo, en relación a los individuos no expuestos.

Empleando la siguiente notación:

	Enfermos	Sanos	Total
Expuestos	a	b	a+b
No expuestos	c	d	c+d
Total	a+c	b+d	a+b+c+d

Tabla 2: Cálculo del riesgo relativo

el RR se puede estimar de la siguiente forma:

$$RR = \frac{I_e}{I_o} = \frac{a/(a+b)}{c/(c+d)} \quad (12)$$

siendo  $I_e$  la incidencia en expuestos e  $I_o$  la incidencia en no expuestos.

Si:

- $RR = 1$ : El riesgo de desarrollar la enfermedad es idéntico en el grupo expuesto y en el no expuesto.
- $RR < 1$ : Se trata de un factor de protección.
- $RR > 1$ : Se trata de un factor de riesgo.

Para que el estudio del RR sea completo se debe calcular su intervalo de confianza, pues si este no contiene el valor 1 se puede concluir que el riesgo relativo es estadísticamente significativo.

Como el RR sigue una distribución lognormal, se emplea el  $\ln(RR)$  para calcular el intervalo de confianza ( $\ln(RR) \sim N(\mu, \sigma)$ ).

El intervalo de confianza será [5]:

$$IC_{1-\alpha} = (RR \times \exp(-z_{\frac{\alpha}{2}} \times EE[\ln(RR)]), RR \times \exp(z_{\frac{\alpha}{2}} \times EE[\ln(RR)])) \quad (13)$$

donde el error estándar del  $\ln(RR)$  ( $EE[\ln(RR)]$ ) se calcula de la siguiente manera:

$$EE[\ln(RR)] = \sqrt{\frac{1}{a} - \frac{1}{a+b} + \frac{1}{c} - \frac{1}{c+d}} \quad (14)$$

En los estudios del riesgo relativo es habitual tomar como nivel de confianza un 95 %, luego  $z_{\frac{\alpha}{2}} = 1,96$ .

En los estudios de casos y controles la incidencia es desconocida, por lo que el riesgo relativo se estima mediante el cálculo del Odds Ratio (OR).

El Odds Ratio es una división entre dos Odds (probabilidad de que se presente el suceso entre probabilidad de que no se presente).

Empleando la siguiente notación:

	Casos	Controles
Expuestos	a	b
No expuestos	c	d

Tabla 3: Cálculo del Odds Ratio

el OR se puede estimar de la siguiente forma:

$$OR = \frac{\text{Odds en los casos}}{\text{Odds en los controles}} = \frac{a/c}{b/d} = \frac{a \times d}{b \times c} \quad (15)$$

En los estudios epidemiológicos es habitual emplear un análisis de regresión logística, pues los parámetros de la ecuación están directamente relacionados con el OR.

Cuando se realiza una regresión logística se busca estimar los parámetros  $(\beta_0, \beta_1, \dots, \beta_k)$  de la función:

$$Z = \beta_0 + \beta_1 \times x_1 + \dots + \beta_k \times x_k \quad (16)$$

siendo  $Z$  el logaritmo neperiano del OR de padecer una enfermedad y  $(x_1, x_2, \dots, x_k)$  las variables independientes o factores de riesgo. Los coeficientes  $\beta_i$  son el logaritmo neperiano del OR para cada factor de riesgo  $x_i$ . Por lo que el Odds Ratio se estima a partir de la siguiente fórmula:

$$OR = e^{\beta_i} \quad (17)$$

Para calcular los intervalos de confianza se asume que los coeficientes tienen una distribución normal.

$$IC_{95\%}(\beta) = [(\beta - 1,96 \times EE), (\beta + 1,96 \times EE)] \quad (18)$$

$$IC_{95\%}(OR) = [exp(\beta - 1,96 \times EE), exp(\beta + 1,96 \times EE)] \quad (19)$$

siendo EE el error estándar.

La inclusión del valor 1 en el intervalo de confianza es indicativo de ausencia de significación de la covariable en el modelo.

Por último, a la hora de estudiar la interacción entre variables hay que tener en cuenta los posibles factores de confusión.

El fenómeno de confusión aparece cuando la asociación observada entre el factor de estudio y la enfermedad puede ser total o parcialmente explicada por otra variable (factor de confusión), o por el contrario, cuando una asociación real queda enmascarada por este factor.

Para que una variable se considere factor de confusión deber cumplir tres condiciones:

1. Ser un factor de riesgo de la enfermedad
2. Estar asociado a la variable de estudio
3. No debe ser un paso intermedio en la cadena causal (en la asociación entre el factor de estudio y la enfermedad)

Por ello, en este trabajo se realizarán análisis de regresión logística univariada y multivariante. En ambos casos se calcularán los OR para cada factor de riesgo y se compararán para detectar posibles factores de confusión.

Nótese que, en el caso de la regresión multivariante, los OR recibirán el nombre de Odds Ratio Ajustado (aOR).

## 6. Diagnóstico de la tuberculosis

### 6.1. Pequeña introducción a la tuberculosis

La tuberculosis es una enfermedad infecciosa, causada por la bacteria *Mycobacterium tuberculosis*, que afecta principalmente a los pulmones. Se debe distinguir entre:

- **Tuberculosis latente (LTBI):** El individuo está infectado por la bacteria mencionada previamente, pero no presenta manifestaciones clínicas. No contagia.
- **Tuberculosis activa (TB):** El individuo presenta síntomas. Sí contagia.

Existen distintas pruebas para diagnosticar si una persona está infectada con las bacterias de la tuberculosis, entre las que destacan la prueba cutánea de la tuberculina (TST) y la prueba IGRA (mediante un análisis de sangre).

Sin embargo, predecir el riesgo de desarrollar tuberculosis activa puede resultar extremadamente complicado, ya que hay que tener en cuenta numerosos factores a la hora de interpretar

los resultados de las pruebas previamente mencionadas. Esto hace necesario recurrir a algoritmos que permitan mejorar la detección de pacientes con LTBI y alto riesgo de desarrollar TB, para así poder evaluar los riesgos de aplicarles tratamientos preventivos.

## 6.2. Cálculo del riesgo cumulativo de desarrollar tuberculosis

Para calcular el riesgo o probabilidad de que un individuo desarrolle tuberculosis a lo largo de su vida (CRD), basta con multiplicar el riesgo anual de desarrollar la enfermedad (ARD) por los años de vida que le quedan al individuo.

Para simplificar los cálculos, se supone que los individuos viven 80 años.

$$CRD = ARD \times (80 - n) \quad (20)$$

siendo  $n$  la edad del paciente.

## 6.3. Cálculo de la incidencia anual acumulada de la tuberculosis

La incidencia anual acumulada (ARD) es la proporción de individuos que desarrollan la enfermedad a lo largo de un año. Esta proporciona una estimación de la probabilidad de que un individuo libre de tuberculosis desarrolle la enfermedad durante este período determinado. Para su cálculo se emplea la siguiente fórmula:

$$ARD = PPV \times NE \times RR \times e \quad (21)$$

donde PPV es el valor predictivo positivo, NE el nivel endémico, RR el riesgo relativo y  $e$  el riesgo relativo debido a la etnia.

## 6.4. Cálculo del valor predictivo positivo de la TST

Para calcular el valor predictivo positivo (PPV) de la TST (probabilidad de que un individuo con un resultado positivo en la TST tenga tuberculosis activa) se debe tener en cuenta que dicha prueba tiene una sensibilidad del 99% y una especificidad del 95% [6], luego:

$$PPV_{TST} \approx \frac{\text{prevalencia}}{\text{prevalencia} + FP} = \frac{pLBTI}{pLBTI + FP} \quad (22)$$

siendo  $pLBTI$  la prevalencia de la tuberculosis latente.

### 6.4.1. Cálculo de la prevalencia de tuberculosis latente

Para calcular la prevalencia se emplea la fórmula propuesta por J. Nyboe [7], donde la prevalencia de infección esperada se calcula a partir de un riesgo constante conocido y la edad para la cual se quiere conocer la prevalencia:

$$pLBTI = 1 - (1 - ARI)^a \quad (23)$$

siendo  $a$  la edad para la cual se quiere conocer la probabilidad de que el individuo desarrolle la enfermedad y ARI el riesgo anual de infección (proporción de personas que se (re)infectan

en un año). Nótese que, de acuerdo a esto,  $(1 - ARI)^a$  es la proporción de personas no infectadas con a años.

No obstante, esta fórmula es válida si el individuo no ha estado previamente en contacto con ningún individuo infectado. Una expresión más general que recoge el contacto ( $c$ ) con personas con TB es:

$$pLBTI = 1 - [(1 - c)(1 - ARI)^a] \quad (24)$$

#### 6.4.1.1. Cálculo del riesgo anual de infección

Para obtener el riesgo anual de infección (ARI) se ha empleado la regla de Styblo [8], la cual asume una relación matemática entre la incidencia de la TB bacilífera (TBC), la prevalencia de la TBC y el ARI. Dicha relación es la siguiente:

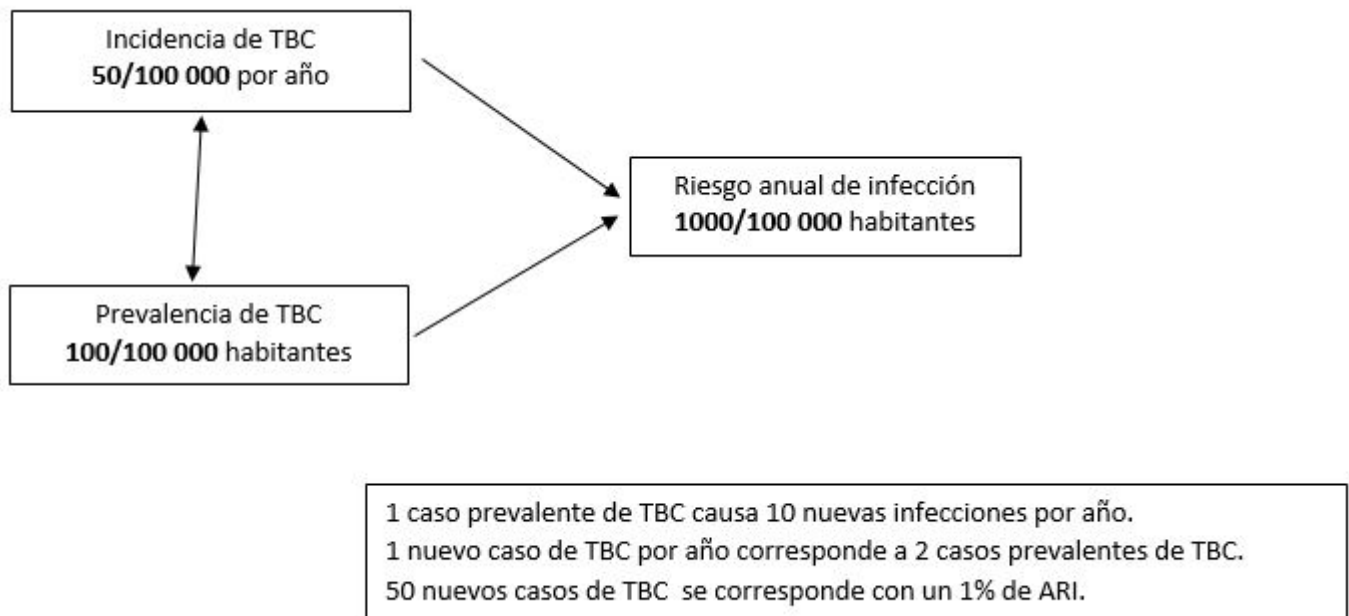


Figura 2: Relación del ARI con la prevalencia e incidencia de TBC (Styblo) [8]

Nótese que los enfermos bacilíferos son aquellos que eliminan la micobacteria a través de la tos, los estornudos y las secreciones respiratorias. Esto se evidencia por presentar una baciloscopia de esputo positiva.

Por tanto, el ARI se calcula dividiendo la incidencia de TBC por cada 100 000 habitantes entre 50, es decir,

$$ARI = \frac{\text{Incidencia de TBC por } 100\ 000 \text{ habitantes}}{50} \quad (25)$$

#### 6.4.1.2. Cálculo de la prevalencia de infección entre contactos

La prevalencia de infección entre contactos ( $c$ ) es [9]:

- $c = 0$ : si el paciente no ha tenido contacto con ninguna persona infectada.



- $c = 8\%$ : si el paciente ha tenido contacto no estrecho con alguna persona infectada.
- $c = 40\%$ : si el paciente ha tenido contacto estrecho con alguna persona infectada.

#### 6.4.2. Cálculo de los falsos positivos

En el caso de la TST existen dos posibles causas por las que se pueden obtener falsos positivos: que el individuo haya sido previamente vacunado con la vacuna BCG (bacilo Calmette-Guérin) y/o que haya estado expuesto a micobacterias no tuberculosas (NTM). Se empleará la siguiente notación:

$$FP = FP - TST_{BCG} + FP - TST_{NTM} \quad (26)$$

##### 6.4.2.1. Falsos positivos por vacunación

Para estudiar los falsos positivos debidos a la vacunación BCG se deben tener en cuenta dos factores: tamaño de la reacción del TST y edad de vacunación.

Gracias a numerosos estudios, se conoce que el factor más influyente es la edad. Esto se debe a que la sensibilidad de la piel a la tuberculina es una manifestación crónica, es decir, los resultados de la TST reflejarán el efecto acumulado de la exposición desde el nacimiento hasta la fecha de la prueba. Por ello, se hará una primera clasificación entre individuos vacunados antes de los dos años, e individuos vacunados después. Posteriormente, dentro de cada uno de los dos grupos, se hará una división en tres categorías en función del tamaño de la reacción: 5-9 mm, 10-14 mm,  $\geq 15$  mm.

Para calcular el número de  $FP - TST_{BCG}$  dentro de cada categoría, se compara la proporción de positivos entre los individuos vacunados con la proporción de positivos entre los individuos no vacunados. La diferencia entre ambas proporciones representará el porcentaje de falsos positivos.

	% positivos y vacunados	% positivos y no vacunados	% de falsos positivos
5-9 mm	$x_1$	$x_2$	$x_1 - x_2$
10-14 mm	$x_3$	$x_4$	$x_3 - x_4$
$\geq 15$ mm	$x_5$	$x_6$	$x_5 - x_6$

Tabla 4: Obtención falsos positivos por BCG

Para conocer el número total de falsos positivos basta con multiplicar la proporción de falsos positivos obtenida por el número total de individuos vacunados que han dado un resultado positivo.

Este procedimiento se aplica a datos extraídos de distintos estudios experimentales y se calcula la media ponderada de los resultados obtenidos en cada caso. Se consigue así la proporción final de falsos positivos para cada una de las categorías.

Los valores obtenidos son:

	% Vacunados con <2 años	% Vacunados con $\geq 2$ años
5-9 mm	4.2	25.8
10-14 mm	5.3	8.7
$\geq 15$ mm	2.8	7.8

Tabla 5: Falsos positivos por vacunación [9]

En caso de que el individuo no haya sido vacunado,  $FP - TST_{BCG} = 0$ .

#### 6.4.2.2. Falsos positivos por exposición a NTM

Para calcular el efecto de la exposición a las NTM en las reacciones a la TST se llevaron a cabo diversos estudios en los que se realizaron pruebas duales con un antígeno de *Mycobacterium tuberculosis* y con un antígeno de NTM. En estos estudios, si la reacción a un antígeno es mayor que al otro, la reacción más pequeña es considerada una reacción cruzada y por ende, se trata de un falso positivo.

Las distintas posibilidades son [10]:

- Si el tamaño de la reacción a la TST se encuentra entre 10-14 mm y la reacción a las NTM es mayor (dominante) entonces la reacción a la TST es una reacción cruzada, es decir, es un falso positivo ( $FP - TST_{NTM}$ ).
- Si el tamaño de la reacción a la TST es superior a 15 mm, se considera un verdadero positivo (independientemente del tamaño de la reacción a las NTM).

Teniendo esto en cuenta, se emplearon estos estudios para calcular la tasa de falsos positivos (FPR por sus siglas en inglés) por cada 100 individuos con una reacción a las NTM superior a 5mm. Es decir, se calculó la proporción de casos negativos que la prueba detecta como positivos.

$$FPR = \frac{FP}{FP + VN} \quad (27)$$

siendo VN es el número de verdaderos negativos.

Esta tasa se multiplica por la prevalencia de NTM para estimar la tasa absoluta de falsos positivos.

### 6.5. Cálculo del valor predictivo positivo

Para poder calcular el PPV necesitamos tener en cuenta el  $PPV_{TST}$  y el  $PPV_{IGRA}$  [9].

- Si el resultado de la prueba IGRA es negativo o no se ha realizado la prueba:

$$PPV = PPV_{TST} \quad (28)$$

- Si el resultado de la prueba IGRA es negativo y no se ha realizado la TST:

$$PPV = 0 \quad (29)$$

- Si el resultado de la prueba IGRA es positivo y no se ha realizado la TST:

$$PPV = PPV_{IGRA} = 0,98 \quad (30)$$

- Si el resultado de la prueba IGRA es positivo y se ha realizado la TST:

$$PPV = 1 - ((1 - PPV_{TST}) \times (1 - PPV_{IGRA})) \quad (31)$$

## 6.6. Cálculo del nivel endémico

El nivel endémico hace referencia a la cantidad de una enfermedad específica que está generalmente presente en una comunidad. Se trata, por tanto, de la prevalencia habitual de esa enfermedad.

En el caso de la TB, este dato se consiguió calculando la prevalencia en una cohorte de militares sanos con TST positivo, los cuales fueron estudiados durante 4 años. El resultado obtenido fue  $NE = 0.1\%$  en personas sanas con radiografías de torax normales y ningún otro factor de riesgo.

## 6.7. Factores de riesgo y riesgo relativo

A continuación, se presentan los factores de riesgo que afectan al desarrollo de la TB entre aquellas personas que presenten tuberculosis latente:

Factor de riesgo	RR
Presencia de granulomas en el pecho	2
Fumador (más de un paquete al día)	3
Peso insuficiente ( < 90 % de peso ideal)	3
Diabetes	3.6
Infliximab	4
Tejido fibronodular	6
Infección tuberculosa reciente (menos de dos años)	15
Contacto cercano con un positivo en TB	15
Carcinoma en cabeza o cuello	16
Fallo renal crónico	25.3
Silicosis pulmonar	30
VIH	50
Trasplantes (que requieran de un tratamiento inmunosupresor)	74

Tabla 6: Factores de riesgo para la TB [9]

## 6.8. RR debido a la etnia

En adición a los factores de riesgo estudiados en el apartado anterior, destaca el RR debido a la etnia del individuo.

En este caso, se denota el  $RR$  por  $e$  y toma los siguientes valores [9]:

- Si el individuo no ha nacido en Estados Unidos o Canadá:  $e = 1$
- Si el individuo ha nacido en Estados Unidos:
  - Si es caucásico:  $e = 1$
  - Si es afroamericano o negro y no latino:  $e = 3,09$
  - Si es latino:  $e = 3,66$
  - Otro caso no considerado previamente:  $e = 1$
- Si el individuo ha nacido en Canadá:
  - Si es aborigen:  $e = 1,55$
  - Si no lo es:  $e = 1$

## 7. Diagnósis del COVID-19

### 7.1. Pequeña introducción al COVID-19

El COVID-19 es una enfermedad infecciosa causada por el virus *Coronavirus SARS-CoV-2*.

La OMS tuvo noticia por primera vez de este nuevo virus en diciembre del 2019, siendo en marzo de 2020 cuando dicha enfermedad fue declarada una pandemia (enfermedad epidémica que se extiende a muchos países o que ataca a casi todos los individuos de una localidad o región).

Se dispone de las siguientes pruebas para detectar esta enfermedad:

- **Pruebas de detección de ácidos nucleicos (PCR):** sirven para detectar la presencia del virus; sin embargo, no pueden detectar si ya se ha superado la enfermedad. Se trata de la prueba con mayor sensibilidad y especificidad (gold standard).
- **Pruebas de detección de antígeno (Ag):** su utilidad es idéntica a la de la PCR. Debe realizarse durante los 7 primeros días desde la infección, momento en el que la carga viral es mayor.
- **Pruebas de detección de anticuerpos (Ac):** IgM/A e IgG. Esta prueba no detecta la presencia del virus en sí, sino los anticuerpos que se hayan generado en el organismo como respuesta a la enfermedad.

Las características de estas pruebas son las siguientes:

	RT-PCR (exudado nasofaríngeo /orofaríngeo)	RT-PCR (saliva)	Test antigénicos rápidos de última generación (exudado nasofaríngeo)
<b>Sensibilidad</b>	85-90% ( <i>Gold standard</i> )	Muy variable (5-91%)	<ul style="list-style-type: none"> <li>• Sintomáticos: &gt;95%</li> <li>• Asintomáticos: escasa evidencia</li> </ul>
<b>Especificidad</b>	99,5% ( <i>Gold standard</i> )	Similar al <i>gold standard</i>	95-99%

Figura 3: Sensibilidad y especificidad pruebas COVID [11]

## 7.2. Datos empleados

Los datos utilizados para todos los cálculos que siguen proceden de un estudio de cohortes retrospectivo de todos los contactos cercanos de casos confirmados de COVID-19 en Singapur, entre el 23 de enero y el 3 de abril de 2020 [15].

## 7.3. Cálculo del valor predictivo positivo

Para calcular el valor predictivo positivo de la PCR (probabilidad de que un individuo con un resultado positivo en la PCR tenga COVID-19) se emplea la fórmula obtenida previamente:

$$PPV_{PCR} = \frac{\text{sensibilidad} \times \text{seroprevalencia}}{\text{sensibilidad} \times \text{seroprevalencia} + FP} \quad (32)$$

Nótese que, al ser menores la sensibilidad y la especificidad que en el caso de la TST, se multiplica por la sensibilidad (no se aproxima a 1).

### 7.3.1. Cálculo de la seroprevalencia

En este caso, para estimar la extensión de la infección se realizará un estudio sero-epidemiológico. Es decir, se calculará el porcentaje de personas que presenten anticuerpos IgG anti SARS-CoV-2 (seroprevalencia); o lo que es lo mismo, el número de personas que hayan estado previamente infectadas con SARS-CoV-2.

$$SP = \frac{\text{Total de personas con anticuerpos}}{\text{Total de individuos}} \quad (33)$$

De esta forma se tendrán en cuenta individuos asintomáticos o individuos con síntomas leves que no hayan acudido a centros médicos, no se hayan realizado ninguna prueba y, por lo tanto, no hayan sido contabilizados como nuevos casos.

### 7.3.2. Cálculo de los falsos positivos

En el caso de la PCR y las Ag, las posibles causas por las que se pueden obtener falsos positivos son: que se produzcan contaminaciones cruzadas entre muestras o que haya un

error de etiquetaje; mientras que en el caso de las Ac, los falsos positivos se producen debido a la reactividad cruzada con otros virus respiratorios.

Para estimar el rango de las tasas de falsos positivos en las PCR se ha procedido de forma análoga al caso de los falsos positivos en la TST debidos a NTM. Para ello, se han tomado datos procedentes de numerosos Controles de Calidad Externos (CCE) de varios laboratorios que llevaron a cabo ensayos clínicos con pruebas PCR para detectar distintos virus ARN (es decir, pruebas similares a la PCR para el COVID-19). Para cada conjunto de datos se ha calculado la tasa de falsos positivos

$$FPR = \frac{FP}{FP + VN} \quad (34)$$

y, en el caso de que se obtuviese un rango, se ha tomado el límite inferior de este. Por último, se ha calculado la mediana y el rango intercuartílico del conjunto de valores obtenidos.

Con todo ello, se ha obtenido el siguiente rango de FPR (0 %, 16,7 %) y el siguiente rango intercuartílico (0,8 %, 4 %) [12].

Nótese que los falsos positivos aumentan cuando la prevalencia es baja. Esto se debe a que cuando la prevalencia disminuye, un resultado negativo permitirá descartar la enfermedad con mayor seguridad; mientras que un resultado positivo no permitirá confirmar el diagnóstico (bajo valor predictivo positivo) [13].

#### 7.4. Tasa de ataque secundario

A diferencia de la tuberculosis, el COVID-19 actualmente es una pandemia, por lo que se necesita calcular la tasa de ataque secundario (TAS).

La tasa de ataque secundario expresa el número de casos de una enfermedad que aparecen dentro del periodo de incubación entre los contactos susceptibles. Varía según las condiciones en las cuales se produzca la transmisión: el tipo y la duración de la exposición, las relaciones entre la persona índice (caso inicial) y los contactos, la ausencia de medidas de prevención, etc. [14].

Nótese que dicha medida se emplea para situaciones en las que la población entera de interés ha sido expuesta a un factor en un período muy corto y es de gran utilidad, pues permite determinar medidas de prevención más concretas.

De acuerdo a la OMS, esta tasa se obtiene a partir de los resultados de las PCR de los contactos cercanos estratificados en contactos convivientes, laborales y sociales. A continuación se especifica qué individuos pertenecen a cada grupo:

- **Contactos convivientes:** individuos que compartieron la misma dirección de residencia que la persona índice, independientemente de la duración y la proximidad.
- **Contactos laborales:** individuos que tuvieron contacto cercano (durante al menos 30 minutos a menos de 2 metros) con la persona índice durante el trabajo, desde 2 días antes al inicio de los síntomas hasta el aislamiento de la persona índice.

- **Contactos sociales:** individuos que tuvieron contacto cercano (durante al menos 30 minutos a menos de 2 metros) con la persona índice durante actividades sociales, desde 2 días antes al inicio de los síntomas hasta el aislamiento de la persona índice.

Para calcular las variables relacionadas con la TAS se ha llevado a cabo un análisis de regresión logística multivariante, en el que se ha relacionado la tasa de ataque secundario con la edad, el sexo, la duración de los síntomas y el tipo de contacto. Se han obtenido los siguientes resultados:

	TAS	aOR	$IC_{95\%}$	p
Hombre	0.025	1.24	(0,83, 1,58)	0.30
Mujer	0.023	1 (REF)		
< 30 años	0.015	0.52	(0,35, 0,78)	<b>0.0014</b>
Entre 30 y 49 años	0.029	1.11	(0,78, 1,60)	0.55
> 50 años	0.031	1 (REF)		
Duración de los síntomas		1	(0,95, 1,06)	0.88
Contacto social	0.059	0.34	(0,17, 0,65)	<b>0.0015</b>
Contacto laboral	0.013	0.09	(0,03, 0,25)	<b>&lt; 0.0001</b>
Contacto conviviente	0.013	1 (REF)		

Tabla 7: Factores de riesgo para la TAS [15]

Nótese que un p-valor  $< 0,05$  indica una diferencia estadísticamente significativa entre los grupos. Recuérdese también que  $aOR = 1$  indica ausencia de asociación entre las variables.

Se puede observar como los contactos sociales y laborales tienen un menor riesgo que los convivientes. Análogamente, las personas menores de 30 años presentan menor riesgo que los mayores de 50 años.

También se ha estudiado la relación entre la TAS y las siguientes interacciones: ser hombre con el tipo de contacto y la duración de los síntomas con el tipo de contacto. Se han obtenido los siguientes resultados:

	aOR	$IC_{95\%}$	p
Hombre x Contacto laboral	2.93	(1,15, 8,51)	<b>0.032</b>
Hombre x Contacto social	0.76	(0,03, 0,25)	0.46
Hombre x Contacto conviviente	1		
Duración síntomas x Contacto laboral	1 (REF)	(0,88, 1,13)	0.99
Duración síntomas x Contacto social	0.94	(0,84, 1,04)	0.24
Duración síntomas x Contacto conviviente	1 (REF)		

Tabla 8: Factores de riesgo para la TAS (interacciones) [15]

Se observa que la única interacción significativa es la de hombre y contacto laboral, siendo el riesgo mayor que para los hombres que sean contacto conviviente.

Por último, se ha estudiado la relación entre la TAS y las siguientes interacciones:

	Odds de TAS entre	aOR	IC <sub>95%</sub>	p
Hombres	C. laboral vs c. conviviente	0.26	(0,11, 0,59)	<b>0.0014</b>
	C.social vs c. conviviente	0.26	(0,13, 0,53)	<b>0.0002</b>
Mujeres	C. laboral vs c.conviviente	0.09	(0,03, 0,26)	<b>&lt;0.0001</b>
	C. social vs c. conviviente	0.34	(0,17, 0,66)	<b>0.0015</b>
Contactos convivientes	Hombre vs mujer	1.24	(0,83, 1,85)	0.29
Contactos sociales	Hombre vs mujer	0.94	(0,52, 1,72)	0.86
Contactos laborales	Hombre vs mujer	3.63	(1,48, 8,94)	<b>0.0049</b>

Tabla 9: Odds de TAS [15]

Nótese que los OR no solo se emplean para calcular el cociente entre los Odds de los expuestos y los de los no expuestos; sino que también permiten calcular el cociente entre los Odds de los expuestos a un factor y los de los expuestos a un factor distinto.

Se puede observar que los contactos sociales y laborales tienen una menor tasa de ataque que los contactos convivientes, tanto en hombres como en mujeres. También se puede advertir que dentro de los contactos laborales, los hombres tienen mayor riesgo de infección.

## 7.5. Análisis de conglomerados

Un conglomerado (o cluster, en inglés) es la aparición de un número de casos de una enfermedad en una cantidad superior a la que se esperaría para un determinado grupo de población, un área geográfica y un período de tiempo determinado. El incremento de casos puede ser percibido y/o real.

La importancia de los conglomerados reside en que permiten la detección temprana de brotes (aumento inusual en el número de casos relacionados epidemiológicamente, de aparición súbita y diseminación localizada en un espacio específico).

Se ha estudiado si hay conglomerados en la tasa de ataque secundario dentro de los distintos tipos de contactos. Para ello, se ha comparado un modelo binomial con uno beta-binomial mediante el uso del estadístico G de razón de verosimilitud.

Para poder entender el análisis llevado a cabo, se necesitan conocer previamente las diferencias entre una distribución binomial y una beta-binomial.

Sea

$$X_{i,j} = \begin{cases} 1 & \text{si el paciente } j - \text{ésimo del grupo } i - \text{ésimo está enfermo} \\ 0 & \text{si el paciente } j - \text{ésimo del grupo } i - \text{ésimo no está enfermo} \end{cases} \quad (35)$$



con  $i \in \{\text{contactos convivientes}, \text{contactos laborales}, \text{contactos sociales}\}$  y  $j \in \{1, 2, \dots, n_i\}$ ; siendo  $n_i$  el número de individuos en el grupo  $i$ -ésimo y  $p_i$  la probabilidad de enfermarse en dicho grupo.

Sea  $N_i$  una variable aleatoria que representa el número de individuos enfermos dentro del grupo  $i$ ,

$$N_i = \sum_{j=1}^{n_i} X_{i,j} \quad (36)$$

existen dos opciones:

1. Que el estatus de una persona enferma no esté relacionado con el estatus de sus vecinos. Luego la probabilidad de enfermarse es constante y las variables  $X_{i,j}$  son independientes. La variable  $N_i$  sigue una distribución binomial.

$$N_i \sim \text{Bin}(n_i, p_i) \quad (37)$$

2. Que el estatus de una persona enferma sí esté relacionado con el estatus de sus vecinos. Luego la probabilidad de enfermarse es una variable aleatoria con distribución beta y las variables  $X_{i,j}$  están correlacionadas. La variable  $N_i$  sigue una distribución beta-binomial.

$$N_i | p_i \sim \text{Bin}(n_i, p_i) \quad \text{y} \quad p_i \sim \text{Beta}(a_i, b_i) \quad (38)$$

o lo que es lo mismo

$$N_i \sim \text{BBin}(n_i, a_i, b_i) \quad (39)$$

En este caso, existe un conglomerado. Es decir, que  $N_i$  siga una distribución beta-binomial implica que existe sobredispersión en los datos (una minoría de individuos positivos son responsables de un porcentaje inesperadamente alto de transmisión).

Nótese que se trata de dos modelos anidados.

Para comprobar si hay conglomerados se plantea el siguiente contraste de hipótesis:

$$\begin{cases} H_0 : N_i \sim \text{Bin}(n_i, p_i) \\ H_1 : N_i \sim \text{BBin}(n_i, a_i, b_i) \end{cases} \quad (40)$$

o lo que es lo mismo

$$\begin{cases} H_0 : \text{No hay cluster} \\ H_1 : \text{Hay cluster} \end{cases} \quad (41)$$

Para determinar si se rechaza o no la hipótesis nula se emplea el estadístico G de máxima verosimilitud o likelihood ratio test (LR). Este estadístico compara la bondad de ajuste de dos modelos anidados de la siguiente forma:

$$LR = -2 \times (L_B - L_{BB}) \quad (42)$$

siendo  $L_B$  la log-verosimilitud, es decir, el logaritmo del máximo de la función de verosimilitud de la distribución binomial; y  $L_{BB}$  el de la distribución beta-binomial.

El teorema de Wilks indica que, asumiendo que la hipótesis nula es cierta, cuando el tamaño de la muestra tiende a  $\infty$ , el LR se distribuye como una chi-cuadrado con número de grados de libertad igual a la diferencia de parámetros entre el espacio de parámetros completo y el subconjunto del espacio de parámetros asociado con  $H_0$ .

Por tanto, en este caso,

$$LR \sim X_1^2 \quad (43)$$

Siguiendo este proceso se ha obtenido evidencia de conglomerados en la tasa de ataque secundario de los contactos laborales y convivientes (con un p-valor  $< 0,0001$  para ambos), pero no en el caso de los contactos sociales ( $p = 0,40$ ).

Además, se ha obtenido una inflación en la desviación típica de la TAS del 15 % en el caso de hogares con tres individuos y del 28 % en hogares con cuatro individuos; respecto de la desviación típica esperada en el caso de que las infecciones fuesen estadísticamente independientes [15].

Recuérdese que las fórmulas para obtener la desviación típica son:

- En el caso binomial:

$$\sigma_B = \sqrt{n_i \times p_i \times (1 - p_i)} \quad (44)$$

- En el caso beta-binomial:

$$\sigma_{BB} = \sqrt{\frac{n_i \times a_i \times b_i \times (a_i + b_i + n_i)}{(a_i + b_i)^2 \times (a_i + b_i + 1)}} \quad (45)$$

## 7.6. Factores de riesgo y riesgo relativo

Para determinar cuáles son los factores de riesgo que contribuyen a la infección del SARS-CoV-2 se ha realizado un estudio de casos y controles, donde los casos son aquellos contactos con una PCR o una serología positiva y los controles son aquellos sujetos que terminaron la cuarentena impuesta a todos los contactos sin ser diagnosticados con COVID-19 y con una serología negativa.

Sobre los datos de estos individuos se ha llevado a cabo un análisis de regresión logística univariada y multivariante, obteniéndose los siguientes resultados:

- Para los contactos convivientes

	Análisis univariado			Análisis multivariado		
	OR	$IC_{95\%}$	p	aOR	$IC_{95\%}$	p
$\leq 35$ años	1 (REF)			1 (REF)		
$> 35$ años	1.30	(0,82, 2,06)	0.26	0.99	(0,54, 1,83)	0.98
Mujer	1 (REF)			1 (REF)		
Hombre	1.15	(0,73, 1,80)	0.56	1.26	(0,73, 2,18)	0.41
Contacto con 1 caso	1 (REF)			1 (REF)		
Contacto con $> 1$ caso	1.89	(1,11, 3,22)	<b>0.019</b>	1.65	(0,86, 3,19)	0.14
Sin vínculo familiar con el caso	1 (REF)			1 (REF)		
Familiar del caso	2.17	(0,83, 5,67)	0.12	1.52	(0,53, 4,32)	0.44
Pareja del caso	9.20	(3,5, 24,17)	<b>&lt; 0.0001</b>	1.63	(0,45, 5,93)	0.46
No tocó algo que el caso acababa de tocar	1 (REF)			1 (REF)		
Tocó algo que el caso acababa de tocar	4.34	(2,25, 8,37)	<b>&lt;0.0001</b>	1.67	(0,77, 3,64)	0.20
No comió con el caso	1 (REF)			1 (REF)		
Comió con el caso pero sin compartir utensilios	2.47	(1,30, 4,72)	<b>0.0060</b>	1.03	(0,48, 2,21)	0.93
Comió con el caso compartiendo utensilios	4.9	(2,69, 8,9)	<b>&lt;0.0001</b>	1.29	(0,60, 2,80)	0.52
No compartió habitación ni baño con el caso	1 (REF)			1 (REF)		
Compartió baño con el caso	1.78	(0,84, 3,79)	0.14	1.11	(0,49, 2,54)	0.8
Compartió habitación con el caso	7.23	(3,23, 16,18)	<b>&lt;0.0001</b>	5.38	(1,82, 15,84)	<b>0.0023</b>
Compartió habitación y baño con el caso	10.62	(5,84, 19,33)	<b>&lt;0.0001</b>	5.05	(1,85, 13,79)	<b>0.0016</b>
No compartió vehículo con el caso	1 (REF)			1 (REF)		
Compartió vehículo con el caso	2.38	(1,48, 3,81)	<b>0.00030</b>	0.84	(0,46, 1,52)	0.56
No habló con el caso	1 (REF)			1 (REF)		
Habló con el caso $<30$ min	4.07	(2,26, 7,32)	<b>&lt;0.0001</b>	3.91	(2,09, 7,34)	<b>&lt;0.0001</b>
Habló con el caso $\geq 30$ min	14.19	(7,55, 26,64)	<b>&lt;0.0001</b>	7.86	(3,86, 16,02)	<b>&lt;0.0001</b>

Tabla 10: Factores de riesgo entre convivientes [15]

De acuerdo a la tabla, se puede concluir que los factores de riesgo son compartir una habitación con el paciente índice y hablar con el mismo, siendo mayor el riesgo si este habla durante más de 30 minutos.

- Para los no convivientes (contactos sociales y laborales)

	Análisis univariado			Análisis multivariado		
	OR	IC <sub>95%</sub>	p	aOR	IC <sub>95%</sub>	p
≤ 35 años	1 (REF)			1 (REF)		
> 35 años	0.89	(0,51, 1,57)	0.69	0.69	(0,37, 1,29)	0.25
Mujer	1 (REF)			1 (REF)		
Hombre	1.21	(0,69, 2,13)	0.50	1.52	(0,82, 2,83)	0.19
Contacto con 1 caso	1 (REF)			1 (REF)		
Contacto con > 1 caso	4.50	(2,51, 8,06)	< <b>0.0001</b>	3.92	(2,07, 7,40)	< <b>0.0001</b>
No tuvo contacto físico directo con un caso	1 (REF)			1 (REF)		
Tuvo contacto físico directo con un caso	1.66	(0,94, 2,94)	0.083	1.10	(0,55, 2,19)	0.79
No tocó algo que el caso acababa de tocar	1 (REF)			1 (REF)		
Tocó algo que el caso acababa de tocar	2.27	(1,27, 4,05)	<b>0.0054</b>	1.24	(0,62, 2,46)	0.55
No comió con el caso	1 (REF)			1 (REF)		
Comió con el caso pero sin compartir utensilios	1.83	(0,87, 3,87)	0.11	1.04	(0,44, 2,46)	0.92
Comió con el caso compartiendo utensilios	2.86	(1,5, 5,46)	<b>0.0015</b>	1.45	(0,63, 3,31)	0.38
No compartió baño con el caso	1 (REF)			1 (REF)		
Compartió baño con el caso	2.03	(1,06, 3,88)	<b>0.033</b>	1.03	(0,48, 2,18)	0.95
No compartió vehículo con el caso	1 (REF)			1 (REF)		
Compartió vehículo con el caso	3.94	(2,21, 7,04)	< <b>0.0001</b>	3.07	(1,55, 6,08)	<b>0.0013</b>
No habló con el caso	1 (REF)			1 (REF)		
Habló con el caso <30 min	2.10	(1,01, 4,34)	<b>0.047</b>	2.50	(1,15, 5,44)	<b>0.021</b>
Habló con el caso ≥30 min	3.39	(1,65, 6,97)	<b>0.0009</b>	2.67	(1,21, 5,88)	<b>0.015</b>
El caso no llevó mascarilla durante el contacto	1 (REF)			1 (REF)		
El caso llevó mascarilla durante el contacto	1.55	(0,70, 3,43)	0.28	No se realizó		
No llevó mascarilla durante el contacto con el caso	1 (REF)			1 (REF)		
Llevó mascarilla durante el contacto con el caso	1.53	(0,62, 3,77)	0.35	No se realizó		

Tabla 11: Factores de riesgo entre no convivientes [15]

De acuerdo a la tabla, se puede concluir que los factores de riesgo son tener contacto con más de un caso de COVID-19, compartir vehículo con algún caso de COVID-19 y hablar con algún caso COVID-19, siendo ligeramente mayor el riesgo si este habla durante más de 30 minutos.

Nótese que para todos los factores de riesgo se ha comprobado la significación estadística a través del estadístico de Wald con un nivel de significación del 5%.

Dadas las hipótesis:

$$\begin{aligned} H_0 : \beta_j &= 0 \\ H_1 : \beta_j &\neq 0 \end{aligned} \tag{46}$$

o lo que es lo mismo,

$H_0$ : la variable independiente j-ésima no está relacionada con la variable dependiente

$H_1$ : la variable independiente j-ésima está relacionada con la variable dependiente

el estadístico de Wald está definido por

$$w = \sqrt{\frac{\beta_j^2}{Var(\beta_j)}} \tag{47}$$

y se distribuye como una normal

$$w \sim N(0, 1) \tag{48}$$

También se puede emplear la prueba de la chi-cuadrado de Wald, que es el estadístico de Wald al cuadrado y que se distribuye como una chi-cuadrado con un grado de libertad.

$$w^2 \sim X_1^2 \tag{49}$$

Por último, se calcula el intervalo de confianza de Wald:

$$IC_{95\%} = (\beta_j - w \times se_{\beta_j}, \beta_j + w \times se_{\beta_j}) \tag{50}$$

siendo  $se_{\beta_j}$  el error estándar.

## 7.7. Análisis de la participación en el estudio serológico

Se ha evaluado si existen diferencias entre aquellos individuos que aceptaron participar en el estudio serológico y contestar a un cuestionario; y aquellos que no.

Para ello se ha llevado a cabo un análisis de regresión logística para cada tipo de contacto (social, laboral y conviviente), relacionando su participación en dicho estudio con su edad, con su sexo y con el hecho de haberse realizado una PCR previamente.

En el caso de contactos que hayan compartido vivienda con la persona índice, haberse sometido a una PCR previamente no está estadísticamente relacionado con aceptar participar en el estudio serológico ( $OR = 0,80$ ,  $p = 0,067$ ). Tampoco lo están la edad ( $p = 0,18$ ) ni el sexo ( $p = 0,95$ ) del contacto.

Entre los contactos sociales, aquellos de mayor edad presentan una mayor predisposición a someterse a un estudio serológico ( $OR = 1,13$  por cada década,  $IC_{95\%} = (1,08, 1,19)$ ,  $p < 0,0001$ ). Sin embargo, ni el sexo ( $p = 0,55$ ) ni haberse realizado una PCR previamente ( $p = 0,33$ ) son significativos.

Por último, entre los contactos laborales, ni el sexo ( $p = 0,29$ ), ni la edad ( $p = 0,86$ ), ni la realización previa de una PCR ( $p = 0,14$ ) son estadísticamente significativos [15].

## 7.8. Cálculo del riesgo de infección y proporción de asintomáticos

Para estimar el riesgo total de infección y la proporción total de individuos asintomáticos se ha ajustado un modelo empleando la inferencia bayesiana. Para ello se ha tenido en cuenta si los individuos se han realizado pruebas PCR y/o serológicas, las sensibilidades de dichas pruebas y la probabilidad de que un individuo desarrolle síntomas en caso de estar, o no, infectado.

Este modelo permite conocer el número total de infecciones entre contactos (sociales, laborales, convivientes y en conjunto) confirmadas por una PCR y/o una serología. Además, permite acomodar a aquellos contactos que no dieron su consentimiento para realizar el estudio serológico; así como el riesgo de los falsos negativos en las pruebas diagnósticas.

Se han dividido los contactos en nueve grupos, y 16 subgrupos, atendiendo a si estaban infectados o no, la presencia de síntomas y los resultados de sus pruebas diagnósticas en caso de haberse realizado:

A	A	Contacto era sintomático. Se le hizo una PCR y resultó positiva
BL		Contacto era sintomático. Se hizo una PCR y resultó negativa. No se hizo una serología. Estaba:
	B	Verdaderamente infectado (falso negativo en la PCR)
	L	No infectado
C	C	Contacto era sintomático. Se hizo una PCR y resultó negativa. Se hizo una serología y resultó positiva.
DK		Contacto era sintomático. Se hizo una PCR y resultó negativa. Se hizo una serología y resultó negativa. Estaba:
	D	Verdaderamente infectado (falso negativo en la PCR y en la serología)
	K	No infectado
E	E	Contacto era sintomático. No se hizo una PCR. Se hizo una serología y resultó positiva.
FM		Contacto era sintomático. No se hizo una PCR. Se hizo una serología y resultó negativa. Estaba:
	F	Verdaderamente infectado (falso negativo en la serología)
	M	No infectado
GJNP		No se hizo ni una PCR ni una serología. Estaba:
	G	Verdaderamente infectado. Sintomático.
	J	Verdaderamente infectado. Asintomático
	N	No infectado. Con síntomas causados por otro patógeno
	P	No infectado. Asintomático.
H	H	Contacto era asintomático. Se hizo una serología y resultó positiva.
IO		Contacto era asintomático. Se hizo una serología y resultó negativa. Estaba:
	I	Verdaderamente infectado (falso negativo en la serología)
	O	No infectado

Tabla 12: Clasificación de los contactos en grupos [15]

Para poder determinar la probabilidad de pertenecer a cada uno de los grupos, se han definido los siguientes parámetros:

$$\pi = P(\text{Estar infectado}) \quad (51)$$

$$s = P(\text{Tener síntomas} | \text{Estar infectado}) \quad (52)$$

$$y = P(\text{Tener síntomas} | \text{No estar infectado}) \quad (53)$$

$$p = P(\text{Haberse hecho una PCR} | \text{Tener síntomas}) \quad (54)$$

$$\nu = P(\text{Tener una PCR positiva} | \text{Haberse hecho una prueba y estar infectado}) \quad (55)$$

$$\sigma = P(\text{Tener una serología positiva} | \text{Haberse hecho una prueba y estar infectado}) \quad (56)$$

$$\tau = P(\text{Haberse hecho una serología} | \text{Tener una PCR negativa}) \quad (57)$$

Observando el siguiente árbol:

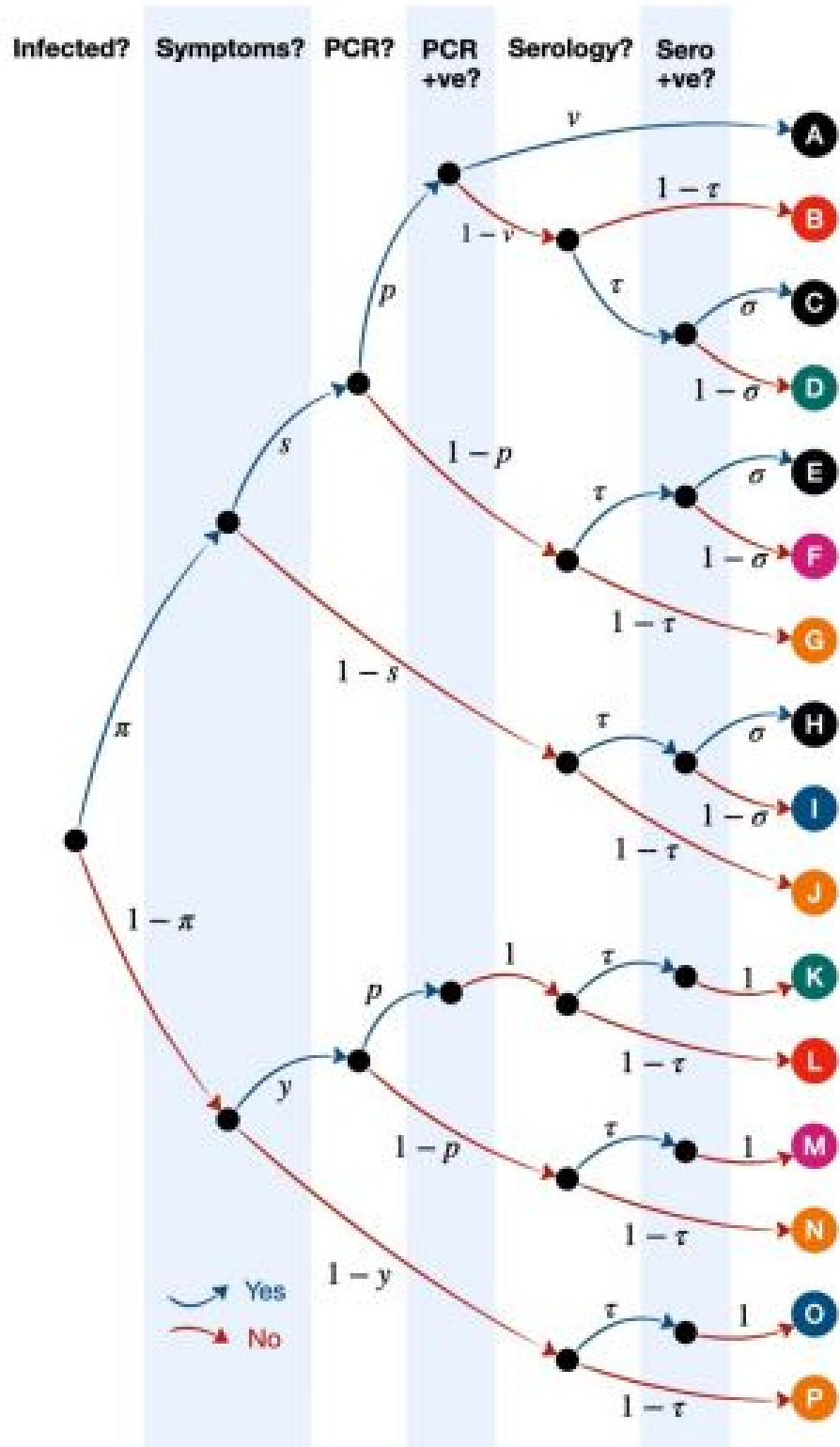


Figura 4: Árbol de probabilidades [15]



se puede ver que la probabilidad de pertenecer a cada subgrupo viene dada por:

Grupo	Subgrupo	Probabilidad del subgrupo	Nº convivientes	Nº sociales	Nº laborales	Total
A	A	$\pi sp\nu$	110	47	31	188
BL	B	$\pi sp(1-\nu)(1-\tau)$	273	374	278	925
BL	L	$(1-\pi)yp(1-\tau)$	-	-	-	-
C	C	$\pi sp(1-\nu)\tau\sigma$	5	2	2	9
DK	D	$\pi sp(1-\nu)\tau(1-\sigma)$	106	58	35	199
DK	K	$(1-\pi)yp\tau$	-	-	-	-
E	E	$\pi s(1-p)\tau\sigma$	7	2	1	10
FM	F	$\pi s(1-p)\tau(1-\sigma)$	20	20	15	55
FM	M	$(1-\pi)y(1-p)\tau$	-	-	-	-
GJNP	G	$\pi s(1-p)(1-\tau)$	956	2748	1803	5507
GJNP	J	$\pi(1-s)(1-\tau)$	-	-	-	-
GJNP	N	$(1-\pi)y(1-p)(1-\tau)$	-	-	-	-
GJNP	P	$(1-\pi)(1-y)(1-\tau)$	-	-	-	-
H	H	$\pi(1-s)\tau\sigma$	17	5	3	25
IO	I	$\pi(1-s)\tau(1-\sigma)$	369	332	151	852
IO	O	$(1-\pi)(1-y)\tau$	-	-	-	-

Tabla 13: Probabilidades de pertenecer a cada subgrupo [15]

Se han empleado los datos de la tabla anterior para ajustar un modelo para cada grupo de contactos (convivientes, laborales y sociales); así como para la combinación de los tres.

Para calcular las distribuciones a posteriori de los parámetros  $(\pi, s, y, p, \nu, \sigma, \tau)$  se empleará el teorema de Bayes.

Sea  $\theta$  un parámetro cualquiera y  $x_1, \dots, x_n$  los datos de los que se dispone,

$$P(\theta|x_1, \dots, x_n) = \frac{P(x_1, \dots, x_n|\theta) \times P(\theta)}{P(x_1, \dots, x_n)} \quad (58)$$

siendo:

- $P(\theta|x_1, \dots, x_n)$  la probabilidad a posteriori de  $\theta$
- $P(x_1, \dots, x_n|\theta) = L(x_1, \dots, x_n; \theta)$  la verosimilitud de  $\theta$  aportada por los datos
- $P(\theta)$  la probabilidad a priori de  $\theta$
- $P(x_1, \dots, x_n)$  la distribución de los datos

Al ser los datos observados fijos,  $P(x_1, \dots, x_n)$  es una constante, por lo que

$$P(\theta|x_1, \dots, x_n) \propto L(x_1, \dots, x_n; \theta) \times P(\theta) \quad (59)$$

Para todos los parámetros se ha asumido una distribución a priori uniforme ( $U(0, 1)$ ), salvo para  $\sigma$  (sensibilidad de la serología), para la que se ha asumido una distribución beta: de 175 individuos infectados, 173 obtuvieron un resultado de la prueba positivo, luego  $\sigma \sim Be(174, 3)$ .

Como no todos los subgrupos quedan perfectamente determinados a partir de los datos (tal y como se ve en la tabla anterior), se han unido algunos de ellos sumando sus probabilidades. De acuerdo a esto, los datos conocidos son:

$$\{n_A, n_{BL}, n_C, n_{DK}, n_E, n_{FM}, n_{GJNP}, n_H, n_{IO}\} \quad (60)$$

Sea  $X_i$  una variable aleatoria que indique el número de personas que pertenecen al grupo  $i = A, BL, C, DK, E, FM, GJNP, H, IO$ , entonces:

$$X = (X_A, X_{BL}, X_C, X_{DK}, X_E, X_{FM}, X_{GJNP}, X_H, X_{IO}) \sim Multinomial(n, p) \quad (61)$$

siendo  $n = n_A + n_{BL} + n_C + n_{DK} + n_E + n_{FM} + n_{GJNP} + n_H + n_{IO}$  y  $p$  el vector con las probabilidades de pertenecer a cada uno de los grupos (indicadas en la tabla 13).

Recuérdese que este modelo se ajustará cuatro veces, luego  $n$  corresponderá al número total de contactos convivientes, al número total de contactos sociales, al número total de contactos laborales y al conjunto de todos ellos, respectivamente.

Con todo ello, se tiene:

$$P(\pi, \dots, \tau | n_A, \dots, n_{IO}) \propto L(n_A, \dots, n_{IO} | \pi, \dots, \tau) \times P(\pi, \dots, \tau) \quad (62)$$

Para calcular las distribuciones a posteriori se ha empleado el método MCMC (Markov Chain Monte Carlo) con 100 000 iteraciones.

Este método consiste en simular una cadena de Markov  $(\theta_1, \dots, \theta_{100000})$  cuya distribución estacionaria sea  $P(\theta | x_1, \dots, x_n)$ , de manera que cada valor simulado  $\theta_t$  dependa únicamente de su predecesor  $\theta_{t-1}$ .

Las medias y los intervalos de confianza al 95 % obtenidos son [15]:

	Definición	C. convivientes	C. sociales	C. laborales	Total
$\pi$	Probabilidad de infección	0.11 (0,09, 0,14)	0.04 (0,03, 0,05)	0.05 (0,03, 0,08)	0.06 (0,05, 0,08)
$s$	Probabilidad de síntomas en caso de estar infectado	0.72 (0,62, 0,82)	0.65 (0,45, 0,84)	0.64 (0,40, 0,87)	0.64 (0,55, 0,73)
$p$	Probabilidad de hacerse una PCR en caso de tener síntomas	0.84 (0,79, 0,89)	0.71 (0,63, 0,79)	0.65 (0,54, 0,75)	0.75 (0,70, 0,79)
$\nu$	Probabilidad de tener una PCR positiva en caso de estar infectado	0.85 (0,74, 0,94)	0.75 (0,52, 0,93)	0.66 (0,41, 0,90)	0.79 (0,68, 0,88)
$\tau$	Probabilidad de hacerse una serología	0.30 (0,28, 0,32)	0.12 (0,11, 0,13)	0.09 (0,08, 0,10)	0.15 (0,14, 0,16)
$\sigma$	Probabilidad de tener una serología positiva en caso de estar infectado	0.98 (0,96, 1,00)	0.98 (0,96, 1,00)	0.98 (0,96, 1,00)	0.98 (0,96, 1,00)
$y$	Probabilidad de tener síntomas en caso de no estar infectado	0.26 (0,24, 0,29)	0.17 (0,15, 0,19)	0.21 (0,18, 0,25)	0.20 (0,18, 0,22)

Tabla 14: Valor de los parámetros a posteriori [15]

a partir de los cuales se han obtenido los siguientes resultados:

	Definición	C. convivientes	C. sociales	C. laborales	Total
$\pi(1 - sp\nu)$	Probabilidad de estar infectado y de que esa infección no sea detectada por una PCR	0.06 (0,04, 0,07)	0.03 (0,01, 0,04)	0.04 (0,02, 0,06)	0.04 (0,03, 0,05)
$1 - sp\nu$	Proporción de infecciones no detectadas por pruebas PCR	0.48 (0,39, 0,57)	0.66 (0,52, 0,77)	0.73 (0,58, 0,84)	0.62 (0,55, 0,69)
$(1 - s)/s$	Ratio asintomáticos/sintomáticos	0.39 (0,23, 0,62)	0.58 (0,20, 1,20)	0.61 (0,15, 1,47)	0.57 (0,36, 0,83)
$(1 - s)$	Proporción de asintomáticos	0.28 (0,18, 0,38)	0.35 (0,16, 0,55)	0.36 (0,13, 0,60)	0.36 (0,27, 0,45)

Tabla 15: Valores derivados de los parámetros a posteriori [15]

Se puede observar que entre los contactos convivientes la proporción de infecciones no detectadas es menor, así como la proporción de asintomáticos.

Para finalizar, se ha comprobando la convergencia del método MCMC mediante la representación de los siguiente trace plots (representando el valor que toma el parámetro en cada una de las iteraciones), donde (a) es para los contactos convivientes, (b) para los laborales, (c) para los sociales y (d) para el conjunto de los tres.

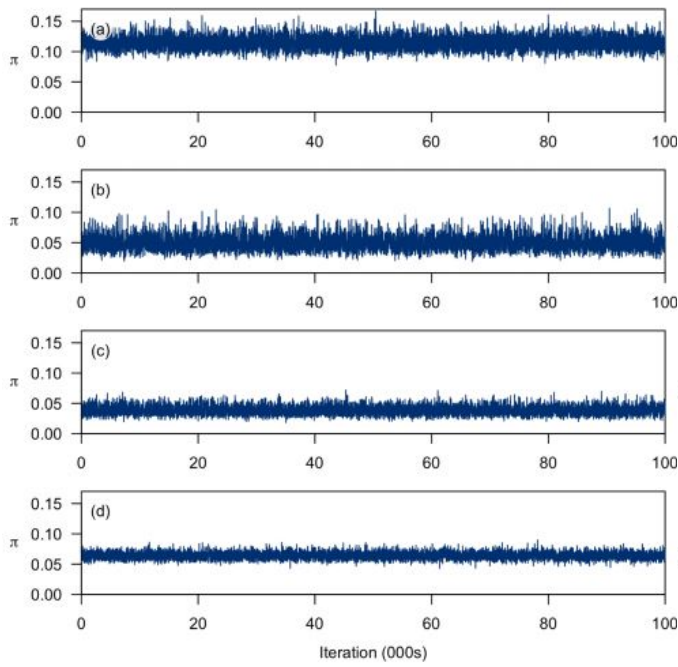


Figura 5: Trace plot para  $\pi$  [15]

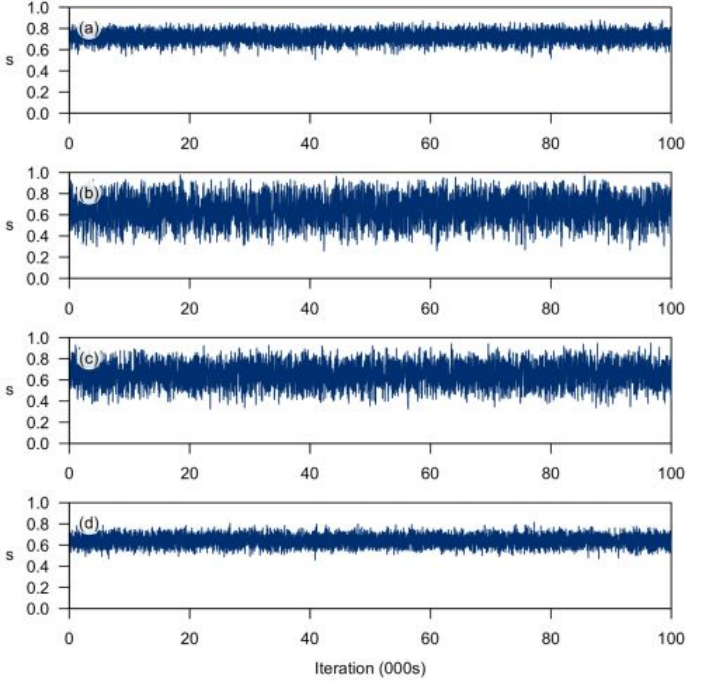


Figura 6: Trace plot para  $s$  [15]

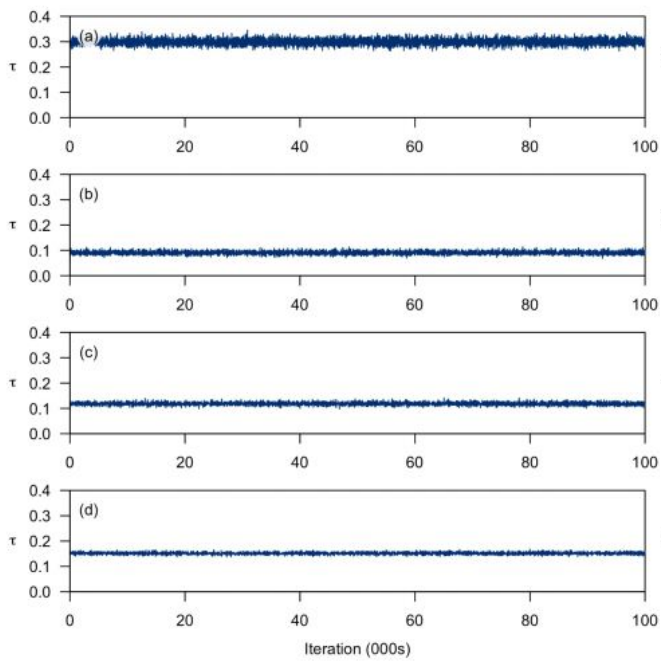


Figura 7: Trace plot para  $\tau$  [15]

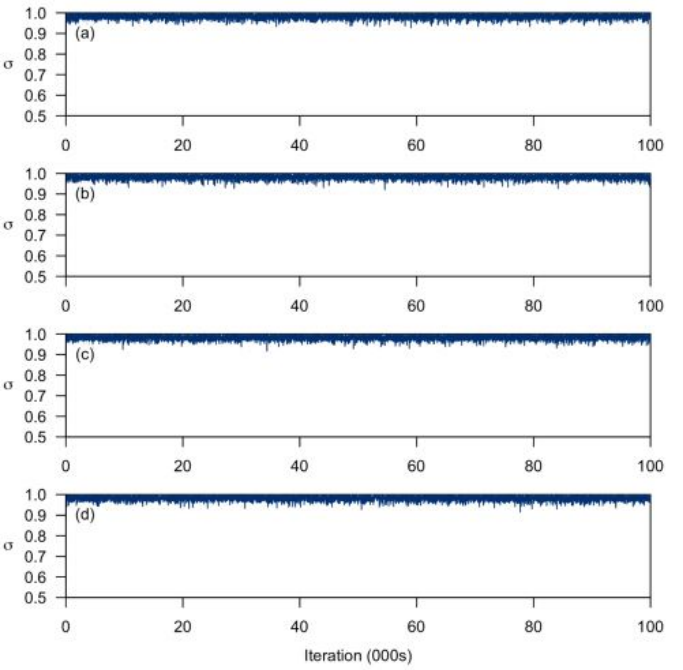


Figura 8: Trace plot para  $\sigma$  [15]

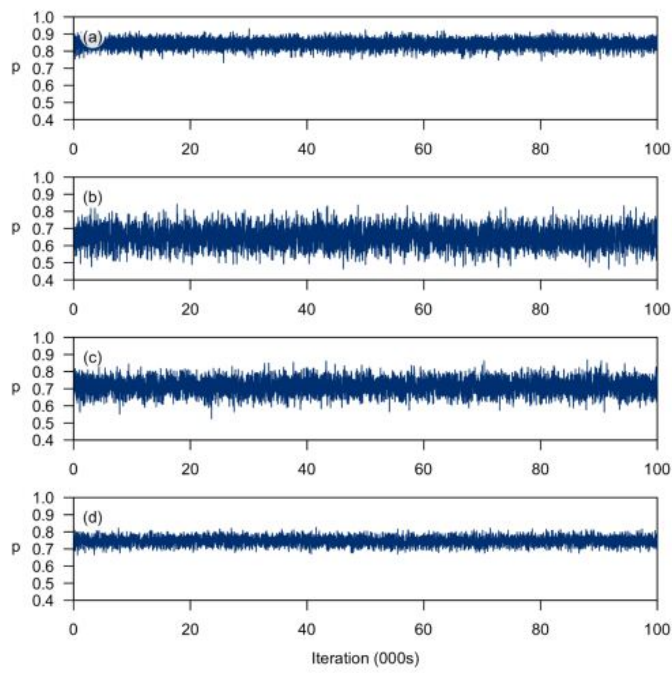


Figura 9: Trace plot para  $p$  [15]

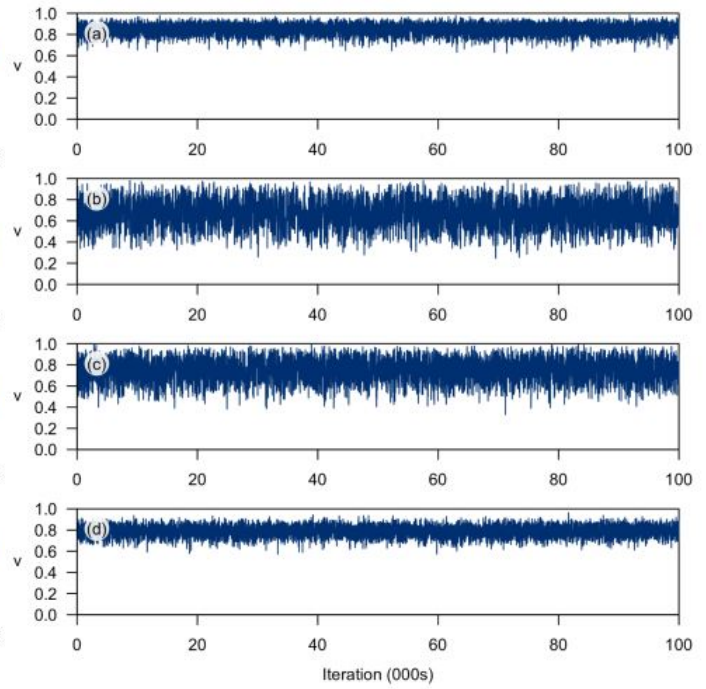


Figura 10: Trace plot para  $\nu$  [15]

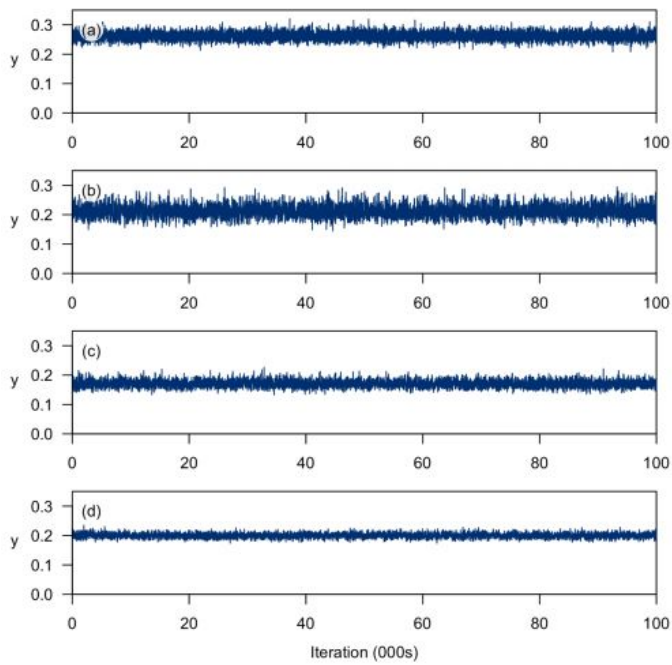


Figura 11: Trace plot para  $y$  [15]

Se puede observar que hay convergencia.

## Siglas

Ac	Pruebas de detección de anticuerpos.
Ag	Pruebas de detección de antígeno.
aOR	Odds Ratio Ajustado.
ARD	Incidencia anual acumulada.
ARI	Riesgo anual de infección.
BCG	Bacilo Calmette-Guérin.
CCE	Control de Calidad Externo.
CRD	Riesgo acumulativo de desarrollar una enfermedad.
DI	Densidad de incidencia.
FN	Falsos negativos.
FP	Falsos positivos.
FPR	Tasa de falsos positivos.
IA	Incidencia acumulada.
LR	Likelihood ratio test.
LTBI	Tuberculosis latente.
NE	Nivel endémico.
NTM	Micobacterias no tuberculosas.
OR	Odds Ratio.
PC	Punto de corte.
PCR	Pruebas de detección de ácidos nucleicos.
PD	Prueba diagnóstica.
pLTBI	Prevalencia de la tuberculosis latente.
PPV	Valor predictivo positivo.
RR	Riesgo relativo.
TAS	Tasa de ataque secundario.
TB	Tuberculosis activa.

TBC Tuberculosis bacilífera.  
TST Prueba cutánea de la tuberculina.

VN Verdaderos negativos.  
VP Verdaderos positivos.

## Referencias

- [1] J. F. A. Murphy, *A correct diagnosis is of increasing importance*, 323 This Month 324 IMJ Commentary, 324, 2016.
- [2] J. P. Kassirer, *Our stubborn quest for diagnostic certainty*, 1989.
- [3] V. Hernández, *Estudios epidemiológicos: tipos, diseño e interpretación*, Enfermedad Inflamatoria Intestinal al Día 16(3), 98-105, 2017.
- [4] J. H. Donis, *Evaluación de la validez y confiabilidad de una prueba diagnóstica*, Avances en biomedicina 1(2), 73-81, 2012.
- [5] J. J. Barendregt, *The effect size in uncertainty analysis*, Value in Health 13(4), 388-391, 2010.
- [6] A. A. Flores-Ibarra & M. D. Ochoa-Vázquez & G. A. S. Tec, *Estrategias diagnósticas aplicadas en la Clínica de Tuberculosis del Hospital General Centro Médico Nacional la Raza*, Revista Médica del Instituto Mexicano del Seguro Social 54(1), 122-127, 2016.
- [7] J. Nyboe, *Interpretation of tuberculosis infection age curves*, Bulletin of the World Health Organization 17(2), 319-339, 1957.
- [8] F. Van Leth & M. J. Van der Werf & M. W. Borgdorff, *Prevalence of tuberculous infection and incidence of tuberculosis: a re-assessment of the Styblo rule*, Bulletin of the World Health Organization 86, 20-26, 2008.
- [9] D. Menzies & G. Gardiner & M. Farhat & C. Greenaway & M. Pai, *Thinking in three dimensions: a web-based algorithm to aid the interpretation of tuberculin skin test results*, The International Journal of Tuberculosis and Lung Disease 12(5), 498-505, 2008.
- [10] M. Farhat & C. Greenway & M. Pai & D. Menzies, *False-positive tuberculin skin tests: what is the absolute effect of BCG and non-tuberculous mycobacteria?*, The International Journal of Tuberculosis and Lung Disease 10(11), 1192-1204, 2006.
- [11] L. S. Langa & L. V. Sallent & S. R. Díez, *Interpretación de las pruebas diagnósticas de la COVID-19*, FMC-Formación Médica Continuada en Atención Primaria 28(3), 167-173, 2021.
- [12] A. N. Cohen & B. Kessel, *False positives in reverse transcription PCR testing for SARS-CoV-2*, medRxiv, 2020.
- [13] B. Healy & A. Khan & H. Metezai & I. Blyth & H. Asad, *The impact of false positive COVID-19 results in an area of low prevalence*, Clinical Medicine 21(1), e54, 2021.
- [14] Ministerio de Sanidad España, *Información científica-técnica. Parámetros epidemiológicos*, 15 de enero de 2021, [https://www.mscbs.gob.es/profesionales/saludPublica/ccayes/alertasActual/nCov/documentos/Documento\\_EPIDEMIOLOGIA.pdf](https://www.mscbs.gob.es/profesionales/saludPublica/ccayes/alertasActual/nCov/documentos/Documento_EPIDEMIOLOGIA.pdf)



- [15] O.T. Ng & K. Marimuthu & V. Koh, J. Pang & K. Z. Linn & J. Sun & ... & V. J. Lee, *SARS-CoV-2 seroprevalence and transmission risk factors among high-risk close contacts: a retrospective cohort study*, *The Lancet infectious diseases* 21(3), 333-343, 2021.