

# Valoración de la prueba de ADN en las identificaciones a gran escala de personas desaparecidas

Autores: Prieto L.<sup>1</sup>, Ruiz Y.<sup>2</sup>, Hernandis E.<sup>3</sup>, Carracedo A<sup>1</sup>.

<sup>1</sup> Grupo de Medicina Xenómica. Instituto de Ciencias Forenses. Universidade de Santiago de Compostela, España.

<sup>2</sup> Comité Internacional de Cruz Roja, México.

<sup>3</sup> Facultad de Ciencias. Universidad Autónoma de Madrid, España.

Corresponding author: Elias Hernandis <elias@hernandis.me>

## Abstract

The evaluation of the DNA test in massive identification cases requires the use of Bayes' Theorem to estimate the probability of identification from *a priori* data together with probabilities obtained from the DNA test itself. To apply it, one needs to specify the prior probabilities of the hypotheses. An interdisciplinary team and an identification coordinator are key stakeholders in this process. The statistical approach can be complex but there exists validated non-commercial software, such as Familias, which aid in estimating the likelihood ratios of the DNA test for the given hypotheses. Next, the posterior probabilities in massive identification events can be estimated using the one to one, PM-driven, AM-driven or Global approaches published recently by Kling et al. (2021) which are discussed in this article.

Keywords: identification, DNA, probability, Disaster Victim Identification (DVI)

## Resumen

La valoración de prueba de ADN en casos de identificación masiva exige el uso por los peritos del teorema de Bayes para estimar la probabilidad de identificación a partir de unos datos *a priori* a los que suman las probabilidades proporcionadas por la prueba de ADN. Para aplicarlo hace falta por una parte especificar la probabilidad *a priori* de las hipótesis de identidad que se pueden plantear de modo que un equipo multidisciplinario y la figura de un coordinador de identificaciones son clave. El abordaje estadístico puede ser complejo pero existen programas validados no comerciales, como el software Familias que facilitan las estimas de las razones de verosimilitud de la prueba de ADN para las hipótesis que se establezcan. A continuación la probabilidad *a posteriori* en eventos de identificación a gran escala se puede estimar a través de las aproximaciones *one to one*, *PM-driven*, *AM-driven* y *Global approach* publicadas recientemente por Kling et al. (2021) y que son descritas en detalle en este artículo.

Palabras clave: Identificación, ADN, Probabilidad, Identificación de víctimas en desastres masivos (DVI)

## 1. Introducción

La prueba de ADN es una herramienta fundamental en la identificación masiva de víctimas, sin embargo, no es la única prueba que se ha de tener en cuenta en estos escenarios. Tradicionalmente, los resultados del análisis genético con fines forenses se han acompañado de una valoración estadística que permite medir su significado. El *likelihood ratio* (LR, cociente de verosimilitud), es el cociente de probabilidades entre dos hipótesis alternativas que habitualmente se utiliza para valorar los resultados de la prueba genética (1) (2) sin tener en cuenta otra información. Su valor por tanto depende sólo de datos genéticos (frecuencias alélicas, grado de compatibilidad entre los perfiles a comparar, calidad de los perfiles genéticos, etc.).

Sin embargo, el hecho de que dos perfiles genéticos sean compatibles (por ejemplo, que compartan un alelo en cada marcador) no significa que tengan que provenir de un padre y su hijo. Cuando nos encontramos ante un escenario con múltiples víctimas, es necesario realizar multitud de comparaciones de perfiles genéticos procedentes de los fallecidos (perfiles *post-mortem*, PM) con perfiles genéticos de familiares de personas desaparecidas (*ante-mortem*, AM). En ese contexto masivo, es posible encontrar compatibilidades genéticas aparentes sin que necesariamente haya una relación familiar entre los donantes de las muestras, principalmente si los perfiles genéticos son parciales. Es necesario contar con información adicional para lograr la identificación, ya que un *match* genético no significa que directamente hayamos identificado a una víctima. En otros escenarios en los que la prueba de ADN es útil (por ejemplo, en la investigación de delitos), es la autoridad judicial la que integra el valor del LR con otras pruebas no genéticas para lograr una conclusión final. Pero en los casos masivos de identificación, suelen ser los peritos de diferentes áreas los que se pronuncian sobre el resultado final; reportar un valor de LR no es suficiente para tal fin, ya que este no informa sobre la probabilidad final de identificación, la llamada probabilidad *a posteriori*. Ésta se obtiene aplicando el teorema de Bayes, que varios autores describen como la única forma coherente de sacar conclusiones sobre las probabilidades de la fuente sobre la que se basa de la evidencia forense (3). Sin embargo, la aplicación de este teorema no es sencilla ni siempre se ha entendido bien, pero la gran ventaja que presenta es que permite integrar información no genética que se tiene sobre el caso con la información genética. Para ello es necesario comenzar con la asignación de probabilidades previas a las proposiciones de interés, es decir, se debe decidir cuál es la probabilidad *a priori* de identificación, antes de realizar el análisis de ADN.

Los procesos de identificación a gran escala ocurren como consecuencia de desapariciones masivas relacionadas con conflictos armados, desastres naturales, violaciones a los derechos humanos o el derecho internacional humanitario entre otros. Estos casos pueden desbordar los servicios forenses locales por lo que se requiere una adaptación de procedimientos y prácticas en la materia. Esta adaptación incluye desde el fortalecimiento de capacidades y la gestión de cadáveres e información forense hasta el planteamiento de hipótesis de identidad, abordaje estadístico y valoración de la prueba de ADN.

Hoy en día, en muchos sistemas médico-legales que realizan identificaciones a gran escala de víctimas y personas desaparecidas, aún se observa que los peritos en genética trabajan

aislados del resto de las especialidades forenses, y siguen usando valores de probabilidades *a priori* establecidas por defecto. Uno de los errores más comunes es aplicar un valor de 0,5 de probabilidad (50%), clásicamente aplicado en pruebas de paternidad aunque de muy discutible uso en pruebas judiciales. Justamente en este artículo abordamos la importancia de entender la identificación como una tarea multidisciplinar y los riesgos de no integrar una adecuada probabilidad *a priori* en los casos masivos de identificación.

## 2. El teorema de Bayes en casos de identificación masiva

Para calcular un LR es necesario enunciar al menos dos hipótesis sobre los hechos. Las hipótesis deben ser excluyentes (si una es cierta, la otra debe ser falsa), por ejemplo:

$H_1$  = los restos óseos pertenecen a un hijo de P y M

$H_2$  = los restos óseos no pertenecen a un hijo de P y M

El LR nos mide la probabilidad de haber obtenido los perfiles genéticos que se obtuvieron en los restos óseos, y en P y M (sea cual sea este resultado, es decir, compartan alelos o no) bajo las dos hipótesis mencionadas. Y se formula de la siguiente forma:

$$LR = \frac{P(E|H_1)}{P(E|H_2)} = \frac{\text{probabilidad de obtener los resultados de la prueba genética suponiendo que } H_1 \text{ es cierta}}{\text{probabilidad de obtener los resultados de la prueba genética suponiendo que } H_2 \text{ es cierta}} \quad (1)$$

siendo  $E$  = evidencia (el resultado genético en las muestras analizadas),  $P$  = probabilidad y el símbolo “|” = suponiendo.

Los resultados obtenidos en el análisis pueden apoyar una hipótesis u otra dependiendo principalmente de si los perfiles genéticos son o no compatibles entre sí y de la frecuencia de los alelos compartidos. El LR puede tomar valores entre 0 e infinito. Si el valor se encuentra por debajo de 1, los resultados genéticos apoyan más la hipótesis del denominador, normalmente la que supone ausencia de relación de parentesco. Si el valor del LR es 1, los resultados genéticos no apoyan más una hipótesis que la otra, es decir, la evidencia genética es neutra. Un ejemplo claro de esta situación se produce cuando no se han obtenido resultados para algún marcador genético. Si el valor del LR es mayor que 1, los resultados genéticos apoyan más la hipótesis del numerador, y más la apoyan cuanto mayor que 1 sea ese valor.

Pero el LR no indica cual es la probabilidad *a posteriori* de identificación. Para estimar esta probabilidad, debemos de tener en cuenta otra información que no es genética. En pocas palabras, el hecho de obtener una coincidencia o compatibilidad genética no significa haber logrado la identificación.

El valor del LR obtenido en la prueba de ADN puede integrarse con otra información no genética para calcular la probabilidad *a posteriori* de identificación, aplicando el teorema de Bayes. Su formulación es sencilla cuando solo hay dos hipótesis que contrastar:

$$P(H_1|E) = \frac{P(E|H_1) P(H_1)}{P(E|H_1) P(H_1) + P(E|H_2) P(H_2)} \quad (2)$$

$P(H_1|E)$  es la probabilidad *a posteriori* de identificación, es decir, la probabilidad que al final interesa al investigador o a la autoridad judicial que entiende de los hechos.  $P(H_1)$  es la probabilidad *a priori* de la hipótesis 1 y  $P(H_2)$  es la probabilidad *a priori* de la hipótesis 2.  $P(E|H_1)$  y  $P(E|H_2)$  se definieron ya anteriormente en el LR. Por tanto, la probabilidad *a posteriori* es una combinación del LR con la probabilidad *a priori*, y da respuesta a cuál es la probabilidad de una hipótesis teniendo en cuenta los datos (la evidencia genética y no genética). El teorema de Bayes también puede expresarse con probabilidades en forma de apuesta (en inglés, *odds*). En el Anexo 1 se puede ver cómo se relacionan las dos formas del teorema.

Volviendo a la forma definida en (2) del teorema, si  $P(H_1)$  es 0,5 (y por tanto  $P(H_2)$  también será 0,5), la fórmula se simplifica:

$$P(H_1|E) = \frac{P(E|H_1)}{P(E|H_1) + P(E|H_2)} \quad (3)$$

Y si dividimos el numerador y el denominador de (3) entre  $P(E|H_2)$ , obtenemos la famosa fórmula que habitualmente se utiliza en los casos de paternidad para calcular la probabilidad *a posteriori*:

$$P(H_1|E) = \frac{LR_1}{LR_1 + 1} \quad (4)$$

Sin embargo, en casos de identificación masiva, nos enfrentamos a muchas más hipótesis; por ejemplo, en el caso de hallazgos de fosas masivas debemos contrastar un conjunto de hipótesis y no solo dos como en los casos aislados de identificación. Kling et al. (2021) (4) ejemplifican un escenario sencillo de una fosa común en la que se encuentran tres cuerpos (V1, V2 y V3) y tres familias que buscan a tres personas desaparecidas (F1, F2 y F3). Las hipótesis a contrastar serían:

- H<sub>1</sub> = V1 pertenece a la familia F1
- H<sub>2</sub> = V2 pertenece a la familia F1
- H<sub>3</sub> = V3 pertenece a la familia F1
- H<sub>4</sub> = otra víctima desconocida pertenece a la familia F1

H<sub>4</sub> contempla que la persona desaparecida que la familia F1 está buscando, no está realmente en el conjunto de restos humanos investigados. En el caso de las familias F2 y F3, las hipótesis serían las mismas que para F1, pero sustituyendo F1 por F2 y F3 respectivamente.

En los casos con múltiples hipótesis, la aplicación del teorema de Bayes para calcular la probabilidad *a posteriori* se formula de la siguiente manera:

$$P(H_i|E) = \frac{P(E|H_i) P(H_i)}{\sum_{j=1}^k P(E|H_j)P(H_j)} \quad (5)$$

donde  $P(H_i)$  es la probabilidad *a priori* de identificación para la hipótesis de interés y el sumatorio  $\sum_{j=1}^k P(E|H_j)P(H_j)$  indica la suma de las hipótesis 1 a  $k$ . En nuestro ejemplo, si la hipótesis de interés fuera  $H_1$ , la fórmula (5) se expresaría:

$$P(H_1|E) = \frac{P(E|H_1) P(H_1)}{P(E|H_1) P(H_1) + P(E|H_2) P(H_2) + P(E|H_3)P(H_3) + P(E|H_4)P(H_4)} \quad (6)$$

Podemos utilizar un pequeño truco para simplificar esta fórmula (6), simplemente dividiendo el numerador y el denominador por el término  $P(E|H_4)$ . Así obtendremos:

$$P(H_1|E) = \frac{LR_1 P(H_1)}{LR_1 P(H_1) + LR_2 P(H_2) + LR_3 P(H_3) + LR_4 P(H_4)} \quad (7)$$

El valor del  $LR_4$  es 1, ya que se obtuvo al dividir  $P(E|H_4) / P(E|H_4)$ . Y si además el valor de la probabilidad *a priori* es igual para todas las hipótesis, aún podemos simplificar más. Supongamos que en este caso  $P(H_1) = P(H_2) = P(H_3) = P(H_4) = 1/4$  (ver siguiente apartado sobre probabilidades *a priori*). Obtendremos entonces:

$$P(H_1|E) = \frac{LR_1}{LR_1 + LR_2 + LR_3 + LR_4} \quad (8)$$

Hasta ahora hemos visto que hay una gran diferencia entre aplicar el teorema de Bayes en escenarios con sólo dos hipótesis y en escenarios con  $k$  hipótesis. Ahora nos centraremos en ver qué ocurre en escenarios masivos cuando se aplica el teorema de Bayes. Para ello nos remitimos de nuevo a Kling et al (2021) (4), quienes describen diferentes aproximaciones para establecer las hipótesis. En este apartado solo describimos las distintas formas de abordar el problema, pero en la sección 4 el lector podrá ver con más detalle cómo se aplica cada una de ellas con un ejemplo sencillo. Las diferentes aproximaciones comprenden:

- a) *One to one*: esta aproximación se aplica habitualmente en escenarios no masivos, cuando existe una presunta identidad y se compara un solo perfil *post-mortem* con un sólo pedigrí. Pero también se aplica en escenarios masivos, cuando se compara el perfil *post-mortem* con todos los pedigríes, y finalmente sólo se tienen en cuenta dos hipótesis cuando se aplica el teorema de Bayes. El hecho de que se contemplen sólo dos hipótesis no implica que el valor de la probabilidad *a priori* tenga que ser

0,5; en la sección 4 el lector puede ver un ejemplo ad hoc. Pero lo que nos interesa ahora son las hipótesis, que serían:

$H_1$  = la víctima 1 está relacionada con la familia F1

$H_2$  = la víctima 1 no está relacionada con la familia F1

- b) *PM-driven*: en esta aproximación las hipótesis se establecen teniendo en cuenta a las víctimas, es decir, se trata de contestar a la pregunta ¿qué familia está relacionada con esta víctima concreta? Esta aproximación intenta proporcionar la familia más probable a cada víctima, y no se tiene en cuenta que dos víctimas distintas podrían asociarse con la misma familia. Las hipótesis por tanto serían:

$H_1$  = la víctima 1 está relacionada con la familia F1

$H_2$  = la víctima 1 está relacionada con la familia F2

...

$H_N$  = la víctima 1 está relacionada con la familia FN, siendo N el número de familias que buscan a sus seres queridos.

$H_{N+1}$  = la víctima 1 no está relacionada con ninguna de las N familias

- c) *AM-driven*: en esta aproximación las hipótesis se establecen teniendo en cuenta a las familias, es decir, se trata de contestar a la pregunta ¿qué víctima está relacionada con esta familia concreta? No se tiene en cuenta que una misma víctima podría asociarse a dos familias distintas. Y las hipótesis se establecen:

$H_1$  = la familia F1 está relacionada con V1

$H_2$  = la familia F1 está relacionada con V2

....

$H_M$  = la familia F1 está relacionada con la víctima M, siendo M en número de víctimas.

$H_{M+1}$  = la familia F1 no está relacionada con ninguna de las M víctimas

El abordaje *one to one* es el más habitual por su sencillez, pero no deja de presentar problemas. Al considerar sólo dos hipótesis, por ejemplo, V1 está relacionada con F1 vs. V1 no está relacionada con F1, no se está teniendo en cuenta la posibilidad de que la víctima 1 esté relacionada con F2, F3, etc. Es decir, aunque la comparación se haya realizado con todos los perfiles, en la valoración estadística sólo estamos teniendo en cuenta uno de los resultados de la comparación.

Por el contrario, en las aproximaciones *PM-driven* y *AM-driven* se tienen en cuenta todas las comparaciones en la valoración estadística de los resultados, pues se consideran varias hipótesis a la vez. Sin embargo, ambas aproximaciones tienen la desventaja de que la evaluación estadística se realiza de forma secuencial, es decir, se va considerando cada hipótesis de una en una en el numerador de la fórmula (5). Esto puede tener una gran influencia en casos en los que hay más de una persona desaparecida en una familia. Por ejemplo, si un padre P y un hijo H han desaparecido y como muestras de referencia sólo se dispone de la madre de H, el padre P no podrá identificarse a no ser que el perfil genético de H se eleve a la categoría de perfil de referencia. Por tanto, la identificación en este caso es secuencial, primero identificando al hijo, y una vez que se da por cierta esa identificación se realiza la del padre.

La solución a este problema la ofrecen los propios autores, a través de la llamada *aproximación global* (“*global approach*”), si bien este abordaje requiere un gran esfuerzo computacional. Con este método se tienen en cuenta todas las posibles soluciones que se pueden obtener en el proyecto de identificación masiva, por ejemplo, que no se identifique a ninguna víctima, que sólo se identifique a una, solo a dos, etc. En la Figura 1 puede verse un ejemplo de posibles soluciones S de un caso mediante este abordaje global. La principal ventaja del método es que no es posible asignar una misma víctima a diferentes familias o que a una misma familia se le asignen diferentes víctimas en la misma posición en el pedigrí, ya que el enfoque tiene en cuenta todos los resultados a la vez y no de forma secuencial como en los otros tres abordajes. El hecho de que existan dos personas desaparecidas dentro de una misma familia no es un problema con el enfoque global, pues en el ejemplo anterior en el que P y H estaban desaparecidos, no es necesario elevar el perfil genético de H a la categoría de perfil *ante-mortem*, sino que ambos pueden identificarse a la vez.

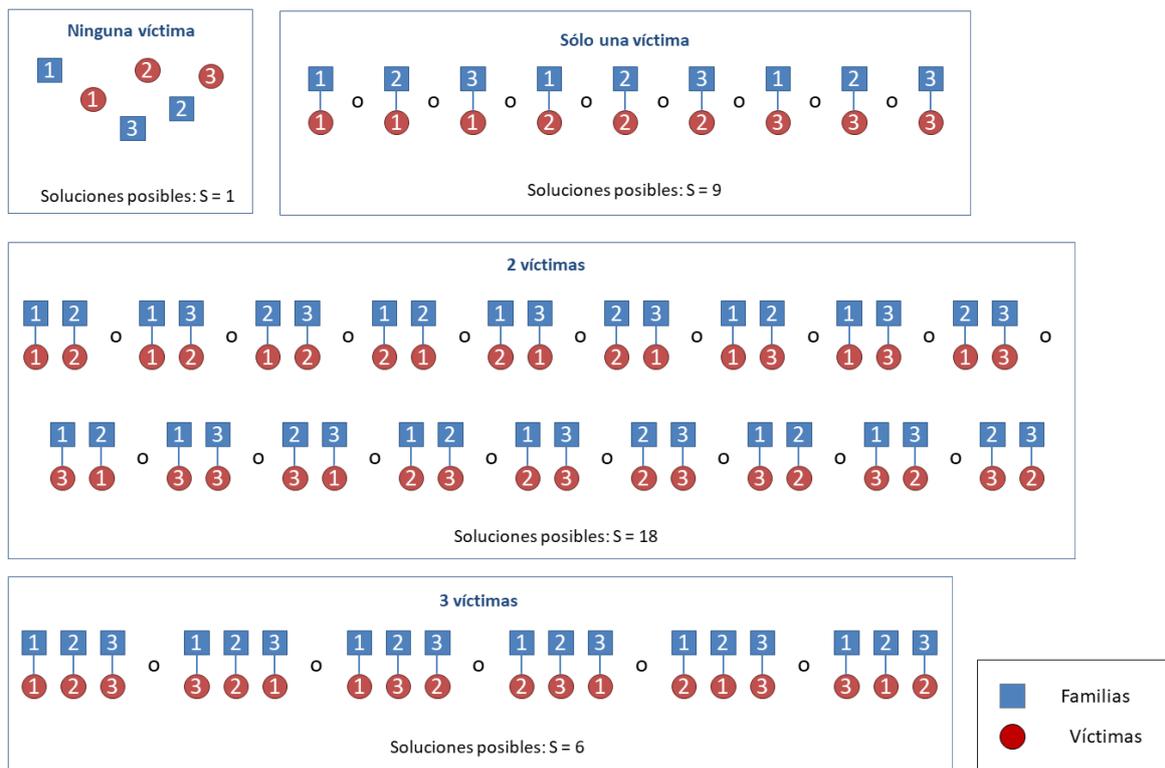


Figura 1. Aproximación global (traducido de Kling et al., 2021) (4). Escenario con 3 víctimas y 3 familias de tres personas desaparecidas. Se muestran todos los resultados que se pueden obtener (S) teniendo en cuenta que es posible que no se obtenga:

1. ningún *match* genético,
2. sólo un *match* (F1 con V1, o F2 con V1, ..., o F3 con V3),
3. dos *matches* genéticos (F1 con V1 y F2 con V2, o F1 con V1 y F3 con V2, ..., o F2 con V3 y F3 con V2)
4. o tres *matches* genéticos (F1 con V1, F2 con V2 y F3 con V3, ..., o F1 con V3, F2 con V1 y F3 con V2).

El número total de posibles soluciones S es  $1 + 9 + 18 + 6 = 34$ .

Afortunadamente, existe software gratuito que nos permite llevar a cabo cualquiera de estas aproximaciones al problema. El programa Familias (<https://familias.no/>) (5) dispone de las herramientas necesarias para definir las hipótesis según los abordajes *AM-driven*, *PM-driven* y *one to one*. Recientemente, Vigeland y Egeland (2021) (6) han desarrollado un

paquete para R llamado *dvir* (<https://cran.r-project.org/web/packages/dvir/index.html>) que permite aproximarse al problema desde una perspectiva global (*global approach*, llamado *joint approach* en su publicación).

### 3. La probabilidad *a priori* de identificación

La probabilidad *a priori* de identificación es la creencia de que una persona fallecida sin identificar pueda pertenecer a una determinada familia. Es una probabilidad subjetiva, pero que puede basarse en ciertos datos.

Uno de los errores más habituales a la hora de establecer el valor de la probabilidad *a priori* de identificación en escenarios masivos es aplicar el mismo valor que se utiliza en las pruebas de paternidad (50%). Habitualmente, el perito genetista no tiene ninguna información no genética del caso de paternidad como para establecer un valor de probabilidad *a priori*, y casi más por tradición que por exactitud, los peritos usan, en ocasiones, el valor de 50%.

Se tiene la creencia de que la probabilidad del 50% tiene carácter no informativo, pues se le da el mismo valor a la probabilidad de paternidad como a la de no paternidad (100% - 50% = 50%). Sin embargo, esto no es así. Al aplicar un valor de probabilidad *a priori* del 50% estamos asignando una probabilidad de paternidad del 50% al presunto padre, y el 50% restante se reparte entre todos los varones de la población de interés, y por tanto se asigna mayor probabilidad al presunto padre que a cada uno del resto de posibles padres. También en casos de paternidades judiciales es el juez el que debería fijar el valor de la probabilidad *a priori*, y en ese sentido el perito está excediéndose en su papel.

Se ha de tener en cuenta también, que si aplicamos una probabilidad *a priori* del 50% y contrastamos dos hipótesis (parentesco vs. no relacionados), en los casos de identificación masiva sólo necesitamos un valor de LR = 1000 para alcanzar una probabilidad *a posteriori* del 99,9%, y con un LR = 10000, alcanzaremos un valor de 99,99%. Estos valores de probabilidad *a posteriori* pueden llevar a un juez a determinar la identificación con gran certeza, y sin embargo se puede alcanzar un valor de LR = 1000 incluso si el *match* genético no es una coincidencia genética real.

Como hemos apuntado anteriormente, la compatibilidad entre perfiles genéticos no es sinónimo de identificación. Imaginemos que en el análisis de un evento masivo encontramos compatibilidad entre el perfil genético *post-mortem* obtenido de unos restos óseos y el perfil genético *ante-mortem* de un niño de 10 años que busca a su padre. Pudiera ocurrir que esa compatibilidad fuera falsa, sobre todo si hay muchos individuos involucrados o los perfiles genéticos *post-mortem* son parciales. Si el antropólogo determina que los restos óseos en cuestión proceden de un individuo sub-adulto. Claramente podemos dar la compatibilidad como falsa, pues la probabilidad *a priori* en este caso tendría un valor de 0 ya que la edad de los individuos descarta que puedan tener una relación padre-hijo.

Tradicionalmente se han clasificado los escenarios masivos en cerrados o abiertos. En casos cerrados, el número de desaparecidos es cuantificable generalmente, y su identidad es conocida. El ejemplo típico es un accidente aéreo donde existe una lista de pasajeros, si bien esta puede no ser fiable al 100%. En casos abiertos, el número real de personas desaparecidas puede ser desconocido, es decir, su estima es más incierta. Pero en la vida real lo más frecuente es encontrarse ante casos mixtos, es decir, casos en los que se conoce la identidad de algunas de las personas desaparecidas, pero no de todas (7).

En ausencia de otra información, suele tenerse en cuenta el número de personas desaparecidas para establecer la probabilidad *a priori* de identificación (8). Así, en el caso mencionado en el apartado anterior, en el que el número de desaparecidos era 3, la probabilidad *a priori* podría establecerse en  $1/4$  ( $1/(n^{\circ} \text{ desaparecidos} + 1)$ ), con el fin de tener en cuenta que la persona que estamos buscando puede no estar en la fosa investigada. Mediante este abordaje se asume que todas las víctimas tienen la misma probabilidad *a priori* de identificación (*flat priors*). Y la probabilidad de no identificación para cada individuo sería  $3/4$ . El lector puede encontrar ejemplos en los ejercicios llevados a cabo por el Grupo de Habla Española y Portuguesa de la International Society for Forensic Genetics (GHEP-ISFG) (9) (10).

La probabilidad *a priori* de identificación puede ser re-definida por otros expertos como los antropólogos, teniendo en cuenta otras características como el sexo o la edad. Si en el caso de las 3 personas desaparecidas, dos de ellas son mujeres, y los restos humanos que se desea identificar todavía presentan características físicas que hacen visible su sexo femenino, se podría entonces re-definir la anterior probabilidad *a priori* a  $1/3$  ( $1 / (n^{\circ} \text{ mujeres desaparecidas} + 1)$ ). Sin embargo, no debemos olvidar que también puede haber incertidumbre a la hora de determinar la edad o el sexo de unos restos humanos, sobre todo si nos encontramos ante un escenario con restos óseos incompletos. Por ejemplo, la estima de la edad en restos *post-mortem* puede ser no muy exacta dependiendo de la edad del individuo, como ocurre cuando se tiene en cuenta la morfología de la sínfisis púbica en individuos a partir de los 40 años (11).

Otras características como tatuajes, marcas, prótesis, pueden ser más difíciles de cuantificar. Las declaraciones de testigos que pudieron estar presentes en los hechos también pueden tenerse en cuenta, pero igualmente son difíciles de cuantificar e incluso pueden no ser fiables (12). En Budowle et al., 2011 (13), el lector puede encontrar una discusión sobre los problemas que pueden surgir cuando se da una credibilidad del 100% a las declaraciones de testigos, a otras informaciones (localización de la fosa, intervalo *post-mortem*, variables demográficas) o cuando no se tiene en cuenta la posible dependencia entre distintas variables utilizadas para establecer la probabilidad *a priori* de identificación.

Para definir la probabilidad *a priori*, Gill et al. (1994) (14) tuvieron en cuenta el trabajo realizado en las piezas dentales de los restos esqueléticos que posteriormente se identificaron como pertenecientes a la familia Romanov. En King et al. 2014 (15), también se puede ver un ejemplo de cómo aplicar el pensamiento bayesiano en la identificación. El lector podrá comprobar que en ocasiones la información no genética es difícilmente cuantificable. La frecuencia de varias características o condiciones no se conoce con precisión, y por tanto el valor de probabilidad *a priori* es difícil de estimar (16).

Por tanto, como conclusión diremos que la estima de la probabilidad *a priori* teniendo en cuenta el número de personas desaparecidas suele ser conservadora y defendible. Incluso cuando el número exacto no se conoce se pueden realizar estimas. Por ejemplo, en el atentado terrorista al *World Trade Center* en 2001, se utilizó una probabilidad *a priori* de 1/3000 porque se estimó que el número de desaparecidos fue algo menos que 3000 (17). Esta estima (número estimado de desaparecidos) es la más adecuada en casos de identificación a gran escala, para que los laboratorios de ADN reporten el listado de posibles identificaciones que podrán corroborarse o descartarse tras la incorporación de la información no genética.

## 4. Cálculo de la probabilidad *a posteriori*: un ejemplo y diferentes abordajes

En este apartado retomamos el sencillo ejemplo de Kling et al., 2021 (4) en el cual nos enfrentamos a un escenario con 3 víctimas (V1, V2 y V3) y tres familias (F1, F2 y F3) que buscan, cada una de ellas, a una persona desaparecida (MP1, MP2 y MP3). Basándonos en el epígrafe sobre el valor de la probabilidad *a priori*, diremos que para cada víctima  $V_i$  el valor de esta probabilidad es de 1/4, teniendo así en cuenta la posibilidad de que la víctima que buscamos no se encuentre en el escenario investigado.

Supongamos que tras el análisis genético y la comparación de los perfiles obtenemos los resultados que se describen en la Tabla 1. El valor del LR se ha calculado teniendo en cuenta cada hipótesis de identidad frente a la hipótesis que considera la ausencia de relación familiar (en nuestro ejemplo,  $H_4$ ). Estrictamente hablando, el valor del LR no debería ser 0 si consideramos la posibilidad de mutación, pero lo aproximamos a 0 para simplificar los cálculos.

Tabla 1. Supuesto práctico de un escenario con tres víctimas y tres personas desaparecidas. Se muestran los resultados obtenidos tras la comparación de los perfiles genéticos *post-mortem* con los perfiles genéticos *ante-mortem*.

Víctima	Familia	LR
V1	F1	$10^6$
V1	F2	10
V1	F3	0
V2	F1	500
V2	F2	0
V2	F3	0
V3	F1	0
V3	F2	0
V3	F3	0

Mostramos ahora cómo se calculan las probabilidades *a posteriori* con las aproximaciones *one to one*, *PM-driven* y *AM-driven*.

### **One to one**

Con este método sólo consideramos dos hipótesis, que a la vista de los resultados serían:

$H_1 = V1$  está relacionada con la familia F1

$H_2 = V1$  no está relacionada con la familia F1

No se consideran los resultados obtenidos en las demás comparaciones, es decir, otros *matches* candidatos. La probabilidad *a posteriori* se calcularía según la fórmula (2), teniendo en cuenta que si la probabilidad *a priori* de identificación tiene un valor de 1/4, la de no identificación será  $1 - 1/4 = 3/4$ . Pero por simplicidad, dado que ya tenemos valores de LR, podemos utilizar la fórmula (7) con solo dos hipótesis:

$$P(H_1|E) = \frac{LR_1 P(H_1)}{LR_1 P(H_1) + LR_2 P(H_2)} = \frac{10^6 * 1/4}{(10^6 * 1/4) + (1 * 3/4)} = 0,999997$$

En el Anexo 2 se puede ver este mismo ejemplo, pero aplicando el teorema de Bayes en forma de apuesta.

### **PM-driven**

En este abordaje se compara una víctima con cada una de las familias. En nuestro caso diríamos, por ejemplo para V1:

$H_1 = V1$  está relacionada con la familia F1

$H_2 = V1$  está relacionada con la familia F2

$H_3 = V1$  está relacionada con la familia F3

$H_4 = V1$  no está relacionada con ninguna de las 3 familias

La probabilidad *a posteriori* se calcularía según la fórmula (5), pero podemos usar la simplificación (8) ya que los valores de probabilidad *a priori* son iguales. Tenemos por tanto:

$$P(H_1|E) = \frac{LR_1}{LR_1 + LR_2 + LR_3 + LR_4} = \frac{10^6}{10^6 + 10 + 0 + 1} = 0,999989$$

En el Anexo 3 se muestran todas las probabilidades *a posteriori* para cada hipótesis y para cada víctima.

### **AM-driven**

En este abordaje se compara una familia con cada una de las víctimas. Las hipótesis de nuestro ejemplo serían, para F1:

$H_1 =$  la familia F1 está relacionada con V1

$H_2 =$  la familia F1 está relacionada con V2

$H_3 =$  la familia F1 está relacionada con V3

$H_4 =$  la familia F1 está relacionada con otra víctima desconocida

La probabilidad *a posteriori* se calcularía según la fórmula (5), pero podemos usar la simplificación (8) ya que los valores de probabilidad *a priori* son iguales. Tenemos por tanto:

$$P(H_1|E) = \frac{LR_1}{LR_1 + LR_2 + LR_3 + LR_4} = \frac{10^6}{10^6 + 500 + 0 + 1} = 0,9995$$

En el Anexo 3 se muestran todas las probabilidades *a posteriori* para cada hipótesis y para cada familia.

### Aproximación global

Como se muestra en la Figura 1, en este caso se consideran todas las posibles soluciones a la vez: que no se identifique a ninguna víctima, sólo a una, a dos y a las tres. En nuestro ejemplo hay un total de 34 posibles soluciones, es decir, 34 hipótesis distintas. Así, tendríamos:

- H<sub>1</sub> = V1, V2 y V3 no son ni MP1, ni MP2, ni MP3
- H<sub>2</sub> = V1 es MP1, V2 y V3 no son ni MP1, ni MP2, ni MP3
- H<sub>3</sub> = V2 es MP1, V1 y V3 no son ni MP1, ni MP2, ni MP3
- ...
- H<sub>11</sub> = V1 es MP1 y V2 es MP2, V3 no es ni MP1, ni MP2, ni MP3
- ...
- H<sub>29</sub> = V1 es MP1, V2 es MP2 y V3 es MP3
- ...
- H<sub>34</sub> = V1 es MP3, V2 es MP2 y V3 es MP1

En la Tabla S3 del Anexo 3 se muestran todas estas posibilidades, respetando los valores de LR utilizados en los demás abordajes y añadiendo valores de LR = 0 para los casos que no se contemplaban en las aproximaciones secuenciales.

La probabilidad *a priori* debe ahora repartirse entre las 34 hipótesis. Si queremos que *a priori* todas las soluciones sean igualmente probables usaremos un valor de 1/34.

Aplicando la simplificación de la fórmula (8), a partir de esta tabla se pueden calcular las probabilidades *a posteriori* de identificación individual, teniendo en cuenta cada hipótesis en la que interviene una víctima concreta. Por ejemplo, para V1, en el numerador se consideran todos los LRs en los que MP1 aparece en la columna V1 de la Tabla. El denominador es la suma de todos los LRs. Así tenemos:

$$P(V1 = MP1|E) = \frac{LR_2 + LR_{11} + LR_{12} + LR_{17} + LR_{18} + LR_{29} + LR_{30}}{LR_1 + \dots + LR_{34}} =$$

$$= \frac{10^6 + 0 + 0 + 0 + 0 + 0 + 0}{1 + 10^6 + 10 + 0 + 500 + 0 + \dots + 0} = 0,999489261$$

En la Tabla S4 del Anexo 3 se muestran todas las probabilidades *a posteriori* individuales para cada víctima.

## 5. Papel del coordinador de identificación en análisis de casos a gran escala

Mientras que en criminalística o casos de paternidad judicial, el juez es quien determina la probabilidad *a priori*, cuando se trata de desapariciones a gran escala, la figura del Coordinador de Identificaciones, debe jugar un rol clave en la formulación de las hipótesis del caso, establecer las probabilidades *a priori*, el umbral de identificación y consolidar el reporte integrado de identificación junto al equipo multidisciplinario a través de la reconciliación del caso.

El umbral de identificación debe establecerse de forma que se encuentre un equilibrio entre maximizar las identificaciones y minimizar las falsas identificaciones (18). Los laboratorios de genética suelen establecer umbrales de probabilidad *a posteriori* de 99,90%, 99,95% o 99,99% para informar una coincidencia genética concreta. Un valor de 99,95% de probabilidad *a posteriori* significa que se está dispuesto a asumir que una de cada 2000 coincidencias genéticas es falsa (16). Del mismo modo, con un valor de 99,90% se asume que uno de cada 1000 *matches* será falso; y para 99,99%, se espera que ocurra 1 *match* falso de cada 10000.

Las discusiones dirigidas por el coordinador de identificaciones sobre la reconciliación de un caso son un paso esencial durante el proceso de identificación donde toda la información correspondiente a la persona fallecida no identificada (*post mortem*, PM) es comparada y evaluada con toda la información de la persona desaparecida (*ante mortem*, AM). Esto, con el fin de poder i) alcanzar una identificación formal o ii) determinar los pasos que sean requeridos para alcanzar una identificación.

En este sentido, previo a que tome lugar la reunión de reconciliación (y especialmente en casos de restos óseos), el Coordinador Identificaciones, una vez que recibe por parte de los especialistas en genética forense un *match* preliminar del caso procede a corroborar la información AM y contextual verificando la consistencia con la hipótesis de identidad arrojada. Así mismo, todos los especialistas involucrados en las fases del proceso de análisis PM, deben confirmar que todos los esfuerzos, visitas a los sitios, investigaciones y colecciones de información del caso y demás diligencias han finalizado para entonces proceder a preparar el plan de reconciliación del caso.

Además de la información arqueológica, antropológica, odontológica y médica que forma parte de la información PM, en las reuniones de reconciliación de casos asociados a desapariciones a gran escala se discute sobre la siguiente información genética: número mínimo de individuos por ADN (en casos mezclados), muestras analizadas que alcanzaron un perfil apto para análisis de parentesco heredo-biológico y/o para asociación intra-esquelética, muestras biológicas de referencia pertenecientes a familiares empleadas en el análisis, razón de verosimilitud obtenida (LR), probabilidad *a priori* empleada durante el caso y su base, probabilidad *a posteriori* obtenida tras el análisis, árbol familiar disponible de la persona desaparecida que no se utilizó en la comparación y evaluación de presencia de otras personas desaparecidas en el mismo árbol familiar.

Si se observan consistencias luego de la evaluación y comparación de datos *post-mortem* obtenidos de las investigaciones correspondientes a cada una de las especialidades (incluyendo la genética) con relación a la información *ante-mortem* (incluyendo las circunstancias de la desaparición y demás información investigativa), y no se encuentran discrepancias inexplicables, entonces se procede a concluir la identificación. Si por el contrario se encuentran discrepancias en la información reconciliada, se establecen recomendaciones adicionales (ej. confirmación con muestras de referencia adicionales o mayor número de marcadores, información investigativa adicional, análisis adicionales antropológicos, etc.).

## 6. Conclusiones y recomendaciones para el análisis genético forense en casos de identificación a gran escala

- Los procesos de identificación a gran escala requieren adaptaciones de procedimientos que incluyen ajustes en la gestión de información. Los peritos genetistas necesitan de la formación especializada en estadística forense y manejo adecuado de software.
- Parte de estos procedimientos en materia de ADN incluyen el planteamiento de múltiples hipótesis de identidad así como la estimación de la probabilidad *a priori* basado en la información de contexto. Se recomienda que en ausencia de mayor información de contexto, la probabilidad *a priori* pueda ser estimada a partir del número de desaparecidos asociados a un evento.
- El teorema de Bayes es una la herramienta que aporta un análisis comprehensivo en identificaciones a gran escala, debido a la posibilidad de integrar múltiples hipótesis incluyendo la posibilidad de que la víctima que es buscada no se encuentre en el escenario investigado. Aplicando este teorema se puede considerar tanto información genética como no genética. La probabilidad *a posteriori* en eventos de identificación a gran escala se puede estimar a través de las aproximaciones *one to one*, *PM-driven*, *AM-driven* y *Global-approach*.
- Para el correcto abordaje estadístico, existen herramientas tecnológicas gratuitas como lo son el software Familias<sup>1</sup> que trabaja en conjunto con paquetes desarrollados en el lenguaje de programación de R<sup>2</sup>.
- En eventos de identificación a gran escala, la figura del Coordinador de Identificaciones, debe jugar un rol clave en la formulación de las hipótesis del caso, establecer las probabilidades *a priori*, el umbral de identificación y consolidar el

---

<sup>1</sup> <https://familias.no>

<sup>2</sup> <https://www.r-project.org>

reporte integrado de identificación junto al equipo multidisciplinario a través de la reconciliación del caso.

## **Agradecimientos**

Agradecemos los valiosos comentarios recibidos de Thore Egeland durante la preparación de este artículo.

## Anexo 1 – Teorema de Bayes en forma de apuesta

Partimos de la fórmula general del teorema de Bayes:

$$P(H_1|E) = \frac{P(E|H_1) P(H_1)}{P(E)} \quad (A1)$$

Las probabilidades en forma de apuesta no son más que los cocientes entre la probabilidad de identificación y la de no identificación. Por tanto, si aplicamos (A1) también a  $H_2$ , tendríamos:

$$P(H_2|E) = \frac{P(E|H_2) P(H_2)}{P(E)} \quad (A2)$$

y si dividimos la ecuación A1 entre A2, se obtiene:

$$\frac{P(H_1|E)}{P(H_2|E)} = \frac{P(E|H_1)P(H_1)}{P(E|H_2)P(H_2)}$$

lo que es lo mismo que:

$$\frac{P(H_1|E)}{P(H_2|E)} = LR * \frac{P(H_1)}{P(H_2)}$$

es decir: *Posterior odds = LR \* prior odds.*

Esta formulación es útil cuando las hipótesis elegidas no son exhaustivas, es decir no suman 1. Por ejemplo, supongamos que las hipótesis del comienzo de la sección 2 las hubiéramos formulado así:

$H_1$  = los restos óseos pertenecen a un hijo de P y M

$H_2$  = los restos óseos no pertenecen a un hijo de P y M, sino a un individuo de la población A, no emparentado con P ni con M.

En este caso, las hipótesis ya no serían exhaustivas, ya que cabe la posibilidad de que los restos óseos pertenezcan a un individuo solo emparentado con M pero no con P, o viceversa. Es importante puntualizar que considerar que las hipótesis no son exhaustivas no sólo pasa por reconocerlo, sino también por asignar valores a las probabilidades *a priori* de manera que sumen menos de 1. En este caso, el teorema en forma de *odds* nos permite dar valor informativo a pesar de no poder calcular los LRs para las hipótesis que no se han enumerado (por ejemplo, en el caso anterior, que el individuo sólo esté emparentado con M). Para profundizar en este tema, el lector puede consultar la presentación de Thore Egeland (2017) alojada en (19).

## Anexo 2 – Aproximación *one to one* - Probabilidad *a posteriori* usando Bayes en forma de apuesta

Retomamos el abordaje *one to one* del ejemplo, en el que calculamos la probabilidad *a posteriori* usando el teorema de Bayes en su forma habitual:

$$P(H_1|E) = \frac{LR_1 P(H_1)}{LR_1 P(H_1) + LR_2 P(H_2)} = \frac{10^6 * 1/4}{(10^6 * 1/4) + (1 * 3/4)} = 0,999997$$

Si queremos aplicar el teorema en forma de apuesta tendremos:

$$\frac{P(H_1|E)}{P(H_2|E)} = LR * \frac{P(H_1)}{P(H_2)} = 10^6 * \frac{0,25}{0,75} = 333.333,3$$

Los *posterior odds* tienen un valor entonces de 333.333,3, es decir aproximadamente 333.333 a 1 a favor de la identificación. Para pasar este valor a porcentaje simplemente tenemos que dividir el valor a favor de la apuesta (333.333,3) entre la suma de los dos valores de la apuesta (333.333,3 + 1), y así obtenemos 0,999997, es decir, el mismo valor que el obtenido usando en teorema de Bayes con la fórmula (2).

Algunos expertos simplifican estos cálculos usando la siguiente secuencia:

$$LR * \text{probabilidad a priori} = 10^6 * \frac{1}{4} = 250.000$$

$$P(H_1|E) = \frac{250.000}{250.000 + 1} = 0,999996$$

Esta fórmula se puede derivar de (2):

$$P(H_1|E) = \frac{P(E|H_1)P(H_1)}{P(E|H_1)P(H_1) + P(E|H_2)P(H_2)} = \frac{LR_1 P(H_1)}{LR_1 P(H_1) + LR_2 P(H_2)} \cong \frac{LR_1 P(H_1)}{LR_1 P(H_1) + 1}$$

El valor de probabilidad *a posteriori* obtenido (0,999996) no coincide exactamente con el valor de la probabilidad *a posteriori* obtenido anteriormente (0,999997). Esto se debe a que en esta secuencia de cálculos se ha supuesto que la probabilidad *a priori* de la hipótesis de no identificación  $P(H_2)$  es 1. Al ser 1 el máximo valor que puede tomar  $P(H_2)$ , es decir siendo conservador, esta fórmula en realidad produce un valor menor que la probabilidad *a posteriori* siempre y cuando haya sólo dos hipótesis.

## Anexo 3 – Material suplementario del ejercicio en la sección 4

### PM-driven

En la Tabla S1 se muestran las probabilidades *a posteriori* para cada una de las hipótesis del abordaje *PM-driven*. Nótese que para cada víctima  $V_i$  la suma de las probabilidades *a posteriori* es 1. No ocurre así para cada familia. En las hipótesis  $H_1$  a  $H_4$ , sólo se tienen en

cuenta los resultados obtenidos de la comparación de V1 con las 3 familias, pero no se tienen en cuenta los resultados de las comparaciones de V2 y V3, es decir, en el denominador de la probabilidad a posteriori, no se tienen en cuenta por ejemplo el  $LR_5 = 500$ , el cual asocia V2 también con F1.

Tabla S1

Hipótesis	Víctima	Familia	LR	Probabilidad <i>a posteriori</i>	Suma
H <sub>1</sub>	V1	F1	10 <sup>6</sup>	$\frac{10^6}{10^6 + 10 + 0 + 1} = 0,999989$	0,999989 + 0,00000999989 + 0 + 0,000000999989 = 1
H <sub>2</sub>	V1	F2	10	$\frac{10}{10^6 + 10 + 0 + 1} = 0,00000999989$	
H <sub>3</sub>	V1	F3	0	$\frac{0}{10^6 + 10 + 0 + 1} = 0$	
H <sub>4</sub>	V1	*	1	$\frac{1}{10^6 + 10 + 0 + 1} = 0,000000999989$	
H <sub>5</sub>	V2	F1	500	$\frac{500}{500 + 0 + 0 + 1} = 0,998004$	0,998004 + 0 + 0 + 0,001996 = 1
H <sub>6</sub>	V2	F2	0	$\frac{0}{500 + 0 + 0 + 1} = 0$	
H <sub>7</sub>	V2	F3	0	$\frac{0}{500 + 0 + 0 + 1} = 0$	
H <sub>8</sub>	V2	*	1	$\frac{1}{500 + 0 + 0 + 1} = 0,001996$	
H <sub>9</sub>	V3	F1	0	$\frac{0}{0 + 0 + 0 + 1} = 0$	0 + 0 + 0 + 1 = 1
H <sub>10</sub>	V3	F2	0	$\frac{0}{0 + 0 + 0 + 1} = 0$	
H <sub>11</sub>	V3	F3	0	$\frac{0}{0 + 0 + 0 + 1} = 0$	
H <sub>12</sub>	V3	*	1	$\frac{1}{0 + 0 + 0 + 1} = 1$	

### AM-driven

En la Tabla S2 se muestran las probabilidades *a posteriori* para cada una de las hipótesis del abordaje *AM-driven*. Nótese que para cada familia  $F_i$  la suma de las probabilidades *a posteriori* es 1. No ocurre así para cada víctima. En las hipótesis  $H_1$  a  $H_4$ , sólo se tienen en cuenta los resultados obtenidos de la comparación de  $F_1$  con las 3 víctimas, y por tanto, a diferencia del abordaje *PM-driven*, sí se está teniendo en cuenta  $F_1$  asocia con dos víctimas ( $V_1$  y  $V_2$ ). Sin embargo, no se tiene en cuenta que una de esas víctimas ( $V_1$ ), también asocia con otra familia ( $F_2$ ,  $LR_5 = 10$ ). Más aún, la suma de las probabilidades obtenidas en las filas 1 y 5 es mayor que 1, y por lo tanto es difícil interpretar estos resultados. ¿Cómo evaluar el hecho de que  $V_1$  puede pertenecer tanto a  $F_1$  como a  $F_2$  con valores de probabilidad nada desdeñables? El lector podrá comprobar que con la aproximación global se soluciona este problema.

Tabla S2

Hipótesis	Familia	Víctima	LR	Probabilidad <i>a posteriori</i>	Suma
$H_1$	F1	V1	$10^6$	$\frac{10^6}{10^6 + 500 + 0 + 1} = 0,9995$	$0,9995 + 0,0005 + 0 + 0,0000009995 = 1$
$H_2$	F1	V2	500	$\frac{500}{10^6 + 500 + 0 + 1} = 0,0005$	
$H_3$	F1	V3	0	$\frac{0}{10^6 + 500 + 0 + 1} = 0$	
$H_4$	F1	*	1	$\frac{1}{10^6 + 500 + 0 + 1} = 0,0000009995$	
$H_5$	F2	V1	10	$\frac{10}{10 + 0 + 0 + 1} = 0,9091$	$0,9091 + 0 + 0 + 0,0909 = 1$
$H_6$	F2	V2	0	$\frac{0}{10 + 0 + 0 + 1} = 0$	
$H_7$	F2	V3	0	$\frac{0}{10 + 0 + 0 + 1} = 0$	
$H_8$	F2	*	1	$\frac{1}{10 + 0 + 0 + 1} = 0,0909$	
$H_9$	F3	V1	0	$\frac{0}{0 + 0 + 0 + 1} = 0$	$0 + 0 + 0 + 1 = 1$
$H_{10}$	F3	V2	0	$\frac{0}{0 + 0 + 0 + 1} = 0$	
$H_{11}$	F3	V3	0	$\frac{0}{0 + 0 + 0 + 1} = 0$	
$H_{12}$	F3	*	1	$\frac{1}{0 + 0 + 0 + 1} = 1$	

## Aproximación global

En la Tabla S3, se muestran las hipótesis a tener en cuenta en el enfoque global. MP denota “persona desaparecida” (missing person). El símbolo “\*” significa que no se puede asignar una víctima concreta a una persona desaparecida concreta (es decir a una familia concreta). H<sub>1</sub> contempla que no se ha encontrado ninguna compatibilidad entre las tres víctimas y las tres familias (no relacionados, hipótesis del denominador en el LR).

Tabla S3

Hipótesis	V1	V2	V3	LR
Ninguna víctima				
H <sub>1</sub>	*	*	*	1
Una sola víctima				
H <sub>2</sub>	MP1	*	*	10 <sup>6</sup>
H <sub>3</sub>	*	MP1	*	10
H <sub>4</sub>	*	*	MP1	0
H <sub>5</sub>	MP2	*	*	500
H <sub>6</sub>	*	MP2	*	0
H <sub>7</sub>	*	*	MP2	0
H <sub>8</sub>	MP3	*	*	0
H <sub>9</sub>	*	MP3	*	0
H <sub>10</sub>	*	*	MP3	0
Dos víctimas				
H <sub>11</sub>	MP1	MP2	*	0
H <sub>12</sub>	MP1	*	MP2	0
H <sub>13</sub>	*	MP1	MP2	0
H <sub>14</sub>	MP2	MP1	*	0
H <sub>15</sub>	MP2	*	MP1	0
H <sub>16</sub>	*	MP2	MP1	0
H <sub>17</sub>	MP1	MP3	*	0
H <sub>18</sub>	MP1	*	MP3	0
H <sub>19</sub>	*	MP1	MP3	0
H <sub>20</sub>	MP3	MP1	*	0
H <sub>21</sub>	MP3	*	MP1	0
H <sub>22</sub>	*	MP3	MP1	0
H <sub>23</sub>	MP2	MP3	*	0
H <sub>24</sub>	MP2	*	MP3	0
H <sub>25</sub>	*	MP2	MP3	0
H <sub>26</sub>	MP3	MP2	*	0
H <sub>27</sub>	MP3	*	MP2	0
H <sub>28</sub>	*	MP3	MP2	0
Tres víctimas				
H <sub>29</sub>	MP1	MP2	MP3	0
H <sub>30</sub>	MP1	MP3	MP2	0
H <sub>31</sub>	MP2	MP1	MP3	0
H <sub>32</sub>	MP2	MP3	MP1	0
H <sub>33</sub>	MP3	MP1	MP2	0
H <sub>34</sub>	MP3	MP2	MP1	0

En la Tabla S4 se muestran todas las probabilidades *a posteriori* individuales para cada una de las víctimas. Nótese que el denominador en la columna de la probabilidad *a posteriori*

tiene en cuenta todas las comparaciones entre  $V_i$  y  $F_i$  a la vez, a diferencia de los dos abordajes anteriores (denominador =  $LR_1 + LR_2 + LR_3 + LR_4 + LR_5 + LR_6 + \dots + LR_{34} = 1 + 10^6 + 10 + 0 + 500 + 0 + \dots + 0 = 1.000.511$ ). La suma de las probabilidades de cada fila de la tabla es 1; esta es la clave de la aproximación global, ya que indica que los resultados son consistentes.

Tabla S4

	MP1	MP2	MP3	*
V1	$\frac{10^6}{1000511} = 0,99949$	$\frac{500}{1000511} = 0,0005$	$\frac{0}{1000511} = 0$	$\frac{1 + 10}{1000511} = 0,00001$
V2	$\frac{10}{1000511} = 0,00001$	$\frac{0}{1000511} = 0$	$\frac{0}{1000511} = 0$	$\frac{1 + 10^6 + 500}{1000511} = 0,99999$
V3	$\frac{0}{1000511} = 0$	$\frac{0}{1000511} = 0$	$\frac{0}{1000511} = 0$	$\frac{1 + 10^6 + 10 + 500}{1000511} = 1$

Finalmente, en la Tabla S5 se muestran las probabilidades *a posteriori* para cada hipótesis, es decir para cada solución. Nótese que la solución más probable es  $H_2$ , en la que sólo se identifica a una de las víctimas (V1 asociada a F1). La suma de todas las probabilidades a posteriori es 1.

Tabla S5

Hipótesis	V1	V2	V3	LR	Probabilidades <i>a posteriori</i>
$H_1$	*	*	*	1	$\frac{1}{1000511} = 9,995e - 07$
$H_2$	MP1	*	*	$10^6$	$\frac{10^6}{1000511} = 0,99949$
$H_3$	*	MP1	*	10	$\frac{10}{1000511} = 9,995e - 06$
$H_4$	*	*	MP1	0	$\frac{0}{1000511} = 0$
$H_5$	MP2	*	*	500	$\frac{500}{1000511} = 0,0005$
$H_6$	*	MP2	*	0	$\frac{0}{1000511} = 0$
$H_7$	*	*	MP2	0	$\frac{0}{1000511} = 0$
$H_8 \dots H_{34}$				0...0	$\frac{0}{1000511} = 0 \dots \frac{0}{1000511} = 0$

# Bibliografía

1. Prieto L, Carracedo A. El valor de la prueba de ADN. In Crespillo M, Barrio P, editors. *Genética Forense: del laboratorio a los tribunales.*: Díaz de Santos; 2019.
2. Carracedo A, Prieto L. Más allá del efecto CSI: claves para una buena comunicación de la genética forense. *Mètode Science Studies Journal*. 2018;(97): p. 61-67.
3. Thompson W, Vuille J, Biedermann A, Taroni. The role of prior probability in forensic assessments. *Frontiers in Genetics*. 2013; 4: p. 220.
4. Kling D, Egeland T, Tillmar A, Prieto L. Mass identifications. *Statistical methods in forensic genetics*, ISBN 9780128184233. 1st ed.: Elsevier, Acedemis Press; 2021.
5. Kling D, Tillmar A, Egeland T. Familias 3 - Extensions and new functionality. *Forensic Sci. Int.: Genet*. 2014 Nov; 13: p. 121-127.
6. Vigeland M, Egeland T. Joint DNA-based Disaster Victim Identification. *Research Square*. 2021, enviado; DOI:10.21203/rs.3.rs-296414/v1.
7. CICR. Directrices para el uso de la genética forense en investigaciones sobre derechos humanos y derecho internacional humanitario Ginebra: CICR; 2021.
8. Prinz M, Carracedo A, Mayr W, Morling N, Parsons T, Sajantila A, et al. DNA Commission of the International Society for Forensic Genetics (ISFG): Recommendations regarding the role of forensic genetics for disaster victim identification (DVI). *Forensic Sci. Int.: Genet*. 2007; 1: p. 3-12.
9. Vullo C, Romero M, Catelli L, Šakić M, Saragoni V, Jimenez M, et al. GHEP-ISFG collaborative simulated exercise for DVI/MPI: Lessons learned about large-scale profile database comparisons. *Forensic Sci Int Genet*. 2016; 21: p. 45-53.
10. Vullo C, Catelli L, Ibarra A, Papaioannou A, Álvarez J, López-Parra A, et al. Second GHEP-ISFG exercise for DVI: "DNA-led" victim's identification in a simulated air crash. *Forensic Sci. Int.: Genet*. 2021;: p. aceptado.
11. Brooks S, Suchey J. Skeletal age determination based on the os pubis: a comparison of the Acsádi-Nemeskéri and Suchey-Brooks methods. *Hum. Evol*. 1990; 5: p. 227-238.
12. Scheck B, Neufeld P, Dwyer J. *Actual Innocence* New York: New American Library; 2000.
13. Budowle B, Jianye G, Chakraborty R, Gill-King H. Use of prior odds for missing persons identifications. *Investigative Genetics*. 2011; 2: p. 15.
14. Gill P, Ivanov P, Kimpton C, Piercy R, Benson N, Tully G, et al. Identification of the remains of the Romanov family by DNA analysis. *Nat Genet*. 1994; 6(2): p. 130-135.
15. King T, Fortes G, Balaesque P, Thomas M, Balding D, Desler Masiano P, et al. Identification of the remains of King Richard III. *Nature Communications*. 2014; 5: p. 5631.
16. Parsons T, Huel R. DNA and missing persons identification: practice, progress and perspectives. In A BBaA. *Handbook of Forensic Genetics*. London: London: World Scientific.; 2017. p. 337-376.
17. Biesecker L, Bailey-Wilson J, Ballantyne J, Baum H, Bieber F, Brenner C, et al. DNA identifications after the 9/11 World Trade Center attack. *Science*. 2005 Nov; 310(5751): p. 1222-3.

18. Brenner C, Weir B. Issues and strategies in the DNA identification of World Trade Center victims. *Theor. Popul. Biol.* 2003; 63: p. 173-178.
19. Egeland T. Properties of Likelihood Ratios. [Online].; 2017 [cited 2021 abril 19. Available from: <https://familias.name/mty/Day1-Part3-TE-LR-mty.pdf>.