

ESTIMACIÓN DEL NÚMERO DE PERSONAS LESIONADAS GRAVES Y FALLECIDAS A 30 DÍAS POR ACCIDENTE DE TRÁFICO



**Observatorio Nacional
de Seguridad Vial**

C S B Consorci Sanitari de Barcelona

 **Agència
de Salut Pública**

Edición del informe:

Ana M Novoa

Catherine Pérez

Elena Santamariña-Rubio

Agradecimientos: Pere Arribas

Institución:

Agència de Salut Pública de Barcelona

Realizado para:

Dirección General de Tráfico

Ministerio del Interior

Diciembre 2009

Índice del documento

ÍNDICE DE TABLAS	7
ÍNDICE DE FIGURAS	13
0. RESUMEN EJECUTIVO.....	14
1. ANTECEDENTES	18
Conexión de fuente de datos policiales y hospitalarios	19
2. OBJETIVOS.....	24
Objetivo general:	24
Objetivos específicos:	24
3. METODOLOGÍA.....	25
3.1. Diseño y población de estudio	25
3.2. Fuentes de información.....	25
3.3. Conexión de las bases de datos	26
Selección de las variables de conexión y homogeneización de los datos	26
Cálculo de los pesos parciales.....	27
Valoración de la viabilidad de la conexión	28
Restricción de los pares a comparar (<i>blocking</i>)	30
Comparación de los registros (<i>matching</i>)	30
Estadio de decisión (<i>linking</i>).....	31
Validación del proceso	32
3.4. Estimación de lesionados graves	33
3.5. Estimación del número de fallecidos por accidente tráfico a 30 días	35
4. RESULTADOS	37
4.1. Conexión probabilística	37

Andalucía	38
Selección de casos del CMBD	38
Representatividad de los datos de Andalucía	38
Variables de conexión	41
Análisis de la viabilidad de la conexión	42
Resultados de la conexión	45
Validación de la conexión	51
Aragón	53
Selección de casos del CMBD	53
Representatividad de los datos de Aragón	53
Variables de conexión	56
Análisis de la viabilidad de la conexión	57
Resultados de la conexión	60
Validación de la conexión	66
Cantabria	68
Selección de casos del CMBD	68
Representatividad de los datos de Cantabria	68
Variables de conexión	71
Análisis de la viabilidad de la conexión	72
Resultados de la conexión	74
Validación de la conexión	81
Castilla la Mancha	83
Selección de casos del CMBD	83
Representatividad de los datos de Castilla la Mancha	83
Variables de conexión	85
Análisis de la viabilidad de la conexión	87

Resultados de la conexión	89
Validación de la conexión	96
Galicia	98
Selección de casos del CMBD	98
Representatividad de los datos de Galicia	98
Variables de conexión	101
Análisis de la viabilidad de la conexión	102
Resultados de la conexión	105
Validación de la conexión	111
Madrid	113
Selección de casos del CMBD	113
Representatividad de los datos de Madrid	113
Variables de conexión	116
Análisis de la viabilidad de la conexión	117
4.2. Estimación de lesionados graves	120
Andalucía	120
Aragón	122
Cantabria	124
Castilla la Mancha	126
Galicia	128
Comparación entre comunidades autónomas	130
Factores de conversión según sexo y edad	132
4.3. Estimación de muertos a 30 días	135
Andalucía	135
Aragón	136
Cantabria	137

Castilla la Mancha.....	138
Galicia	139
Comparación entre comunidades autónomas.....	140
Factores de conversión según sexo y edad.....	143
5. DISCUSIÓN.....	146
Conexión probabilística	147
Validación de la conexión.....	150
Estimación de lesionados graves y muertos.....	152
Limitaciones y fortalezas	154
6. CONCLUSIONES Y RECOMENDACIONES	156
Recomendaciones.....	157
7. REFERENCIAS	159

ÍNDICE DE TABLAS

Tabla 1.	Características de la conexión de registros entre bases de datos policiales y hospitalarias utilizados en diferentes estudios. Conexión determinística o probabilística.	22
Tabla 2.	Número de lesionados de tráfico presentes en la base de datos policial, hospitalaria o ambas según los resultados de la conexión probabilística.	34
Tabla 3.	Número de lesionados de tráfico presentes en la base de datos policial, hospitalaria o ambas según los resultados de la conexión probabilística en función de los días de estancia hospitalaria.	35
Tabla 4.	Número de lesionados de tráfico presentes en la base de datos policial, hospitalaria o ambas según los resultados de la conexión probabilística.	36
Tabla 5.	Características de los lesionados de tráfico en Andalucía y el resto de España según gravedad. Base de datos policial. España 2005-2007.	39
Tabla 6.	Letalidad (número de muertos por 1.000 lesionados) en Andalucía y el resto de España. Base de datos policial. España 2005-2007.	40
Tabla 7.	Características de los lesionados de tráfico en Andalucía y el resto de España según gravedad. Base de datos hospitalaria. España 2005-2007.	40
Tabla 8.	Descripción de las variables utilizadas para la conexión probabilística. Andalucía 2005-2007.	41
Tabla 9.	Pesos de coincidencia de las variables de conexión, peso mínimo alcanzado según distintas combinaciones de variables y viabilidad de la conexión según el peso mínimo. Andalucía 2005-2007.	44
Tabla 10.	Distribución del número de coincidencias y su correspondiente rango de pesos (n=19.728). Andalucía 2005-2007.	45
Tabla 11.	Resultados de la conexión probabilística. Andalucía 2005-2007.	47
Tabla 12.	Comparación de las características entre los pares conectados y no conectados de la base de datos de la policía. Análisis bivariado. Andalucía 2005-2007.	48
Tabla 13.	Comparación de las características entre los pares conectados y no conectados de la base de datos hospitalaria. Análisis bivariado. Andalucía 2005-2007.	50

Tabla 14.	Verdaderos positivos y negativos y falsos positivos y negativos de la validación de la conexión probabilística. Andalucía 2005-2007.	52
Tabla 15.	Características de los lesionados de tráfico en Aragón y el resto de España según gravedad. Base de datos policial. España 2005-2007.	54
Tabla 16.	Letalidad (número de muertos por 1.000 lesionados) en Aragón y el resto de España. Base de datos policial. España 2005-2007.	55
Tabla 17.	Características de los lesionados de tráfico en Aragón y el resto de España según gravedad. Base de datos hospitalaria. España 2005-2007.	55
Tabla 18.	Descripción de las variables utilizadas para la conexión probabilística. Aragón 2005-2007.	56
Tabla 19.	Pesos de coincidencia de las variables de conexión, peso mínimo alcanzado según distintas combinaciones de variables y viabilidad de la conexión según el peso mínimo. Aragón 2005-2007.	59
Tabla 20.	Distribución del número de coincidencias y su correspondiente rango de pesos (n=3.375). Aragón 2005-2007.	60
Tabla 21.	Resultados de la conexión probabilística. Aragón 2005-2007.	62
Tabla 22.	Comparación de las características entre los pares conectados y no conectados de la base de datos de la policía. Análisis bivariado. Aragón 2005-2007.	63
Tabla 23.	Comparación de las características entre los pares conectados y no conectados de la base de datos hospitalaria. Análisis bivariado. Aragón 2005-2007.	65
Tabla 24.	Verdaderos positivos y negativos y falsos positivos y negativos de la validación de la conexión probabilística. Aragón 2005-2007.	67
Tabla 25.	Características de los lesionados de tráfico en Cantabria y el resto de España según gravedad. Base de datos policial. España 2005-2007.	69
Tabla 26.	Letalidad (número de muertos por 1.000 lesionados) en Cantabria y el resto de España. Base de datos policial. España 2005-2007.	70
Tabla 27.	Características de los lesionados de tráfico en Cantabria y el resto de España según gravedad. Base de datos hospitalaria. España 2005-2007.	70
Tabla 28.	Descripción de las variables utilizadas para la conexión probabilística. Cantabria 2005-2007.	71

Tabla 29.	Pesos de coincidencia de las variables de conexión, peso mínimo alcanzado según distintas combinaciones de variables y viabilidad de la conexión según el peso mínimo. Cantabria 2005-2007.....	74
Tabla 30.	Distribución del número de coincidencias y su correspondiente rango de pesos (n=937). Cantabria 2005-2007.....	75
Tabla 31.	Resultados de la conexión probabilística. Cantabria 2005-2007.....	77
Tabla 32.	Comparación de las características entre los pares conectados y no conectados de la base de datos de la policía. Análisis bivariado. Cantabria 2005-2007.....	78
Tabla 33.	Comparación de las características entre los pares conectados y no conectados de la base de datos hospitalaria. Análisis bivariado. Cantabria 2005-2007.	80
Tabla 34.	Verdaderos positivos y negativos y falsos positivos y negativos de la validación de la conexión probabilística. Cantabria 2005-2007.	82
Tabla 35.	Características de los lesionados de tráfico en Castilla la Mancha y el resto de España según gravedad. Base de datos policial. España 2005-2007.	84
Tabla 36.	Letalidad (número de muertos por 1.000 lesionados) en Castilla la Mancha y el resto de España. Base de datos policial. España 2005-2007.....	85
Tabla 37.	Características de los lesionados de tráfico en Castilla la Mancha y el resto de España según gravedad. Base de datos hospitalaria. España 2005-2007.....	85
Tabla 38.	Descripción de las variables utilizadas para la conexión probabilística. Castilla la Mancha 2005-2007.	86
Tabla 39.	Pesos de coincidencia de las variables de conexión, peso mínimo alcanzado según distintas combinaciones de variables y viabilidad de la conexión según el peso mínimo. Castilla la Mancha 2005-2007.....	89
Tabla 40.	Distribución del número de coincidencias y su correspondiente rango de pesos (n=5.638). Castilla la Mancha 2005-2007.	90
Tabla 41.	Resultados de la conexión probabilística. Castilla la Mancha 2005-2007.	92
Tabla 42.	Comparación de las características entre los pares conectados y no conectados de la base de datos de la policía. Análisis bivariado. Castilla la Mancha 2005-2007.	93
Tabla 43.	Comparación de las características entre los pares conectados y no conectados de la base de datos hospitalaria. Análisis bivariado. Castilla la Mancha 2005-2007.....	95
Tabla 44.	Verdaderos positivos y negativos y falsos positivos y negativos de la validación de la conexión probabilística. Castilla la Mancha 2005-2007.....	97

Tabla 45.	Características de los lesionados de tráfico en Galicia y el resto de España según gravedad. Base de datos policial. España 2005-2007.....	99
Tabla 46.	Letalidad (número de muertos por 1.000 lesionados) en Galicia y el resto de España. Base de datos policial. España 2005-2007.....	100
Tabla 47.	Características de los lesionados de tráfico en Galicia y el resto de España según gravedad. Base de datos hospitalaria. España 2005-2007.....	100
Tabla 48.	Descripción de las variables utilizadas para la conexión probabilística. Galicia 2005-2007.....	101
Tabla 49.	Pesos de coincidencia de las variables de conexión, peso mínimo alcanzado según distintas combinaciones de variables y viabilidad de la conexión según el peso mínimo. Galicia 2005-2007.....	104
Tabla 50.	Distribución del número de coincidencias y su correspondiente rango de pesos (n=5.183). Galicia 2005-2007.....	105
Tabla 51.	Resultados de la conexión probabilística. Galicia 2005-2007.....	107
Tabla 52.	Comparación de las características entre los pares conectados y no conectados de la base de datos de la policía. Análisis bivariado. Galicia 2005-2007.....	108
Tabla 53.	Comparación de las características entre los pares conectados y no conectados de la base de datos hospitalaria. Análisis bivariado. Galicia 2005-2007.....	110
Tabla 54.	Verdaderos positivos y negativos y falsos positivos y negativos de la validación de la conexión probabilística. Galicia 2005-2007.....	112
Tabla 55.	Características de los lesionados de tráfico en Madrid y el resto de España según gravedad. Base de datos policial. España 2005-2007.....	114
Tabla 56.	Letalidad (número de muertos por 1.000 lesionados) en Madrid y el resto de España. Base de datos policial. España 2005-2007.....	115
Tabla 57.	Características de los lesionados de tráfico en Madrid y el resto de España según gravedad. Base de datos hospitalaria. España 2005-2007.....	115
Tabla 58.	Descripción de las variables utilizadas para la conexión probabilística. Madrid 2005-2007.....	116
Tabla 59.	Pesos de coincidencia de las variables de conexión, peso mínimo alcanzado según distintas combinaciones de variables y viabilidad de la conexión según el peso mínimo. Madrid 2005-2007.....	119

Tabla 60.	Número de lesionados de tráfico presentes en la base de datos policial, hospitalaria o ambas según los resultados de la conexión probabilística. Andalucía 2005-2007.	120
Tabla 61.	Número de lesionados de tráfico presentes en la base de datos policial, hospitalaria o ambas según los resultados de la conexión probabilística en función de los días de estancia hospitalaria. Andalucía 2005-2007.	121
Tabla 62.	Número de lesionados de tráfico presentes en la base de datos policial, hospitalaria o ambas según los resultados de la conexión probabilística. Aragón 2005-2007.....	122
Tabla 63.	Número de lesionados de tráfico presentes en la base de datos policial, hospitalaria o ambas según los resultados de la conexión probabilística en función de los días de estancia hospitalaria. Aragón 2005-2007.....	123
Tabla 64.	Número de lesionados de tráfico presentes en la base de datos policial, hospitalaria o ambas según los resultados de la conexión probabilística. Cantabria 2005-2007.	124
Tabla 65.	Número de lesionados de tráfico presentes en la base de datos policial, hospitalaria o ambas según los resultados de la conexión probabilística en función de los días de estancia hospitalaria. Cantabria 2005-2007.	125
Tabla 66.	Número de lesionados de tráfico presentes en la base de datos policial, hospitalaria o ambas según los resultados de la conexión probabilística. Castilla la Mancha 2005-2007.....	126
Tabla 67.	Número de lesionados de tráfico presentes en la base de datos policial, hospitalaria o ambas según los resultados de la conexión probabilística en función de los días de estancia hospitalaria. Castilla la Mancha 2005-2007.....	127
Tabla 68.	Número de lesionados de tráfico presentes en la base de datos policial, hospitalaria o ambas según los resultados de la conexión probabilística. Galicia 2005-2007.	128
Tabla 69.	Número de lesionados de tráfico presentes en la base de datos policial, hospitalaria o ambas según los resultados de la conexión probabilística en función de los días de estancia hospitalaria. Galicia 2005-2007.	129
Tabla 70.	Número de lesionados graves observados y estimados, diferencia entre ambos y factores de conversión según comunidad autónoma. Andalucía, Aragón, Cantabria, Castilla la Mancha y Galicia 2005-2007.....	130
Tabla 71.	Número de lesionados graves observados, estimados y diferencia entre ambos en España 2005-2007 según los factores de conversión de cada comunidad autónoma y según la media de los factores de conversión. Proyección de los resultados para España.....	131

Tabla 72.	Factores de conversión para el cálculo de lesionados graves según sexo y edad. Andalucía, Aragón, Cantabria, Castilla la Mancha y Galicia 2005-2007.....	133
Tabla 73.	Estimación del número de lesionados graves según sexo y edad y diferencia respecto al número estimado por la DGT. España 2005-2007.	134
Tabla 74.	Número de fallecidos por accidente de tráfico presentes en la base de datos policial, hospitalaria o ambas según los resultados de la conexión probabilística. Andalucía 2005-2007.....	135
Tabla 75.	Número de fallecidos por accidente de tráfico presentes en la base de datos policial, hospitalaria o ambas según los resultados de la conexión probabilística. Aragón 2005-2007.....	136
Tabla 76.	Número de fallecidos por accidente de tráfico presentes en la base de datos policial, hospitalaria o ambas según los resultados de la conexión probabilística. Cantabria 2005-2007.....	137
Tabla 77.	Número de fallecidos por accidente de tráfico presentes en la base de datos policial, hospitalaria o ambas según los resultados de la conexión probabilística. Castilla la Mancha 2005-2007.....	138
Tabla 78.	Número de fallecidos por accidente de tráfico presentes en la base de datos policial, hospitalaria o ambas según los resultados de la conexión probabilística. Galicia 2005-2007.....	139
Tabla 79.	Número de muertos a 30 días observados y estimados, diferencia entre ambos y factores de conversión según comunidad autónoma. Andalucía, Aragón, Cantabria, Castilla la Mancha y Galicia 2005-2007. Proyección de los resultados para España.....	140
Tabla 80.	Número de muertos a 30 días observados, estimados y diferencia entre ambos en España 2005-2007 según los factores de conversión de cada comunidad autónoma y según la media de los factores de conversión.....	142
Tabla 81.	Factores de conversión para el cálculo de muertos de tráfico a 30 días según sexo y edad. Andalucía, Aragón, Cantabria, Castilla la Mancha y Galicia 2005-2007.	144
Tabla 82.	Estimación del número de muertos de tráfico según sexo y edad y diferencia respecto al número de muertos a 24h según datos de la DGT. España 2005-2007.	145

ÍNDICE DE FIGURAS

Figura 1.	Distribución del los pesos de concordancia y punto de corte visual. Andalucía 2005-2007.....	46
Figura 2.	Distribución del los pesos de concordancia en función del número de coincidencias y punto de corte visual. Andalucía 2005-2007.....	46
Figura 3.	Distribución del los pesos de concordancia y punto de corte visual. Aragón 2005-2007.....	61
Figura 4.	Distribución del los pesos de concordancia en función del número de coincidencias y punto de corte visual. Aragón 2005-2007.....	61
Figura 5.	Distribución del los pesos de concordancia y punto de corte visual. Cantabria 2005-2007.....	76
Figura 6.	Distribución del los pesos de concordancia en función del número de coincidencias y punto de corte visual. Cantabria 2005-2007.....	76
Figura 7.	Distribución del los pesos de concordancia y punto de corte visual. Castilla la Mancha 2005-2007.....	91
Figura 8.	Distribución del los pesos de concordancia en función del número de coincidencias y punto de corte visual. Castilla la Mancha 2005-2007.....	91
Figura 9.	Distribución del los pesos de concordancia y punto de corte visual. Galicia 2005.2007.....	106
Figura 10.	Distribución del los pesos de concordancia en función del número de coincidencias y punto de corte visual. Galicia 2005-2007.....	106
Figura 11.	Estimación de heridos graves según diferentes factores de conversión derivados Andalucía, Aragón, Cantabria, Castilla la Mancha, Galicia y media de factores de conversión. España 2005-2007.....	132
Figura 12.	Estimación de muertos a 30 días según diferentes factores de conversión derivados Andalucía, Aragón, Cantabria, Castilla la Mancha, Galicia y media de factores de conversión. España 2005-2008.....	143

0. RESUMEN EJECUTIVO

El trabajo que se presenta se enmarca en un proyecto de investigación europeo cuyo objetivo consiste en establecer una metodología común a nivel europeo para estimar el número de heridos graves por accidente de tráfico.

El objetivo de este trabajo consiste en estimar el número lesionados graves de tráfico y el número de muertos de tráfico a 30 días en España a partir de la conexión probabilística realizada en cada comunidad autónoma entre los registros de la Dirección General de Tráfico (DGT) y el Conjunto Mínimo Básico de Datos de altas hospitalarias (CMBD).

La población de estudio fueron los lesionados y muertos de tráfico en España entre los años 2005 y 2007. Se utilizaron dos bases de datos: 1) la base de datos de accidentes y lesionados de tráfico de la DGT y; 2) el CMBD. Se definieron ingresos hospitalarios por lesión de tráfico aquellos ingresos urgentes, con un diagnóstico de lesión (códigos 800-959.9 del CIE9-MC) y que presentaran un código de causa externa E810-E819 o E826 y/o una “Aseguradora de accidente de tráfico” como régimen de financiación. Se excluyeron los reingresos.

Se tuvo acceso a los datos hospitalarios de seis comunidades autónomas: Andalucía, Aragón, Cantabria, Castilla la Mancha, Galicia y Madrid.

Previamente a la realización de la conexión se evaluó la representatividad de los datos de cada comunidad autónoma, comparándolos con respecto a los datos del resto de España.

Para realizar la conexión, se llevaron a cabo los siguientes pasos:

- 1) Selección de las variables de conexión y homogeneización de los datos. Las variables utilizadas en la conexión fueron: fecha de colisión y de ingreso, provincia, edad, sexo, vehículo implicado y posición respecto al vehículo y localización de la lesión principal. Las categorías de dichas variables se homogeneizaron entre las bases de datos a conectar.

- 2) Valoración de la viabilidad de la conexión.
- 3) Restricción de los pares a comparar (*blocking*). Se dividieron las bases de datos según coincidieran la fecha de la colisión y la fecha de ingreso hospitalaria con una tolerancia de +3 días.
- 4) Comparación de los registros (*matching*). Los registros de las dos bases de datos se conectaron mediante un software creado por la Agencia de Salud Pública de Barcelona, utilizando el método de Fellegi-Sunter modificado.
- 5) Estadio de decisión (*linking*). A partir de los pesos de conexión y de un algoritmo creado por los investigadores, se decidieron los pares reales entre los registros conectados.
- 6) Validación del proceso de conexión. Mediante dos técnicas distintas se calcularon la sensibilidad y especificidad de la conexión.

La conexión resultó viable en 5 de las 6 comunidades autónomas (resultó inviable para la Comunidad de Madrid). La conexión probabilística asignó un registro policial a un 15,6% de los registros hospitalarios en Andalucía, a un 36,2% en Aragón, a un 21,9% en Cantabria y a un 27,4% en Castilla la Mancha y a un 31,0% en Galicia. Entre éstos, en Andalucía el 36,0% correspondió a un lesionado leve según la policía, el 34,8% en Aragón, el 45,9% en Cantabria, el 33,1% en Castilla la Mancha y el 30,4% en Galicia. Este dato refleja la mala clasificación de la gravedad en la base de datos policial, pues estos casos se deberían de haber clasificado como lesionados graves.

En Andalucía se conectaron un 4,5% del total de registros policiales (un 16,8% de los lesionados graves), un 8,9% y 27,5%, respectivamente, en Aragón, un 4,9% y 21,5% en Cantabria, un 8,0% y 25,3% en Castilla la Mancha y un 7,8% y 21,3% en Galicia. En cuanto a los registros no conectados el 14,9% correspondieron a lesionados graves en Andalucía, el 16,8% en Aragón, el 10,3% en Cantabria, el 17,2% en Castilla la Mancha y el 21,7% en Galicia. Estos casos, que se deberían de haber conectado, es de esperar que no lo hayan hecho en la mayoría de los casos debido a la falta de suficientes variables discriminatorias para identificar de forma eficaz todos los pares de registros reales.

La sensibilidad de la conexión fue del 98,3% en Andalucía, del 99,9% en Aragón y Castilla la Mancha y del 100% en Cantabria y Galicia. La especificidad osciló entre un

89,7% en Galicia hasta un 94,7% en Andalucía. A pesar de los elevados valores de sensibilidad y especificidad obtenidos, éstos deben interpretarse con cautela, debido a las limitaciones de los métodos de validación.

Se estimaron los lesionados graves de tráfico a partir de los resultados de la conexión: se calcularon dos factores de conversión (N1 y N2) que se aplicaron al número de lesionados leves y graves de la base de datos de la policía. Los resultados muestran que los datos de la policía infraestiman el número de lesionados graves en un 19,5% en Andalucía, en un 23,2% en Aragón, en un 26,4% en Cantabria, en un 25,2% en Castilla la Mancha y en un 21,0% en Galicia.

Se estimaron los muertos de tráfico a partir de los resultados de la conexión: se calcularon tres factores de conversión (N1, N2 y N3) que se aplicaron al número de lesionados leves, graves y muertos a 24 horas de la base de datos de la policía. Los resultados muestran que el número de muertos a 24 horas de policía infraestima el número de muertos a 30 días en un 19,1% en Andalucía, en un 18,4% en Aragón, en un 15,9% en Cantabria, en un 20,1% en Castilla la Mancha y en un 17,8% en Galicia.

El número de lesionados graves y muertos de tráfico estimados deberían interpretarse con cautela, pues el porcentaje de registros conectados difiere considerablemente del porcentaje esperado. Por lo tanto, es de esperar que los factores de conversión obtenidos sean superiores a los reales, pues deberían haberse conectado más registros policiales y hospitalarios.

De este trabajo se extraen las siguiente recomendaciones:

1. Las conexiones realizadas muestran la importante limitación de no disponer de identificadores personales. Con el fin de poder realizar la conexión de forma válida y fiable sería necesario disponer de datos individuales, tal como los nombres y apellidos, como ya se dispone en varios países, o sino alguna otra información individual como el código postal de residencia del lesionado o la fecha de nacimiento.
2. Sería conveniente poder llevar a cabo una validación de la conexión con nombres y apellidos de manera que permita valorar con un mayor grado de fiabilidad el alcance de las conexiones llevadas a cabo.

3. La calidad del código E de causa externa de la base de datos hospitalaria se debería mejorar.
4. Dadas las diferencias entre las características de los registros conectados y los no conectados, los factores de conversión se deberían calcular por separado según dichas características.
5. Para realizar una estimación del número de lesionados graves y muertos a nivel estatal sería necesario poder disponer de datos hospitalarios de más comunidades autónomas, así como de una mayor representatividad de las lesiones de tráfico en zona urbana.

1. ANTECEDENTES

Los registros de accidentes y víctimas basados exclusivamente en fuentes policiales no son suficientes para conocer de forma fiable el número de lesionados y fallecidos por accidente de tráfico. Hay cierto consenso en que el número de fallecidos a 24h suele estar bien notificado, pero resulta más difícil conocer el número de fallecidos a 30 días. Por otra parte, existe una infranotificación importante en el número total de heridos, así como en la valoración que hace la policía de la gravedad de las lesiones. Además el nivel de infranotificación varía según el perfil de los lesionados. Se infranotifican en mayor medida las personas mayores, los lesionados en ámbito urbano, los lesionados leves, los usuarios de vehículos de dos ruedas y los ocupantes de turismo (Pérez, 2006, 2004) (Amoros, 2006).

Por ello, en los últimos años se han empezado a complementar los datos obtenidos por fuentes policiales con otras fuentes de origen sanitario como pueden ser los registros de ingresos hospitalarios o los registros forenses. La utilización de estas fuentes permite disponer de cifras válidas de víctimas de tráfico, de modo que se puedan llevar a cabo comparaciones internacionales, además de estimar los costes reales que suponen para la sociedad y de planificar los recursos sanitarios. Asimismo, estas fuentes proporcionan información detallada sobre el tipo de lesiones, la región anatómica afectada y su gravedad.

En España durante varios años se ha venido trabajando con la estimación de factores de conversión que se derivaban de estudios de seguimiento en una muestra de heridos. Es una metodología poco precisa, y recientemente ha puesto de manifiesto que además es poco válida. La opción más útil, informativa y sostenible vendría sin duda por la posibilidad de conectar los datos del registro de víctimas por accidente de tráfico con las altas hospitalarias.

Con el objetivo de valorar la posibilidad de utilizar datos de altas hospitalarias, como complemento a los datos policiales, para estimar el número de muertos de tráfico a 30 días se puso en marcha el estudio EMAT-30 (Estudio de la Mortalidad por Accidentes

de Tráfico a 30 días) (Pérez, 2004, 2006). Este estudio constató que la aplicación de los factores de conversión para estimar los muertos a 30 días infraestima el número real de muertos entre un 3 y un 6,6%, en especial entre los grupos de mayor edad.

Como segunda fase de este trabajo se planteó estudiar la viabilidad llevar a cabo una conexión probabilística (sin identificadores personales, y basada en la probabilidad de coincidencia de las variables) de ambas fuentes con las variables disponibles. Se concluyó que con la información disponible no era viable. (Informe Conecta, no publicado). Posteriormente, se analizó la viabilidad de la conexión probabilística si ésta se realiza a nivel de comunidad autónoma. Se concluyó que, con los datos disponibles, a pesar de que la conexión no es viable a nivel estatal, sí que es viable en determinadas comunidades autónomas, si se asume que el hospital de ingreso se encuentra en la misma provincia en la que tuvo lugar la colisión (Cirera, 2006).

Así, ya en el marco de un proyecto de investigación europeo se propuso conectar los datos por CCAA asumiendo que éstas proporcionarían algunos campos más de información como el hospital. Este proyecto tenía como objetivo establecer una metodología común a nivel europeo para estimar el número de heridos graves por accidente de tráfico y establecer una definición consensuada sobre de herido grave. En España, se llevó a cabo una conexión para el segundo semestre del año 2005 de la Comunidad de Castilla-León. Se valoró la validez del proceso y se constató que en este caso la conexión era viable. La validez de la conexión demostró ser buena (sensibilidad y especificidad superiores al 95%), y los factores de conversión obtenidos plausibles. Como limitación importante destaca la falta de representatividad de los resultados obtenidos (Broughton, 2008). Ante estos resultados, la Dirección General de Tráfico se planteó repetir el proceso, incluyendo varios años y varias CCAA, todas si fuera posible.

Conexión de fuente de datos policiales y hospitalarios

Existen principalmente tres métodos para realizar la conexión de dos bases de datos: la conexión manual, la determinística y la probabilística. La conexión manual consiste en determinar de forma visual la concordancia de cada uno de los registros de una base de datos con todos los registros de la otra base de datos. Este método precisa que las bases

de datos a conectar contengan pocos registros y poca información. Además, está sujeto a error, el cual aumenta de forma proporcional al volumen de datos a conectar.

La conexión determinística se basa en la existencia de un identificador único – o combinación de variables - entre ambas bases de datos a conectar. Este método conecta registros a partir de la concordancia exacta de las variables de identificación, por lo tanto está sujeto a la calidad de las variables identificadoras.

En los casos en que el volumen de las bases de datos sea elevado, no se disponga de un identificador único y de calidad, o que el número de variables identificadoras disponibles sea limitado o que éstas tengan poco poder discriminatorio, se recomienda la utilización de la conexión probabilística. Este método se basa en la conexión de registros con la máxima probabilidad de pertenecer a un mismo individuo.

A nivel internacional, son varios los países que han realizado conexión de bases de datos policiales y sanitarias (OECD, 2003), algunos de los cuales llevan a cabo la conexión periódicamente, como es el caso de algunos estados de los EEUU (NHTSA, 1996, 2009) o de una región de Australia (Ferrante, 1993). Sin embargo, existe gran heterogeneidad en la metodología empleada. Algunos países conectan sus datos de forma manual (Cryer, 2001) (Aptel, 1999), mientras que otros han utilizado la conexión determinística, ya sea considerando como caso conectado únicamente aquellos en los que coincidan las variables identificadoras (González, 2006) (Amoros, 2007) (Lai, 2006) o bien permitiendo cierto nivel de tolerancia en las variables (ej. edad +/- 5 años) (Petridou, 2009) (Jarvis, 2000) (TRL 420, 1999) (TRL 173, 1996) (Morrison, 2000) (Tercero, 2004) (Van, 2006).

Existen diversas experiencias sobre conexión probabilística de datos policiales y hospitalarios. Básicamente, se distinguen dos métodos para calcular la probabilidad de que dos registros provenientes de dos fuentes de datos diferentes provengan del mismo individuo: 1) mediante el cálculo de pesos para las variables de conexión (Ferrante, 1993); 2) mediante el cálculo de distancias para las variables de conexión (SafetyNet, 2008). Ambos métodos pretenden reflejar la similitud entre ambos registros para cada variable de conexión teniendo en cuenta la distribución de frecuencias de dicha variable. También existen diferencias entre países en cuanto a las variables de conexión disponibles: mientras que algunos países tienen acceso a las iniciales y apellidos de los registros (ej. Australia, Nueva Zelanda, EEUU), otros no disponen de dicha información (ej. Austria, España), principalmente debido a razones de confidencialidad de datos. En

la Tabla 1 se muestran las características de los estudios que han llevado a cabo una conexión determinística o probabilística.

Tabla 1. Características de la conexión de registros entre bases de datos policiales y hospitalarias utilizados en diferentes estudios. Conexión determinística o probabilística.

Primer autor y año	País y periodo	Tipo de conexión	Variables de conexión y criterio de conexión	Validación de la conexión
Gonzalez, 2006	7 regiones de Alabama, Estados Unidos Octubre 2001 – Mayo 2003	Conexión determinística	Edad, sexo, fecha y hora de la colisión e ingreso, región donde ha tenido lugar la colisión y del hospital y tipo de usuario	--
Tecl, 2008 (SafetyNet)	Distrito de Kromeriz, República Checa 2003 – 2005	Conexión determinística con rango de tolerancia	Año de nacimiento (+/-5), sexo, fecha de colisión e ingreso (+1) y tipo de usuario	-- (revisión manual de los registros conectados... pero no disponen de variables de calidad para ello)
Merényi, 2008 (SafetyNet)	Hospital Károlyi Sándor en Budapest, Hungría 1 agosto 2004 – 31 enero 2006	Conexión determinística con rango de tolerancia	Edad (+/-1), sexo, fecha y hora (+/-1h) de la colisión, localización de la colisión, tipo de vehículo, usuario y mecanismo de la colisión	--
Broughton, 2008 (SafetyNet)	Escocia, Reino Unido 1997 – 2005	Conexión determinística con rango de tolerancia	Edad (+/-1), sexo, gravedad del lesionado (categoría vecina), fecha de colisión e ingreso (+1), tipo de usuario (unión de peatón y ciclista), área de referencia del hospital (permite conexión con áreas vecinas) y código del hospital	--
Jarvis, 2000	Northumbria, Reino Unido Abril 1990 – Marzo 1995 (lesionados graves menores de 15 años)	Conexión determinística con rango de tolerancia	Sexo, edad (+/-1), fecha de colisión e ingreso (+1) y código postal	--
Tercero, 2004	Municipalidad de León, Nicaragua 1993	Conexión determinística con rango de tolerancia	Nombre, edad, sexo y fecha de colisión e ingreso	--
Van, 2006	3 hospitales de la ciudad de Thai Nguyen, Vietnam 2000 – 2004	Conexión determinística con rango de tolerancia	Nombre y apellidos, sexo, edad (+/-2), fecha de colisión e ingreso (+/-2), dirección de la residencia del lesionado	--
Amoros, 2008 (SafetyNet)	Región de Rhône, Francia 1996 – 2001	Conexión determinística con rango de tolerancia y con revisión manual	Mes y año de nacimiento, sexo, fecha (+2) y hora de colisión e ingreso, tipo de usuario La localización de la colisión (incluye la dirección) se utilizó para la revisión manual	Se conectó el año 2001 con el código fonético de los nombres y apellidos – recuperados de la base de datos policial – y se calcularon falsos positivos y negativos
Yannis, 2008 (SafetyNet) Petridou, 2009	Isla de Corfu, Grecia 1996 – 2003	Conexión determinística con rango de tolerancia y con revisión manual de los casos dudosos	Edad (+/-5), sexo, nacionalidad, fecha de colisión e ingreso (+2), tipo de usuario y modo de transporte (unión de motocicleta y ciclomotor)	--

Alsop, 2008	Nueva Zelanda 1995	Conexión probabilística cálculo de pesos	con	Fecha de nacimiento, sexo, gravedad del lesionado, iniciales, apellido, código fonético del apellido, fecha de colisión e ingreso, -- localización geográfica de la colisión y del hospital	
Johnson, 1996 (CODES)	Algunos estados de los Estados Unidos	Conexión probabilística cálculo de pesos	con	Varían en función del estado. La mayoría disponen de fecha de nacimiento, sexo, código de residencia, fecha y hora de colisión e ingreso, localización de la colisión y área de referencia del hospital. Aproximadamente la mitad disponen del nombre y apellidos	Conexión manual de una muestra de registros y/o mediante una revisión de todos aquellos registros que indicaran la presencia de una lesión
Karlson, 1996	Wisconsin, Estados Unidos 1993	Conexión probabilística cálculo de pesos	con	Fecha de nacimiento, sexo, código postal, fecha de colisión e ingreso y localización geográfica de la colisión y del hospital	Cálculo del porcentaje de falsos positivos (aunque los autores no explican cómo lo calculan)
Ferrante, 1993	Western Australia, Australia Octubre 1987 – Diciembre 1988	Conexión probabilística cálculo de pesos	con	Año de nacimiento, sexo, nombre, inicial del segundo nombre, apellido, fecha de colisión e ingreso y tipo de usuario	Revisión manual de una muestra de registros en busca de falsos positivos y negativos
Lujic, 2008	New South Wales, Australia 1 Julio 2000 – 30 Junio 2001	Conexión probabilística con cálculo de pesos y con revisión manual de los casos dudosos	con	Fecha de nacimiento, edad, sexo, iniciales, apellido, código fonético del apellido, código postal y fecha de colisión e ingreso	--
Pérez, 2008 (SafetyNet)	Castilla-León, España Julio 2005 – Diciembre 2005	Conexión probabilística con cálculo de pesos y con revisión manual de casos	con	Edad, sexo, fecha de colisión e ingreso y provincia de la colisión y del hospital	Se conectó la base de datos hospitalaria con una base de datos ficticia que contiene un 70% aleatorio de los registros hospitalarios y 4 veces más casos de otros años a los que se les ha modificado el año de colisión. Cálculo de sensibilidad y especificidad
Hoeglinger, 2008 (SafetyNet)	Austria 2001	Conexión probabilística cálculo de distancias	con	Edad, sexo, nacionalidad, fecha de colisión e ingreso, estado federal del lugar de colisión y del hospital	1) análisis de la distribución de las lesiones (es diferente entre motoristas y ocupantes de vehículo) 2) conexión de registros de años distintos, modificando la fecha para que correspondan al mismo año
Bos, 2008 (SafetyNet)	Holanda 1997 – 2003	Conexión probabilística cálculo de distancias	con	Fecha de nacimiento, sexo, fecha y hora de la colisión e ingreso y provincia de la colisión y del hospital	Los autores argumentan que no se puede hacer validación adecuada, pues no se disponen información para saber con certeza qué registros son correctos. Realizaron una revisión manual del modo de transporte entre los pares conectados

2. OBJETIVOS

Objetivo general:

- Estimar el número de lesionados graves y el número de muertos a 30 días por accidente de tráfico en España a partir de la conexión probabilística realizada en cada comunidad autónoma entre los registros de la Dirección General de Tráfico (DGT) y el Conjunto Mínimo Básico de altas hospitalarias (CMBD).

Objetivos específicos:

- Analizar la viabilidad de la conexión probabilística entre las bases de datos de accidentes y lesionados de tráfico de la DGT y la del CMBD por comunidad autónoma.
- Realizar una conexión probabilística de las bases de datos de la DGT y del CMBD por comunidad autónoma en aquellas comunidades donde la conexión sea viable.
- Estimar el número de lesionados graves de tráfico en España mediante el cálculo de factores de conversión por comunidad autónoma obtenidos a partir de la conexión de las bases de datos de la DGT y del CMBD.
- Estimar el número de muertos por tráfico a 30 días en España mediante el cálculo de factores de conversión por comunidad autónoma obtenidos a partir de la conexión de las bases de datos de la DGT y del CMBD.

3. METODOLOGÍA

3.1. Diseño y población de estudio

El diseño consistió en un estudio transversal. La población de estudio fueron los lesionados y muertos de tráfico en España. El periodo a analizar depende de los años disponibles para cada base de datos. En el caso de la DGT, se dispuso de los años 2000 a 2008, que fueron analizados por separado para cada comunidad autónoma. Los datos del CMBD se envían a nivel de comunidad autónoma, de manera que los años disponibles varían según la región.

3.2. Fuentes de información

Se utilizaron dos bases de datos:

La base de datos de accidentes y lesionados de tráfico de la Dirección General de Tráfico (DGT), que incluye los lesionados de tráfico en España y contiene información sobre las características de la colisión, del vehículo y de los individuos implicados.

El conjunto mínimo básico de datos altas hospitalarias (CMBD), que contiene información sobre el paciente: el sexo, la edad, la fecha de ingreso y los diagnósticos (codificados según la Clasificación Internacional de Enfermedades, novena versión, Modificación Clínica - CIE9-MC -). El tipo de vehículo y la posición en el vehículo se recogen en el cuarto dígito del código E de causa externa, aunque no están disponibles en la mayoría de los casos dado que esta variable no se recoge con exhaustividad.

Se consideraron lesionados de tráfico a aquellos individuos (1) con un ingreso urgente, (2) con un diagnóstico de lesión en cualquier campo de diagnóstico (principal o secundario) (códigos 800-959.9 del CIE9-MC) y (3) en los que conste un código de causa externa E810-E819 o E826 y/o que conste “Aseguradora de accidente de tráfico” como régimen de financiación. Se eliminaron los reingresos, es decir, aquellos

ingresos de un individuo que ingresa durante los 30 días posteriores a la fecha de alta de un ingreso previo.

Previamente a realizar la conexión, se evaluó la representatividad de los datos de cada comunidad autónoma, comparándolos con respecto a los datos del resto de España. Con los datos policiales, se comparó la distribución de las variables sexo, edad, tipo de usuario y tipo de carretera, según gravedad del lesionado, en cada comunidad autónoma respecto al resto de España. También se comparó la letalidad (número de muertos dividido por el número de lesionados de tráfico por 1.000 lesionados) en zona urbana y en carretera. Con los datos hospitalarios se comparó la distribución de las variables sexo, edad y MAIS (Maximum Abbreviated Injury Scale) en cada comunidad autónoma respecto al resto de España.

3.3. Conexión de las bases de datos

En una primera fase del estudio, se llevó a cabo una conexión probabilística entre la base de datos de la DGT a nivel de comunidad autónoma y cada una de las bases de datos autonómicas del CMBD. Se utilizó el método de Fellegi-Sunter modificado, en el cual se escaló utilizando pesos basados en las frecuencias de las variables de conexión (Fellegi, 1969) (Zhu, 2009). Este método aumenta la especificidad del método de Fellegi-Sunter con una mínima disminución de la sensibilidad.

A continuación se detalla el método utilizado para llevar a cabo la conexión de dichas bases de datos:

Selección de las variables de conexión y homogeneización de los datos

Las variables disponibles en la base de datos de la DGT que se han considerado para realizar la conexión son:

- Fecha de la colisión
- Provincia donde se produjo la colisión. Se asume que el hospital se encuentra en la misma provincia donde se produjo la colisión
- Edad (no se dispone de la fecha de nacimiento) de la persona lesionada

- Sexo de la persona lesionada
- Vehículo implicado en la colisión y posición de la persona con respecto al vehículo (conductor, pasajero o peatón) o tipo de usuario (combinación de vehículo y posición)
- Localización de la lesión principal. En un principio, esta variable no se utilizará como variable de conexión porque su validez es muy cuestionable, dado que se recoge bajo criterio policial y no bajo criterio médico. Sin embargo, en el caso de que las variables disponibles no sean suficientes para realizar la conexión, se planteará la utilización de esta variable

Las variables de conexión que se han considerado en el CMBD son:

- Fecha de ingreso hospitalario
- Provincia donde se localiza el hospital
- Edad (extraída a partir de la fecha de nacimiento) de la persona ingresada
- Sexo de la persona ingresada
- Vehículo implicado y posición de la persona con respecto al vehículo (extraídos a partir del cuarto dígito del código E) o tipo de usuario (combinación de vehículo y posición)
- Localización de la lesión del diagnóstico principal

Se homogeneizaron las categorías de las variables de conexión entre ambas bases de datos.

Cálculo de los pesos parciales

Una vez se han seleccionado las variables de conexión, se procede a calcular los pesos parciales de dichas variables, lo cual nos permitirá posteriormente valorar la viabilidad de la conexión y realizar el proceso de conexión en el caso de que éste se haya demostrado viable.

El cálculo de los pesos parciales se ha llevado a cabo a partir de la base de datos del CMBD.

Siendo m la probabilidad de que las variables coincidan sabiendo que el par revisado es un par coincidente y u la probabilidad de que los pares coincidan por azar, se calculan los pesos parciales para cada variable en cada registro conectado:

$W_p = \log_2 (m/u)$ para variables coincidentes entre ambos registros

$W_p = \log_2 (1-m/1-u)$ si las variables no coinciden

$W_p = 0$ si el valor es desconocido en alguno de los registros

El valor de W_p cuando las variables son coincidentes es positivo y, cuando no son coincidentes, es negativo.

Se calcula m restando de 1 la proporción de error asociada a la variable (ej. proporción de valores desconocidos). Para las variables con menos del 10% de datos desconocidos, se considera un 10% de error.

Se calcula u a partir de la distribución de categorías de la variable: es la proporción que representa la categoría del registro que se está considerando sobre el total de categorías de la variable. Dicho cálculo confiere pesos mayores a aquellas categorías menos frecuentes.

Valoración de la viabilidad de la conexión

La viabilidad de la conexión se realizó a partir de dos métodos distintos:

- a) Según las posibles combinaciones de las variables conexión

El número de posibles combinaciones entre las variables de conexión debe de ser superior al de la suma de registros de las 2 bases de datos. Por lo tanto, para cada comunidad autónoma se multiplicará el número de categorías de cada una de las variables de conexión. Este número debe de ser superior al total de registros de ambas bases de datos.

- b) Según el peso mínimo (W_{\min}) alcanzado por las variables de conexión

Este método se basa en el cálculo de dos pesos: 1) el peso de conexión más pequeño (W_{\min}) alcanzado cuando dos registros coinciden en todas sus variables de conexión y éstas están formadas por las categorías más frecuentes de dichas variables y; 2) el peso mínimo (W_t) que será necesario para alcanzar una determinada probabilidad de que dos registros conectados sean un par real, es decir, el peso que se deberá alcanzar para que el proceso de conexión considere conectados dos registros con un error por debajo de un determinado valor.

Por lo tanto, el cálculo de W_{\min} consiste en sumar los pesos parciales de todas las variables de conexión cuando éstas son coincidentes y presentan la categoría más frecuente de cada variable (es decir, para cada variable de conexión, la suma de sus pesos parciales de coincidencia más pequeños).

Para calcular W_t , si queremos conectar los registros con una probabilidad p de que sean correctos, a dicha probabilidad, hay que restarle aquella de encontrar dos registros al azar. Para ello se aplica la siguiente fórmula:

$$W_t = \log_2 [p/(1-p)] - \log_2 [E/(A*B-E)]$$

Siendo:

p = probabilidad de que dos registros sean correctos

A = tamaño de base de datos 1

B = tamaño de base de datos 2

E = número de pares correctos

El número de pares correctos representa el número potencial de pares que deberían conectarse si el proceso de conexión fuera perfecto. Se consideraron dos escenarios. (1) Teóricamente, todos los ingresos hospitalarios por lesión de tráfico deberían estar registrados en el registro de accidentes y lesionados de la DGT, por tanto, el número de pares correctos se calculó como el 100% de los registros de la base de datos del CMBD. (2) Sin embargo, se ha estimado que un 32% de los registros hospitalarios no constan como lesionados graves en la base de datos de la DGT (Pérez, 2006). Algunos de éstos puede que aparezcan como lesionados leves, mientras que otros puede que incluso no

aparezcan en la base de datos. Considerando el peor de los casos, el número de pares correctos se calculó como el 68% de los registros de la base de datos del CMBD.

Así pues, para valorar la viabilidad de la conexión se comparan los valores de W_{\min} y W_t , para determinar si se puede alcanzar W_t con la información contenida en las variables de conexión (W_{\min}).

Se consideraron viables (sensibilidad y especificidad por encima del 90%) aquellas conexiones que cumplieran el siguiente supuesto (Cook, 2001):

$$W_{\min} \geq W_t - 3$$

Restricción de los pares a comparar (*blocking*)

Este paso del proceso consiste en crear bloques de registros en cada una de las bases de datos que cumplan una determinada condición. El proceso de conexión únicamente comparará los registros de las dos bases de datos que pertenezcan a un mismo bloque. De este modo se reduce el número de conexiones que es necesario realizar y no se permite que se conecten dos registros con pocas probabilidades de ser un par real. En el trabajo que se presenta la condición utilizada para la creación de los bloques fue que la fecha de ingreso hospitalario fuera la misma o 3 días posterior a la fecha de la colisión.

Comparación de los registros (*matching*)

Los registros de las dos bases de datos se conectaron mediante un software creado por la Agencia de Salud Pública de Barcelona (ASPB), basado en programación con lenguaje SQL.

Se compararon los registros de cada bloque de comparación de las bases de datos de la DGT y el CMBD (base de datos primaria). Para cada par de registros se asignó un peso parcial a cada variable de conexión (en función de la coincidencia de las variables y de los valores específicos que presentan), los cuales posteriormente se sumaron para obtener el peso total de comparación (W).

Finalmente, dentro de cada bloque de comparación, se conectaron aquellos pares de registros con el peso total de comparación (W) más elevado.

Estadio de decisión (*linking*)

Se consideraron registros concordantes:

- Los registros en los que coinciden todas las variables de conexión
- Los que tienen un W_t superior al W_t mínimo de los registros en los que coinciden todas las variables de conexión

De los restantes, mediante revisión semi-manual, se consideraron registros concordantes:

- En aquellas conexiones que se realicen con las variables vehículo y posición en lugar de usuario, dado que los peatones presentan un valor perdido en la variable vehículo, se considerarán conectados aquellos registros que correspondan a peatones en los que coincidan todas las variables de conexión excepto vehículo
- Los registros en los que coinciden todas las variables de conexión excepto una, siempre que ésta sea un valor perdido en uno o ambos registros
- Los registros en los que coinciden todas las variables de conexión excepto una, siempre y cuando la localización de la lesión principal coincida (con un valor distinto a un valor perdido)
- Los registros en los que coinciden todas las variables de conexión excepto la edad, siempre que ésta difiera en +/-1 año
- Los registros en los que coinciden todas las variables de conexión excepto la edad, siempre que ésta difiera en +/-5 años

Finalmente, el número de registros conectados obtenidos con estos criterios se comparó con el derivado de utilizar un punto de corte visual a partir del histograma de pesos. De los registros enlazados, los pares reales suelen presentar un peso de concordancia elevado, mientras que los pares no-reales un peso más bajo. Esto ocasiona que la distribución de los pesos de concordancia de los registros enlazados suela ser bimodal. Por lo tanto, una manera de discriminar los registros correctos de los incorrectos es estableciendo un punto de corte en el histograma que se corresponde al punto que separa ambas campanas de la curva (Cirera, 2006).

Validación del proceso

La conexión se validó mediante dos técnicas distintas:

- a) Se conectaron dos bases de datos ficticias A y B (Cook, 2001) (Cirera, 2006).

La base de datos A contiene todos los registros hospitalarios del año 2005. Dado que se considera que se pueden conectar el 68% de los registros de la base de datos hospitalaria con la policial, se seleccionaron de forma aleatoria un 68% de los registros de la base de datos A y se incluyeron en la base de datos B, con un número identificador que coincide con el asignado en la base de datos A. A la base de datos B se le añadieron los registros de los años 2006 a 2008, a los cuales se les modificó el año de colisión (a 2005) para que coincidiera con el de la base de datos A.

Se conectaron ambas bases de datos utilizando los mismos criterios comentados en el apartado del estadio de decisión. A partir de los casos conectados y no conectados y utilizando los identificadores de cada registro se pueden conocer los verdaderos positivos (VP), falsos positivos (FP), verdaderos negativos (VN) y falsos negativos (FN), con los cuales se calcularon la sensibilidad y especificidad:

- $\text{Sensibilidad} = \text{VP} / (\text{VP} + \text{FN})$
→ Probabilidad de identificar un par realmente correcto.
- $\text{Especificidad} = \text{VN} / (\text{FP} + \text{VN})$
→ Probabilidad de identificar un par realmente incorrecto

- b) De forma similar a la utilizada en Austria, se conectaron registros no concordantes de años distintos (SafetyNet, 2008):

Se crearon dos bases de datos ficticias A y B. La base de datos A contiene todos los registros hospitalarios ingresados en el año 2006. La base de datos B contiene registros policiales del año 2007 a los que se les ha modificado el año de colisión para que aparezca el año 2006. Se conectaron ambas bases de datos. Dado que los registros no son pares correctos, aquellos que se hayan conectado miden los falsos positivos.

Finalmente, tanto para la base de datos policial como para la hospitalaria, se compararon las características de los pares conectados y los no conectados con el fin de comprobar si existen registros con mayor probabilidad de conectarse.

3.4. Estimación de lesionados graves

Para el cálculo de la estimación de los lesionados graves de tráfico se asume que la conexión entre la base de datos de la DGT y la del CMBD se ha realizado correctamente.

Según el criterio de la DGT, se consideró lesionado grave de tráfico todo aquel lesionado de tráfico que ha sido ingresado en un hospital durante 24 horas o más. Se estima el número total de lesionados graves a partir de la aplicación de factores de conversión (N1 y N2). Estos factores que se aplicaron a los datos de lesionados informados por la policía para obtener el número real de lesionados graves. Para calcular los factores de conversión, se tabularon los individuos de la base de datos policial presentes y ausentes en el CMBD – según los resultados de la conexión –, separando los graves y leves (según policía) y se añadió otra columna para los individuos presentes en el CMBD no conectados (aquellos que la policía no ha notificado) (Tabla 2).

Tabla 2. Número de lesionados de tráfico presentes en la base de datos policial, hospitalaria o ambas según los resultados de la conexión probabilística.

	Policía		CMBD no-policía	Total
	Graves	Leves		
En CMBD				
No en CMBD				
Total				

CMBD: Conjunto Mínimo Básico de Datos de Altas Hospitalarias

Para el cálculo del factor de conversión N1 se estimaron previamente los lesionados graves teniendo en cuenta los individuos presentes en la base de datos hospitalaria y ausentes en la policial a partir de la siguiente fórmula:

$$\begin{aligned} & \text{Lesionados graves corregidos (total)} = \\ & = \text{lesionados graves policía} / \text{total lesionados policía} * \text{total lesionados (policía + CMBD)} \end{aligned}$$

Seguidamente, se calcularon los factores de conversión a partir de las siguientes fórmulas:

$$\text{Factor de conversión N1} = \text{lesionados graves corregidos} / \text{lesionados graves policía}$$

$$\text{Factor de conversión N2} = \text{lesionados leves presentes en CMBD} / \text{lesionados leves policía}$$

Por lo tanto, el número total de lesionados graves se estimó a partir de la siguiente fórmula:

$$\begin{aligned} & \text{Número total estimado de lesionados graves} = \\ & = \text{N1} * \text{lesionados graves según policía} + \text{N2} * \text{lesionados leves según policía} \end{aligned}$$

Los factores de conversión se calcularon para cada comunidad autónoma. Se valoró la utilización de los factores de conversión por comunidad autónoma para estimar el número de muertos a 30 días a nivel nacional.

También se tabularon los lesionados presentes y ausentes en las bases de datos policial y hospitalarias en función de los días de estancia hospitalaria (<1 día, 1-3 días, >3 días). La metodología para calcular los lesionados estimados y los factores de conversión fue similar a la utilizada según los valores de MAIS (Tabla 3).

Tabla 3. Número de lesionados de tráfico presentes en la base de datos policial, hospitalaria o ambas según los resultados de la conexión probabilística en función de los días de estancia hospitalaria.

		Policía		CMBD no-policía	Total
		Graves	Leves		
En CMBD	<1 día				
	1-3 días				
	>3 días				
No en CMBD					
Total					

CMBD: Conjunto Mínimo Básico de Datos de Altas Hospitalarias

3.5. Estimación del número de fallecidos por accidente tráfico a 30 días

Para el cálculo de la estimación del número de fallecidos por tráfico a 30 días se asume que la conexión entre la base de datos de la DGT y la del CMBD se ha realizado correctamente.

El número de fallecidos 30 días se estimó a partir de factores de conversión (N1, N2 y N3) que se aplicaron a la base de datos de la DGT.

Para calcular los factores de conversión, de la base de datos conectada, se tabularon los individuos conectados en función de si fueron dados de alta o murieron antes o después de los 30 días de la colisión, separando los graves y leves (según policía) y se añadió otra columna para los individuos en el CMBD no conectados (aquellos que la policía no ha notificado) (Tabla 4).

Tabla 4. Número de lesionados de tráfico presentes en la base de datos policial, hospitalaria o ambas según los resultados de la conexión probabilística.

	Policía			CMBD no-policía	Total
	Muertos	Graves	Leves		
CMBD-muertos <=30d					
CMBD-alta o muertos >30d					
No en CMBD					
TOTAL					

CMBD: Conjunto Mínimo Básico de Datos de Altas Hospitalarias

Para el cálculo del factor de conversión N1 se estimaron previamente los muertos de tráfico teniendo en cuenta los individuos presentes en la base de datos hospitalaria y ausentes en la policial a partir de la siguiente fórmula:

$$\begin{aligned} & \text{Muertos de tráfico corregidos} = \\ & = \text{muertos policía} / \text{total lesionados policía} * \text{total lesionados (policía + CMBD)} \end{aligned}$$

Seguidamente, se calcularon los pesos a partir de las siguientes fórmulas:

$$\text{Factor de conversión N1} = \text{muertos corregidos} / \text{muertos policía}$$

$$\text{Factor de conversión N2} = \text{graves según policía muertos } \leq 30 \text{ días en CMBD} / \text{graves policía}$$

$$\text{Factor de conversión N3} = \text{leves según policía muertos } \leq 30 \text{ días en CMBD} / \text{leves policía}$$

Por lo tanto,

$$\begin{aligned} & \text{Número estimado de muertos de tráfico} = \\ & = \text{N1} * \text{muertos según policía} + \text{N2} * \text{lesionados graves según policía} + \text{N3} * \text{lesionados} \\ & \quad \text{leves según policía} \end{aligned}$$

Los factores de conversión se calcularon por comunidad autónoma. Se valoró la utilización de los factores de conversión por comunidad autónoma para estimar el número de muertos a 30 días a nivel nacional.

4. RESULTADOS

Se recibieron las bases de datos del CMBD de 6 comunidades autónomas: Andalucía, Aragón, Cantabria, Castilla la Mancha, Galicia y Madrid.

4.1. Conexión probabilística

Se ha sugerido que la eficiencia y efectividad del proceso de conexión aumenta si se eliminan de las bases de datos aquellos registros que se espera que no se encuentren en ambas bases de datos (CODES, 1996b). Por lo tanto, para realizar la conexión, se eliminaron los muertos de la base de datos de policía, pues es de esperar que la mayoría muera en el lugar del accidente o durante su traslado al hospital, y no lleguen a ingresar. Por lo tanto, estos casos no se encontrarán en la base de datos del CMBD.

Dado que el criterio de *blocking* permite conectar casos ingresados hasta 3 días después de la fecha de colisión, se incluyeron los registros de la base de datos policial de los días 29, 30 y 31 de diciembre del año anterior al primer año disponible en la base de datos hospitalaria, para cada comunidad autónoma.

Andalucía

Selección de casos del CMBD

La base de datos del CMBD de Andalucía incluyó las altas hospitalarias entre los años 2005 y 2008. Se seleccionaron las personas con fecha de ingreso en los años 2005 a 2007 y según los criterios comentados en el apartado de metodología.

Representatividad de los datos de Andalucía

La Tabla 5 muestra las características de los lesionados de tráfico en Andalucía en comparación con las de los lesionados en el resto de España, según gravedad. Cabe destacar una menor proporción de muertos usuarios de motocicleta en Andalucía, aunque la proporción fue similar entre lesionados graves y leves, y una mayor proporción de lesionados usuarios de ciclomotor. La proporción de lesionados como peatones también fue menor en Andalucía. La proporción de lesionados leves, graves y muertos en zona urbana fue menor en Andalucía que en el resto de España, en especial en los lesionados graves.

Tabla 5. Características de los lesionados de tráfico en Andalucía y el resto de España según gravedad. Base de datos policial. España 2005-2007.

	Muertos			Graves			Leves		
	Andalucía	Resto de España	Valor de p	Andalucía	Resto de España	Valor de p	Andalucía	Resto de España	Valor de p
N	1.847	8.850		11.832	52.376		57.447	298.797	
Sexo			0,000			0,000			0,146
Hombre	63,6	62,8		75,6	72,1		80,6	78,8	
Mujer	35,8	36,6		24,1	27,4		19,3	21,0	
Desconocido	0,5	0,7		0,3	0,4		0,2	0,3	
Edad			0,000			0,000			0,000
0-13 años	3,7	3,6		2,7	2,9		2,1	2,0	
14-17 años	7,3	5,2		7,1	6,3		3,5	3,3	
18-29 años	37,3	33,9		35,1	31,3		29,2	26,4	
30-44 años	26,3	27,4		29,2	26,7		29,2	27,6	
45-64 años	15,6	16,0		16,9	18,1		22,9	22,0	
65-74 años	3,2	3,8		4,4	5,5		7,6	8,6	
75 y más años	1,5	2,6		2,3	4,9		5,0	9,0	
Desconocido	5,2	7,6		2,3	4,2		0,6	1,1	
Usuario			0,000			0,000			0,000
Turismo	57,4	56,8		45,8	45,4		51,7	52,5	
Motocicleta	7,6	11,4		15,2	14,7		13,1	12,9	
Ciclomotor	19,8	12,8		19,3	13,4		8,8	5,4	
Camión o autobús	7,5	8,6		8,8	8,7		9,8	9,8	
Otros	2,4	2,9		3,4	4,5		4,6	4,6	
Peatón	5,3	7,5		7,5	13,2		12,1	14,6	
Desconocido	0,0	0,1		0,0	0,2		0,0	0,2	
Tipo de carretera			0,000			0,000			0,000
Zona urbana	46,1	50,2		23,6	33,3		12,9	16,5	
Carretera	53,9	49,8		76,4	66,7		87,1	83,5	
Desconocido	0,0	0,0		0,0	0,0		0,0	0,0	

En la Tabla 6 se muestra la letalidad en Andalucía y el resto de España para el total de muertos y según el tipo de carretera donde tuvo lugar la colisión. La letalidad en Andalucía fue similar a la observada en el resto de España, tanto en zona urbana como en carretera.

Tabla 6. Letalidad (número de muertos por 1.000 lesionados) en Andalucía y el resto de España. Base de datos policial. España 2005-2007.

	Andalucía	Resto de España
Zona urbana	8,1	8,7
Carretera	38,7	38,7
Total	26,0	24,6

Finalmente, en la Tabla 7 se comparan las características de los lesionados según datos hospitalarios. No se observaron diferencias destacables entre los lesionados ingresados en Andalucía y el resto de España, a excepción de un mayor porcentaje de individuos con edades comprendidas entre los 14 y los 29 años de edad.

Tabla 7. Características de los lesionados de tráfico en Andalucía y el resto de España según gravedad. Base de datos hospitalaria. España 2005-2007.

	Andalucía	Resto de España	Valor de p
N	15.544	69.235	
Sexo			0,000
Hombre	77,3	72,3	
Mujer	22,7	27,7	
Desconocido	0,0	0,0	
Edad			0,000
0-13 años	6,1	7,8	
14-17 años	12,9	9,4	
18-29 años	32,4	27,1	
30-44 años	23,5	22,6	
45-64 años	15,6	18,4	
65-74 años	5,4	7,4	
75 y más años	4,2	7,3	
Desconocido	0,1	0,0	
MAIS			0,000
1 Leve	7,4	9,6	
2 Moderado	53,9	51,6	
3-6 Grave	38,1	38,0	
Desconocido	0,6	0,8	

MAIS: Maximum Abbreviated Injury Scale

Variables de conexión

En la Tabla 8 se describen las variables de conexión.

Tabla 8. Descripción de las variables utilizadas para la conexión probabilística. Andalucía 2005-2007.

VARIABLES CONEXIÓN	DGT	CMBD
Identificador	<u>Nombre:</u> p_id <u>Formato:</u> %8s <u>Valores perdidos:</u> --	<u>Nombre:</u> h_id <u>Formato:</u> %8s <u>Valores perdidos:</u> --
Sexo	<u>Nombre:</u> p_sexo <u>Formato:</u> float %8.0g <u>Categorías:</u> 1 hombres 2 mujeres <u>Valores perdidos:</u> (.) (1,5%)	<u>Nombre:</u> h_sexo <u>Formato:</u> float %8.0g <u>Categorías:</u> 1 hombres 2 mujeres <u>Valores perdidos:</u> (.) (0,02%)
Edad	<u>Nombre:</u> p_edad <u>Formato:</u> float %8.0g <u>Categorías:</u> 1-99 <u>Valores perdidos:</u> (.) (4,7%)	<u>Nombre:</u> h_edad <u>Formato:</u> float %8.0g <u>Categorías:</u> 1-107 <u>Valores perdidos:</u> --
Provincia	<u>Nombre:</u> p_provin <u>Formato:</u> float %8.0g <u>Categorías:</u> 04 Almería 11 Cádiz 14 Córdoba 18 Granada 21 Huelva 23 Jaén 29 Málaga 41 Sevilla <u>Valores perdidos:</u> --	<u>Nombre:</u> h_provin <u>Formato:</u> float %8.0g <u>Categorías:</u> 04 Almería 11 Cádiz 14 Córdoba 18 Granada 21 Huelva 23 Jaén 29 Málaga 41 Sevilla <u>Valores perdidos:</u> --
Vehículo	<u>Nombre:</u> p_vehiculo <u>Formato:</u> float %8.0g <u>Categorías:</u> 1 No moto-ciclo 2 Moto-ciclo 3 Vehículo tracción animal 4 Bici <u>Valores perdidos:</u> (.) (6,7%)	<u>Nombre:</u> h_vehículo <u>Formato:</u> float %8.0g <u>Categorías:</u> 1 No moto-ciclo 2 Moto-ciclo 3 Vehículo tracción animal 4 Bici <u>Valores perdidos:</u> (.) (62,4%)

Cont.

Posición en vehículo	<u>Nombre: p_posicion</u> <u>Formato: float %8.0g</u> Categorías: 1 Conductor 2 Pasajero 3 Ocupante carro/bici 4 Peatón Valores perdidos: (.) (1,1%)	<u>Nombre: h_posición</u> <u>Formato: float %8.0g</u> Categorías: 1 Conductor 2 Pasajero 3 Ocupante carro/bici 4 Peatón Valores perdidos: (.) (61,4%)
Usuario	<u>Nombre: p_posveh</u> <u>Formato: float %8.0g</u> Categorías: 0 Conductor no moto-ciclo 1 Pasajero no moto-ciclo 2 Conductor moto-ciclo 3 Pasajero no moto-ciclo 5 Ocupante carro 6 Ciclista 7 Peatón Valores perdidos: (.) (1,9%)	<u>Nombre: h_posveh</u> <u>Formato: float %8.0g</u> Categorías: 0 Conductor no moto-ciclo 1 Pasajero no moto-ciclo 2 Conductor moto-ciclo 3 Pasajero no moto-ciclo 5 Ocupante carro 6 Ciclista 7 Peatón Valores perdidos: (.) (52,3%)
Fecha colisión	<u>Nombre: p_fecha_acc</u> <u>Formato: flota %d</u> Valores perdidos: --	<u>Nombre: h_fecha_acc</u> <u>Formato: float %d</u> Valores perdidos: --
Localización lesión (diagnóstico principal) (Para la revisión semi-manual)	<u>Nombre: p_locles</u> <u>Formato: float %8.0g</u> Categorías: 1 Cabeza/cara 2 Cuello 3 Pecho 4 Espalda 5 Abdomen 6 EESS 7 EEII Valores perdidos: (.) (26,7%)	<u>Nombre: h_locles</u> <u>Formato: float %8.0g</u> Categorías: 1 Cabeza/cara 2 Cuello 3 Pecho 4 Espalda 5 Abdomen 6 EESS 7 EEII Valores perdidos: (.) (6,4%)

Análisis de la viabilidad de la conexión

1) Según las posibles combinaciones de las variables conexión:

Tamaño de la base de datos de la DGT 2005-2007 sin muertos = 69.433

Tamaño de la base de datos del CMBD ingresos 2005-2007 = 19.916

Suma de registros: 69.433 + 19.916 = 89.349

VARIABLES DE CONEXIÓN:

- Sexo: 2 categorías
- Edad: 99 categorías
- Provincia: 8 categorías
- Vehículo: 4 categorías
- Posición: 4 categorías
- Usuario: 7 categorías
- Fecha de colisión: 365 días * 3 años = 1.095 categorías

Para la combinación de las variables sexo, edad, provincia, vehículo, posición y fecha colisión:

Combinaciones posibles = $2 * 99 * 8 * 4 * 4 * 1095 = 27.751.680$

Para la combinación de las variables sexo, edad, provincia, usuario y fecha colisión:

Combinaciones posibles = $2 * 99 * 8 * 7 * 1095 = 12.141.360$

Por tanto, según este análisis, la conexión sería viable porque:

$27.751.680 > 89.349$ y $12.141.360 > 89.349$

2) Según el peso mínimo (W_{min}) alcanzado por las variables de conexión:

$$W_t = \log_2(p/(1-p)) - \log_2(E/(A*B-E))$$

P (probabilidad de que dos registros sean correctos) = 0,95

A (tamaño de base de datos 1) = 69.433

B (tamaño de base de datos 2) = 19.916

E (número de pares correctos) = 19.916 (100% según un escenario optimista)
= 13.542 (68% según un escenario pesimista)

Para E=100%: $W_t = \log_2 (0,95/(1-0,95)) - \log_2 (19.916/(69.433*19.916-19.916)) = 20,33$

Para E=68%: $W_t = \log_2 (0,95/(1-0,95)) - \log_2 (19.916/(69.433*19.916-13.542)) = 20,89$

Por lo tanto, para que la conexión sea viable, el peso mínimo deberá ser superior a:

$W_t - 3 = 20,33 - 3 = 17,33$ (para el escenario optimista)

$= 20,88 - 3 = 17,88$ (para el escenario pesimista)

En la Tabla 9 se muestran los cálculos de los pesos de coincidencia y el peso mínimo para distintas combinaciones de variables.

Tabla 9. Pesos de coincidencia de las variables de conexión, peso mínimo alcanzado según distintas combinaciones de variables y viabilidad de la conexión según el peso mínimo. Andalucía 2005-2007.

Variables	Peso de coincidencia	Combinación de variables 1	Combinación de variables 2	Combinación de variables 3	Combinación de variables 4
Fecha de colisión	8,81	X	X	X	X
Provincia de colisión y hospital	2,14	X	X	X	X
Edad	4,41	X	X	X	X
Sexo	0,21	X	X	X	X
Vehículo implicado	0,68	X		X	
Posición respecto el vehículo	0,85	X		X	
Usuario	1,07		X		X
Localización de la lesión principal	1,37			X	X
PESO MÍNIMO		17,10	16,64	18,47	18,01
Viabilidad de la conexión para E=100% (17,33)		No	No	Sí	Sí
Viabilidad de la conexión para E=68% (17,88)		No	No	Sí	Sí

Los resultados de la viabilidad de la conexión muestran que ésta es viable únicamente si se utiliza la localización de la lesión principal como variable de conexión (combinaciones 3 ó 4).

Resultados de la conexión

Las variables utilizadas para la conexión en Andalucía fueron fecha de la colisión, provincia, edad, sexo, vehículo, posición y localización de la lesión. Dado que en esta comunidad autónoma la conexión sin la localización de la lesión principal no era viable, esta variable no se pudo utilizar para la selección semi-manual de pares concordantes.

De los 19.916 ingresos hospitalarios entre los años 2005 y 2007, se encontró un posible registro de conexión de la base de datos policial para 19.728 casos (99,1%) (registros enlazados). De éstos, tan sólo en un 2,5% de registros coincidieron las 7 variables de conexión, y en 8,5% 6 variables (Tabla 10). En la Tabla 10 se muestran, además, los pesos de concordancia mínimos y máximos de estos registros en función del número de variables coincidentes. La distribución de pesos se puede observar en la Figura 1, que se muestra en función del número de coincidencias en la Figura 2. En el caso de Andalucía, el histograma de pesos muestra una distribución bimodal. En estas figuras también se muestra el lugar en el que se sitúa el punto de corte visual, equivalente a un peso de 18,0.

Tabla 10. Distribución del número de coincidencias y su correspondiente rango de pesos (n=19.728). Andalucía 2005-2007.

	n	%	Peso mínimo	Peso máximo
1	38	0,2	0,12	7,55
2	1.308	6,6	0,98	14,10
3	6.340	32,1	2,13	20,05
4	6.522	33,1	4,24	25,05
5	3.363	17,0	5,89	28,11
6	1.673	8,5	7,10	31,78
7	484	2,5	18,99	34,89

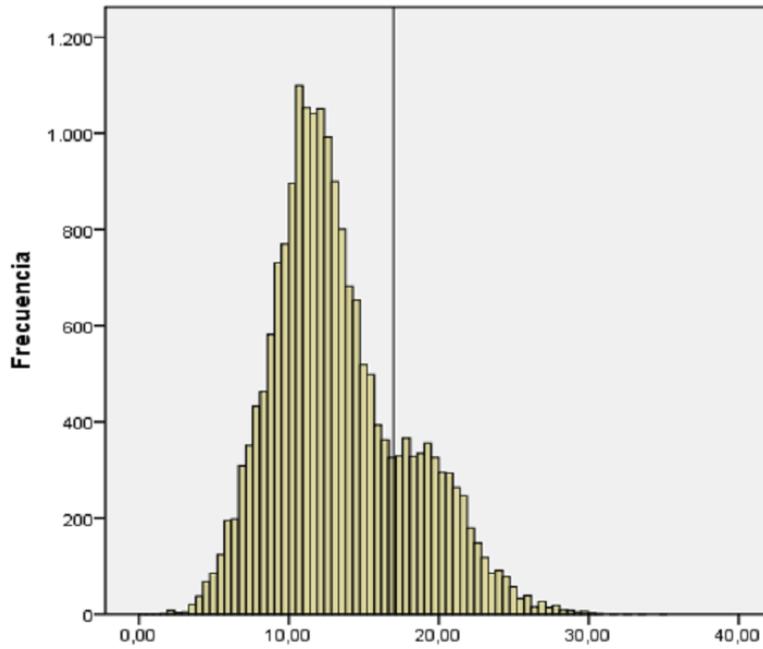


Figura 1. Distribución del los pesos de concordancia y punto de corte visual. Andalucía 2005-2007.

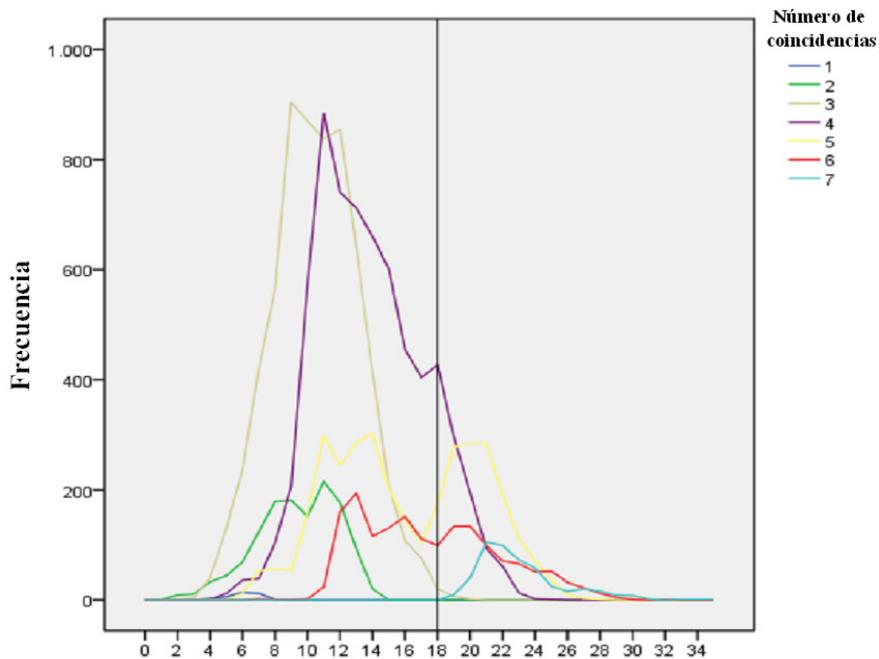


Figura 2. Distribución del los pesos de concordancia en función del número de coincidencias y punto de corte visual. Andalucía 2005-2007.

Según los criterios establecidos en el apartado de metodología, se consideraron conectados 3.112 pares de registros (15,6% de la base de datos inicial; 23,0% del número esperado de registros conectados). De éstos, la mayoría (2.754 registros) se conectaron según el criterio del peso mínimo de los registros en los que coinciden todas las variables de conexión, 180 se conectaron según los criterios establecidos para la revisión semi-manual, 62 registros presentaban una diferencia de edad de sólo un año y 116 registros de hasta 5 años (Tabla 11).

Estos resultados fueron similares a los observados aplicando un punto de corte visual. Según este criterio se conectaron 3.463 registros, que representan un 17,6% de los registros conectables y un 25,6% de los registros que se esperan conectar.

Tabla 11. Resultados de la conexión probabilística. Andalucía 2005-2007.

	n	% conectables	% enlazados	% esperados
Registros conectables	19.916			
Registros enlazados	19.728			
Registros conectados esperados^a	13.542			
Registros no conectados	16.616	83,4	84,2	
Registros conectados	3.112	15,6	15,8	23,0
Según criterio¹^b	2.754	13,8	14,0	20,3
+criterio²^c	180	0,9	0,9	1,3
+criterio³^d	62	0,3	0,3	0,5
+criterio⁴^e	116	0,6	0,6	0,9
Registros no enlazados	188	0,9		

a. Se espera que se conecte un 68% de los registros conectables.

b. Cuando coinciden todas las variables de conexión o cuando el peso de concordancia es superior al peso mínimo de los registros en los que coinciden todas las variables de conexión.

c. Según los criterios de la revisión semi-manual.

d. Se permite una variación de +/-1 año de edad.

e. Se permite una variación de +/-5 años de edad.

Se comprobó si las características de los pares conectados difieren de las de los pares no conectados en la base de datos policial y en la hospitalaria (Tabla 12 y Tabla 13). De los 69.433 registros policiales, los 3.112 casos conectados representaron un 4,5% del total de registros (un 16,8% de los lesionados graves). Se observó una mayor proporción de

hombres que mujeres con registros conectados, así como de menores de 18 años y de mayores de 65 años. En cuanto a la gravedad de los lesionados, aunque la mayoría de los casos conectados fueron identificados por la policía como graves, el 36,0% se identificaron como lesionados leves, que representaron un 1,9% de los lesionados leves de la base de datos policial. En relación a la localización de la lesión principal, la proporción de registros conectados fue mayor entre los individuos con lesiones en la cabeza o la cara y en las extremidades inferiores, y menor en aquellos con lesiones en el cuello. Los usuarios de vehículos de dos ruedas también presentaron una mayor proporción de registros conectados, al igual que los peatones, mientras que para pasajeros fue menor. En cuanto a las características de la colisión, los lesionados en zona urbana presentaron una probabilidad de ser conectados ligeramente superior a los lesionados en carretera. Además, los lesionados en un día laborable en horario nocturno presentaron una menor probabilidad de conectarse. Finalmente, los individuos implicados en una colisión con una sola víctima presentaron una mayor probabilidad de conectarse (Tabla 12).

Tabla 12. Comparación de las características entre los pares conectados y no conectados de la base de datos de la policía. Análisis bivariado. Andalucía 2005-2007.

	Pares no conectados			Pares conectados			Valor de p
	n	% fila	% columna	n	% fila	% columna	
Sexo							0,000
Hombre	42.875	95,0	64,6	2.266	5,0	72,8	
Mujer	22.413	96,4	33,8	839	3,6	27,0	
Desconocido	1.033	99,3	1,6	7	0,7	0,2	
Edad							0,000
0-13 años	2.262	93,0	3,4	170	7,0	5,5	
14-17 años	4.744	94,2	7,2	291	5,8	9,4	
18-29 años	24.594	96,0	37,1	1.015	4,0	32,6	
30-44 años	17.816	95,8	26,9	789	4,2	25,4	
45-64 años	10.436	95,0	15,7	550	5,0	17,7	
65-74 años	2.177	92,8	3,3	170	7,2	5,5	
75 y más años	1.052	92,6	1,6	84	7,4	2,7	
Desconocida	3.240	98,7	4,9	43	1,3	1,4	

Cont.

Gravedad del lesionado							0,000
Leve	56.451	98,1	85,1	1.120	2,0	36,0	
Grave	9.870	83,2	14,9	1.992	16,8	64,0	
Desconocida	--			--			
Localización de la lesión principal							0,000
Cabeza o cara	8.066	93,0	12,2	610	7,0	19,6	
Cuello	14.819	99,3	22,3	106	0,7	3,4	
Pecho	4.317	97,0	6,5	133	3,0	4,3	
Espalda	2.865	96,3	4,3	111	3,7	3,6	
Abdomen	930	94,8	1,4	51	5,2	1,6	
Extremidades superiores	8.685	95,8	13,1	377	4,2	12,1	
Extremidades inferiores	8.821	90,0	13,3	978	10,0	31,4	
Desconocida	17.818	96,0	26,9	746	4,0	24,0	
Vehículo							0,000
Vehículo a motor más de 2 ruedas	42.693	96,8	64,4	1.411	3,2	45,3	
Vehículo a motor de 2 ruedas	18.545	93,4	28,0	1.322	6,7	42,5	
Vehículo tracción animal	26	96,3		1	3,7		
Bicicleta	724	92,0	1,1	63	8,0	2,0	
Desconocido	4.333	93,2	6,5	315	6,8	10,1	
Posición							0,000
Conductor/a	40.004	95,2	60,3	1.999	4,8	64,2	
Pasajero/a	21.234	96,7	32,0	734	3,3	23,6	
Ocupante carro/bici	750	92,1	1,1	64	7,9	2,1	
Peatón/a	3.637	93,0	5,5	275	7,0	8,8	
Desconocida	696	94,6	1,0	40	5,4	1,3	
Tipo de vía							0,000
Carretera	28.318	96,5	42,7	1.031	3,5	33,1	
Zona urbana	38.003	94,8	57,3	2.081	5,2	66,9	
Desconocida	--			--			
Momento de colisión							0,000
Laborable diurno	34.507	95,7	52,0	1.537	4,3	49,4	
Laborable nocturno	9.480	96,3	14,3	365	3,7	11,7	
Fin de semana diurno	11.776	94,3	17,8	713	5,7	22,9	
Fin de semana nocturno	10.558	95,5	15,9	497	4,5	16,0	
Número de víctimas							0,000
1	27.430	94,5	41,4	1.613	5,6	51,8	
2	19.539	96,0	29,5	812	4,0	26,1	
≥3	19.352	96,6	29,2	687	3,4	22,1	

Al comparar las características de los pares conectados y no conectados de los datos hospitalarios, las mujeres presentaron una probabilidad ligeramente superior de conectarse que los hombres. Además, los más jóvenes y los mayores tuvieron una menor probabilidad de ser conectados. En cuanto a la estancia hospitalaria, aquellos individuos con un tiempo más corto de estancia hospitalaria presentaron menos posibilidades de ser conectados. También se observaron algunas diferencias según la localización de la lesión principal. Finalmente, los usuarios de bicicleta presentaron una menor probabilidad de conectarse, mientras que la probabilidad fue mayor en usuarios de vehículo a motor de más de 2 ruedas (Tabla 13).

Tabla 13. Comparación de las características entre los pares conectados y no conectados de la base de datos hospitalaria. Análisis bivariado. Andalucía 2005-2007.

	Pares no conectados			Pares conectados			Valor de p
	n	% fila	% columna	n	% fila	% columna	
Sexo							0,000
Hombre	13.166	85,1	78,4	2.307	14,9	74,1	
Mujer	3.636	81,9	21,6	804	18,1	25,8	
Desconocido	2	66,7	0,0	1	33,3	0,0	
Edad							0,000
0-13 años	1.416	89,1	8,4	173	10,9	5,6	
14-17 años	2.284	88,1	13,6	309	11,9	9,9	
18-29 años	5.142	83,5	30,6	1.014	16,5	32,6	
30-44 años	3.806	82,8	22,6	793	17,2	25,5	
45-64 años	2.542	81,9	15,1	560	18,1	18,0	
65-74 años	901	83,7	5,4	176	16,3	5,7	
75 y más años	713	89,1	4,2	87	10,9	2,8	
Desconocida	0		0,0	0		0,0	
Estancia hospitalaria							0,000
1er cuartil	4.815	86,9	28,7	729	13,1	23,4	
2º cuartil	4.649	85,9	27,7	761	14,1	24,5	
3er cuartil	3.389	84,3	20,2	629	15,7	20,2	
4o cuartil	3.951	79,9	23,5	993	20,1	31,9	

Cont.

Localización de la lesión principal							0,000
Cabeza o cara	3.563	83,1	21,2	727	16,9	23,4	
Cuello	314	81,3	1,9	72	18,7	2,3	
Pecho	1.368	86,5	8,1	214	13,5	6,9	
Espalda	942	85,9	5,6	154	14,1	4,9	
Abdomen	712	90,1	4,2	78	9,9	2,5	
Extremidades superiores	3.078	86,6	18,3	475	13,4	15,3	
Extremidades inferiores	5.757	83,0	34,3	1.180	17,0	37,9	
Desconocida	1.070	83,5	6,4	212	16,5	6,8	
Vehículo							0,000
Vehículo a motor más de 2 ruedas	746	60,8	4,4	480	39,2	15,4	
Vehículo a motor de 2 ruedas	3.812	81,3	22,7	874	18,7	28,1	
Vehículo tracción animal	3	50,0		3	50,0		
Bicicleta	1.513	96,4	9,0	56	3,6	1,8	
Desconocido	10.730	86,3	63,9	1.699	13,7	54,6	
Posición							0,000
Conductor/a	4.161	79,1	24,8	1.099	20,9	35,3	
Pasajero/a	397	60,9	2,4	255	39,1	8,2	
Ocupante carro/bici	1.516	96,3	9,0	59	3,7	1,9	
Peatón/a	1.783	88,8	10,6	226	11,2	7,3	
Desconocida	8.947	85,9	53,2	1.473	14,1	47,3	

Validación de la conexión

Se validó la conexión mediante dos métodos. En primer lugar, se conectó la base de datos A, que corresponde a los registros hospitalarios de Andalucía del año 2005 (n=7.106), con la base de datos B, que contiene el 68% de los registros de la base A y los registros hospitalarios de los años 2006-2008 (n=22.740).

En número de falsos positivos obtenidos mediante la conexión fue de 124 registros (1,7%), y el de falsos negativos fue de 82 (1,2%). En la Tabla 14 se observa la distribución de los registros en función de si son verdaderos positivos o negativos y falsos positivos o negativos. A partir de ello se calculó la sensibilidad y la especificidad de la conexión.

Tabla 14. Verdaderos positivos y negativos y falsos positivos y negativos de la validación de la conexión probabilística. Andalucía 2005-2007.

		Conectados		
		No	Sí	Total
Par real	No	2.197	124	2.321
	Sí	82	4.703	4.785
	Total	2.279	4.827	7.106

Sensibilidad = $VP / (VP + FN) = 4.703 / (4.703 + 82) = 98,3\%$

Especificidad = $VN / (FP + VN) = 2.197 / (124 + 2.197) = 94,7\%$

El segundo método para validar consistió en conectar dos bases de datos de registros no concordantes: la base de datos A, que contiene los registros hospitalarios de Andalucía del año 2006 (n=7.091) y la base de datos B, que contiene los registros policiales de Andalucía del año 2007 (n=23.165). Esta validación nos permitió calcular mediante otro método los falsos positivos: 231 registros (3,3%), lo cual supone una especificidad del 96,7%.

Aragón

Selección de casos del CMBD

La base de datos del CMBD de Aragón incluyó las altas hospitalarias entre los años 2005 y 2008. Se seleccionaron las personas con fecha de ingreso en los años 2005 a 2007 y según los criterios comentados en el apartado de metodología.

Representatividad de los datos de Aragón

La Tabla 15 muestra las características de los lesionados de tráfico en Aragón en comparación con las de los lesionados en el resto de España, según gravedad. Cabe destacar una menor proporción de lesionados usuarios de motocicleta o ciclomotor en Aragón, a expensas de una mayor proporción de usuarios de camión o autobús. Asimismo, la proporción de lesionados en zona urbana fue menor en Aragón que en el resto de España, en especial en los muertos.

Tabla 15. Características de los lesionados de tráfico en Aragón y el resto de España según gravedad. Base de datos policial. España 2005-2007.

	Muertos			Graves			Leves		
	Aragón	Resto de España	Valor de p	Aragón	Resto de España	Valor de p	Aragón	Resto de España	Valor de p
N	497	10.200		2.895	61.313		10.870	345.374	
Sexo			0,125			0,083			0,000
Hombre	63,9	62,9		71,7	72,8		77,9	79,1	
Mujer	35,5	36,5		27,7	26,8		20,9	20,6	
Desconocido	0,6	0,6		0,6	0,4		1,2	0,2	
Edad			0,000			0,000			0,000
0-13 años	4,3	3,6		2,1	2,9		3,6	2,0	
14-17 años	5,2	5,6		4,5	6,6		1,8	3,4	
18-29 años	31,9	34,5		29,5	32,1		15,3	27,4	
30-44 años	26,8	27,2		27,9	27,1		30,4	27,8	
45-64 años	16,4	15,9		20,7	17,8		28,2	21,8	
65-74 años	4,7	3,7		6,2	5,3		8,5	8,4	
75 y más años	3,4	2,4		5,6	4,4		11,3	8,1	
Desconocido	7,5	7,2		3,5	3,9		1,0	1,0	
Usuario			0,000			0,000			0,000
Turismo	57,6	56,9		49,8	45,3		54,5	52,2	
Motocicleta	6,1	10,9		10,4	15,0		5,6	13,3	
Ciclomotor	10,2	14,0		9,0	14,8		3,0	6,1	
Camión o autobús	15,2	8,2		15,0	8,4		17,5	9,5	
Otros	3,1	2,8		4,8	4,3		8,5	4,5	
Peatón	7,9	7,2		11,0	12,2		10,9	14,3	
Desconocido	0,0	0,1		0,1	0,1		0,0	0,2	
Tipo de carretera			0,000			0,000			0,031
Zona urbana	39,7	49,9		27,2	31,7		11,7	16,1	
Carretera	60,3	50,2		72,8	68,3		88,3	83,9	
Desconocido	0,0	0,0		0,0	0,0		0,0	0,0	

En la Tabla 16 se muestra la letalidad en Aragón y el resto de España para el total de muertos y según el tipo de carretera donde tuvo lugar la colisión. La letalidad en zona urbana en Aragón fue similar a la observada en el resto de España, y algo superior en carretera.

Tabla 16. Letalidad (número de muertos por 1.000 lesionados) en Aragón y el resto de España. Base de datos policial. España 2005-2007.

	Aragón	Resto de España
Zona urbana	11,2	8,5
Carretera	48,2	38,3
Total	34,8	24,5

Finalmente, en la Tabla 17 se comparan las características de los lesionados según datos hospitalarios. No se observaron diferencias destacables entre los lesionados ingresados en Aragón y el resto de España, a excepción de la mayor proporción de individuos mayores de 64 años en Aragón.

Tabla 17. Características de los lesionados de tráfico en Aragón y el resto de España según gravedad. Base de datos hospitalaria. España 2005-2007.

	Aragón	Resto de España	Valor de p
N	3.298	80.481	
Sexo			0,091
Hombre	71,9	73,2	
Mujer	28,1	26,8	
Desconocido	0,0	0,0	
Edad			0,000
0-13 años	8,9	7,4	
14-17 años	7,6	10,1	
18-29 años	24,7	28,1	
30-44 años	22,1	22,8	
45-64 años	20,4	17,8	
65-74 años	20,4	7,0	
75 y más años	8,5	6,7	
Desconocido	0,1	0,0	
MAIS			0,000
1 Leve	11,0	9,1	
2 Moderado	54,1	51,9	
3-6 Grave	34,7	38,2	
Desconocido	0,2	0,8	

MAIS: Maximum Abbreviated Injury Scale

Variables de conexión

En la Tabla 18 se describen las variables de conexión.

Tabla 18. Descripción de las variables utilizadas para la conexión probabilística. Aragón 2005-2007.

Variables conexión	DGT	CMBD
Identificador	<u>Nombre:</u> p_id <u>Formato:</u> %8s <u>Valores perdidos:</u> --	<u>Nombre:</u> h_id <u>Formato:</u> %8s <u>Valores perdidos:</u> --
Sexo	<u>Nombre:</u> p_sexo <u>Formato:</u> float %8.0g <u>Categorías:</u> 1 hombres 2 mujeres <u>Valores perdidos:</u> (.) (1,1%)	<u>Nombre:</u> h_sexo <u>Formato:</u> float %8.0g <u>Categorías:</u> 1 hombres 2 mujeres <u>Valores perdidos:</u> (.) (0,03%)
Edad	<u>Nombre:</u> p_edad <u>Formato:</u> float %8.0g <u>Categorías:</u> 1-99 <u>Valores perdidos:</u> (.) (6,7%)	<u>Nombre:</u> h_edad <u>Formato:</u> float %8.0g <u>Categorías:</u> 1-99 <u>Valores perdidos:</u> (.) (0,03%)
Provincia	<u>Nombre:</u> p_provin <u>Formato:</u> float %8.0g <u>Categorías:</u> 22 Huesca 44 Teruel 50 Zaragoza <u>Valores perdidos:</u> --	<u>Nombre:</u> h_provin <u>Formato:</u> float %8.0g <u>Categorías:</u> 22 Huesca 44 Teruel 50 Zaragoza <u>Valores perdidos:</u> --
Vehículo	<u>Nombre:</u> p_vehiculo <u>Formato:</u> float %8.0g <u>Categorías:</u> 1 No moto-ciclo 2 Moto-ciclo 3 Vehículo tracción animal 4 Bici <u>Valores perdidos:</u> (.) (10,1%)	<u>Nombre:</u> h_vehículo <u>Formato:</u> float %8.0g <u>Categorías:</u> 1 No moto-ciclo 2 Moto-ciclo 3 Vehículo tracción animal 4 Bici <u>Valores perdidos:</u> (.) (61,7%)

Cont.

<p>Posición en vehículo</p>	<p>Nombre: p_posicion Formato: float %8.0g Categorías: 1 Conductor 2 Pasajero 3 Ocupante carro/bici 4 Peatón Valores perdidos: (.) (1,6%)</p>	<p>Nombre: h_posición Formato: float %8.0g Categorías: 1 Conductor 2 Pasajero 3 Ocupante carro/bici 4 Peatón Valores perdidos: (.) (50,4%)</p>
<p>Usuario</p>	<p>Nombre: p_posveh Formato: float %8.0g Categorías: 0 Conductor no moto-ciclo 1 Pasajero no moto-ciclo 2 Conductor moto-ciclo 3 Pasajero no moto-ciclo 5 Ocupante carro 6 Ciclista 7 Peatón Valores perdidos: (.) (1,6%)</p>	<p>Nombre: h_posveh Formato: float %8.0g Categorías: 0 Conductor no moto-ciclo 1 Pasajero no moto-ciclo 2 Conductor moto-ciclo 3 Pasajero no moto-ciclo 5 Ocupante carro 6 Ciclista 7 Peatón Valores perdidos: (.) (50,4%)</p>
<p>Fecha colisión</p>	<p>Nombre: p_fecha_acc Formato: flota %d Valores perdidos: --</p>	<p>Nombre: h_fecha_acc Formato: float %d Valores perdidos: --</p>
<p>Localización lesión (diagnóstico principal) (Para la revisión semi-manual)</p>	<p>Nombre: p_locles Formato: float %8.0g Categorías: 1 Cabeza/cara 2 Cuello 3 Pecho 4 Espalda 5 Abdomen 6 EESS 7 EEII Valores perdidos: (.) (38,5%)</p>	<p>Nombre: h_locles Formato: float %8.0g Categorías: 1 Cabeza/cara 2 Cuello 3 Pecho 4 Espalda 5 Abdomen 6 EESS 7 EEII Valores perdidos: (.) (9,8%)</p>

Análisis de la viabilidad de la conexión

1) Según las posibles combinaciones de las variables conexión:

Tamaño de la base de datos de la DGT 2005-2007 sin muertos = 13.795

Tamaño de la base de datos del CMBD ingresos 2005-2007 = 3.377

Suma de registros: $13.795 + 3.377 = 17.172$

Variables de conexión:

- Sexo: 2 categorías
- Edad: 99 categorías
- Provincia: 3 categorías
- Vehículo: 4 categorías
- Posición: 4 categorías
- Usuario: 7 categorías
- Fecha de colisión: 365 días * 3 años = 1.095 categorías

Para la combinación de las variables sexo, edad, provincia, vehículo, posición y fecha colisión:

Combinaciones posibles = $2*99*3*4*4*1095 = 10.406.880$

Para la combinación de las variables sexo, edad, provincia, usuario y fecha colisión:

Combinaciones posibles = $2*99*3*7*1095 = 4.553.010$

Por tanto, según este análisis, la conexión sería viable porque $10.406.880 > 17.172$ y $4.553.010 > 17.172$

2) Según el peso mínimo (W_{min}) alcanzado por las variables de conexión:

$$W_t = \log_2(p/(1-p)) - \log_2(E/(A*B-E))$$

P (probabilidad de que dos registros sean correctos) = 0,95

A (tamaño de base de datos 1) = 13.795

B (tamaño de base de datos 2) = 3.377

E (número de pares correctos) = 3.377 (100% según un escenario optimista)
= 2.296 (68% según un escenario pesimista)

Para E=100%: $W_t = \log_2(0,95/(1-0,95)) - \log_2(3.377/(13.795*3.377-3.377)) = 18,00$

Para E=68%: $W_t = \log_2(0,95/(1-0,95)) - \log_2(2.296/(13.795*3.377-2.296)) = 18,56$

Por lo tanto, para que la conexión sea viable, el peso mínimo deberá ser superior a:

$W_t - 3 = 18,00 - 3 = 15,00$ (para el escenario optimista)

$= 18,56 - 3 = 15,56$ (para el escenario pesimista)

En la Tabla 19 se muestran los cálculos de los pesos de coincidencia y el peso mínimo para distintas combinaciones de variables.

Tabla 19. Pesos de coincidencia de las variables de conexión, peso mínimo alcanzado según distintas combinaciones de variables y viabilidad de la conexión según el peso mínimo. Aragón 2005-2007.

Variables	Peso de coincidencia	Combinación de variables 1	Combinación de variables 2	Combinación de variables 3	Combinación de variables 4
Fecha de colisión	7,98	X	X	X	X
Provincia de colisión y hospital	0,53	X	X	X	X
Edad	5,18	X	X	X	X
Sexo	0,32	X	X	X	X
Vehículo implicado	1,32	X		X	
Posición respecto el vehículo	1,17	X		X	
Usuario	1,74		X		X
Localización de la lesión principal	1,94			X	X
PESO MÍNIMO		16,50	15,75	18,44	17,69
Viabilidad de la conexión para E=100% (15,00)		Sí	Sí	Sí	Sí
Viabilidad de la conexión para E=68% (15,56)		Sí	Sí	Sí	Sí

Los resultados de la viabilidad de la conexión muestran que ésta es viable con cualquier combinación de variables, tanto para el escenario optimista como para el pesimista.

Resultados de la conexión

Las variables utilizadas para la conexión en Aragón fueron: fecha de la colisión, provincia, edad, sexo y usuario. Dado que la conexión resultó viable sin la variable localización de la lesión principal, ésta se utilizó únicamente en la selección semi-manual de pares concordantes.

De los 3.377 ingresos hospitalarios entre los años 2005 y 2007, se encontró un posible registro de conexión de la base de datos policial para 3.375 casos (99,9%) (registros enlazados). De éstos, tan sólo en un 11,4% de registros coincidieron las 5 variables de conexión, y en 31,7% 4 variables (Tabla 20). En esta tabla se muestran, además, los pesos de concordancia mínimos y máximos de estos registros en función del número de variables coincidentes. La distribución de pesos se puede observar en la Figura 3, que se muestra en función del número de coincidencias en la Figura 4. En el caso de Aragón, el histograma de pesos muestra una distribución bimodal. En estas figuras también se muestra el lugar en el que se sitúa el punto de corte visual, equivalente a un peso de 14,0.

Tabla 20. Distribución del número de coincidencias y su correspondiente rango de pesos (n=3.375). Aragón 2005-2007.

	n	%	Peso mínimo	Peso máximo
0	0	0,0	--	--
1	36	1,1	-6,85	8,06
2	418	12,4	-6,62	13,65
3	1.466	43,4	2,41	18,15
4	1.071	31,7	4,52	24,16
5	384	11,4	16,34	24,72

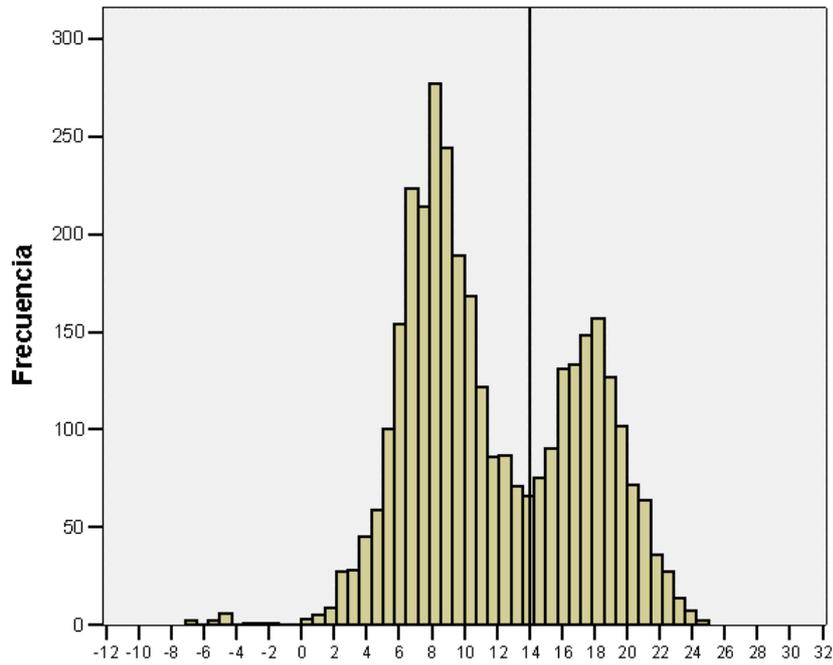


Figura 3. Distribución de los pesos de concordancia y punto de corte visual. Aragón 2005-2007.

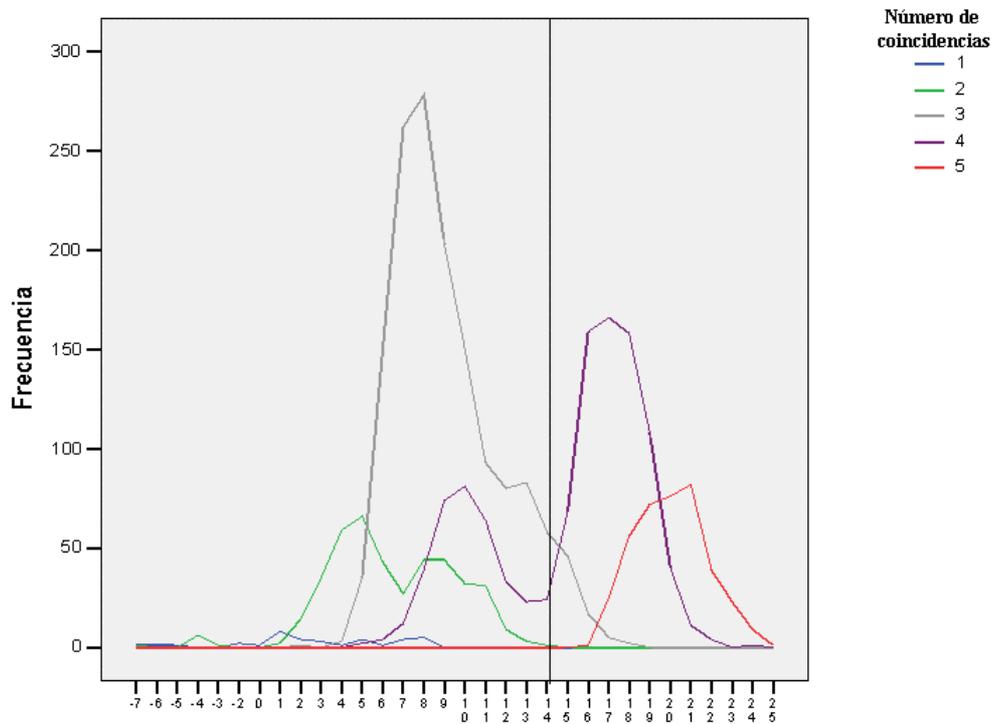


Figura 4. Distribución de los pesos de concordancia en función del número de coincidencias y punto de corte visual. Aragón 2005-2007.

Según los criterios establecidos en el apartado de metodología, se consideraron conectados 1.224 pares de registros (36,2% de la base de datos inicial; 53,3% del número esperado de registros conectados). De éstos, la mayoría (909 registros) se conectaron según el criterio del peso mínimo de los registros en los que coinciden todas las variables de conexión, 232 se conectaron según los criterios establecidos para la revisión semi-manual, 51 registros presentaban una diferencia de edad de sólo un año y 32 registros de hasta 5 años (Tabla 21).

Estos resultados fueron similares a los observados aplicando un punto de corte visual. Según este criterio se conectaron 1.210 registros, que representan un 35,8% de los registros conectables y un 52,7% de los registros que se esperan conectar.

Tabla 21. Resultados de la conexión probabilística. Aragón 2005-2007.

	n	% conectables	% enlazados	% esperados
Registros conectables	3.377			
Registros enlazados	3.375			
Registros conectados esperados^a	2.296			
Registros no conectados	2.151	63,7	63,7	
Registros conectados	1.224	36,2	36,3	53,3
Según criterio¹^b	909	26,9	26,9	39,6
+criterio ² ^c	232	6,9	6,9	10,1
+criterio ³ ^d	51	1,5	1,5	2,2
+criterio ⁴ ^e	32	0,9	0,9	1,4
Registros no enlazados	2	0,1		

a. Se espera que se conecte un 68% de los registros conectables.

b. Cuando coinciden todas las variables de conexión o cuando el peso de concordancia es superior al peso mínimo de los registros en los que coinciden todas las variables de conexión.

c. Según los criterios de la revisión semi-manual.

d. Se permite una variación de +/-1 año de edad.

e. Se permite una variación de +/-5 años de edad.

Se comprobó si las características de los pares conectados difieren de las de los pares no conectados en la base de datos policial y en la hospitalaria (Tabla 22 y Tabla 23). De los 13.795 registros policiales de Aragón, los 1.224 casos conectados representan un 8,9% del total de registros (un 27,5% de los lesionados graves). La proporción de registros

conectados fue ligeramente superior en hombres que en mujeres. También fue superior en menores de 18 años y mayores de 45 años. En cuanto a la gravedad de los lesionados, aunque la mayoría de los casos conectados fueron identificados por la policía como graves, el 34,8% se identificaron como lesionados leves, que representan un 3,9% de los lesionados leves de la base de datos policial. En relación a la localización de la lesión principal, la proporción de registros conectados es mayor entre los individuos con lesiones en la cabeza o la cara y en las extremidades inferiores, y menor en aquellos con lesiones en el cuello. Los usuarios de vehículos de dos ruedas también presentaron una mayor proporción de registros conectados, al igual que los peatones, mientras que para pasajeros fue menor. En cuanto a las características de la colisión, la proporción de registros conectados fue ligeramente superior entre los individuos lesionados en zona urbana, en comparación con carretera ($p > 0,05$), y en aquellas colisiones ocurridas en horario diurno. Además, aquellos individuos implicados en una colisión con una sola víctima tienen una mayor probabilidad de haberse conectado (Tabla 22).

Tabla 22. Comparación de las características entre los pares conectados y no conectados de la base de datos de la policía. Análisis bivariado. Aragón 2005-2007.

	Pares no conectados			Pares conectados			Valor de p
	n	% fila	% columna	n	% fila	% columna	
Sexo							<0,001
Hombre	8.140	90,6	64,8	848	9,4	69,3	
Mujer	4.282	92,0	34,1	374	8,0	30,6	
Desconocido	149	98,7	1,2	2	1,3	0,2	
Edad							<0,001
0-13 años	450	86,0	3,6	73	14,0	6,0	
14-17 años	613	88,7	4,9	78	11,3	6,4	
18-29 años	3.991	92,2	31,7	336	7,8	27,5	
30-44 años	3.433	92,2	27,3	292	7,8	23,9	
45-64 años	2.125	89,1	16,9	259	10,9	21,2	
65-74 años	610	88,7	4,9	78	11,3	6,4	
75 y más años	451	84,6	3,6	82	15,4	6,7	
Desconocida	898	97,2	7,1	26	2,8	2,1	

Cont.

Gravedad del lesionado							<0,001
Leve	10.464	96,1	83,2	426	3,9	34,8	
Grave	2.107	72,5	16,8	798	27,5	65,2	
Desconocida	--			--			
Localización de la lesión principal							<0,001
Cabeza o cara	1.332	86,9	10,6	200	13,1	16,3	
Cuello	2.218	97,2	17,6	64	2,8	5,2	
Pecho	821	91,3	6,5	78	8,7	6,4	
Espalda	602	89,6	4,8	70	10,4	5,7	
Abdomen	178	82,4	1,4	38	17,6	3,1	
Extremidades superiores	1.170	91,8	9,3	105	8,2	8,6	
Extremidades inferiores	1.361	84,6	10,8	248	15,4	20,3	
Desconocida	4.889	92,1	38,9	421	7,9	34,4	
Vehículo							<0,001
Vehículo a motor más de 2 ruedas	9.128	92,8	72,6	708	7,2	57,8	
Vehículo a motor de 2 ruedas	2.044	87,8	16,3	285	12,2	23,3	
Vehículo tracción animal	--			--			
Bicicleta	202	84,5	1,6	37	15,5	3,0	
Desconocido	1.197	86,1	9,5	194	13,9	15,8	
Posición							<0,001
Conductor/a	6.922	91,2	55,1	671	8,8	54,8	
Pasajero/a	4.250	93,0	33,8	322	7,0	26,3	
Ocupante carro/bici	202	84,5	1,6	37	15,5	3,0	
Peatón/a	1.002	85,2	8,0	174	14,8	14,2	
Desconocida	195	90,7	1,6	20	9,3	1,6	
Tipo de vía							0,213
Carretera	4.684	93,2	37,3	434	6,8	35,5	
Zona urbana	7.887	90,9	62,7	790	9,1	64,5	
Desconocida	--			--			
Momento de colisión							<0,001
Laborable diurno	6.780	91,1	53,9	666	8,9	54,4	
Laborable nocturno	1.547	93,5	12,3	107	6,5	8,7	
Fin de semana diurno	2.601	88,9	20,7	324	11,1	26,5	
Fin de semana nocturno	1.643	92,8	13,1	127	7,2	10,4	
Número de víctimas							<0,001
1	5.739	90,0	45,7	638	10,0	52,1	
2	3.239	92,6	25,8	260	7,4	21,2	
>=3	3.593	91,7	28,6	326	8,3	26,6	

Al comparar las características de los pares conectados y no conectados de los datos hospitalarios, no se presentaron diferencias entre hombres y mujeres en la probabilidad de conectarse. Sin embargo, sí que se observaron diferencias en función de la edad, de manera que los más jóvenes y los mayores tuvieron una menor probabilidad de ser conectados. No se observaron diferencias significativas en función de la estancia hospitalaria ni en función de la localización de la lesión principal. Finalmente, los usuarios de bicicleta y peatones presentaron una menor probabilidad de conectarse, mientras que la probabilidad fue mayor en usuarios de vehículo a motor de más de 2 ruedas. Aquellas variables asociadas a una menor probabilidad de que los individuos sean conectados señalan las características de los individuos con una menor probabilidad de estar presentes en la base de datos policial (Tabla 23).

Tabla 23. Comparación de las características entre los pares conectados y no conectados de la base de datos hospitalaria. Análisis bivariado. Aragón 2005-2007.

	Pares no conectados			Pares conectados			Valor de p
	n	% fila	% columna	n	% fila	% columna	
Sexo							0,065
Hombre	1.578	65,1	73,3	854	34,9	69,8	
Mujer	574	60,8	26,7	370	39,2	30,2	
Desconocido	1			--			
Edad							<0,001
0-13 años	235	74,1	10,9	82	25,9	6,7	
14-17 años	180	69,8	8,4	78	30,2	6,4	
18-29 años	503	60,1	23,4	334	39,9	27,3	
30-44 años	446	60,0	20,7	297	40,0	24,3	
45-64 años	421	61,5	19,6	263	38,5	21,5	
65-74 años	172	66,4	8,0	87	33,6	7,1	
75 y más años	196	70,3	9,1	83	29,7	6,8	
Desconocida	--			--			
Estancia hospitalaria							0,391
1er cuartil	584	64,9	27,1	316	35,1	25,8	
2º cuartil	599	64,1	27,8	336	35,9	27,5	
3er cuartil	475	64,7	22,1	259	35,3	21,2	
4o cuartil	495	61,3	23,0	313	38,7	25,6	

Cont.

Localización de la lesión principal							0,118
Cabeza o cara	500	64,4	23,2	277	35,6	22,6	
Cuello	67	65,7	3,1	35	34,3	2,9	
Pecho	255	59,4	11,8	174	40,6	14,2	
Espalda	204	65,4	9,5	108	34,6	8,8	
Abdomen	81	70,4	3,8	34	29,6	2,8	
Extremidades superiores	342	66,3	15,9	174	33,7	14,2	
Extremidades inferiores	483	60,9	22,4	310	39,1	25,3	
Desconocida	221	66,4	10,3	112	33,6	9,2	
Vehículo							<0,001
Vehículo a motor más de 2 ruedas	186	42,5	8,6	252	57,5	20,6	
Vehículo a motor de 2 ruedas	306	59,0	14,2	213	41,0	17,4	
Vehículo tracción animal	--			--			
Bicicleta	301	89,1	14,0	37	10,9	3,0	
Desconocido	1.360	65,3	63,2	722	34,7	59,0	
Posición							<0,001
Conductor/a	391	52,6	18,2	353	47,4	28,8	
Pasajero/a	101	47,4	4,7	112	52,6	9,2	
Ocupante carro/bici	301	89,1	14,0	37	10,9	3,0	
Peatón/a	226	59,3	10,5	155	40,7	12,7	
Desconocida	1.134	66,7	52,7	567	33,3	46,3	

Validación de la conexión

Se validó la conexión mediante dos métodos. En primer lugar, se conectó la base de datos A, que corresponde a los registros hospitalarios de Aragón del año 2005 (n=1.167), con la base de datos B, que contiene el 68% de los registros de la base A y los registros hospitalarios de los años 2006-2008 (n=3.917).

En número de falsos positivos obtenidos mediante la conexión fue de 35 registros (3,0%). Se obtuvo únicamente un falso negativo. En la Tabla 24 se observa la distribución de los registros en función de si son verdaderos positivos o negativos y falsos positivos o negativos. A partir de ello se calculó la sensibilidad y la especificidad de la conexión.

Tabla 24. Verdaderos positivos y negativos y falsos positivos y negativos de la validación de la conexión probabilística. Aragón 2005-2007.

		Conectados		
		No	Sí	Total
Par real	No	345	35	380
	Sí	1	786	787
	Total	346	821	1.167

$$\text{Sensibilidad} = \text{VP} / (\text{VP} + \text{FN}) = 786 / (786 + 1) = 99,9\%$$

$$\text{Especificidad} = \text{VN} / (\text{FP} + \text{VN}) = 345 / (35 + 345) = 90,8\%$$

El segundo método para validar consistió en conectar dos bases de datos de registros no concordantes: la base de datos A, que contiene los registros hospitalarios de Aragón del año 2006 (n=1.114) y la base de datos B, que contiene los registros policiales de Aragón del año 2007 (n=4.691). Esta validación nos permitió calcular mediante otro método los falsos positivos: 105 registros (6,3%), lo cual supone una especificidad del 90,6%.

Cantabria

Selección de casos del CMBD

La base de datos del CMBD de Cantabria incluyó las altas hospitalarias entre los años 2005 y 2008. Se seleccionaron las personas con fecha de ingreso en los años 2005 a 2007 y según los criterios comentados en el apartado de metodología.

Representatividad de los datos de Cantabria

La Tabla 25 muestra las características de los lesionados de tráfico en Cantabria en comparación con las de los lesionados en el resto de España, según gravedad. Cabe destacar una mayor proporción de muertos usuarios de turismo en Cantabria, a expensas de una menor proporción de usuarios de motocicleta o ciclomotor. Asimismo, la proporción de muertos en zona urbana es mucho menor en Cantabria (7,4%) que en el resto de España (50,0%). Estas diferencias se mantienen, aunque de forma menos evidente, entre los lesionados graves y leves.

Tabla 25. Características de los lesionados de tráfico en Cantabria y el resto de España según gravedad. Base de datos policial. España 2005-2007.

	Muertos			Graves			Leves		
	Cantabria	Resto de España	Valor de p	Cantabria	Resto de España	Valor de p	Cantabria	Resto de España	Valor de p
N	106	10.591		521	63.687		3.673	352.571	
Sexo			0,000			0,498			0,769
Hombre	60,3	63,0		74,7	72,8		81,1	79,0	
Mujer	39,6	36,4		25,1	26,8		18,9	20,7	
Desconocido	0,2	0,6		0,2	0,4		0,0	0,3	
Edad			0,000			0,479			0,557
0-13 años	4,4	3,6		2,5	2,9		0,9	2,0	
14-17 años	5,0	5,6		6,0	6,5		5,7	3,3	
18-29 años	34,6	34,5		33,8	32,0		29,3	26,9	
30-44 años	28,6	27,2		27,8	27,2		30,2	27,9	
45-64 años	17,4	15,9		18,4	17,9		16,0	22,2	
65-74 años	4,7	3,7		5,0	5,3		8,5	8,4	
75 y más años	1,6	2,4		4,6	4,4		9,4	8,3	
Desconocido	3,8	7,2		1,9	3,9		0,0	1,0	
Usuario			0,000			0,002			0,369
Turismo	73,3	56,7		48,9	45,4		52,8	52,3	
Motocicleta	5,1	10,8		16,9	14,8		19,8	12,8	
Ciclomotor	6,0	14,0		10,8	14,5		5,7	6,0	
Camión o autobús	8,9	8,4		7,3	8,8		5,7	9,9	
Otros	3,7	2,8		6,7	4,3		3,8	4,7	
Peatón	2,9	7,2		9,4	12,1		12,3	14,2	
Desconocido	0,0	0,0		0,0	0,1		0,0	0,2	
Tipo de carretera			0,000			0,000			0,295
Zona urbana	7,4	50,0		8,3	31,7		10,4	15,9	
Carretera	92,7	50,0		91,8	68,3		89,6	84,1	
Desconocido	0,0	0,0		0,0	0,0		0,0	0,0	

En la Tabla 26 se muestra la letalidad en Cantabria y el resto de España para el total de muertos y según el tipo de carretera donde tuvo lugar la colisión. La letalidad en Cantabria es mucho más elevada que la observada en el resto de España, en especial en zona urbana.

Tabla 26. Letalidad (número de muertos por 1.000 lesionados) en Cantabria y el resto de España. Base de datos policial. España 2005-2007.

	Cantabria	Resto de España
Zona urbana	179,0	8,3
Carretera	110,4	37,4
Total	115,6	23,9

Finalmente, en la Tabla 27 se comparan las características de los lesionados según datos hospitalarios. No se observan diferencias destacables entre los lesionados ingresados en Cantabria y el resto de España.

Tabla 27. Características de los lesionados de tráfico en Cantabria y el resto de España según gravedad. Base de datos hospitalaria. España 2005-2007.

	Cantabria	Resto de España	Valor de p
N	919	82.860	
Sexo			0,929
Hombre	72,8	73,2	
Mujer	27,2	26,8	
Desconocido	0,0	0,0	
Edad			0,002
0-13 años	9,4	7,5	
14-17 años	7,1	10,0	
18-29 años	25,8	28,0	
30-44 años	21,6	22,8	
45-64 años	19,7	17,9	
65-74 años	7,6	7,0	
75 y más años	8,9	6,8	
Desconocido	0,0	0,0	
MAIS			0,311
1 Leve	9,8	9,2	
2 Moderado	54,3	52,0	
3-6 Grave	35,0	38,0	
Desconocido	0,9	0,8	

MAIS: Maximum Abbreviated Injury Scale

Variables de conexión

En la Tabla 28 se describen las variables de conexión.

Tabla 28. Descripción de las variables utilizadas para la conexión probabilística. Cantabria 2005-2007.

Variables conexión	DGT	CMBD
Identificador	<u>Nombre:</u> p_id <u>Formato:</u> %8s <u>Valores perdidos:</u> --	<u>Nombre:</u> h_id <u>Formato:</u> %8s <u>Valores perdidos:</u> --
Sexo	<u>Nombre:</u> p_sexo <u>Formato:</u> float %8.0g <u>Categorías:</u> 1 hombres 2 mujeres <u>Valores perdidos:</u> (.) (0,24%)	<u>Nombre:</u> h_sexo <u>Formato:</u> float %8.0g <u>Categorías:</u> 1 hombres 2 mujeres <u>Valores perdidos:</u> --
Edad	<u>Nombre:</u> p_edad <u>Formato:</u> float %8.0g <u>Categorías:</u> 1-99 <u>Valores perdidos:</u> (.) (3,6%)	<u>Nombre:</u> h_edad <u>Formato:</u> float %8.0g <u>Categorías:</u> 1-93 <u>Valores perdidos:</u> --
Vehículo	<u>Nombre:</u> p_vehiculo <u>Formato:</u> float %8.0g <u>Categorías:</u> 1 No moto-ciclo 2 Moto-ciclo 3 Vehículo tracción animal 4 Bici <u>Valores perdidos:</u> (.) (5,2%)	<u>Nombre:</u> h_vehículo <u>Formato:</u> float %8.0g <u>Categorías:</u> 1 No moto-ciclo 2 Moto-ciclo 3 Vehículo tracción animal 4 Bici <u>Valores perdidos:</u> (.) (50,4%)
Posición en vehículo	<u>Nombre:</u> p_posicion <u>Formato:</u> float %8.0g <u>Categorías:</u> 1 Conductor 2 Pasajero 3 Ocupante carro/bici 4 Peatón <u>Valores perdidos:</u> (.) (1,5%)	<u>Nombre:</u> h_posición <u>Formato:</u> float %8.0g <u>Categorías:</u> 1 Conductor 2 Pasajero 3 Ocupante carro/bici 4 Peatón <u>Valores perdidos:</u> (.) (35,5%)

Cont.

Usuario	<u>Nombre:</u> p_posveh <u>Formato:</u> float %8.0g <u>Categorías:</u> 0 Conductor no moto-ciclo 1 Pasajero no moto-ciclo 2 Conductor moto-ciclo 3 Pasajero no moto-ciclo 5 Ocupante carro 6 Ciclista 7 Peatón Valores perdidos: (.) (1,5%)	<u>Nombre:</u> h_posveh <u>Formato:</u> float %8.0g <u>Categorías:</u> 0 Conductor no moto-ciclo 1 Pasajero no moto-ciclo 2 Conductor moto-ciclo 3 Pasajero no moto-ciclo 5 Ocupante carro 6 Ciclista 7 Peatón Valores perdidos: (.) (35,5%)
Fecha colisión	<u>Nombre:</u> p_fecha_acc <u>Formato:</u> flota %d Valores perdidos: --	<u>Nombre:</u> h_fecha_acc <u>Formato:</u> float %d Valores perdidos: --
Localización lesión (diagnóstico principal) (Para la revisión semi-manual)	<u>Nombre:</u> p_locles <u>Formato:</u> float %8.0g <u>Categorías:</u> 1 Cabeza/cara 2 Cuello 3 Pecho 4 Espalda 5 Abdomen 6 EESS 7 EEII Valores perdidos: (.) (27,4%)	<u>Nombre:</u> h_locles <u>Formato:</u> float %8.0g <u>Categorías:</u> 1 Cabeza/cara 2 Cuello 3 Pecho 4 Espalda 5 Abdomen 6 EESS 7 EEII Valores perdidos: (.) (16,2%)

Análisis de la viabilidad de la conexión

1) Según las posibles combinaciones de las variables conexión:

Tamaño de la base de datos de la DGT 2005-2007 sin muertos = 4.206

Tamaño de la base de datos del CMBD ingresos 2005-2007 = 947

Suma de registros: $4.206 + 947 = 5.153$

Variables de conexión:

- Sexo: 2 categorías
- Edad: 99 categorías

- Vehículo: 4 categorías
- Posición: 4 categorías
- Usuario: 7 categorías
- Fecha de colisión: 365 días * 3 años = 1.095 categorías

Para la combinación de las variables sexo, edad, provincia, vehículo, posición y fecha colisión:

$$\text{Combinaciones posibles} = 2 * 99 * 4 * 4 * 1.095 = 3.468.960$$

Para la combinación de las variables sexo, edad, provincia, usuario y fecha colisión:

$$\text{Combinaciones posibles} = 2 * 99 * 7 * 1.095 = 1.517.670$$

Por tanto, según este análisis, la conexión es viable porque:

$$3.468.960 > 5.153 \text{ y } 1.517.670 > 5.153$$

2) Según el peso mínimo (W_{min}) alcanzado por las variables de conexión:

$$W_t = \log_2 (p/(1-p)) - \log_2 (E/(A*B-E))$$

P (probabilidad de que dos registros sean correctos) = 0,95

A (tamaño de base de datos 1) = 4.206

B (tamaño de base de datos 2) = 947

E (número de pares correctos) = 947 (100% según un escenario optimista)
= 644 (68% según un escenario pesimista)

Para E=100%: $W_t = \log_2 (0,95/(1-0,95)) - \log_2 (947/(4.206*947-947)) = 16,29$

Para E=68%: $W_t = \log_2 (0,95/(1-0,95)) - \log_2 (644/(4.206*947-644)) = 16,84$

Por lo tanto, para que la conexión sea viable, el peso mínimo deberá ser superior a:

$W_t - 3 = 16,29 - 3 = 13,29$ (para el escenario optimista)

$= 16,84 - 3 = 13,84$ (para el escenario pesimista)

En la Tabla 29 se muestran los cálculos de los pesos de coincidencia y el peso mínimo para distintas combinaciones de variables.

Tabla 29. Pesos de coincidencia de las variables de conexión, peso mínimo alcanzado según distintas combinaciones de variables y viabilidad de la conexión según el peso mínimo. Cantabria 2005-2007.

Variables	Peso de coincidencia	Combinación de variables 1	Combinación de variables 2	Combinación de variables 3	Combinación de variables 4
Fecha de colisión	7,41	X	X	X	X
Provincia de colisión y hospital	--	X	X	X	X
Edad	4,61	X	X	X	X
Sexo	0,30	X	X	X	X
Vehículo implicado	0,93	X		X	
Posición respecto el vehículo	1,07	X		X	
Usuario	1,35		X		X
Localización de la lesión principal	1,31			X	X
PESO MÍNIMO		14,32	13,67	15,63	14,98
Viabilidad de la conexión para E=100% (13,29)		Sí	Sí	Sí	Sí
Viabilidad de la conexión para E=68% (13,84)		Sí	No	Sí	Sí

Los resultados de la viabilidad de la conexión muestran que ésta es viable con cualquier combinación de variables para el escenario optimista, mientras que para el escenario pesimista la combinación de fecha de colisión, provincia, edad, sexo y usuario no es viable.

Resultados de la conexión

Las variables utilizadas para la conexión en Cantabria fueron fecha de la colisión, edad, sexo, vehículo y posición. Dado que la conexión resultó viable sin la variable localización de la lesión principal, ésta se utilizó únicamente en la selección semi-manual de pares concordantes.

De los 947 ingresos hospitalarios entre los años 2005 y 2007, se encontró un posible registro de conexión de la base de datos policial para 937 casos (98,9%) (registros enlazados). De éstos, tan sólo en un 14,0% de registros coincidieron las 5 variables de conexión, y en 22,0% 4 variables (Tabla 30). En esta tabla se muestran, además, los pesos de concordancia mínimos y máximos de estos registros en función del número de variables coincidentes. La distribución de pesos se puede observar en la Figura 5, que se muestra en función del número de coincidencias en la Figura 6. En el caso de Cantabria, el histograma de pesos muestra una distribución trimodal. En estas figuras también se muestra el lugar en el que se sitúa el punto de corte visual, equivalente a un peso de 12,0.

Tabla 30. Distribución del número de coincidencias y su correspondiente rango de pesos (n=937). Cantabria 2005-2007.

	n	%	Peso mínimo	Peso máximo
0	6	0,6	-9,50	-3,32
1	127	13,6	-8,42	8,74
2	408	43,5	-5,82	13,66
3	206	22,0	-4,32	17,90
4	131	14,0	4,47	21,35
5	59	6,3	15,12	24,55

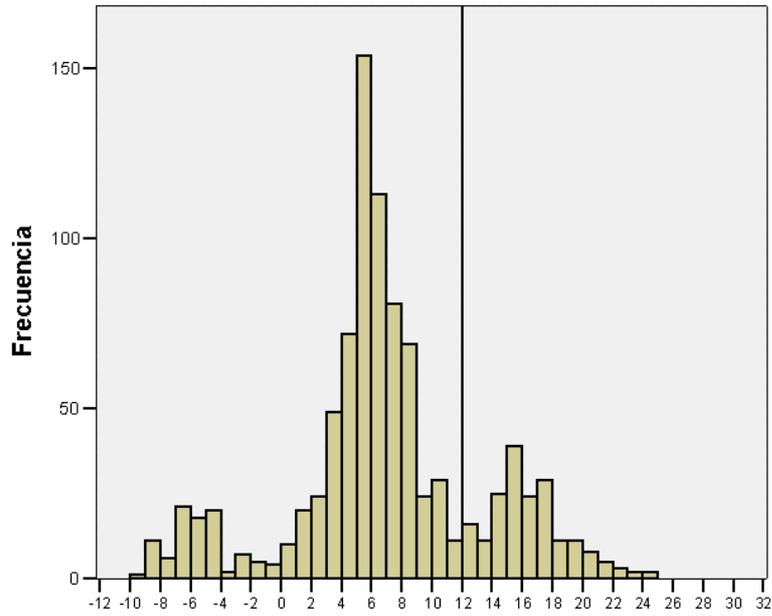


Figura 5. Distribución del los pesos de concordancia y punto de corte visual. Cantabria 2005-2007.

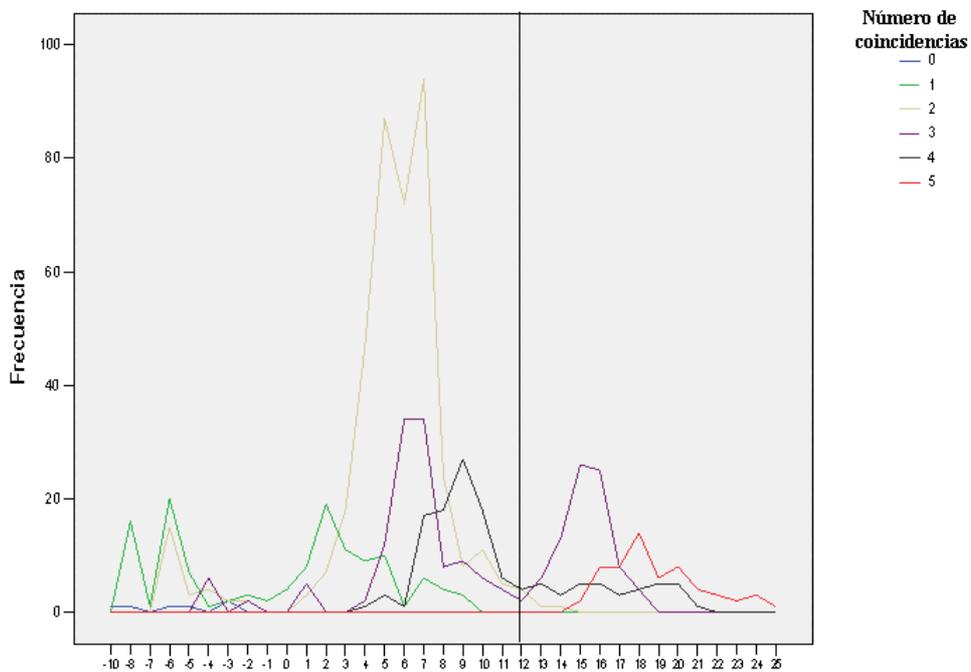


Figura 6. Distribución del los pesos de concordancia en función del número de coincidencias y punto de corte visual. Cantabria 2005-2007.

Según los criterios establecidos en el apartado de metodología, se consideraron conectados 207 pares de registros (21,9% de la base de datos inicial; 32,1% del número esperado de registros conectados). De éstos, la mayoría (130 registros) se conectaron según el criterio del peso mínimo de los registros en los que coinciden todas las variables de conexión, 52 se conectaron según los criterios establecidos para la revisión semi-manual, 6 registros presentaban una diferencia de edad de sólo un año y 19 registros de hasta 5 años (Tabla 31). Estos resultados fueron similares a los observados aplicando un punto de corte visual. Según este criterio se conectaron 186 registros, que representan un 19,6% de los registros conectables y un 28,9% de los registros que se esperan conectar.

Tabla 31. Resultados de la conexión probabilística. Cantabria 2005-2007.

	n	% conectables	% enlazados	% esperados
Registros conectables	947			
Registros enlazados	937			
Registros conectados esperados^a	644			
Registros no conectados	730	77,1	77,9	
Registros conectados	207	21,9	22,1	32,1
Según criterio1^b	130	13,7	13,9	20,2
+criterio2^c	52	5,5	5,5	8,1
+criterio3^d	6	0,6	0,6	0,9
+criterio4^e	19	2,0	2,0	3,0
Registros no enlazados	10	1,1		

a. Se espera que se conecte un 68% de los registros conectables.

b. Cuando coinciden todas las variables de conexión o cuando el peso de concordancia es superior al peso mínimo de los registros en los que coinciden todas las variables de conexión.

c. Según los criterios de la revisión semi-manual.

d. Se permite una variación de +/-1 año de edad.

e. Se permite una variación de +/-5 años de edad.

Se comprobó si las características de los pares conectados difieren de las de los pares no conectados en la base de datos policial y en la hospitalaria (Tabla 32 y Tabla 33). De los 4.206 registros policiales de Cantabria, los 207 casos conectados representaron un 4,9% del total de registros (un 21,5% de los lesionados graves). Se observó una mayor proporción de hombres que mujeres con registros conectados, así como de menores de 18 años y de mayores de 75 años. En cuanto a la gravedad de los lesionados, aunque la

mayoría de los casos conectados fueron identificados por la policía como graves, el 45,9% se identificaron como lesionados leves, que representaron un 2,6% de los lesionados leves de la base de datos policial. En relación a la localización de la lesión principal, la proporción de registros conectados es mayor entre los individuos con lesiones en la cabeza o la cara y en las extremidades inferiores, y menor en aquellos con lesiones en el cuello. Los usuarios de vehículos de dos ruedas también presentaron una mayor proporción de registros conectados, al igual que los peatones, mientras que para pasajeros fue menor. En cuanto a las características de la colisión, la proporción de registros conectados fue ligeramente superior entre los individuos lesionados en zona urbana, en comparación con carretera ($p>0,05$), y en aquellas colisiones ocurridas en horario diurno ($p>0,05$). Además, aquellos individuos implicados en una colisión con una sola víctima presentaron una mayor probabilidad de haberse conectado (Tabla 32).

Tabla 32. Comparación de las características entre los pares conectados y no conectados de la base de datos de la policía. Análisis bivariado. Cantabria 2005-2007.

	Pares no conectados			Pares conectados			Valor de p
	n	% fila	% columna	n	% fila	% columna	
Sexo							0,009
Hombre	2.458	94,3	61,5	149	5,7	72,0	
Mujer	1.531	96,3	38,3	58	3,7	28,0	
Desconocido	10		0,3	--			
Edad							<0,001
0-13 años	161	90,4	4,0	17	9,6	8,2	
14-17 años	197	92,5	4,9	16	7,5	7,7	
18-29 años	1.382	95,2	34,6	70	4,8	33,8	
30-44 años	1.152	96,2	28,8	46	3,8	22,2	
45-64 años	695	94,7	17,4	39	5,3	18,8	
65-74 años	190	96,4	4,8	7	3,6	3,4	
75 y más años	73	88,0	1,8	10	12,0	4,8	
Desconocida	149	98,6	3,7	2	1,3	1,0	
Gravedad del lesionado							<0,001
Leve	3.589	97,4	89,7	95	2,6	45,9	
Grave	410	78,5	10,3	112	21,5	54,1	
Desconocida	--			--			

Cont.

Localización de la lesión principal							<0,001
Cabeza o cara	423	91,4	10,6	40	8,6	19,3	
Cuello	1.140	98,7	28,5	15	1,3	7,2	
Pecho	335	97,1	8,4	10	2,9	4,8	
Espalda	235	95,9	5,9	10	4,1	4,8	
Abdomen	56	84,8	1,4	10	15,2	4,8	
Extremidades superiores	345	93,2	8,6	25	6,8	12,1	
Extremidades inferiores	352	86,1	8,8	57	13,9	27,5	
Desconocida	1.113	96,5	27,8	40	3,5	19,3	
Vehículo							<0,001
Vehículo a motor más de 2 ruedas	3.242	97,0	81,1	101	3,0	48,8	
Vehículo a motor de 2 ruedas	483	87,5	12,1	69	12,5	33,3	
Vehículo tracción animal	1			--			
Bicicleta	86	93,5	2,2	6	6,5	2,9	
Desconocido	187	85,8	4,7	31	14,2	15,0	
Posición							<0,001
Conductor/a	2.336	94,7	58,4	132	5,3	63,8	
Pasajero/a	1.389	97,3	34,7	38	2,7	18,4	
Ocupante carro/bici	87	93,5	2,2	6	6,5	2,9	
Peatón/a	131	84,0	3,3	25	16,0	12,1	
Desconocida	56	90,3	1,4	6	9,7	2,9	
Tipo de vía							0,675
Carretera	297	94,6	7,4	17	5,4	8,2	
Zona urbana	3.702	95,1	92,6	190	4,9	91,8	
Desconocida	--			--			
Momento de colisión							0,102
Laborable diurno	2.080	94,7	52,0	116	5,3	56,0	
Laborable nocturno	467	96,1	11,7	19	3,9	9,2	
Fin de semana diurno	844	94,3	21,1	51	5,7	24,6	
Fin de semana nocturno	608	96,7	15,2	21	3,3	10,1	
Número de víctimas							<0,001
1	1.570	92,7	39,3	123	7,3	59,4	
2	1.130	95,9	28,3	48	4,1	23,2	
>=3	1.299	97,3	32,5	36	2,7	17,4	

Al comparar las características de los pares conectados y no conectados de los datos hospitalarios, no se presentaron diferencias entre hombres y mujeres en la probabilidad de conectarse. Sin embargo, sí que se observaron diferencias en función de la edad, de

manera que los más jóvenes y los mayores tuvieron una menor probabilidad de ser conectados. En cuanto a la estancia hospitalaria, aquellos individuos con un tiempo más corto de estancia hospitalaria presentaron menos posibilidades de ser conectados. No se observaron diferencias en cuanto a la localización de la lesión principal. Finalmente, los usuarios de bicicleta y peatones presentaron una menor probabilidad de conectarse, mientras que la probabilidad fue mayor en usuarios de vehículo a motor de más de 2 ruedas (Tabla 33).

Tabla 33. Comparación de las características entre los pares conectados y no conectados de la base de datos hospitalaria. Análisis bivariado. Cantabria 2005-2007.

	Pares no conectados			Pares conectados			Valor de p
	n	% fila	% columna	n	% fila	% columna	
Sexo							0,763
Hombre	544	78,4	73,5	150	21,6	72,5	
Mujer	196	77,5	26,5	57	22,5	27,5	
Desconocido	--			--			
Edad							0,003
0-13 años	85	82,5	11,5	18	17,5	8,7	
14-17 años	53	73,6	7,2	19	26,4	9,2	
18-29 años	168	70,6	22,7	70	29,4	33,8	
30-44 años	155	77,9	20,9	44	22,1	21,3	
45-64 años	144	79,1	19,5	38	20,9	18,4	
65-74 años	63	91,3	8,5	6	8,7	2,9	
75 y más años	72	85,7	9,7	12	14,3	5,8	
Desconocida	--			--			
Estancia hospitalaria							0,016
1er cuartil	195	80,9	26,4	46	19,1	22,2	
2º cuartil	222	82,2	30,0	48	17,8	23,2	
3er cuartil	163	77,3	22,0	48	22,7	23,2	
4o cuartil	160	71,1	21,6	65	28,9	31,4	

Cont.

Localización de la lesión principal							0,672
Cabeza o cara	118	73,8	15,9	42	26,3	20,3	
Cuello	16	76,2	2,2	5	23,8	2,4	
Pecho	68	82,9	9,2	14	17,1	6,8	
Espalda	51	85,0	6,9	9	15,0	4,3	
Abdomen	20	74,1	2,7	7	25,9	3,4	
Extremidades superiores	96	78,0	13,0	27	22,0	13,0	
Extremidades inferiores	252	78,5	34,1	69	21,5	33,3	
Desconocida	119	77,8	16,1	34	22,2	16,4	
Vehículo							<0,001
Vehículo a motor más de 2 ruedas	60	58,8	8,1	42	41,2	20,3	
Vehículo a motor de 2 ruedas	180	73,2	24,3	66	26,8	31,9	
Vehículo tracción animal	--			--			
Bicicleta	114	93,4	15,4	8	6,6	3,9	
Desconocido	386	80,9	52,2	91	19,1	44,0	
Posición							<0,001
Conductor/a	208	71,2	28,1	84	28,8	40,6	
Pasajero/a	32	57,1	4,3	24	42,9	11,6	
Ocupante carro/bici	114	93,4	15,4	8	6,6	3,9	
Peatón/a	121	85,8	16,4	20	14,2	9,7	
Desconocida	265	78,9	35,8	71	21,1	34,3	

Validación de la conexión

Se validó la conexión mediante dos métodos. En primer lugar, se conectó la base de datos A, que corresponde a los registros hospitalarios de Cantabria del año 2005 (n=342), con la base de datos B, que contiene el 68% de los registros de la base A y los registros hospitalarios de los años 2006-2008 (n=1.156).

En número de falsos positivos obtenidos mediante la conexión fue de 8 registros (2,3%). No se obtuvo ningún falso negativo. En la Tabla 34 se observa la distribución de los registros en función de si son verdaderos positivos o negativos y falsos positivos o negativos. A partir de ello se calculó la sensibilidad y la especificidad de la conexión.

Tabla 34. Verdaderos positivos y negativos y falsos positivos y negativos de la validación de la conexión probabilística. Cantabria 2005-2007.

		Conectados		
		No	Sí	Total
Par real	No	102	8	110
	Sí	0	232	232
	Total	102	240	342

$$\text{Sensibilidad} = \text{VP} / (\text{VP} + \text{FN}) = 232 / (232 + 0) = 100\%$$

$$\text{Especificidad} = \text{VN} / (\text{FP} + \text{VN}) = 102 / (8 + 102) = 92,7\%$$

El segundo método para validar consistió en conectar dos bases de datos de registros no concordantes: la base de datos A, que contiene los registros hospitalarios de Cantabria del año 2006 (n=324) y la base de datos B, que contiene los registros policiales de Cantabria del año 2007 (n=1.331). Esta validación nos permitió calcular mediante otro método los falsos positivos: 13 registros (4,0%), lo cual supone una especificidad del 96,0%.

Castilla la Mancha

Selección de casos del CMBD

La base de datos del CMBD de Castilla la Mancha incluyó las altas hospitalarias entre los años 2005 y 2008. Se seleccionaron las personas con fecha de ingreso en los años 2005 a 2007 y según los criterios comentados en el apartado de metodología.

Representatividad de los datos de Castilla la Mancha

La Tabla 35 muestra las características de los lesionados de tráfico en Castilla la Mancha en comparación con las de los lesionados en el resto de España, según gravedad. Cabe destacar una menor proporción de lesionados como peatones y como usuarios de motocicleta y ciclomotor en Castilla la Mancha, mientras que la proporción de usuarios de camión o autobús es superior. Asimismo, la proporción de lesionados leves, graves y muertos en zona urbana es menor en Castilla la Mancha que en el resto de España.

Tabla 35. Características de los lesionados de tráfico en Castilla la Mancha y el resto de España según gravedad. Base de datos policial. España 2005-2007.

	Muertos			Graves			Leves		
	Castilla la Mancha	Resto de España	Valor de p	Castilla la Mancha	Resto de España	Valor de p	Castilla la Mancha	Resto de España	Valor de p
N	873	9.824		4.125	60.083		15.381	340.863	
Sexo			0,808			0,025			0,146
Hombre	63,9	62,9		72,7	72,8		79,7	79,0	
Mujer	35,7	36,5		27,1	26,8		19,9	20,7	
Desconocido	0,4	0,6		0,2	0,4		0,3	0,3	
Edad			0,000			0,000			0,005
0-13 años	5,0	3,5		3,4	2,9		3,0	1,9	
14-17 años	6,4	5,5		5,9	6,5		2,6	3,4	
18-29 años	6,4	34,5		31,1	32,0		24,3	27,1	
30-44 años	25,6	27,3		27,8	27,1		29,4	27,8	
45-64 años	17,2	15,9		20,2	17,7		20,6	22,3	
65-74 años	5,3	3,6		6,5	5,3		11,2	8,2	
75 y más años	2,9	2,4		3,8	4,5		8,3	8,3	
Desconocido	3,7	7,3		3,8	4,1		0,6	1,1	
Usuario			0,000			0,000			0,000
Turismo	65,6	56,5		56,5	44,7		62,3	51,4	
Motocicleta	3,4	11,1		8,6	15,2		7,3	13,4	
Ciclomotor	8,6	14,1		9,1	14,9		2,8	6,2	
Camión o autobús	14,6	8,2		14,7	8,3		14,6	9,4	
Otros	3,2	2,8		5,4	4,2		4,6	4,6	
Peatón	4,7	7,3		5,8	12,5		8,5	14,7	
Desconocido	0,0	0,1		0,0	0,1		0,0	0,2	
Tipo de carretera			0,000			0,000			0,000
Zona urbana	25,8	50,6		15,8	32,6		8,4	16,6	
Carretera	74,2	49,4		84,2	67,4		91,6	83,4	
Desconocido	0,0	0,0		0,0	0,0		0,0	0,0	

En la Tabla 36 se muestra la letalidad en Castilla la Mancha y el resto de España para el total de muertos y según el tipo de carretera donde tuvo lugar la colisión. La letalidad en Castilla la Mancha es ligeramente superior a la observada en el resto de España, tanto en zona urbana como en carretera.

Tabla 36. Letalidad (número de muertos por 1.000 lesionados) en Castilla la Mancha y el resto de España. Base de datos policial. España 2005-2007.

	Castilla la Mancha	Resto de España
Zona urbana	15,6	8,4
Carretera	51,0	37,8
Total	42,8	23,9

Finalmente, en la Tabla 37 se comparan las características de los lesionados según datos hospitalarios. No se observan diferencias destacables entre los lesionados ingresados en Castilla la Mancha y el resto de España.

Tabla 37. Características de los lesionados de tráfico en Castilla la Mancha y el resto de España según gravedad. Base de datos hospitalaria. España 2005-2007.

	Castilla la Mancha	Resto de España	Valor de p
N	5.563	78.216	
Sexo			0,000
Hombre	76,8	72,9	
Mujer	23,2	27,1	
Desconocido	0,0	0,0	
Edad			0,011
0-13 años	8,1	7,4	
14-17 años	10,5	10,0	
18-29 años	26,6	28,1	
30-44 años	23,8	22,7	
45-64 años	17,2	18,0	
65-74 años	7,4	7,0	
75 y más años	6,4	6,8	
Desconocido	0,0	0,0	
MAIS			0,000
1 Leve	8,7	9,2	
2 Moderado	49,7	52,2	
3-6 Grave	41,1	37,8	
Desconocido	0,5	0,8	

MAIS: Maximum Abbreviated Injury Scale

Variables de conexión

En la Tabla 38 se describen las variables de conexión.

Tabla 38. Descripción de las variables utilizadas para la conexión probabilística. Castilla la Mancha 2005-2007.

Variables conexión	DGT	CMBD
Identificador	<u>Nombre:</u> p_id <u>Formato:</u> %8s <u>Valores perdidos:</u> --	<u>Nombre:</u> h_id <u>Formato:</u> %8s <u>Valores perdidos:</u> --
Sexo	<u>Nombre:</u> p_sexo <u>Formato:</u> float %8.0g <u>Categorías:</u> 1 hombres 2 mujeres <u>Valores perdidos:</u> (.) (1,0%)	<u>Nombre:</u> h_sexo <u>Formato:</u> float %8.0g <u>Categorías:</u> 1 hombres 2 mujeres <u>Valores perdidos:</u> --
Edad	<u>Nombre:</u> p_edad <u>Formato:</u> float %8.0g <u>Categorías:</u> 1-99 <u>Valores perdidos:</u> (.) (3,2%)	<u>Nombre:</u> h_edad <u>Formato:</u> float %8.0g <u>Categorías:</u> 1-99 <u>Valores perdidos:</u> --
Provincia	<u>Nombre:</u> p_provin <u>Formato:</u> float %8.0g <u>Categorías:</u> 02 Albacete 13 Ciudad Real 16 Cuenca 19 Guadalajara 45 Toledo <u>Valores perdidos:</u> --	<u>Nombre:</u> h_provin <u>Formato:</u> float %8.0g <u>Categorías:</u> 02 Albacete 13 Ciudad Real 16 Cuenca 19 Guadalajara 45 Toledo <u>Valores perdidos:</u> --
Vehículo	<u>Nombre:</u> p_vehiculo <u>Formato:</u> float %8.0g <u>Categorías:</u> 1 No moto-ciclo 2 Moto-ciclo 3 Vehículo tracción animal 4 Bici <u>Valores perdidos:</u> (.) (6,8%)	<u>Nombre:</u> h_vehículo <u>Formato:</u> float %8.0g <u>Categorías:</u> 1 No moto-ciclo 2 Moto-ciclo 3 Vehículo tracción animal 4 Bici <u>Valores perdidos:</u> (.) (70,1%)
Posición en vehículo	<u>Nombre:</u> p_posicion <u>Formato:</u> float %8.0g <u>Categorías:</u> 1 Conductor 2 Pasajero 3 Ocupante carro/bici 4 Peatón <u>Valores perdidos:</u> (.) (1,9%)	<u>Nombre:</u> h_posición <u>Formato:</u> float %8.0g <u>Categorías:</u> 1 Conductor 2 Pasajero 3 Ocupante carro/bici 4 Peatón <u>Valores perdidos:</u> (.) (61,4%)

Cont.

Usuario	<u>Nombre:</u> p_posveh <u>Formato:</u> float %8.0g <u>Categorías:</u> 0 Conductor no moto-ciclo 1 Pasajero no moto-ciclo 2 Conductor moto-ciclo 3 Pasajero no moto-ciclo 5 Ocupante carro 6 Ciclista 7 Peatón Valores perdidos: (.) (1,9%)	<u>Nombre:</u> h_posveh <u>Formato:</u> float %8.0g <u>Categorías:</u> 0 Conductor no moto-ciclo 1 Pasajero no moto-ciclo 2 Conductor moto-ciclo 3 Pasajero no moto-ciclo 5 Ocupante carro 6 Ciclista 7 Peatón Valores perdidos: (.) (61,4%)
Fecha colisión	<u>Nombre:</u> p_fecha_acc <u>Formato:</u> flota %d Valores perdidos: --	<u>Nombre:</u> h_fecha_acc <u>Formato:</u> float %d Valores perdidos: --
Localización lesión (diagnóstico principal) (Para la revisión semi-manual)	<u>Nombre:</u> p_locles <u>Formato:</u> float %8.0g <u>Categorías:</u> 1 Cabeza/cara 2 Cuello 3 Pecho 4 Espalda 5 Abdomen 6 EESS 7 EEII Valores perdidos: (.) (32,9%)	<u>Nombre:</u> h_locles <u>Formato:</u> float %8.0g <u>Categorías:</u> 1 Cabeza/cara 2 Cuello 3 Pecho 4 Espalda 5 Abdomen 6 EESS 7 EEII Valores perdidos: (.) (8,3%)

Análisis de la viabilidad de la conexión

1) Según las posibles combinaciones de las variables conexión:

Tamaño de la base de datos de la DGT 2005-2007 sin muertos = 19.544

Tamaño de la base de datos del CMBD ingresos 2005-2007 = 5.693

Suma de registros: $19.544 + 5.693 = 25.237$

Variables de conexión:

- Sexo: 2 categorías
- Edad: 99 categorías

- Provincia: 4 categorías
- Vehículo: 4 categorías
- Posición: 4 categorías
- Usuario: 7 categorías
- Fecha de colisión: 365 días * 3 años = 1.095 categorías

Para la combinación de las variables sexo, edad, provincia, vehículo, posición y fecha colisión:

$$\text{Combinaciones posibles} = 2 * 99 * 5 * 4 * 4 * 1095 = 17.344.800$$

Para la combinación de las variables sexo, edad, provincia, usuario y fecha colisión:

$$\text{Combinaciones posibles} = 2 * 99 * 5 * 7 * 1095 = 7.588.350$$

Por tanto, según este análisis, la conexión sería viable porque:

$$17.344.800 > 25.237 \text{ y } 7.588.350 > 25.237$$

2) Según el peso mínimo (W_{min}) alcanzado por las variables de conexión:

$$W_t = \log_2(p/(1-p)) - \log_2(E/(A*B-E))$$

P (probabilidad de que dos registros sean correctos) = 0,95

A (tamaño de base de datos 1) = 19.544

B (tamaño de base de datos 2) = 5.693

E (número de pares correctos) = 5.693 (100% según un escenario optimista)

= 3.871 (68% según un escenario pesimista)

Para E=100%: $W_t = \log_2(0,95/(1-0,95)) - \log_2(5.693/(19.544*5.693-5.693)) = 18,50$

Para E=68%: $W_t = \log_2(0,95/(1-0,95)) - \log_2(3.871/(19.544*5.693-3.871)) = 19,06$

Por lo tanto, para que la conexión sea viable, el peso mínimo deberá ser superior a:

$W_t - 3 = 18,50 - 3 = 15,50$ (para el escenario optimista)

= $19,06 - 3 = 16,06$ (para el escenario pesimista)

En la Tabla 39 se muestran los cálculos de los pesos de coincidencia y el peso mínimo para distintas combinaciones de variables.

Tabla 39. Pesos de coincidencia de las variables de conexión, peso mínimo alcanzado según distintas combinaciones de variables y viabilidad de la conexión según el peso mínimo. Castilla la Mancha 2005-2007.

Variables	Peso de coincidencia	Combinación de variables 1	Combinación de variables 2	Combinación de variables 3	Combinación de variables 4
Fecha de colisión	8,08	X	X	X	X
Provincia de colisión y hospital	1,47	X	X	X	X
Edad	4,69	X	X	X	X
Sexo	0,23	X	X	X	X
Vehículo implicado	0,99	X		X	
Posición respecto el vehículo	1,09	X		X	
Usuario	1,04		X		X
Localización de la lesión principal	1,74			X	X
PESO MÍNIMO		16,55	15,51	18,29	17,25
Viabilidad de la conexión para E=100% (15,50)		Sí	Sí	Sí	Sí
Viabilidad de la conexión para E=68% (16,06)		Sí	No	Sí	Sí

Los resultados de la viabilidad de la conexión muestran que ésta es viable con cualquier combinación de variables para el escenario optimista, mientras que para el escenario pesimista la combinación de fecha de colisión, provincia, edad, sexo y usuario no es viable.

Resultados de la conexión

Las variables utilizadas para la conexión en Castilla la Mancha fueron fecha de la colisión, provincia, edad, sexo, vehículo y posición. Dado que la conexión resultó viable sin la variable localización de la lesión principal, ésta se utilizó únicamente en la selección semi-manual de pares concordantes.

De los 5.693 ingresos hospitalarios entre los años 2005 y 2007, se encontró un posible registro de conexión de la base de datos policial para 5.638 casos (99,0%) (registros enlazados). De éstos, tan sólo en un 7,9% de registros coincidieron las 5 variables de conexión, y en 27,1% 4 variables (Tabla 40). En la Tabla 40 se muestran, además, los pesos de concordancia mínimos y máximos de estos registros en función del número de variables coincidentes. La distribución de pesos se puede observar en la Figura 7, que se muestra en función del número de coincidencias en la Figura 8. En el caso de Castilla la Mancha, el histograma de pesos muestra una distribución bimodal. En estas figuras también se muestra el lugar en el que se sitúa el punto de corte visual, equivalente a un peso de 15,0.

Tabla 40. Distribución del número de coincidencias y su correspondiente rango de pesos (n=5.638). Castilla la Mancha 2005-2007.

	n	%	Peso mínimo	Peso máximo
1	53	0,9	-8,18	8,14
2	953	16,9	-5,23	13,90
3	2.398	42,5	-5,16	19,55
4	1.528	27,1	-2,53	23,58
5	448	7,9	5,15	25,15
6	258	4,6	17,32	30,38

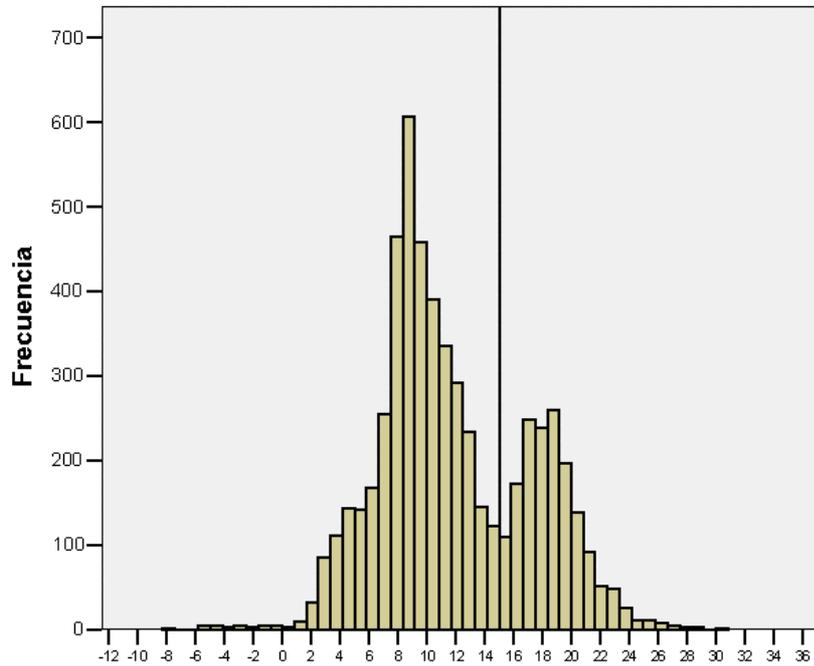


Figura 7. Distribución del los pesos de concordancia y punto de corte visual. Castilla la Mancha 2005-2007.

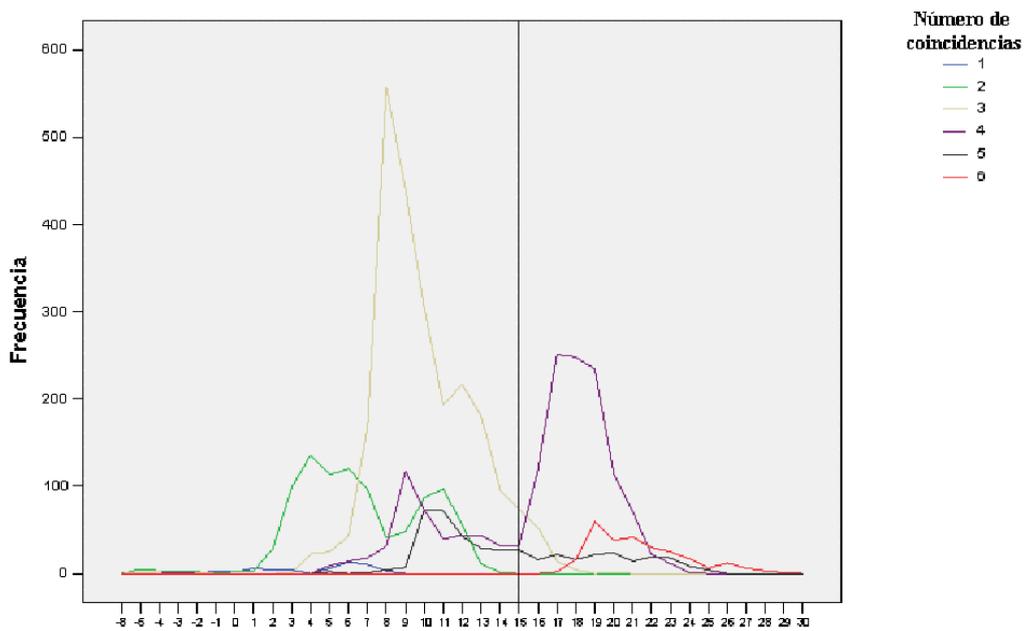


Figura 8. Distribución del los pesos de concordancia en función del número de coincidencias y punto de corte visual. Castilla la Mancha 2005-2007.

Según los criterios establecidos en el apartado de metodología, se consideraron conectados 1.562 pares de registros (27,4% de la base de datos inicial; 40,3% del número esperado de registros conectados). De éstos, la mayoría (1.151 registros) se conectaron según el criterio del peso mínimo de los registros en los que coinciden todas las variables de conexión, 338 se conectaron según los criterios establecidos para la revisión semi-manual, 28 registros presentaban una diferencia de edad de sólo un año y 45 registros de hasta 5 años (Tabla 41).

Estos resultados fueron similares a los observados aplicando un punto de corte visual. Según este criterio se conectaron 1.618 registros, que representan un 28,4% de los registros conectables y un 41,8% de los registros que se esperan conectar.

Tabla 41. Resultados de la conexión probabilística. Castilla la Mancha 2005-2007.

	n	% conectables	% enlazados	% esperados
Registros conectables	5.693			
Registros enlazados	5.638			
Registros conectados esperados^a	3.871			
Registros no conectados	4.076	71,6	72,3	
Registros conectados	1.562	27,4	27,7	40,3
Según criterio1^b	1.151	20,2	20,4	29,7
+criterio2^c	338	5,9	6,0	8,7
+criterio3^d	28	0,5	0,5	0,7
+criterio4^e	45	0,8	0,8	1,2
Registros no enlazados	55	1,0		

a. Se espera que se conecte un 68% de los registros conectables.

b. Cuando coinciden todas las variables de conexión o cuando el peso de concordancia es superior al peso mínimo de los registros en los que coinciden todas las variables de conexión.

c. Según los criterios de la revisión semi-manual.

d. Se permite una variación de +/-1 año de edad.

e. Se permite una variación de +/-5 años de edad.

Se comprobó si las características de los pares conectados difieren de las de los pares no conectados en la base de datos policial y en la hospitalaria (Tabla 42 y Tabla 43). De los 19.544 registros policiales, los 1.562 casos conectados representaron un 8,0% del total

de registros (un 25,3% de los lesionados graves). Se observó una mayor proporción de hombres que mujeres con registros conectados, así como de menores de 18 años y de mayores de 65 años. En cuanto a la gravedad de los lesionados, aunque la mayoría de los casos conectados fueron identificados por la policía como graves, el 33,1% se identificaron como lesionados leves, que representaron un 3,4% de los lesionados leves de la base de datos policial. En relación a la localización de la lesión principal, la proporción de registros conectados fue mayor entre los individuos con lesiones en la cabeza o la cara y en las extremidades inferiores, y menor en aquellos con lesiones en el cuello. Los usuarios de vehículos de dos ruedas también presentaron una mayor proporción de registros conectados, al igual que los peatones, mientras que para pasajeros fue menor. En cuanto a las características de la colisión, no se observaron diferencias en función del lugar de la colisión (carretera o zona urbana) aunque sí se observó una mayor probabilidad de ser un registro conectado para las colisiones ocurridas en horario nocturno. También, aquellos individuos implicados en una colisión con una sola víctima presentaron una mayor probabilidad de haberse conectado (Tabla 42).

Tabla 42. Comparación de las características entre los pares conectados y no conectados de la base de datos de la policía. Análisis bivariado. Castilla la Mancha 2005-2007.

	Pares no conectados			Pares conectados			Valor de p
	n	% fila	% columna	n	% fila	% columna	
Sexo							<0,001
Hombre	11.596	90,8	64,5	1.171	9,2	75,0	
Mujer	6.192	94,1	34,4	390	5,9	25,0	
Desconocido	194	99,5	1,1	1	0,5	0,1	
Edad							<0,001
0-13 años	800	88,3	4,4	106	11,7	6,8	
14-17 años	1.106	89,6	6,2	128	10,4	8,2	
18-29 años	6.050	93,0	33,6	456	7,0	29,2	
30-44 años	4.692	92,0	26,1	406	8,0	26,0	
45-64 años	3.203	91,9	17,8	282	8,1	18,1	
65-74 años	982	90,1	5,5	108	9,9	6,9	
75 y más años	546	90,2	3,0	59	9,8	3,8	
Desconocida	603	96,8	3,4	20	3,2	1,3	

Cont.

Gravedad del lesionado							<0,001
Leve	14.898	96,6	82,8	517	3,4	33,1	
Grave	3.084	74,7	17,2	1.045	25,3	66,9	
Desconocida	--			--			
Localización de la lesión principal							<0,001
Cabeza o cara	2.928	90,3	16,3	316	9,7	20,2	
Cuello	2.737	97,1	15,2	82	2,9	5,2	
Pecho	1.895	93,4	10,5	134	6,6	8,6	
Espalda	814	93,9	4,5	53	6,1	3,4	
Abdomen	278	86,3	1,5	44	13,7	2,8	
Extremidades superiores	1.791	92,0	10,0	155	8,0	9,9	
Extremidades inferiores	1.557	82,3	8,7	336	17,7	21,5	
Desconocida	5.982	93,1	33,3	442	6,9	28,3	
Vehículo							<0,001
Vehículo a motor más de 2 ruedas	14.357	93,3	79,8	1.024	6,7	65,6	
Vehículo a motor de 2 ruedas	2.244	87,1	12,5	331	12,9	21,2	
Vehículo tracción animal	1			--			
Bicicleta	228	86,0	1,3	37	14,0	2,4	
Desconocido	1.152	87,1	6,4	170	12,9	10,9	
Posición							<0,001
Conductor/a	9.628	91,3	53,5	923	8,7	59,1	
Pasajero/a	6.973	94,2	38,8	432	5,8	27,7	
Ocupante carro/bici	229	86,1	1,3	37	13,9	2,4	
Peatón/a	831	86,9	4,6	125	13,1	8,0	
Desconocida	321	87,7	1,8	45	12,3	2,9	
Tipo de vía							0,193
Carretera	4.280	92,5	23,8	349	7,5	22,3	
Zona urbana	13.702	91,9	76,2	1.213	8,1	77,7	
Desconocida	--			--			
Momento de colisión							<0,001
Laborable diurno	8.706	91,8	48,4	780	8,2	49,9	
Laborable nocturno	2.418	94,8	13,4	133	5,2	8,5	
Fin de semana diurno	3.966	90,2	22,1	432	9,8	27,7	
Fin de semana nocturno	2.892	93,0	16,1	217	7,0	13,9	
Número de víctimas							<0,001
1	6.637	89,9	36,9	742	10,1	47,5	
2	5.011	93,3	27,9	360	6,7	23,0	
>=3	6.334	93,2	35,2	460	6,8	29,4	

Al comparar las características de los pares conectados y no conectados de los datos hospitalarios, no se presentaron diferencias entre hombres y mujeres en la probabilidad de conectarse. Sin embargo, sí que se observaron diferencias en función de la edad, de manera que los más jóvenes y los mayores tuvieron una menor probabilidad de ser conectados. En cuanto a la estancia hospitalaria, aquellos individuos con un tiempo más corto de estancia hospitalaria presentaron menos posibilidades de ser conectados. También se observaron algunas diferencias según la localización de la lesión principal. Finalmente, los usuarios de bicicleta presentaron una menor probabilidad de conectarse, mientras que la probabilidad fue mayor en usuarios de vehículo a motor de más de 2 ruedas (Tabla 43).

Tabla 43. Comparación de las características entre los pares conectados y no conectados de la base de datos hospitalaria. Análisis bivariado. Castilla la Mancha 2005-2007.

	Pares no conectados			Pares conectados			Valor de p
	n	% fila	% columna	n	% fila	% columna	
Sexo							0,221
Hombre	3.192	73,0	77,3	1.183	27,0	75,7	
Mujer	939	71,2	22,7	379	28,8	24,3	
Desconocido	--			--			
Edad							<0,001
0-13 años	382	77,5	9,2	111	22,5	7,1	
14-17 años	481	78,1	11,6	135	21,9	8,6	
18-29 años	1.037	69,4	25,1	457	30,6	29,3	
30-44 años	928	69,7	22,5	404	30,3	25,9	
45-64 años	690	70,4	16,7	290	29,6	18,6	
65-74 años	318	74,3	7,7	110	25,7	7,0	
75 y más años	295	84,3	7,1	55	15,7	3,5	
Desconocida	--			--			
Estancia hospitalaria							<0,001
1er cuartil	1.123	76,9	27,2	337	23,1	21,6	
2º cuartil	1.117	73,6	27,0	401	26,4	25,7	
3er cuartil	927	69,1	22,4	414	30,9	26,5	
4o cuartil	964	70,2	23,3	410	29,8	26,2	

Cont.

Localización de la lesión principal							0,006
Cabeza o cara	880	73,8	21,3	313	26,2	20,0	
Cuello	214	78,4	5,2	59	21,6	3,8	
Pecho	462	68,3	11,2	214	31,7	13,7	
Espalda	357	75,2	8,6	118	24,8	7,6	
Abdomen	165	72,4	4,0	63	27,6	4,0	
Extremidades superiores	624	74,4	15,1	215	25,6	13,8	
Extremidades inferiores	1.076	70,1	26,0	458	29,9	29,3	
Desconocida	353	74,3	8,5	122	25,7	7,8	
Vehículo							<0,001
Vehículo a motor más de 2 ruedas	193	51,5	4,7	182	48,5	11,7	
Vehículo a motor de 2 ruedas	614	71,6	14,9	243	28,4	15,6	
Vehículo tracción animal	8			--			
Bicicleta	428	92,6	10,4	34	7,4	2,2	
Desconocido	2.888	72,4	69,9	1.103	27,6	70,6	
Posición							<0,001
Conductor/a	699	67,9	16,9	331	32,1	21,2	
Pasajero/a	108	53,5	2,6	94	46,5	6,0	
Ocupante carro/bici	436	93,4	10,6	31	6,6	2,0	
Peatón/a	398	79,9	9,6	100	20,1	6,4	
Desconocida	2.490	71,3	60,3	1.003	28,7	64,2	

Validación de la conexión

Se validó la conexión mediante dos métodos. En primer lugar, se conectó la base de datos A, que corresponde a los registros hospitalarios de Castilla la Mancha del año 2005 (n=2.035), con la base de datos B, que contiene el 68% de los registros de la base A y los registros hospitalarios de los años 2006-2008 (n=6.673).

En número de falsos positivos obtenidos mediante la conexión fue de 62 registros (3,0%). No se obtuvo ningún falso negativo. En la Tabla 44 se observa la distribución de los registros en función de si son verdaderos positivos o negativos y falsos positivos o negativos. A partir de ello se calculó la sensibilidad y la especificidad de la conexión.

Tabla 44. Verdaderos positivos y negativos y falsos positivos y negativos de la validación de la conexión probabilística. Castilla la Mancha 2005-2007.

		Conectados		
		No	Sí	Total
Par real	No	605	62	667
	Sí	1	1.367	1.368
	Total	606	1.429	2.035

$$\text{Sensibilidad} = \text{VP} / (\text{VP} + \text{FN}) = 1.367 / (1.367 + 1) = 99,9\%$$

$$\text{Especificidad} = \text{VN} / (\text{FP} + \text{VN}) = 605 / (62 + 605) = 90,7\%$$

El segundo método para validar consistió en conectar dos bases de datos de registros no concordantes: la base de datos A, que contiene los registros hospitalarios de Castilla la Mancha del año 2006 (n=1.873) y la base de datos B, que contiene los registros policiales de Castilla la Mancha del año 2007 (n=6.221). Esta validación nos permitió calcular mediante otro método los falsos positivos: 129 registros (4,0%), lo cual supone una especificidad del 93,1%.

Galicia

Selección de casos del CMBD

La base de datos del CMBD de Galicia incluyó las altas hospitalarias entre los años 2005 y 2008. Se seleccionaron las personas con fecha de ingreso en los años 2005 a 2007 (en el momento del estudio, los datos del año 2008 de la DGT no estaban disponibles).

Cabe destacar que en la comunidad de Galicia la codificación del código E es obligatoria por ley, por lo tanto no existe ningún valor desconocido en este campo.

Representatividad de los datos de Galicia

La Tabla 45 muestra las características de los lesionados de tráfico en Galicia en comparación con las de los lesionados en el resto de España, según gravedad. Cabe destacar una mayor proporción de muertos usuarios de turismo en Galicia, a expensas de una menor proporción de usuarios de motocicleta o ciclomotor. Asimismo, la proporción de muertos en zona urbana fue mucho menor en Galicia (19,4%) que en el resto de España (50,9%). Estas diferencias fueron menos pronunciadas entre los lesionados graves, y prácticamente inexistentes entre los lesionados leves.

Tabla 45. Características de los lesionados de tráfico en Galicia y el resto de España según gravedad. Base de datos policial. España 2005-2007.

	Muertos			Graves			Leves		
	Galicia	Resto de España	Valor de p	Galicia	Resto de España	Valor de p	Galicia	Resto de España	Valor de p
N	897	9.800		5.291	58.917		15.510	340.734	
Sexo			0,000			0,000			0,005
Hombre	61,7	63,0		71,7	72,9		75,7	79,4	
Mujer	38,2	36,4		28,2	26,7		24,3	20,3	
Desconocido	0,2	0,7		0,1	0,4		0,0	0,3	
Edad			0,000			0,000			0,000
0-13 años	3,8	3,6		2,9	2,9		1,9	2,0	
14-17 años	5,3	5,6		6,7	6,5		3,8	3,3	
18-29 años	36,0	34,4		34,0	31,8		29,4	26,6	
30-44 años	25,0	27,3		23,0	27,5		22,3	28,4	
45-64 años	18,1	15,8		20,0	17,7		20,6	22,3	
65-74 años	5,3	3,6		6,1	5,3		10,0	8,3	
75 y más años	3,4	2,4		5,4	4,3		11,7	8,0	
Desconocido	3,0	7,4		2,0	4,0		0,2	1,1	
Usuario			0,000			0,000			0,000
Turismo	72,3	56,2		53,2	44,8		52,4	52,3	
Motocicleta	4,9	11,0		9,4	15,3		10,5	13,1	
Ciclomotor	6,7	14,2		12,2	14,7		6,1	5,9	
Camión o autobús	8,0	8,5		7,7	8,8		6,1	10,2	
Otros	2,4	2,8		4,6	4,2		5,6	4,6	
Peatón	5,7	7,2		12,9	12,0		19,3	13,7	
Desconocido	0,0	0,1		0,0	0,1		0,0	0,2	
Tipo de carretera			0,000			0,000			0,001
Zona urbana	19,4	50,9		14,4	33,0		11,4	16,3	
Carretera	80,6	49,1		85,7	67,0		88,6	83,7	
Desconocido	0,0	0,0		0,0	0,0		0,0	0,0	

En la Tabla 46 se muestra la letalidad en Galicia y el resto de España para el total de muertos y según el tipo de carretera donde tuvo lugar la colisión. La letalidad en Galicia fue superior a la observada en el resto de España, en especial en zona urbana.

Tabla 46. Letalidad (número de muertos por 1.000 lesionados) en Galicia y el resto de España. Base de datos policial. España 2005-2007.

	Galicia	Resto de España
Zona urbana	26,3	8,2
Carretera	44,6	38,2
Total	41,3	23,9

Finalmente, en la Tabla 47 se comparan las características de los lesionados según datos hospitalarios. No se observaron diferencias destacables entre los lesionados ingresados en Galicia y el resto de España.

Tabla 47. Características de los lesionados de tráfico en Galicia y el resto de España según gravedad. Base de datos hospitalaria. España 2005-2007.

	Galicia	Resto de España	Valor de p
N	6.606	77.173	
Sexo			0,699
Hombre	73,4	73,2	
Mujer	26,6	26,8	
Desconocido	0,0	0,0	
Edad			0,000
0-13 años	9,7	7,3	
14-17 años	8,2	10,1	
18-29 años	29,1	27,9	
30-44 años	20,2	23,0	
45-64 años	18,0	17,9	
65-74 años	7,7	7,0	
75 y más años	7,2	6,8	
Desconocido	0,0	0,0	
MAIS			0,000
1 Leve	11,5	9,0	
2 Moderado	47,8	52,4	
3-6 Grave	40,1	37,8	
Desconocido	0,7	0,8	

MAIS: Maximum Abbreviated Injury Scale

Variables de conexión

En la Tabla 48 se describen las variables de conexión.

Tabla 48. Descripción de las variables utilizadas para la conexión probabilística. Galicia 2005-2007.

Variables conexión	DGT	CMBD
Identificador	Nombre: p_id Formato: %8s Valores perdidos: --	Nombre: h_id Formato: %8s Valores perdidos: --
Sexo	Nombre: p_sexo Formato: float %8.0g Categorías: 1 hombres 2 mujeres Valores perdidos: (.) (0,6%)	Nombre: h_sexo Formato: float %8.0g Categorías: 1 hombres 2 mujeres Valores perdidos: --
Edad	Nombre: p_edad Formato: float %8.0g Categorías: 1-99 Valores perdidos: (.) (2,7%)	Nombre: h_edad Formato: float %8.0g Categorías: 1-102 Valores perdidos: --
Provincia	Nombre: p_provin Formato: float %8.0g Categorías: 15 Coruña 27 Lugo 32 Orense 36 Pontevedra Valores perdidos: --	Nombre: h_provin Formato: float %8.0g Categorías: 15 Coruña 27 Lugo 32 Orense 36 Pontevedra Valores perdidos: --
Vehículo	Nombre: p_vehiculo Formato: float %8.0g Categorías: 1 No moto-ciclo 2 Moto-ciclo 3 Vehículo tracción animal 4 Bici Valores perdidos: (.) (9,1%)	Nombre: h_vehiculo Formato: float %8.0g Categorías: 1 No moto-ciclo 2 Moto-ciclo 3 Vehículo tracción animal 4 Bici Valores perdidos: (.) (61,7%)

Cont.

<p>Posición en vehículo</p>	<p>Nombre: p_posicion Formato: float %8.0g Categorías: 1 Conductor 2 Pasajero 3 Ocupante carro/bici 4 Peatón Valores perdidos: (.) (1,6%)</p>	<p>Nombre: h_posición Formato: float %8.0g Categorías: 1 Conductor 2 Pasajero 3 Ocupante carro/bici 4 Peatón Valores perdidos: (.) (48,1%)</p>
<p>Usuario</p>	<p>Nombre: p_posveh Formato: float %8.0g Categorías: 0 Conductor no moto-ciclo 1 Pasajero no moto-ciclo 2 Conductor moto-ciclo 3 Pasajero no moto-ciclo 5 Ocupante carro 6 Ciclista 7 Peatón Valores perdidos: (.) (1,6%)</p>	<p>Nombre: h_posveh Formato: float %8.0g Categorías: 0 Conductor no moto-ciclo 1 Pasajero no moto-ciclo 2 Conductor moto-ciclo 3 Pasajero no moto-ciclo 5 Ocupante carro 6 Ciclista 7 Peatón Valores perdidos: (.) (48,1%)</p>
<p>Fecha colisión</p>	<p>Nombre: p_fecha_acc Formato: flota %d Valores perdidos: --</p>	<p>Nombre: h_fecha_acc Formato: float %d Valores perdidos: --</p>
<p>Localización lesión (diagnóstico principal) (Para la revisión semi-manual)</p>	<p>Nombre: p_locles Formato: float %8.0g Categorías: 1 Cabeza/cara 2 Cuello 3 Pecho 4 Espalda 5 Abdomen 6 EESS 7 EEII Valores perdidos: (.) (37,2%)</p>	<p>Nombre: h_locles Formato: float %8.0g Categorías: 1 Cabeza/cara 2 Cuello 3 Pecho 4 Espalda 5 Abdomen 6 EESS 7 EEII Valores perdidos: (.) (10,3%)</p>

Análisis de la viabilidad de la conexión

1) Según las posibles combinaciones de las variables conexión:

Tamaño de la base de datos de la DGT 2005-2007 sin muertos = 20.824

Tamaño de la base de datos del CMBD ingresos 2005-2007 = 5.226

Suma de registros: 20.824 + 5.226 = 26.050

Variables de conexión:

- Sexo: 2 categorías
- Edad: 99 categorías
- Provincia: 4 categorías
- Vehículo: 4 categorías
- Posición: 4 categorías
- Usuario: 7 categorías
- Fecha de colisión: 365días * 3 años = 1.095 categorías

Para la combinación de las variables sexo, edad, provincia, vehículo, posición y fecha colisión:

$$\text{Combinaciones posibles} = 2 * 99 * 4 * 4 * 4 * 1095 = 13.875.840$$

Para la combinación de las variables sexo, edad, provincia, usuario y fecha colisión:

$$\text{Combinaciones posibles} = 2 * 99 * 4 * 7 * 1095 = 6.070.680$$

Por tanto, según este análisis, la conexión sería viable porque:

$$13.875.840 > 26.050 \text{ y } 6.070.680 > 26.050$$

2) Según el peso mínimo (W_{min}) alcanzado por las variables de conexión:

$$W_t = \log_2(p/(1-p)) - \log_2(E/(A*B-E))$$

P (probabilidad de que dos registros sean correctos) = 0,95

A (tamaño de base de datos 1) = 20.824

B (tamaño de base de datos 2) = 5.226

E (número de pares correctos) = 5.226 (100% según un escenario optimista)
= 3.553 (68% según un escenario pesimista)

Para E=100%: $W_t = \log_2(0,95/(1-0,95)) - \log_2(5.226/(20.824*5.226-5.226)) = 18,59$

Para E=68%: $W_t = \log_2(0,95/(1-0,95)) - \log_2(3.553/(20.824*5.226-3.553)) = 19,15$

Por lo tanto, para que la conexión sea viable, el peso mínimo deberá ser superior a:

$$W_t - 3 = 18,59 - 3 = 15,59 \text{ (para el escenario optimista)}$$

$$= 19,15 - 3 = 16,15 \text{ (para el escenario pesimista)}$$

En la Tabla 49 se muestran los cálculos de los pesos de coincidencia y el peso mínimo para distintas combinaciones de variables.

Tabla 49. Pesos de coincidencia de las variables de conexión, peso mínimo alcanzado según distintas combinaciones de variables y viabilidad de la conexión según el peso mínimo. Galicia 2005-2007.

Variables	Peso de coincidencia	Combinación de variables 1	Combinación de variables 2	Combinación de variables 3	Combinación de variables 4
Fecha de colisión	8,42	X	X	X	X
Provincia de colisión y hospital	0,80	X	X	X	X
Edad	4,95	X	X	X	X
Sexo	0,27	X	X	X	X
Vehículo implicado	1,29	X		X	
Posición respecto el vehículo	1,29	X		X	
Usuario	1,78		X		X
Localización de la lesión principal	1,71			X	X
PESO MÍNIMO		17,02	16,22	18,73	17,93
Viabilidad de la conexión para E=100% (15,59)		Sí	Sí	Sí	Sí
Viabilidad de la conexión para E=68% (16,15)		Sí	Sí	Sí	Sí

Los resultados de la viabilidad de la conexión muestran que ésta es viable con cualquier combinación de variables, tanto para el escenario optimista como para el pesimista.

Resultados de la conexión

Las variables utilizadas para la conexión en Galicia fueron fecha de la colisión, provincia, edad, sexo y usuario. Dado que la conexión resultó viable sin la variable localización de la lesión principal, ésta se utilizó únicamente en la selección semi-manual de pares concordantes.

De los 5.226 ingresos hospitalarios entre los años 2005 y 2007, se encontró un posible registro de conexión de la base de datos policial para 5.183 casos (99,2%) (registros enlazados). De éstos, tan sólo en un 9,5% de registros coincidieron las 5 variables de conexión, y en 28,9% 4 variables (Tabla 50). En esta tabla se muestran, además, los pesos de concordancia mínimos y máximos de estos registros en función del número de variables coincidentes. La distribución de pesos se puede observar en la Figura 9, que se muestra en función del número de coincidencias en la Figura 10. En el caso de Galicia, el histograma de pesos muestra una distribución bimodal. En estas figuras también se muestra el lugar en el que se sitúa el punto de corte visual, equivalente a un peso de 14,5.

Tabla 50. Distribución del número de coincidencias y su correspondiente rango de pesos (n=5.183). Galicia 2005-2007.

	n	%	Peso mínimo	Peso máximo
0	0	0,0	--	--
1	29	0,6	-4,75	7,93
2	780	15,0	-5,27	14,65
3	2.384	46,0	-0,77	19,03
4	1.497	28,9	5,88	24,86
5	493	9,5	16,51	28,50

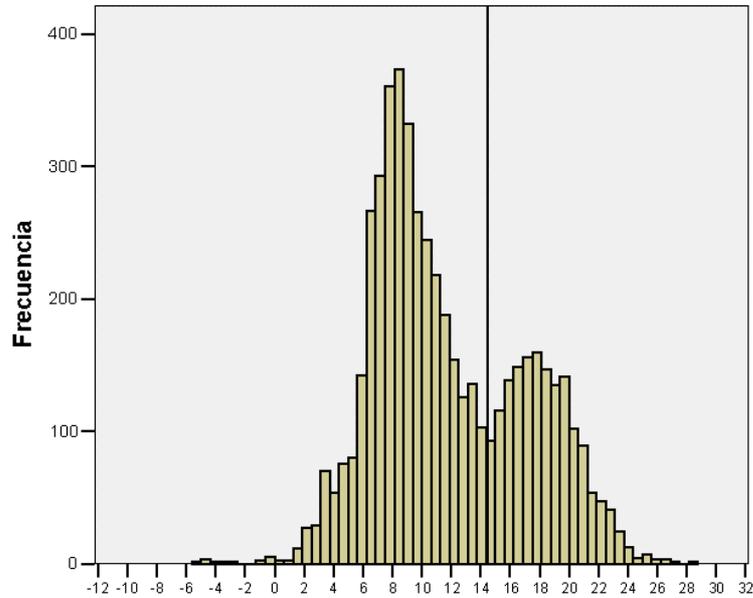


Figura 9. Distribución del los pesos de concordancia y punto de corte visual. Galicia 2005.2007.

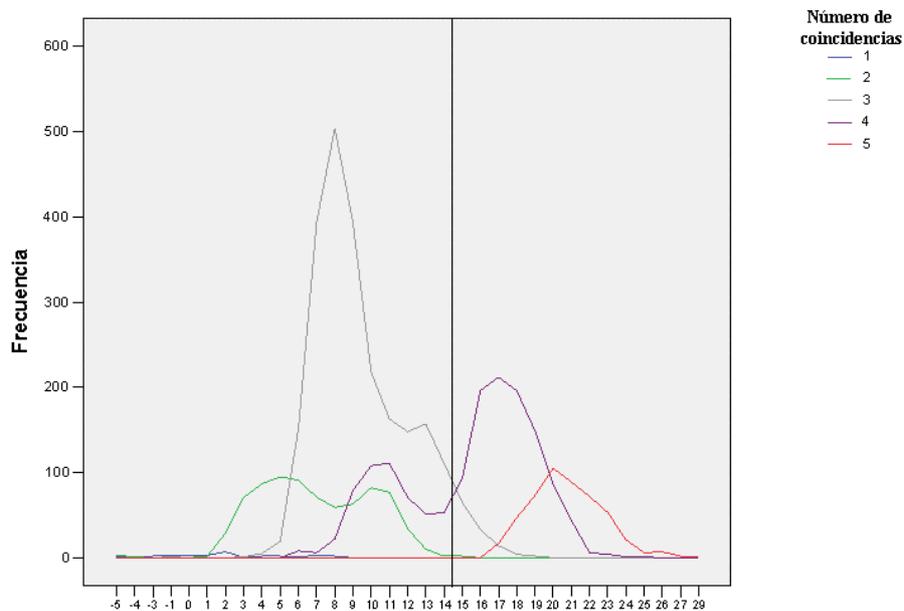


Figura 10. Distribución del los pesos de concordancia en función del número de coincidencias y punto de corte visual. Galicia 2005-2007.

Según los criterios establecidos en el apartado de metodología, se consideraron conectados 1.621 pares de registros (31,0% de la base de datos inicial; 45,6% del número esperado de registros conectados). De éstos, la mayoría (1.209 registros) se

conectaron según el criterio del peso mínimo de los registros en los que coinciden todas las variables de conexión, 341 se conectaron según los criterios establecidos para la revisión semi-manual, 41 registros presentaban una diferencia de edad de sólo un año y 30 registros de hasta 5 años (Tabla 51).

Estos resultados fueron similares a los observados aplicando un punto de corte visual. Según este criterio se conectaron 1.601 registros, que representan un 30,6% de los registros conectables y un 45,0% de los registros que se esperan conectar.

Tabla 51. Resultados de la conexión probabilística. Galicia 2005-2007.

	n	% conectables	% enlazados	% esperados
Registros conectables	5.226			
Registros enlazados	5.183			
Registros conectados esperados^a	3.554			
Registros no conectados	3.562	68,2	68,7	
Registros conectados	1.621	31,0	31,3	45,6
Según criterio1^b	1.209	23,1	23,3	34,0
+criterio2^c	341	6,5	6,6	9,6
+criterio3^d	41	0,8	0,8	1,2
+criterio4^e	30	0,6	0,6	0,8
Registros no enlazados	43	0,8		

a. Se espera que se conecte un 68% de los registros conectables.

b. Cuando coinciden todas las variables de conexión o cuando el peso de concordancia es superior al peso mínimo de los registros en los que coinciden todas las variables de conexión.

c. Según los criterios de la revisión semi-manual.

d. Se permite una variación de +/-1 año de edad.

e. Se permite una variación de +/-5 años de edad.

Se comprobó si las características de los pares conectados difieren de las de los pares no conectados en la base de datos policial y en la hospitalaria (Tabla 52 y Tabla 53). De los 20.824 registros policiales de Galicia, los 1.621 casos conectados representaron un 7,8% del total de registros (un 21,3% de los lesionados graves). Se observó una mayor proporción de hombres que mujeres con registros conectados, así como de menores de 14 años y de mayores de 64 años. En cuanto a la gravedad de los lesionados, aunque la mayoría de los casos conectados fueron identificados por la policía como graves, el 30,4% se identificaron como lesionados leves, que representaron un 3,2% de los

lesionados leves de la base de datos policial. En relación a la localización de la lesión principal, la proporción de registros conectados fue mayor entre los individuos con lesiones en la cabeza o la cara y en las extremidades inferiores, y menor en aquellos con lesiones en el cuello. Los usuarios de vehículos de dos ruedas también presentaron una mayor proporción de registros conectados, al igual que los peatones, mientras que para pasajeros fue menor. En cuanto a las características de la colisión, la proporción de registros conectados fue ligeramente superior entre los individuos lesionados en zona urbana, en comparación con carretera, y en aquellas colisiones ocurridas en horario diurno. Además, aquellos individuos implicados en una colisión con una sola víctima presentaron una mayor probabilidad de haberse conectado (Tabla 52).

Tabla 52. Comparación de las características entre los pares conectados y no conectados de la base de datos de la policía. Análisis bivariado. Galicia 2005-2007.

	Pares no conectados			Pares conectados			Valor de p
	n	% fila	% columna	n	% fila	% columna	
Sexo							<0,001
Hombre	12.113	91,0	63,1	1.203	9,0	74,2	
Mujer	6.979	94,4	36,3	414	5,6	25,5	
Desconocido	111	96,5	0,6	4	3,5	0,2	
Edad							<0,001
0-13 años	642	85,3	3,3	111	14,7	6,8	
14-17 años	1.060	89,8	5,5	121	10,2	7,5	
18-29 años	6.833	92,4	35,6	562	7,6	34,7	
30-44 años	4.766	93,6	24,8	328	6,4	20,2	
45-64 años	3.582	92,6	18,7	288	7,4	17,8	
65-74 años	1.047	91,1	5,5	102	8,9	6,3	
75 y más años	734	89,8	3,8	83	10,2	5,1	
Desconocida	539	95,4	2,8	26	4,6	1,6	
Gravedad del lesionado							<0,001
Leve	15.040	96,8	78,3	492	3,2	30,4	
Grave	4.163	78,7	21,7	1.129	21,3	69,6	
Desconocida	--			--			

Cont.

Localización de la lesión principal							<0,001
Cabeza o cara	2.515	89,1	13,1	309	10,9	19,1	
Cuello	3.361	97,9	17,5	72	2,1	4,4	
Pecho	1.561	92,3	8,1	130	7,7	8,0	
Espalda	744	91,3	3,9	71	8,7	4,4	
Abdomen	227	85,7	1,2	38	14,3	2,3	
Extremidades superiores	1.494	92,0	7,8	132	8,0	8,1	
Extremidades inferiores	2.068	85,2	10,8	358	14,8	22,1	
Desconocida	7.233	93,4	37,7	511	6,6	31,5	
Vehículo							<0,001
Vehículo a motor más de 2 ruedas	14.828	93,9	77,2	964	6,1	59,5	
Vehículo a motor de 2 ruedas	2.549	86,7	13,3	392	13,3	24,2	
Vehículo tracción animal	2	50,0		2	50,0		
Bicicleta	167	84,3	0,9	31	15,7	1,9	
Desconocido	1.657	87,7	8,6	232	12,3	14,3	
Posición							<0,001
Conductor/a	11.066	91,9	57,6	975	8,1	60,1	
Pasajero/a	6.311	94,3	32,9	381	5,7	23,5	
Ocupante carro/bici	169	80,8	0,9	33	19,2	2,0	
Peatón/a	1.358	86,7	7,1	208	13,3	12,8	
Desconocida	299	92,6	1,6	24	7,4	1,5	
Tipo de vía							<0,001
Carretera	3.546	93,9	18,5	232	6,1	14,3	
Zona urbana	15.657	91,9	81,5	1.389	8,1	85,7	
Desconocida	--			--			
Momento de colisión							0,002
Laborable diurno	8.843	92,0	46,1	774	8,0	47,7	
Laborable nocturno	2.860	93,4	14,9	201	6,6	12,4	
Fin de semana diurno	3.734	91,3	19,4	357	8,7	22,0	
Fin de semana nocturno	3.766	92,9	19,6	289	7,1	17,8	
Número de víctimas							<0,001
1	7.742	90,2	40,3	841	9,8	51,9	
2	5.511	93,7	28,7	373	6,3	23,0	
>=3	5.950	93,6	31,0	407	6,4	25,1	

En la Tabla 53 podemos observar que, en la base de datos hospitalaria, no se presentaron diferencias entre hombres y mujeres en la probabilidad de conectarse. Sin embargo, sí que se observaron diferencias en función de la edad, de manera que los más

jóvenes y los mayores tuvieron una menor probabilidad de ser conectados. En cuanto a la estancia hospitalaria, aquellos individuos con un tiempo más corto de estancia hospitalaria presentaron menos posibilidades de ser conectados. También se observaron algunas diferencias según la localización de la lesión principal. Finalmente, los usuarios de bicicleta y peatones presentaron una menor probabilidad de conectarse, mientras que la probabilidad fue mayor en usuarios de vehículo a motor de más de 2 ruedas. Aquellas variables asociadas a una menor probabilidad de que los individuos sean conectados señalan las características de los individuos con una menor probabilidad de estar presentes en la base de datos policial.

Tabla 53. Comparación de las características entre los pares conectados y no conectados de la base de datos hospitalaria. Análisis bivariado. Galicia 2005-2007.

	Pares no conectados			Pares conectados			Valor de p
	n	% fila	% columna	n	% fila	% columna	
Sexo							0,246
Hombre	2.654	68,5	73,6	1.218	31,5	75,1	
Mujer	951	70,2	26,4	403	29,8	24,9	
Desconocido	--			--			
Edad							<0,001
0-13 años	472	80,4	13,1	115	19,6	7,1	
14-17 años	324	71,8	9,0	127	28,2	7,8	
18-29 años	936	62,6	26,0	560	37,4	34,5	
30-44 años	719	68,5	19,9	331	31,5	20,4	
45-64 años	611	67,5	16,9	294	32,5	18,1	
65-74 años	279	72,7	7,7	105	27,3	6,5	
75 y más años	264	74,8	7,3	89	25,2	5,5	
Desconocida	--			--			
Estancia hospitalaria							<0,001
1er cuartil	1.045	77,8	29,0	299	22,2	18,4	
2º cuartil	929	69,7	25,8	404	30,3	24,9	
3er cuartil	875	67,6	24,3	420	32,4	25,9	
4o cuartil	756	60,3	21,0	498	39,7	30,7	

Cont.

Localización de la lesión principal							<0,001
Cabeza o cara	1.006	69,2	27,9	447	30,8	27,6	
Cuello	112	69,6	3,1	49	30,4	3,0	
Pecho	421	66,4	11,7	213	33,6	13,1	
Espalda	219	72,0	6,1	85	28,0	5,2	
Abdomen	187	61,3	5,2	118	38,7	7,3	
Extremidades superiores	464	74,2	12,9	162	25,8	10,0	
Extremidades inferiores	800	65,6	22,2	419	34,4	25,8	
Desconocida	396	75,6	11,0	128	24,4	7,9	
Vehículo							<0,001
Vehículo a motor más de 2 ruedas	246	46,8	6,8	280	53,2	17,3	
Vehículo a motor de 2 ruedas	534	65,4	14,8	283	34,6	17,5	
Vehículo tracción animal	1	50,0		1	50,0		
Bicicleta	617	93,8	17,1	41	6,2	2,5	
Desconocido	2.207	68,5	61,2	1.016	31,5	62,7	
Posición							<0,001
Conductor/a	664	59,9	18,4	444	40,1	27,4	
Pasajero/a	116	49,4	3,2	119	50,6	7,3	
Ocupante carro/bici	618	93,6	17,1	42	6,4	2,6	
Peatón/a	510	71,8	14,1	200	28,2	12,3	
Desconocida	1.697	67,5	47,1	816	32,5	50,3	

Validación de la conexión

Se validó la conexión mediante dos métodos. En primer lugar, se conectó la base de datos A, que corresponde a los registros hospitalarios de Galicia del año 2005 (n=1.858), con la base de datos B, que contiene el 68% de los registros de la base A y los registros hospitalarios de los años 2006-2008 (n=6.138).

En número de falsos positivos obtenidos mediante la conexión fue de 64 registros (3,4%). No se obtuvo ningún falso negativo. En la Tabla 54 se observa la distribución de los registros en función de si son verdaderos positivos o negativos y falsos positivos o negativos. A partir de ello se calculó la sensibilidad y la especificidad de la conexión.

Tabla 54. Verdaderos positivos y negativos y falsos positivos y negativos de la validación de la conexión probabilística. Galicia 2005-2007.

		Conectados		
		No	Sí	Total
Par real	No	559	64	623
	Sí	0	1235	1.235
	Total	559	1.299	1.858

$$\text{Sensibilidad} = \text{VP} / (\text{VP} + \text{FN}) = 1235 / (1235 + 0) = 100\%$$

$$\text{Especificidad} = \text{VN} / (\text{FP} + \text{VN}) = 559 / (64 + 559) = 89,7\%$$

El segundo método para validar consistió en conectar dos bases de datos de registros no concordantes: la base de datos A, que contiene los registros hospitalarios de Galicia del año 2006 (n=1.686) y la base de datos B, que contiene los registros policiales de Galicia del año 2007 (n=6.696). Esta validación nos permitió calcular mediante otro método los falsos positivos: 120 registros (7,2%), lo cual supone una especificidad del 92,9%.

Madrid

Selección de casos del CMBD

La base de datos del CMBD de Madrid incluyó las altas hospitalarias entre los años 2006 y 2008. Los casos fueron previamente seleccionados desde la institución que envió los datos según los siguientes criterios: (1) presencia de un diagnóstico de lesión traumática y (2) “Aseguradora de accidente de tráfico” como régimen de financiación. Dichos criterios son ligeramente distintos a los propuestos y pueden haber afectado la selección de casos, pues se conoce que no siempre consta una “Aseguradora de accidente de tráfico” como régimen de financiación en todos los individuos con código E de accidentes de tráfico.

Representatividad de los datos de Madrid

La Tabla 55 muestra las características de los lesionados de tráfico en Madrid en comparación con las de los lesionados en el resto de España, según gravedad. Cabe destacar que en Madrid se observó una mayor proporción de peatones y de lesionados en motocicleta que en el resto de España, y una menor proporción de lesionados en ciclomotor. Además, la proporción de lesionados leves, graves y muertos en zona urbana fue superior que en el resto de España.

Tabla 55. Características de los lesionados de tráfico en Madrid y el resto de España según gravedad. Base de datos policial. España 2005-2007.

	Muertos			Graves			Leves		
	Madrid	Resto de España	Valor de p	Madrid	Resto de España	Valor de p	Madrid	Resto de España	Valor de p
N	696	10.001		5.356	58.852		53.218	303.026	
Sexo			0,000			0,000			0,009
Hombre	62,1	63,1		70,4	73,0		79,6	79,0	
Mujer	36,2	36,5		27,7	26,7		19,5	20,7	
Desconocido	1,7	0,4		1,9	0,3		0,9	0,2	
Edad			0,000			0,000			0,000
0-13 años	3,3	3,6		3,0	2,9		1,7	2,1	
14-17 años	3,4	5,9		4,8	6,6		3,2	3,3	
18-29 años	33,2	34,7		29,6	32,2		31,5	26,6	
30-44 años	29,4	26,8		27,8	27,1		29,0	27,8	
45-64 años	14,8	16,1		14,6	18,2		19,8	22,3	
65-74 años	3,0	3,8		4,2	5,4		5,0	8,7	
75 y más años	2,0	2,5		4,9	4,4		6,2	8,4	
Desconocido	10,9	6,5		11,2	3,2		3,6	0,8	
Usuario			0,000			0,000			0,000
Turismo	58,2	56,7		36,9	46,2		49,9	52,5	
Motocicleta	13,9	10,2		20,1	14,3		15,4	12,7	
Ciclomotor	8,8	14,8		10,0	14,9		2,7	6,2	
Camión o autobús	7,7	8,6		6,4	8,9		7,9	10,0	
Otros	2,0	2,9		2,5	4,4		1,9	4,8	
Peatón	9,5	6,8		24,1	11,0		22,3	13,6	
Desconocido	0,0	0,1		0,0	0,1		0,0	0,2	
Tipo de carretera			0,000			0,000			0,000
Zona urbana	69,2	46,1		67,2	28,3		25,1	15,2	
Carretera	30,8	53,9		32,8	71,8		74,9	84,8	
Desconocido	0,0	0,0		0,0	0,0		0,0	0,0	

En la Tabla 56 se muestra la letalidad en Madrid y el resto de España para el total de muertos y según el tipo de carretera donde tuvo lugar la colisión. La letalidad en Madrid fue inferior a la observada en el resto de España.

Tabla 56. Letalidad (número de muertos por 1.000 lesionados) en Madrid y el resto de España. Base de datos policial. España 2005-2007.

	Madrid	Resto de España
Zona urbana	4,3	9,7
Carretera	27,9	39,6
Total	11,7	26,9

Finalmente, en la Tabla 57 se comparan las características de los lesionados según datos hospitalarios. No se observan diferencias destacables entre los lesionados ingresados en Madrid y el resto de España.

Tabla 57. Características de los lesionados de tráfico en Madrid y el resto de España según gravedad. Base de datos hospitalaria. España 2005-2007.

	Madrid	Resto de España	Valor de p
N	7.110	76.669	
Sexo			0,000
Hombre	70,9	73,4	
Mujer	29,1	26,6	
Desconocido	0,0	0,0	
Edad			0,000
0-13 años	8,3	7,4	
14-17 años	7,4	10,2	
18-29 años	26,6	28,1	
30-44 años	25,7	22,5	
45-64 años	17,3	18,0	
65-74 años	7,1	7,0	
75 y más años	7,5	6,7	
Desconocido	0,0	0,0	
MAIS			0,000
1 Leve	6,2	9,5	
2 Moderado	50,7	52,2	
3-6 Grave	42,4	37,6	
Desconocido	0,8	0,8	

MAIS: Maximum Abbreviated Injury Scale

Variables de conexión

En la Tabla 58 se describen las variables de conexión.

Tabla 58. Descripción de las variables utilizadas para la conexión probabilística. Madrid 2005-2007.

Variables conexión	DGT	CMBD
Identificador	<u>Nombre: p_id</u> Formato: %8s Valores perdidos: --	<u>Nombre: h_id</u> Formato: %8s Valores perdidos: --
Sexo	<u>Nombre: p_sexo</u> Formato: float %8.0g Categorías: 1 hombres 2 mujeres Valores perdidos: (.) (0,7%)	<u>Nombre: h_sexo</u> Formato: float %8.0g Categorías: 1 hombres 2 mujeres Valores perdidos: --
Edad	<u>Nombre: p_edad</u> Formato: float %8.0g Categorías: 1-99 Valores perdidos: (.) (10,6%)	<u>Nombre: h_edad</u> Formato: float %8.0g Categorías: 1-107 Valores perdidos: --
Vehículo	<u>Nombre: p_vehiculo</u> Formato: float %8.0g Categorías: 1 No moto-ciclo 2 Moto-ciclo 3 Vehículo tracción animal 4 Bici Valores perdidos: (.) (12,8%)	<u>Nombre: h_vehiculo</u> Formato: float %8.0g Categorías: 1 No moto-ciclo 2 Moto-ciclo 3 Vehículo tracción animal 4 Bici Valores perdidos: (.) (76,3%)
Posición en vehículo	<u>Nombre: p_posicion</u> Formato: float %8.0g Categorías: 1 Conductor 2 Pasajero 3 Ocupante carro/bici 4 Peatón Valores perdidos: (.) (1,9%)	<u>Nombre: h_posición</u> Formato: float %8.0g Categorías: 1 Conductor 2 Pasajero 3 Ocupante carro/bici 4 Peatón Valores perdidos: (.) (57,0%)

Cont.

Usuario	Nombre: p_posveh Formato: float %8.0g Categorías: 0 Conductor no moto-ciclo 1 Pasajero no moto-ciclo 2 Conductor moto-ciclo 3 Pasajero no moto-ciclo 5 Ocupante carro 6 Ciclista 7 Peatón Valores perdidos: (.) (1,9%)	Nombre: h_posveh Formato: float %8.0g Categorías: 0 Conductor no moto-ciclo 1 Pasajero no moto-ciclo 2 Conductor moto-ciclo 3 Pasajero no moto-ciclo 5 Ocupante carro 6 Ciclista 7 Peatón Valores perdidos: (.) (57,0%)
Fecha colisión	Nombre: p_fecha_acc Formato: flota %d Valores perdidos: --	Nombre: h_fecha_acc Formato: float %d Valores perdidos: --
Localización lesión (diagnóstico principal) (Para la revisión semi-manual)	Nombre: p_locles Formato: float %8.0g Categorías: 1 Cabeza/cara 2 Cuello 3 Pecho 4 Espalda 5 Abdomen 6 EESS 7 EEII Valores perdidos: (.) (88,6%)	Nombre: h_locles Formato: float %8.0g Categorías: 1 Cabeza/cara 2 Cuello 3 Pecho 4 Espalda 5 Abdomen 6 EESS 7 EEII Valores perdidos: (.) (8,3%)

Análisis de la viabilidad de la conexión

1) Según las posibles combinaciones de las variables conexión:

Tamaño de la base de datos de la DGT 2006-2007 sin muertos = 41.940

Tamaño de la base de datos del CMBD ingresos 2006-2007 = 3.356

Suma de registros: $41.940 + 3.356 = 45.296$

Variables de conexión:

- Sexo: 2 categorías
- Edad: 99 categorías

- Vehículo: 4 categorías
- Posición: 4 categorías
- Fecha de colisión: 365 días * 2 años = 730 categorías

Para la combinación de edad, vehículo, posición y fecha colisión:

Combinaciones posibles = $2*99*4*4*730 = 2.312.640$

Para la combinación de sexo, edad, usuario y fecha colisión:

Combinaciones posibles = $2*99*7*730 = 1.011.780$

Por tanto, según este análisis, la conexión sería viable porque:

$2.312.640 > 45.296$ y $1.011.780 > 45.296$

2) Según el peso mínimo (W_{min}) alcanzado por las variables de conexión:

$$W_t = \log_2(p/(1-p)) - \log_2(E/(A*B-E))$$

P (probabilidad de que dos registros sean correctos) = 0,95

A (tamaño de base de datos 1) = 41.940

B (tamaño de base de datos 2) = 3.356

E (número de pares correctos) = 3.356 (100% según un escenario optimista)
= 2.282 (68% según un escenario pesimista)

Para E=100%: $W_t = \log_2(0,95/(1-0,95)) - \log_2(3.356/(41.940*3.356-3.356)) = 19,60$

Para E=68%: $W_t = \log_2(0,95/(1-0,95)) - \log_2(2.282/(41.940*3.356-2.282)) = 20,16$

Por lo tanto, para que la conexión sea viable, el peso mínimo deberá ser superior a:

$W_t - 3 = 19,60 - 3 = 16,60$ (para el escenario optimista)

$= 20,16 - 3 = 17,16$ (para el escenario pesimista)

En la Tabla 59 se muestran los cálculos de los pesos de coincidencia y el peso mínimo para distintas combinaciones de variables.

Tabla 59. Pesos de coincidencia de las variables de conexión, peso mínimo alcanzado según distintas combinaciones de variables y viabilidad de la conexión según el peso mínimo. Madrid 2005-2007.

Variables	Peso de coincidencia	Combinación de variables 1	Combinación de variables 2	Combinación de variables 3	Combinación de variables 4
Fecha de colisión	7,10	X	X	X	X
Provincia de colisión y hospital	--	X	X	X	X
Edad	4,92	X	X	X	X
Sexo	0,40	X	X	X	X
Vehículo implicado	0,66	X		X	
Posición respecto el vehículo	1,16	X		X	
Usuario	1,16		X		X
Localización de la lesión principal	1,41			X	X
PESO MÍNIMO		14,24	13,58	15,65	14,99
Viabilidad de la conexión para E=100% (16,60)		No	No	No	No
Viabilidad de la conexión para E=68% (17,16)		No	No	No	No

Por lo tanto, los resultados de la viabilidad de la conexión muestran que ésta no es viable en el caso de Madrid, para cualquier combinación de variables y cualquier escenario. Por tanto, no fue posible llevar a cabo la conexión probabilística en esta comunidad autónoma.

4.2. Estimación de lesionados graves

Andalucía

Número de registros hospitalarios Andalucía 2005-2007: 19.916

Número de registros policiales Andalucía 2005-2007: 69.433

Gravedad según policía

Tabla 60. Número de lesionados de tráfico presentes en la base de datos policial, hospitalaria o ambas según los resultados de la conexión probabilística. Andalucía 2005-2007.

	Policía		CMBD no-policía	Total
	Graves	Leves		
En CMBD	1.992	1.120	16.804	19.916
No en CMBD	9.870	56.451		66.321
Total	11.862	57.571	16.804	86.237

CMBD: Conjunto Mínimo Básico de Datos de Altas Hospitalarias

$$\begin{aligned}
 & \text{Lesionados graves corregidos} = \\
 & = \text{lesionados graves policía} / \text{total lesionados policía} * \text{total lesionados (policía + CMBD)} \\
 & = 11.862 / 69.433 * 86.237 = \mathbf{14.733}
 \end{aligned}$$

$$\begin{aligned}
 \text{Factor de conversión N1} & = \text{lesionados graves corregidos} / \text{lesionados graves policía} \\
 & = 14.733 / 11.862 = \mathbf{1,2420}
 \end{aligned}$$

$$\begin{aligned}
 \text{Factor de conversión N2} & = \text{lesionados leves presentes en CMBD} / \text{lesionados leves} \\
 & \text{policía} = 1.120 / 57.571 = \mathbf{0,0195}
 \end{aligned}$$

$$\begin{aligned}
 & \mathbf{\text{Número estimado de lesionados graves}} = \\
 & = \text{N1} * \text{lesionados graves según policía} + \text{N2} * \text{lesionados leves según policía} \\
 & = 1,2420 * 11.862 + 0,0195 * 57.571 = \mathbf{15.853}
 \end{aligned}$$

Distribución de los registros conectados según tiempo de estancia hospitalaria

Tabla 61. Número de lesionados de tráfico presentes en la base de datos policial, hospitalaria o ambas según los resultados de la conexión probabilística en función de los días de estancia hospitalaria. Andalucía 2005-2007.

		Policía		CMBD no-policía	Total
		Graves	Leves		
En CMBD	<1 día	37	13	212	262
	1-3 días	357	322	4.603	5.282
	>3 días	1.598	785	11.989	14.372
No en CMBD		9.870	56.451		66.321
Total		11.862	57.571	16.804	86.237

CMBD: Conjunto Mínimo Básico de Datos de Altas Hospitalarias

Aragón

Número de registros hospitalarios Aragón 2005-2007: 3.377

Número de registros policiales Aragón 2005-2007: 13.795

Gravedad según policía

Tabla 62. Número de lesionados de tráfico presentes en la base de datos policial, hospitalaria o ambas según los resultados de la conexión probabilística. Aragón 2005-2007.

	Policía		CMBD no-policía	Total
	Graves	Leves		
En CMBD	798	426	2.153	3.377
No en CMBD	2.107	10.464		12.571
Total	2.905	10.890	2.153	15.948

CMBD: Conjunto Mínimo Básico de Datos de Altas Hospitalarias

$$\begin{aligned}
 & \text{Lesionados graves corregidos} = \\
 & = \text{lesionados graves policía} / \text{total lesionados policía} * \text{total lesionados (policía + CMBD)} \\
 & = 2.905 / 13.795 * 15.948 = \mathbf{3.358}
 \end{aligned}$$

$$\begin{aligned}
 \text{Factor de conversión N1} & = \text{lesionados graves corregidos} / \text{lesionados graves policía} \\
 & = 3.358 / 2.905 = \mathbf{1,1561}
 \end{aligned}$$

$$\begin{aligned}
 \text{Factor de conversión N2} & = \text{lesionados leves presentes en CMBD} / \text{lesionados leves} \\
 & \text{policía} = 426 / 10.890 = \mathbf{0,0391}
 \end{aligned}$$

$$\begin{aligned}
 & \mathbf{\text{Número estimado de lesionados graves}} = \\
 & = \text{N1} * \text{lesionados graves según policía} + \text{N2} * \text{lesionados leves según policía} \\
 & = 1,1561 * 2.905 + 0,0391 * 10.890 = \mathbf{3.784}
 \end{aligned}$$

Distribución de los registros conectados según tiempo de estancia hospitalaria

Tabla 63. Número de lesionados de tráfico presentes en la base de datos policial, hospitalaria o ambas según los resultados de la conexión probabilística en función de los días de estancia hospitalaria. Aragón 2005-2007.

		Policía		CMBD no-policía	Total
		Graves	Leves		
En CMBD	<1 día	18	16	69	103
	1-3 días	207	169	709	1.085
	>3 días	573	241	1.373	2.187
No en CMBD		2.107	10.464		12.571
Total		2.905	10.890	2.153	15.948

CMBD: Conjunto Mínimo Básico de Datos de Altas Hospitalarias

Cantabria

Número de registros hospitalarios Cantabria 2005-2007: 947

Número de registros policiales Cantabria 2005-2007: 4.206

Gravedad según policía

Tabla 64. Número de lesionados de tráfico presentes en la base de datos policial, hospitalaria o ambas según los resultados de la conexión probabilística. Cantabria 2005-2007.

	Policía		CMBD no-policía	Total
	Graves	Leves		
En CMBD	112	95	740	947
No en CMBD	410	3.589		3.999
Total	522	3.684	740	4.946

CMBD: Conjunto Mínimo Básico de Datos de Altas Hospitalarias

$$\begin{aligned}
 & \text{Lesionados graves corregidos} = \\
 & = \text{lesionados graves policía} / \text{total lesionados policía} * \text{total lesionados (policía + CMBD)} \\
 & = 522 / 4.206 * 4.946 = \mathbf{614}
 \end{aligned}$$

$$\begin{aligned}
 \text{Factor de conversión N1} & = \text{lesionados graves corregidos} / \text{lesionados graves policía} \\
 & = 614 / 522 = \mathbf{1,1759}
 \end{aligned}$$

$$\begin{aligned}
 \text{Factor de conversión N2} & = \text{lesionados leves presentes en CMBD} / \text{lesionados leves} \\
 & \text{policía} = 95 / 3.684 = \mathbf{0,0258}
 \end{aligned}$$

$$\begin{aligned}
 & \mathbf{\text{Número estimado de lesionados graves}} = \\
 & = \text{N1} * \text{lesionados graves según policía} + \text{N2} * \text{lesionados leves según policía} \\
 & = 1,1759 * 522 + 0,0258 * 3.684 = \mathbf{709}
 \end{aligned}$$

Distribución de los registros conectados según tiempo de estancia hospitalaria

Tabla 65. Número de lesionados de tráfico presentes en la base de datos policial, hospitalaria o ambas según los resultados de la conexión probabilística en función de los días de estancia hospitalaria. Cantabria 2005-2007.

		Policía		CMBD no-policía	Total
		Graves	Leves		
En CMBD	<1 día	5	5	30	40
	1-3 días	12	36	228	276
	>3 días	95	54	472	621
No en CMBD		410	3.589		3.999
Total		522	3.684	740	4.946

CMBD: Conjunto Mínimo Básico de Datos de Altas Hospitalarias

Castilla la Mancha

Número de registros hospitalarios Castilla la Mancha 2005-2007: 5.693

Número de registros policiales Castilla la Mancha 2005-2007: 19.544

Gravedad según policía

Tabla 66. Número de lesionados de tráfico presentes en la base de datos policial, hospitalaria o ambas según los resultados de la conexión probabilística. Castilla la Mancha 2005-2007.

	Policía		CMBD no-policía	Total
	Graves	Leves		
En CMBD	1.045	517	4.131	5.693
No en CMBD	3.084	14.898		17.982
Total	4.129	15.415	4.131	23.675

CMBD: Conjunto Mínimo Básico de Datos de Altas Hospitalarias

$$\begin{aligned}
 &\text{Lesionados graves corregidos} = \\
 &= \text{lesionados graves policía} / \text{total lesionados policía} * \text{total lesionados (policía + CMBD)} \\
 &= 4.129 / 19.544 * 23.675 = \mathbf{5.002}
 \end{aligned}$$

$$\begin{aligned}
 \text{Factor de conversión N1} &= \text{lesionados graves corregidos} / \text{lesionados graves policía} \\
 &= 5.002 / 4.129 = \mathbf{1,2114}
 \end{aligned}$$

$$\begin{aligned}
 \text{Factor de conversión N2} &= \text{lesionados leves presentes en CMBD} / \text{lesionados leves policía} \\
 &= 517 / 15.415 = \mathbf{0,0335}
 \end{aligned}$$

$$\begin{aligned}
 &\mathbf{\text{Número estimado de lesionados graves}} = \\
 &= \text{N1} * \text{lesionados graves según policía} + \text{N2} * \text{lesionados leves según policía} \\
 &= 1,2114 * 4.129 + 0,0335 * 15.415 = \mathbf{5.519}
 \end{aligned}$$

Distribución de los registros conectados según tiempo de estancia hospitalaria

Tabla 67. Número de lesionados de tráfico presentes en la base de datos policial, hospitalaria o ambas según los resultados de la conexión probabilística en función de los días de estancia hospitalaria. Castilla la Mancha 2005-2007.

		Policía		CMBD no-policía	Total
		Graves	Leves		
En CMBD	<1 día	16	7	126	149
	1-3 días	276	212	1.438	1.926
	>3 días	753	298	2.512	3.563
No en CMBD		3.084	14.898		17.982
Total		4.129	15.415	4.131	23.675

CMBD: Conjunto Mínimo Básico de Datos de Altas Hospitalarias

Galicia

Número de registros hospitalarios Galicia 2005-2007: 5.226

Número de registros policiales Galicia 2005-2007: 20.824

Gravedad según policía

Tabla 68. Número de lesionados de tráfico presentes en la base de datos policial, hospitalaria o ambas según los resultados de la conexión probabilística. Galicia 2005-2007.

	Policía		CMBD no-policía	Total
	Graves	Leves		
En CMBD	1.129	492	3.605	5.226
No en CMBD	4.163	15.040		19.203
Total	5.292	15.532	3.605	24.429

CMBD: Conjunto Mínimo Básico de Datos de Altas Hospitalarias

$$\begin{aligned}
 & \text{Lesionados graves corregidos (total)} = \\
 & = \text{lesionados graves policía} / \text{total lesionados policía} * \text{total lesionados (policía + CMBD)} \\
 & = 5.292 / 20.824 * 24.429 = \mathbf{6.208}
 \end{aligned}$$

$$\begin{aligned}
 \text{Factor de conversión N1} & = \text{lesionados graves corregidos} / \text{lesionados graves policía} \\
 & = 6.208 / 5.292 = \mathbf{1,1731}
 \end{aligned}$$

$$\begin{aligned}
 \text{Factor de conversión N2} & = \text{lesionados leves presentes en CMBD} / \text{lesionados leves} \\
 & \text{policía} = 492 / 15.532 = \mathbf{0,0317}
 \end{aligned}$$

$$\begin{aligned}
 & \mathbf{\text{Número estimado de lesionados graves}} = \\
 & = \text{N1} * \text{lesionados graves según policía} + \text{N2} * \text{lesionados leves según policía} \\
 & = 1,1731 * 5.292 + 0,0317 * 15.532 = \mathbf{6.700}
 \end{aligned}$$

Distribución de los registros conectados según tiempo de estancia hospitalaria

Tabla 69. Número de lesionados de tráfico presentes en la base de datos policial, hospitalaria o ambas según los resultados de la conexión probabilística en función de los días de estancia hospitalaria. Galicia 2005-2007.

		Policía		CMBD no-policía	Total
		Graves	Leves		
En CMBD	<1 día	20	13	98	131
	1-3 días	137	129	935	1.201
	>3 días	972	350	2.529	3.851
No en CMBD		4.163	15.040		19.203
Total		5.292	15.532	3.605	24.429

CMBD: Conjunto Mínimo Básico de Datos de Altas Hospitalarias

Comparación entre comunidades autónomas

En la Tabla 70 se muestran los resultados de las estimaciones de lesionados graves para las cinco comunidades autónomas disponibles. Los factores de conversión son ligeramente diferentes en cada comunidad autónoma, lo cual conlleva que la diferencia entre el número de lesionados graves estimado y el notificado por la policía también varíe en función de la región geográfica, entre un 21,0% en Galicia y un 26,4% en Cantabria.

Tabla 70. Número de lesionados graves observados y estimados, diferencia entre ambos y factores de conversión según comunidad autónoma. Andalucía, Aragón, Cantabria, Castilla la Mancha y Galicia 2005-2007.

	Andalucía	Aragón	Cantabria	Castilla la Mancha	Galicia
Lesionados graves (ingresos CMBD)	19.916	3.377	947	5.693	5.226
Lesionados graves (policía)	11.862	2.905	522	4.129	5.292
Lesionados graves (estimados)	14.733	3.784	709	5.519	6.700
Diferencia (%)^a	19,5	23,2	26,4	25,2	21,0
Factor de conversión N1	1,2420	1,1561	1,1759	1,2114	1,1731
Factor de conversión N2	0,0195	0,0391	0,0258	0,0335	0,0317

a. Calculado como $[(\text{lesionados graves (estimados)} - \text{lesionados graves (policía)}) / \text{lesionados graves (estimados)}] * 100$.

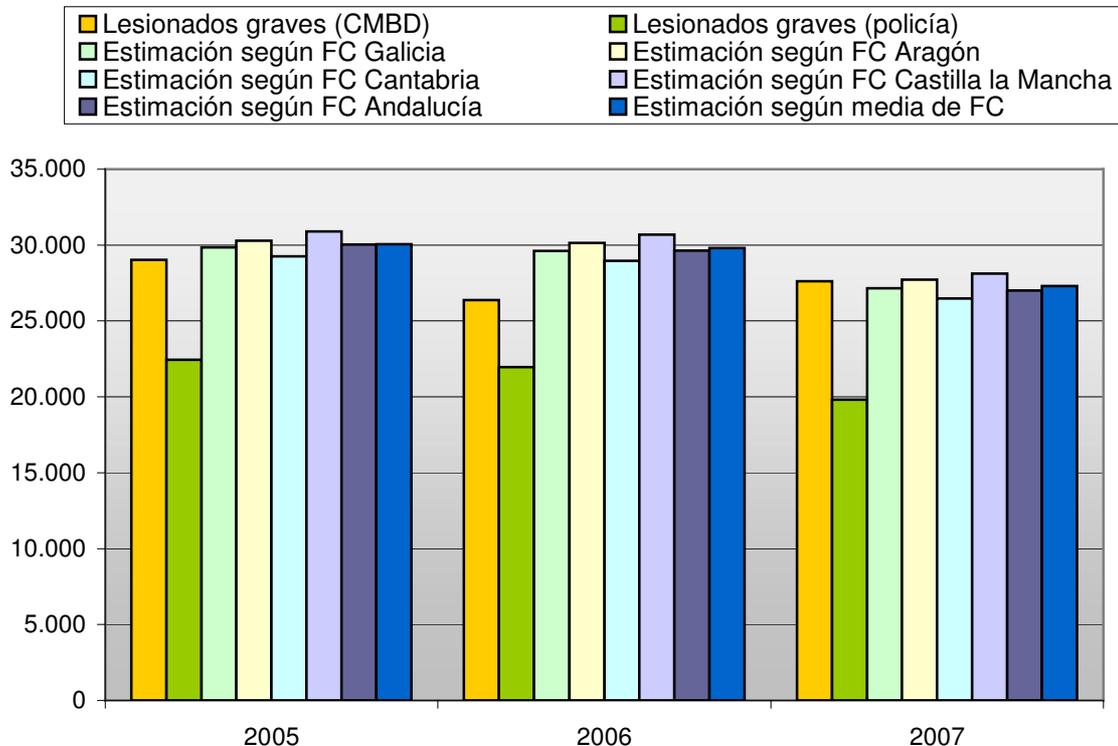
Se proyectaron los resultados por comunidad autónoma para estimar los lesionados graves en España. En la Tabla 71 y en la Figura 12 se muestran las estimaciones según se utilicen los factores de conversión de cada comunidad autónoma o una media de los factores de conversión de las cuatro comunidades autónomas.

Tabla 71. Número de lesionados graves observados, estimados y diferencia entre ambos en España 2005-2007 según los factores de conversión de cada comunidad autónoma y según la media de los factores de conversión. Proyección de los resultados para España.

	2005	2006	2007
Lesionados graves (CMBD)	29.017	26.382	27.611
Lesionados graves (DGT)	21.859	21.382	19.295
Lesionados leves (DGT)	110.950	122.068	123.226
Factores de conversión Andalucía			
Lesionados graves (estimados)	30.039	29.643	27.011
Diferencia (%) ^a	27,2	27,9	28,6
Factores de conversión Aragón			
Lesionados graves (estimados)	30.286	30.150	27.724
Diferencia (%) ^a	27,8	29,1	30,4
Factores de conversión Cantabria			
Lesionados graves (estimados)	29.254	28.962	26.477
Diferencia (%) ^a	25,3	26,2	27,1
Factores de conversión Castilla la Mancha			
Lesionados graves (estimados)	30.905	30.681	28.130
Diferencia (%) ^a	29,3	30,3	31,4
Factores de conversión Galicia			
Lesionados graves (estimados)	29.846	29.620	27.149
Diferencia (%) ^a	26,8	27,8	28,9
Media de los factores de conversión			
Lesionados graves (estimados)	30.066	29.811	27.298
Diferencia (%) ^a	27,3	28,3	29,3

a. Calculado como $[(\text{lesionados graves (estimados)} - \text{lesionados graves (policía)}) / \text{lesionados graves (estimados)}] * 100$.

Figura 11. Estimación de heridos graves según diferentes factores de conversión derivados Andalucía, Aragón, Cantabria, Castilla la Mancha, Galicia y media de factores de conversión. España 2005-2007.



Factores de conversión según sexo y edad

En la Tabla 72 se muestran los factores de conversión por sexo y edad para cada comunidad autónoma. Los factores de conversión difieren ligeramente entre comunidades autónomas, de manera que Andalucía es donde los factores que se deben aplicar presentan los valores más elevados. Se observan ciertas variaciones en los factores de conversión por sexo, aunque éstos varían especialmente por edad. Los factores de conversión muestran que los lesionados graves de 75 y más años son los que están más infraestimados. Además, se observa que la variación según grupos de edad en los factores de conversión no es homogénea entre comunidades autónomas. Por ejemplo, para los grupos de edad hasta los 75 años, mientras que en Andalucía los factores de conversión sugieren que la infranotificación es mayor para el grupo de edad

entre los 14 y 17 años, en Aragón este grupo de edad se encuentra entre los que presentan una mejor notificación.

En la Tabla 73 se muestra el número estimado de heridos graves a nivel estatal aplicando la media de los factores de conversión de las cinco comunidades autónomas conectadas, por sexo y edad. Se muestra también la diferencia que existe entre dicho número y el número de heridos graves calculado por la DGT.

Tabla 72. Factores de conversión para el cálculo de lesionados graves según sexo y edad. Andalucía, Aragón, Cantabria, Castilla la Mancha y Galicia 2005-2007.

	Factores de conversión									
	Andalucía		Aragón		Cantabria		Castilla la Mancha		Galicia	
	N1	N2	N1	N2	N1	N2	N1	N2	N1	N2
Sexo										
Hombres	1,2917	0,0216	1,1549	0,0433	1,1853	0,0293	1,2127	0,0388	1,1762	0,0379
Mujeres	1,1564	0,0164	1,1469	0,0325	1,1517	0,0206	1,1922	0,0251	1,1580	0,0215
Edad										
0-13 años	1,1912	0,0345	1,1434	0,0907	1,1067	0,0667	1,1435	0,0653	1,1315	0,0719
14-17 años	1,2912	0,0269	1,1274	0,0535	1,1784	0,0440	1,2350	0,0516	1,1787	0,0556
18-29 años	1,2301	0,0172	1,1391	0,0328	1,1550	0,0290	1,2047	0,0281	1,1608	0,0316
30-44 años	1,2181	0,0170	1,1302	0,0628	1,1761	0,0171	1,1949	0,0309	1,1690	0,0253
45-64 años	1,1989	0,0209	1,1388	0,0410	1,1730	0,0219	1,1908	0,0309	1,1496	0,0239
65-74 años	1,2177	0,0301	1,1265	0,0552	1,1574	0,0175	1,1908	0,0425	1,1549	0,0266
75 años y más	1,5037	0,0168	1,3329	0,0252	1,3803	0,0200	1,4178	0,0295	1,5973	0,0470

Tabla 73. Estimación del número de lesionados graves según sexo y edad y diferencia respecto al número estimado por la DGT. España 2005-2007.

	Media FC ^a		Heridos graves (DGT)	Heridos graves (estimación FC)	Diferencia (%) ^b
	N1	N2			
Sexo					
Hombres	1,2042	0,0342	46.475	63.573	26,9
Mujeres	1,1610	0,0232	17.129	22.876	25,1
Edad					
0-13 años	1,1433	0,0658	1.852	2.960	37,4
14-17 años	1,2021	0,0463	4.158	5.914	29,7
18-29 años	1,1779	0,0277	20.522	27.573	25,6
30-44 años	1,1777	0,0306	17.435	23.498	25,8
45-64 años	1,1702	0,0277	11.493	15.019	23,5
65-74 años	1,1695	0,0344	3.422	4.454	23,2
75 años y más	1,4464	0,0277	2.841	4.348	34,7

FC: Factores de conversión

a. Media de los factores de conversión de Andalucía, Aragón, Cantabria, Castilla la Mancha y Galicia.

b. Calculado como $[(\text{lesionados graves (estimados)} - \text{lesionados graves (policía)}) / \text{lesionados graves (estimados)}] * 100$.

4.3. Estimación de muertos a 30 días

Andalucía

Tabla 74. Número de fallecidos por accidente de tráfico presentes en la base de datos policial, hospitalaria o ambas según los resultados de la conexión probabilística. Andalucía 2005-2007.

	Policía			CMBD no-policía	Total
	Muertos	Graves	Leves		
CMBD-muertos <=30d	--	97	17	402	516
CMBD-alta o muertos >30d	--	1.895	1.103	16.402	19.400
No en CMBD	1.847	9.870	56.451		68.168
TOTAL	1.847	11.862	57.571	16.804	88.084

CMBD: Conjunto Mínimo Básico de Datos de Altas Hospitalarias

$$\begin{aligned}
 &\text{Muertos de tráfico corregidos} = \\
 &= \text{muertos policía} / \text{total lesionados policía} * \text{total lesionados (policía + CMBD)} \\
 &= 1.847 / 71.280 * 88.084 = \mathbf{2.282}
 \end{aligned}$$

$$\begin{aligned}
 \text{Factor de conversión N1} &= \text{muertos corregidos} / \text{muertos policía} \\
 &= 2.282 / 1.847 = \mathbf{1,2355}
 \end{aligned}$$

$$\begin{aligned}
 \text{Factor de conversión N2} &= \text{graves según policía muertos } \leq 30 \text{ días en CMBD} / \text{graves} \\
 &\text{policía} = 97 / 11.862 = \mathbf{0,0082}
 \end{aligned}$$

$$\begin{aligned}
 \text{Factor de conversión N3} &= \text{leves según policía muertos } \leq 30 \text{ días en CMBD} / \text{leves} \\
 &\text{policía} = 17 / 57.571 = \mathbf{0,0003}
 \end{aligned}$$

$$\begin{aligned}
 &\mathbf{\text{Número estimado de muertos de tráfico}} = \\
 &= \text{N1} * \text{muertos según policía} + \text{N2} * \text{lesionados graves según policía} + \text{N3} * \text{lesionados} \\
 &\text{leves según policía} = 1,2355 * 1.847 + 0,0082 * 11.862 + 0,0003 * 57.571 = \mathbf{2.397}
 \end{aligned}$$

Aragón

Tabla 75. Número de fallecidos por accidente de tráfico presentes en la base de datos policial, hospitalaria o ambas según los resultados de la conexión probabilística. Aragón 2005-2007.

	Policía			CMBD no-policía	Total
	Muertos	Graves	Leves		
CMBD-muertos <=30d	--	30	7	47	84
CMBD-alta o muertos >30d	--	768	419	2.106	3.293
No en CMBD	497	2.107	10.464		13.068
TOTAL	497	2.905	10.890	2.153	16.445

CMBD: Conjunto Mínimo Básico de Datos de Altas Hospitalarias

$$\begin{aligned}
 &\text{Muertos de tráfico corregidos} = \\
 &= \text{muertos policía} / \text{total lesionados policía} * \text{total lesionados (policía + CMBD)} \\
 &= 497 / 14.292 * 16.445 = \mathbf{572}
 \end{aligned}$$

$$\begin{aligned}
 &\text{Factor de conversión N1} = \text{muertos corregidos} / \text{muertos policía} \\
 &= 572 / 497 = \mathbf{1,1506}
 \end{aligned}$$

$$\begin{aligned}
 &\text{Factor de conversión N2} = \text{graves según policía muertos } \leq 30 \text{ días en CMBD} / \text{graves} \\
 &\text{policía} = 30 / 2.905 = \mathbf{0,0103}
 \end{aligned}$$

$$\begin{aligned}
 &\text{Factor de conversión N3} = \text{leves según policía muertos } \leq 30 \text{ días en CMBD} / \text{leves} \\
 &\text{policía} = 7 / 10.890 = \mathbf{0,0006}
 \end{aligned}$$

$$\begin{aligned}
 &\mathbf{\text{Número estimado de muertos de tráfico}} = \\
 &= \text{N1} * \text{muertos según policía} + \text{N2} * \text{lesionados graves según policía} + \text{N3} * \text{lesionados} \\
 &\text{leves según policía} = 1,1506 * 497 + 0,0103 * 2.905 + 0,0006 * 10.890 = \mathbf{608}
 \end{aligned}$$

Cantabria

Tabla 76. Número de fallecidos por accidente de tráfico presentes en la base de datos policial, hospitalaria o ambas según los resultados de la conexión probabilística. Cantabria 2005-2007.

	Policía			CMBD no-policía	Total
	Muertos	Graves	Leves		
CMBD-muertos <=30d	--	2	0	14	16
CMBD-alta o muertos >30d	--	110	95	726	931
No en CMBD	106	410	3.589		4.105
TOTAL	106	522	3.684	740	5.052

CMBD: Conjunto Mínimo Básico de Datos de Altas Hospitalarias

$$\begin{aligned}
 &\text{Muertos de tráfico corregidos} = \\
 &= \text{muertos policía} / \text{total lesionados policía} * \text{total lesionados (policía + CMBD)} \\
 &= 106 / 4.312 * 5.052 = \mathbf{124}
 \end{aligned}$$

$$\begin{aligned}
 &\text{Factor de conversión N1} = \text{muertos corregidos} / \text{muertos policía} \\
 &= 124 / 106 = \mathbf{1,1716}
 \end{aligned}$$

$$\begin{aligned}
 &\text{Factor de conversión N2} = \text{graves según policía muertos } \leq 30 \text{ días en CMBD} / \text{graves} \\
 &\text{policía} = 2 / 522 = \mathbf{0,0038}
 \end{aligned}$$

$$\begin{aligned}
 &\text{Factor de conversión N3} = \text{leves según policía muertos } \leq 30 \text{ días en CMBD} / \text{leves} \\
 &\text{policía} = 0 / 3.684 = \mathbf{0}
 \end{aligned}$$

$$\begin{aligned}
 &\mathbf{\text{Número estimado de muertos de tráfico}} = \\
 &= \text{N1} * \text{muertos según policía} + \text{N2} * \text{lesionados graves según policía} + \text{N3} * \text{lesionados} \\
 &\text{leves según policía} = 1,1716 * 106 + 0,0038 * 522 + 0 * 3.684 = \mathbf{126}
 \end{aligned}$$

Castilla la Mancha

Tabla 77. Número de fallecidos por accidente de tráfico presentes en la base de datos policial, hospitalaria o ambas según los resultados de la conexión probabilística. Castilla la Mancha 2005-2007.

	Policía			CMBD no-policía	Total
	Muertos	Graves	Leves		
CMBD-muertos <=30d	--	31	11	102	144
CMBD-alta o muertos >30d	--	1.014	506	4.029	5.549
No en CMBD	873	3.084	14.898		18.855
TOTAL	873	4.129	15.415	4.131	24.548

CMBD: Conjunto Mínimo Básico de Datos de Altas Hospitalarias

$$\begin{aligned}
 &\text{Muertos de tráfico corregidos} = \\
 &= \text{muertos policía} / \text{total lesionados policía} * \text{total lesionados (policía + CMBD)} \\
 &= 873 / 20.417 * 24.548 = \mathbf{1.049}
 \end{aligned}$$

$$\begin{aligned}
 &\text{Factor de conversión N1} = \text{muertos corregidos} / \text{muertos policía} \\
 &= 1.049 / 873 = \mathbf{1,2023}
 \end{aligned}$$

$$\begin{aligned}
 &\text{Factor de conversión N2} = \text{graves según policía muertos } \leq 30 \text{ días en CMBD} / \text{graves} \\
 &\text{policía} = 31 / 4.129 = \mathbf{0,0075}
 \end{aligned}$$

$$\begin{aligned}
 &\text{Factor de conversión N3} = \text{leves según policía muertos } \leq 30 \text{ días en CMBD} / \text{leves} \\
 &\text{policía} = 11 / 15.415 = \mathbf{0,0007}
 \end{aligned}$$

$$\begin{aligned}
 &\mathbf{\text{Número estimado de muertos de tráfico}} = \\
 &= \text{N1} * \text{muertos según policía} + \text{N2} * \text{lesionados graves según policía} + \text{N3} * \text{lesionados} \\
 &\text{leves según policía} = 1,2023 * 873 + 0,0075 * 4.129 + 0,0007 * 15.415 = \mathbf{1.091}
 \end{aligned}$$

Galicia

Tabla 78. Número de fallecidos por accidente de tráfico presentes en la base de datos policial, hospitalaria o ambas según los resultados de la conexión probabilística. Galicia 2005-2007.

	Policía			CMBD no-policía	Total
	Muertos	Graves	Leves		
CMBD-muertos <=30d	--	38	8	96	142
CMBD-alta o muertos >30d	--	1.091	484	3.509	5.084
No en CMBD	897	4.163	15.040		20.100
TOTAL	897	5.292	15.532	3.605	25.326

CMBD: Conjunto Mínimo Básico de Datos de Altas Hospitalarias

$$\begin{aligned}
 &\text{Muertos de tráfico corregidos} = \\
 &= \text{muertos policía} / \text{total lesionados policía} * \text{total lesionados (policía + CMBD)} \\
 &= 897 / 21.721 * 25.326 = \mathbf{1.046}
 \end{aligned}$$

$$\begin{aligned}
 &\text{Factor de conversión N1} = \text{muertos corregidos} / \text{muertos policía} \\
 &= 1.046 / 897 = \mathbf{1,1660}
 \end{aligned}$$

$$\begin{aligned}
 &\text{Factor de conversión N2} = \text{graves según policía muertos} \leq 30 \text{ días en CMBD} / \text{graves} \\
 &\text{policía} = 38 / 5.292 = \mathbf{0,0072}
 \end{aligned}$$

$$\begin{aligned}
 &\text{Factor de conversión N3} = \text{leves según policía muertos} \leq 30 \text{ días en CMBD} / \text{leves} \\
 &\text{policía} = 8 / 15.532 = \mathbf{0,0005}
 \end{aligned}$$

$$\begin{aligned}
 &\mathbf{\text{Número estimado de muertos de tráfico}} = \\
 &= \text{N1} * \text{muertos según policía} + \text{N2} * \text{lesionados graves según policía} + \text{N3} * \text{lesionados} \\
 &\text{leves según policía} = 1,1660 * 897 + 0,0072 * 5.292 + 0,0005 * 15.532 = \mathbf{1.092}
 \end{aligned}$$

Comparación entre comunidades autónomas

En la Tabla 79 se muestran los resultados de las estimaciones de muertos a 30 días para las cinco comunidades autónomas disponibles. Los factores de conversión son ligeramente diferentes en cada comunidad autónoma, lo cual conlleva que la diferencia entre el número de muertos a 30 días estimado y el número de muertos a 24 horas notificado por la policía también varíe en función de la región geográfica, entre un 15,9% en Cantabria y un 20,0% en Castilla la Mancha. Las estimaciones de muertos a 30 días en Galicia son un 4,2% superior a la estimada según los cálculos de la DGT, un 5,6% superior en Aragón y en Cantabria, un 6,0% superior en Andalucía y un 9,8% superior en Castilla la Mancha.

Tabla 79. Número de muertos a 30 días observados y estimados, diferencia entre ambos y factores de conversión según comunidad autónoma. Andalucía, Aragón, Cantabria, Castilla la Mancha y Galicia 2005-2007. Proyección de los resultados para España.

	Andalucía	Aragón	Cantabria	Castilla la Mancha	Galicia
Muertos a 24h (policía)	1.847	497	106	873	897
Muertos a 30 días (CMBD)	516	84	16	144	142
Muertos a 30 días (policía 24h+CMBD)	2.363	581	122	1.017	1.039
Muertos a 30 días (estimación DGT)	2.146	574	119	984	1.046
Muertos a 30 días (estimados)	2.282	608	126	1.091	1.092
Diferencia vs. estimación 30 días DGT (%)^a	6,0	5,6	5,6	9,8	4,2
Diferencia vs. muertos 24 horas policía (%)^b	19,1	18,3	15,9	20,0	17,9
Factor de conversión N1	1,2355	1,1506	1,1716	1,2023	1,1660
Factor de conversión N2	0,0082	0,0103	0,0038	0,0075	0,0072
Factor de conversión N3	0,0003	0,0006	0,0000	0,0007	0,0005

a. Calculado como [(muertos a 30 días (estimados)-muertos a 30 días (estimación DGT))/ muertos a 30 días (estimados)]*100.

b. Calculado como [(muertos a 30 días (estimados)-muertos a 24h (policía))/ muertos a 30 días (estimados)]*100.

Se proyectaron los resultados por comunidad autónoma para estimar los muertos a 30 días en España. En la Tabla 80 y en la Figura 12 se muestran las estimaciones según se utilicen los factores de conversión de cada comunidad autónoma o una media de los

factores de conversión de las cuatro comunidades autónomas. En todos los casos las estimaciones obtenidas son superiores a las estimadas por la DGT con la metodología convencional.

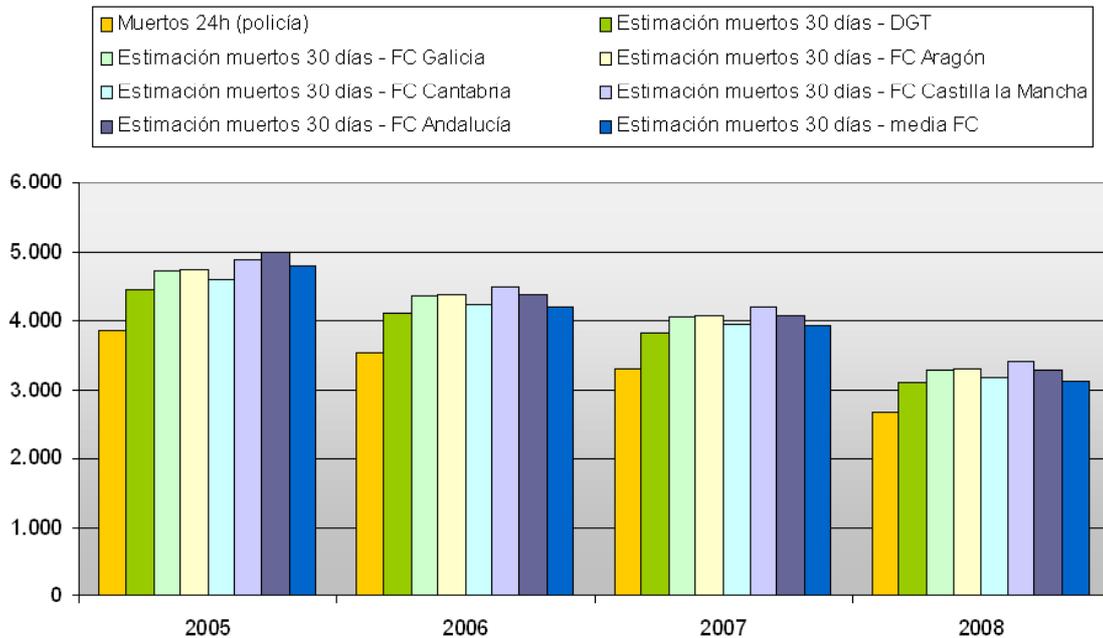
Tabla 80. Número de muertos a 30 días observados, estimados y diferencia entre ambos en España 2005-2007 según los factores de conversión de cada comunidad autónoma y según la media de los factores de conversión.

	2005	2006	2007	2008
Lesionados leves (DGT)	110.950	122.068	123.226	114.459
Lesionados graves (DGT)	21.859	21.382	19.295	16.488
Muertos a 24 horas (DGT)	3.857	3.535	3.305	2.656
Muertos a 30 días (estimación DGT)	4.442	4.104	3.823	3.100
Factores de conversión Andalucía				
Muertos a 30 días (estimados)	4.983	4.367	4.083	3.281
Diferencia vs. estimación 30 días DGT (%)^a	10,9	6,0	6,4	5,5
Diferencia vs. muertos 24 horas policía (%)^b	22,6	19,1	19,1	19,1
Factores de conversión Aragón				
Muertos a 30 días (estimados)	4.736	4.367	4.081	3.299
Diferencia vs. estimación 30 días DGT (%)^a	6,2	6,0	6,3	6,0
Diferencia vs. muertos 24 horas policía (%)^b	18,6	19,0	19,0	19,5
Factores de conversión Cantabria				
Muertos a 30 días (estimados)	4.604	4.225	3.947	3.176
Diferencia vs. estimación 30 días DGT (%)^a	3,5	2,9	3,1	2,4
Diferencia vs. muertos 24 horas policía (%)^b	16,2	16,3	16,3	16,4
Factores de conversión Castilla la Mancha				
Muertos a 30 días (estimados)	4.883	4.500	4.208	3.400
Diferencia vs. estimación 30 días DGT (%)^a	9,0	8,8	9,1	8,8
Diferencia vs. muertos 24 horas policía (%)^b	21,0	21,4	21,5	21,9
Factores de conversión Galicia				
Muertos a 30 días (estimados)	4.714	4.341	4.058	3.276
Diferencia vs. estimación 30 días DGT (%)^a	5,8	5,5	5,8	5,4
Diferencia vs. muertos 24 horas policía (%)^b	18,2	18,6	18,6	18,9
Media de los factores de conversión				
Muertos a 30 días (estimados)	4.784	4.190	3.917	3.148
Diferencia vs. estimación 30 días DGT (%)^a	7,1	2,1	2,4	1,5
Diferencia vs. muertos 24 horas policía (%)^b	19,4	15,6	15,6	15,6

a. Calculado como [(muertos a 30 días (estimados)-muertos a 30 días (estimación DGT))/ muertos a 30 días (estimados)]*100.

b. Calculado como [(muertos a 30 días (estimados)-muertos a 24h (policía))/ muertos a 30 días (estimados)]*100.

Figura 12. Estimación de muertos a 30 días según diferentes factores de conversión derivados Andalucía, Aragón, Cantabria, Castilla la Mancha, Galicia y media de factores de conversión. España 2005-2008.



Factores de conversión según sexo y edad

En la Tabla 81 se muestran los factores de conversión por sexo y edad para cada comunidad autónoma. Los factores de conversión difieren ligeramente entre comunidades autónomas, de manera que Aragón es donde los factores que se deben aplicar presentan los valores más elevados. Se observan ciertas variaciones en los factores de conversión por sexo, e importantes variaciones por edad. Los factores de conversión muestran que los muertos de tráfico de 75 y más años son los que están más infraestimados. Además, se observa que la variación según grupos de edad en los factores de conversión no es homogénea entre comunidades autónomas. Por ejemplo, para los grupos de edad hasta los 75 años, mientras que en Aragón los factores de conversión sugieren que la infranotificación es mayor para el grupo de edad entre los 18 y 29 años, en Cantabria este grupo de edad se encuentra entre los que presentan una mejor notificación.

Tabla 81. Factores de conversión para el cálculo de muertos de tráfico a 30 días según sexo y edad. Andalucía, Aragón, Cantabria, Castilla la Mancha y Galicia 2005-2007.

	Factores de conversión														
	Galicia			Aragón			Cantabria			Castilla la Mancha			Andalucía		
	N1	N2	N3	N1	N2	N3	N1	N2	N3	N1	N2	N3	N1	N2	N3
Sexo															
Hombres	1,2384	0,0084	0,0003	1,5299	0,0183	0,0051	1,1794	0,0051	0,0000	1,2017	0,0084	0,0010	1,1676	0,0079	0,0005
Mujeres	1,2111	0,0077	0,0003	1,5886	0,0375	0,0000	1,1498	0,0000	0,0000	1,1872	0,0054	0,0002	1,1535	0,0054	0,0005
Edad															
0-13 años	1,1883	0,0063	0,0005	1,4518	0,0263	0,0000	1,1061	0,0000	0,0000	1,1395	0,0143	0,0000	1,1286	0,0065	0,0000
14-17 años	1,2890	0,0084	0,0007	1,5029	0,0469	0,0098	1,1735	0,0000	0,0000	1,2307	0,0122	0,0000	1,1737	0,0000	0,0000
18-29 años	1,2254	0,0063	0,0001	1,5937	0,0230	0,0017	1,1517	0,0057	0,0000	1,1983	0,0078	0,0008	1,1552	0,0072	0,0004
30-44 años	1,2120	0,0058	0,0003	1,5226	0,0068	0,0021	1,1715	0,0000	0,0000	1,1855	0,0017	0,0003	1,1626	0,0025	0,0003
45-64 años	1,1915	0,0105	0,0003	1,4534	0,0208	0,0033	1,1691	0,0104	0,0000	1,1814	0,0048	0,0008	1,1428	0,0038	0,0007
65-74 años	1,2055	0,0211	0,0011	1,4203	0,0267	0,0111	1,1505	0,0000	0,0000	1,1751	0,0112	0,0024	1,1437	0,0311	0,0012
75 años y más	1,4923	0,0180	0,0005	1,7439	0,0503	0,0046	1,3648	0,0000	0,0000	1,3931	0,0330	0,0020	1,3277	0,0180	0,0020

En la Tabla 82 se muestra el número estimado de muertos de tráfico a 30 días a nivel estatal aplicando la media de los factores de conversión de las cinco comunidades autónomas conectadas, por sexo y edad. Se muestra también la diferencia que existe entre dicho número y el número de muertos a 24 horas según datos de la DGT.

Tabla 82. Estimación del número de muertos de tráfico según sexo y edad y diferencia respecto al número de muertos a 24h según datos de la DGT. España 2005-2007.

	Media FC ^a			Muertos de tráfico 24h (DGT)	Muertos de tráfico (estimación FC)	Diferencia (%) ^b
	N1	N2	N3			
Sexo						
Hombres	1,2634	0,0096	0,0014	8.444	11.422	26,1
Mujeres	1,2580	0,0112	0,0002	2.206	2.993	26,3
Edad						
0-13 años	1,2029	0,0107	0,0001	217	282	23,1
14-17 años	1,2740	0,0135	0,0021	356	551	35,4
18-29 años	1,2649	0,0100	0,0006	2.875	3.915	26,6
30-44 años	1,2508	0,0034	0,0006	2.985	3.850	22,5
45-64 años	1,2276	0,0101	0,0010	2.368	3.080	23,1
65-74 años	1,2190	0,0180	0,0032	901	1.202	25,0
75 años y más	1,4644	0,0239	0,0018	886	1.381	35,8

FC: Factores de conversión

a. Media de los factores de conversión de Andalucía, Aragón, Cantabria, Castilla la Mancha y Galicia.

b. Calculado como [(muertos 30d (estimados)- muertos 24h (policía))/muertos 30d (estimados)]*100.

5. DISCUSIÓN

La conexión de registros de bases de datos policiales y hospitalarias es de elevado interés en la planificación y prevención de las lesiones por tráfico. Por una parte, permite ampliar el conocimiento disponible sobre dichas lesiones y los mecanismos que las han originado, pues conecta una base de datos en la que existe amplia información sobre las características de la colisión y el vehículo implicado con otra en la que se dispone de información detallada de las lesiones padecidas por el usuario. Asimismo, al conocer el número de registros conectados y sin conectar, es posible calcular factores de conversión para estimar el número real de lesionados graves y muertos a nivel estatal y regional, corrigiendo la infranotificación existente (SafetyNet, 2008).

En este trabajo, de forma similar a como se realizó previamente en Castilla-León en el marco del estudio de SafetyNet, se ha llevado a cabo una conexión probabilística entre las bases de datos policiales e hospitalarias de diversas comunidades autónomas de España. El objetivo inicial pretendía conectar todas, o la mayoría, de las comunidades autónomas para estimar el número de lesionados graves y muertos a 30 días en España. Sin embargo, únicamente se recibieron las bases de datos hospitalarias de 6 de las 17 comunidades autónomas españolas, con lo cual el estudio se ha llevado a cabo sólo con esta información.

En este informe se han presentado los resultados de la conexión y de su validación para las 6 comunidades autónomas de las cuales se disponen los datos de ingresos hospitalarios, así como el cálculo de factores de conversión para estimar el número de lesionados graves y muertos a 30 días en las 5 comunidades autónomas donde ha resultado viable la conexión. Asimismo se han intentado extrapolar dichas estimaciones a nivel estatal, aunque su validez es muy limitada dada la poca representatividad de los datos y las limitaciones de la propia metodología.

Conexión probabilística

La conexión se pudo realizar en 5 de las 6 comunidades autónomas disponibles. No fue posible conectar los datos hospitalarios y policiales en la Comunidad de Madrid dado que, según el estudio de viabilidad, con las variables disponibles, la conexión no alcanzaría los valores mínimos requeridos de sensibilidad y especificidad.

En las restantes comunidades autónomas se llevó a cabo la conexión probabilística utilizando la fecha de colisión e ingreso, el sexo, la edad, la provincia (excepto en Cantabria) y usuario (en Cantabria y Castilla la Mancha, en su lugar se utilizaron las variables vehículo y posición respecto el vehículo para alcanzar el peso mínimo requerido). En Andalucía, fue necesario utilizar también la localización principal de la lesión, para alcanzar el peso mínimo requerido. En el resto de comunidades autónomas no fue necesario utilizar esta variable como variable de conexión, reservando la información en ella contenida para la revisión semi-manual de registros.

Cabe destacar el elevado porcentaje de valores perdidos para algunas de las variables de conexión. Las variables individuales (sexo y edad) fueron considerablemente exhaustivas, en especial en la base de datos hospitalaria. En la base de datos policial destaca un 10,6% de valores perdidos para la variable edad en Madrid, y un 6,7% en Aragón. No se observó ningún valor perdido en las variables provincia y fecha de colisión e ingreso. Sin embargo, el porcentaje de valores perdidos para la variable usuario fue considerablemente elevado en los datos hospitalarios, oscilando entre un 35,5% de valores perdidos en Cantabria hasta un 61,4% en Castilla la Mancha. En Galicia, a pesar de que el código E de causa externa es de obligada cumplimentación por ley esta variable presentó un 48,1% de valores perdidos. Esto se debe a que con frecuencia el cuarto dígito (que identifica el tipo de usuario) o bien no se rellena o bien es inespecífico (valores 8 y 9 del cuarto dígito del código E). Finalmente, la localización de la lesión principal también presentó un elevado porcentaje de valores perdidos en los datos policiales, desde un 26,7% en Andalucía hasta un 88,6% en la Comunidad de Madrid.

La calidad del proceso de conexión depende del poder discriminatorio de las variables de conexión (Newcombe, 1998). La proporción de valores perdidos de las variables afecta la viabilidad de la conexión y su calidad, pues afecta los pesos asignados a los registros y disminuye el poder discriminatorio de las variables de conexión, de manera que aumenta la probabilidad de que 2 o más registros presenten los mismos valores en las variables de conexión. Es probable que la Comunidad de Madrid no fuera viable debido a que presenta un elevado porcentaje de valores perdidos en sus variables de conexión y a que el número de variables de conexión es menor, pues no presenta distintas provincias. Esto último también ocurre con Cantabria, pero la conexión en esta comunidad autónoma sí que es viable probablemente debido a que el pequeño porcentaje de valores perdidos de sus variables compensa el menor número de variables de conexión disponibles.

La conexión probabilística asignó un registro policial a un 15,6% de los registros hospitalarios en Andalucía, a un 36,2% en Aragón, a un 21,9% en Cantabria y a un 27,4% en Castilla la Mancha y a un 31,0% en Galicia. De los registros conectados, aquellos considerados conectados mediante la revisión semi-manual y posteriores criterios supusieron entre un 11,5% y un 37,4%. Cabe destacar que el porcentaje de registros conectados aplicando un punto de corte visual fue parecido al obtenido a partir de los distintos criterios que se aplicaron, lo cual da consistencia a los resultados de la conexión, aunque no se puede asegurar que los casos conectados sean exactamente los mismos. Asimismo, tal y como se observa en las figuras que muestran los pesos de cada registro en función del número de variables de conexión concordantes, mientras que todos los registros en los que coinciden todas las variables tienen un peso superior al punto de corte visual, aquellos en los que al menos una variable no coincide, algunos tienen un peso que se sitúa por debajo del punto de corte. Por lo tanto, si los criterios de conexión hubiesen sido más permisivos en cuanto al número requerido de variables de conexión coincidentes se habría producido un número considerable de falsos positivos.

Al comparar las características de los registros conectados con las de los no conectados de la base de datos policial se observa que tienen una mayor probabilidad de haberse conectado los hombres, los grupos de edad más jóvenes y los más mayores, los individuos cuya principal lesión se sitúa en la cabeza, en la cara, en el abdomen o en las

extremidades inferiores, los usuarios de vehículo a motor de dos ruedas y los peatones, los lesionados en carretera, en una colisión ocurrida de día y aquellas colisiones con una sola víctima. Por el contrario, tienen una menor probabilidad de haberse conectado los individuos lesionados en el cuello y los que circulan como pasajeros. Estas características reflejan principalmente aquellas características asociadas a una mayor gravedad de la persona lesionada y, por lo tanto, una mayor probabilidad de haber implicado un ingreso hospitalario.

En cuanto a las características asociadas a que un registro hospitalario se haya conectado, los grupos de edad más jóvenes y aquellos más mayores tienen una menor probabilidad de conectarse, así como los individuos con una estancia hospitalaria más corta y los usuarios de bicicleta y los peatones. Por el contrario, los usuarios de vehículos a motor de 2 ruedas tienen una mayor probabilidad de conectarse. Lujic y colaboradores identificaron las siguientes características asociadas a una mayor probabilidad de conectarse: conductores de coche, casos con derecho a una compensación económica – que para recibirla deben de haber notificado la colisión a la policía - y lesionados con un diagnóstico de lesión en el diagnóstico principal (Lujic, 2008). Ferrante y colaboradores observaron que los lesionados de vehículo único, en especial de motocicleta, tienen menor probabilidad de conectarse con un registro policial (Ferrante, 1993). Las variables comentadas reflejan las características de los individuos que tienen una mayor probabilidad de haber sido incorporados en la base de datos policial (Rosman, 1994). En Francia, Amoros y colaboradores identificaron las siguientes variables como relacionadas con las subnotificación por parte de la policía: ser un lesionado leve, ser usuario de motocicleta o bicicleta, colisiones de vehículo único, carretera secundaria, zona rural (Amoros, 2006).

En un estudio en Australia se mostró que la probabilidad de conectarse fue superior en mujeres, aumentó con la edad, varió según la raza, fue menor en colisiones de vehículo único, fue mayor en conductores que pasajeros, menor en motocicletas en comparación con cualquier otro vehículo y aumentó con la gravedad de la lesión (Rosman, 1994). Un estudio posterior realizado en la misma región con el registro de traumatología, que recoge los lesionados ingresados más graves, confirmó estos hallazgos (Lopez, 2000).

Cabe destacar que, entre los registros hospitalarios conectados con un registro policial, en Andalucía el 36,0% correspondió a un lesionado leve según la policía, el 34,8% en Aragón, el 45,9% en Cantabria, el 33,1% en Castilla la Mancha y el 30,4% en Galicia. Este dato refleja la mala clasificación de la gravedad en la base de datos policial, pues estos casos se deberían de haber clasificado como lesionados graves.

Por el contrario, si analizamos los registros policiales, en Andalucía el 14,9% de los no conectados corresponden a lesionados graves, el 16,8% en Aragón, el 10,3% en Cantabria, el 17,2% en Castilla la Mancha y el 21,7% en Galicia. Estos casos, por lo tanto, se deberían de haber conectado. Es de esperar que la mayoría de estos registros no se hayan conectado debido a la falta de suficientes variables discriminatorias para identificar de forma eficaz todos los pares de registros reales. Sin embargo, un porcentaje de éstos, pueden haber sido ingresados en otra provincia, incluso en otra comunidad autónoma, o que finalmente no hayan sido ingresados a pesar de que así conste en la base de datos policial.

Validación de la conexión

Son pocos los países que validan la calidad de la conexión entre bases de datos (Tabla 1). La mayoría asumen que las bases de datos policiales y hospitalarias han sido conectadas correctamente.

Siguiendo la metodología propuesta por Cook y colaboradores, la sensibilidad de la conexión fue del 98,3% en Andalucía, del 99,9% en Aragón y Castilla la Mancha y del 100% en Cantabria y Galicia. La especificidad osciló entre un 89,7% en Galicia hasta un 94,7% en Andalucía. La especificidad de la conexión calculada según la metodología utilizada en Austria fue similar (SafetyNet, 2008).

A pesar de estos resultados, se deben interpretar con precaución, en especial en cuanto a los valores de sensibilidad. El método utilizado para validar la conexión consiste en seleccionar un 68% de registros de la base de datos hospitalaria e incluirlos en otra base de datos a la que además se añaden registros de otros años (Cook, 2001). Por lo tanto, en primer lugar, la metodología utilizada evalúa la conexión entre dos bases de datos

hospitalarias, en lugar de una hospitalaria y una policial. Además, en este tipo de validación, que se basa en conectar registros duplicados, las variables de conexión son exactamente iguales, lo cual en la realidad no siempre es así, pues las variables pueden verse afectadas por falta de información (valores perdidos) o por errores en el contenido originados al recoger la información o al entrarla en la base de datos (Ferrante, 1993) (Rosman, 1994). La metodología de validación no tiene en cuenta estos factores, lo cual es consistente con los elevados valores de sensibilidad obtenidos. Al coincidir la distribución de las variables de conexión de los registros duplicados - precisamente el criterio principal según el cual se considera concordante un par de registros - todos los registros duplicados son detectados en la validación. En otras palabras, lo que este tipo de validación está evaluando es la sensibilidad y especificidad de un proceso de conexión basado en unas variables de conexión concretas y para registros idénticos en dos bases de datos. Sin embargo, este tipo de validación no permite conocer la capacidad de detectar pares de registros entre las bases de datos policiales y hospitalarias. Rosman destaca que sin la disponibilidad de nombres en las bases de datos no es posible determinar la sensibilidad y especificidad del proceso de conexión, mientras que los nombres permiten identificar los falsos positivos y negativos mediante la revisión manual (Rosman, 1996). También se ha referido que no es posible validar el proceso de conexión correctamente si no se dispone de una herramienta que permita conocer con certeza los registros correctos (SafetyNet, 2008).

A pesar de no poder conocer con exactitud la sensibilidad de la conexión presentada en este trabajo, en un estudio previo realizado por Rosman se demuestra la importancia de disponer de datos con nombres para realizar la conexión probabilística. La autora comprobó que la conexión probabilística sin nombres entre datos policiales y hospitalarios – utilizando sexo, edad, fechas de colisión e ingreso y usuario - conectaba sólo el 52% de los pares de registros verdaderos en comparación con la conexión utilizando los nombres de los individuos (Rosman, 1996). Según el estudio de Rosman la sensibilidad de la conexión que se ha presentado en este trabajo se situaría alrededor del 50%, lo cual es consistente con el porcentaje de registros conectados (rango: 15,6%-36,2%), que es aproximadamente el 50% del porcentaje esperado (68%). Estos resultados sugieren que la información contenida en las variables de conexión es insuficiente para derivar estimadores fiables. En Austria, con unas variables de

conexión similares a las presentadas - fecha de colisión e ingreso, edad, sexo, nacionalidad, estado federal del hospital y del lugar de colisión -, los autores concluyeron que la información disponible en estas variables no fue suficiente para obtener resultados fiables y válidos (SafetyNet, 2008).

La validez de los resultados presentados también puede haberse visto afectada por la falta de representatividad de las lesiones de tráfico en zona urbana en la base de datos de policía, así como por la infranotificación de los lesionados leves, tanto en zona urbana como en carretera. Las elevadas tasas de letalidad, observadas tanto en zona urbana como en carretera, sugieren que el número de lesionados leves que se están notificando es considerablemente menor de lo que cabría esperar. Además, el porcentaje de lesionados graves es muy superior en carretera que en zona urbana, lo cual sugiere que las lesiones de tráfico se notifican en mayor medida en este tipo de vías que en zona urbana. Estos fenómenos afectan de forma desigual a unas y otras comunidades, de manera que Cantabria es la comunidad que presenta una peor representatividad de los lesionados leves y en zona urbana.

Estimación de lesionados graves y muertos

A pesar de las limitaciones metodológicas de la conexión se han calculado los factores de conversión y se ha estimado el número de heridos graves y de muertos a 30 días para cada comunidad autónoma estudiada, y para el conjunto del Estado. Sin embargo, dado que el porcentaje de registros conectados difiere considerablemente del porcentaje esperado, estos números deberían interpretarse con cautela. Es de esperar que los factores de conversión obtenidos sean superiores a los reales, pues deberían haberse conectado más registros policiales y hospitalarios.

Los resultados muestran una notable diferencia entre el número de lesionados graves notificados según la policía y el estimado según la metodología propuesta por SafetyNet (SafetyNet, 2008). Sin embargo, cabe destacar que, del mismo modo que se aplica un factor de conversión a los lesionados leves según policía – pues un número de lesionados leves se sabe que han sido ingresados según los resultados de la conexión -,

también debería aplicarse un factor de conversión a los lesionados graves según policía, restándolo de la fórmula para corregir aquellos individuos notificados como lesionados graves que no han sido ingresados (no se ha hallado un registro coincidente en el CMBD). Por lo tanto, y asumiendo que la conexión se ha producido de forma correcta, el número estimado de lesionados graves debería de ser menor que el obtenido según la metodología utilizada.

En cuanto al número estimado de muertos a 30 días, el número estimado según la metodología del SafetyNet es bastante superior al notificado por la policía a 24 horas, aunque la diferencia es menor cuando se compara con las estimaciones a 30 días realizadas por la DGT. Durante el proceso de conexión se eliminaron de la base de datos policial los muertos a 24 horas, pues se espera que la mayoría no lleguen a ingresar en el hospital y mueran o bien en el acto o durante el traslado o en urgencias. De este modo se reduce la probabilidad de que un registro del CMBD se conecte con un registro de la policía que con alta probabilidad no se encuentra en la base de datos hospitalaria. Sin embargo, como contrapartida, no es posible saber si entre los muertos a 30 días del CMBD existe alguno de los muertos a 24 horas según policía. Esto podría haber afectado las estimaciones y los factores de conversión, aunque es de esperar que el efecto sea prácticamente nulo.

Se han aplicado los factores de conversión para estimar el número de lesionados graves y muertos a 30 días en España. Sin embargo, sólo se disponen de cinco comunidades autónomas, lo cual pone en duda que dichos factores sean generalizables a nivel estatal. De hecho, se observa una variación no despreciable en las estimaciones obtenidas al aplicar los factores de conversión de una u otra comunidad autónoma. Además, se desconocen los factores de conversión que se observarían en otras comunidades autónomas, que podrían variar considerablemente de los obtenidos para las comunidades analizadas en este trabajo.

Dada la mayor infranotificación en zona urbana, hubiera sido deseable calcular distintos factores de corrección para zona urbana y para carretera. Sin embargo, dado que la base de datos hospitalaria no dispone de información sobre el tipo de carretera en que se produjo la colisión de tráfico, el cálculo de estos factores de corrección no fue posible.

También cabe remarcar que los factores de conversión calculados pueden cambiar con el tiempo debido a diversos factores, como variaciones en la calidad de la notificación de los lesionados por parte de la policía. Por lo tanto, es necesario calcularlos de forma anual.

Limitaciones y fortalezas

El cálculo de las estimaciones del número de lesionados graves y muertos se basa en la calidad de la conexión de las bases de datos policial y hospitalaria. Una conexión deficiente conlleva la obtención de factores de conversión incorrectos. Varios factores podrían afectar los resultados de la conexión y, por lo tanto, las estimaciones del número de lesionados graves y muertos.

En primer lugar, tal y como se ha comentado previamente, el pequeño número de variables de conexión disponibles y su bajo poder discriminatorio limita considerablemente la capacidad de la conexión de identificar los pares de registros reales.

Por otra parte, la conexión se basó en la asunción de que todas las colisiones ocurridas en una provincia son atendidas en un hospital de la misma provincia. Sin embargo, es posible que los lesionados sean atendidos en otra provincia, incluso en otra comunidad autónoma, de manera que a estos registros policiales no se les dará la opción de conectarse con su registro hospitalario. Sin embargo, es de esperar que la mayoría de lesionados graves sean atendidos en un hospital próximo que, por lo tanto, probablemente estará situado en la misma provincia en la que tuvo lugar la colisión.

Asimismo los registros contenidos en la base de datos hospitalaria pertenecen principalmente a hospitales públicos. Aunque es posible que algún lesionado ingrese en un hospital privado no incluido en la base de datos, es de esperar que la mayoría de los lesionados graves de tráfico sea atendido en un hospital público.

Es posible que el número de ingresos por lesión por tráfico seleccionados para ser conectados con la base de datos policial puede haberse visto afectado por la calidad del

código E de causa externa y el código de financiación. De forma similar, es posible, aunque poco probable, que se haya incluido alguna lesión no ocasionada por tráfico. También puede haber casos duplicados que correspondan a pacientes trasladados a otro hospital.

Finalmente, algunos lesionados de tráfico puede que no se encuentren en la base de datos policial, en especial aquellas colisiones en las que no existe implicación de un vehículo a motor (ej. usuarios de bicicletas lesionados en la vía pública por una colisión con un peatón o sin implicación de otro usuario).

La principal fortaleza de este trabajo consiste en la metodología utilizada. La conexión probabilística es más eficaz y eficiente que la conexión manual, pues es mucho menos laboriosa y no está sujeta al error humano. Asimismo, la conexión probabilística mediante el cálculo de pesos obtiene similares resultados en comparación con la conexión probabilística mediante el cálculo de distancias, tanto para variables categóricas (Domingo-Ferrer, 2002) como para variables numéricas (Domingo-Ferrer, 2001). Además, el tipo de conexión probabilística utilizada asigna mayores pesos a los valores poco frecuentes. En comparación con la conexión determinística, varios autores sugieren que la conexión probabilística obtiene un mayor número de registros conectados correctos, en especial cuando no se disponen de identificadores únicos (Gill, 1993) (Grannis, 2004). Se ha estimado que cuando se disponen de pocas variables de conexión o cuando éstas son poco discriminatorias la conexión probabilística puede llegar a obtener hasta 3 veces el número de registros conectados que la conexión determinística (Chini, 2008). Así pues, la conexión probabilística tiene una mayor sensibilidad, aunque una menor especificidad que la determinística (Gomatam, 2002) (Clark, 1995).

6. CONCLUSIONES Y RECOMENDACIONES

En este trabajo se ha llevado a cabo la conexión probabilística entre las bases de datos policial y hospitalaria para las comunidades autónomas que han proporcionado datos y en las que la valoración de la conexión se ha mostrado viable: Andalucía, Aragón, Cantabria, Castilla la Mancha y Galicia. A pesar de que la Comunidad de Madrid ha proporcionado datos hospitalarios, en este caso no se han podido conectar ambos registros ya la exploración de los datos ha mostrado que la conexión no era viable.

Se han conectado entre un 15,6% y un 36,2% de los registros hospitalarios con un margen de confianza elevado por el cual se asume que los pares conectados corresponden realmente al mismo individuo. Los resultados de la conexión ponen en entredicho la validez de las estimaciones. Ello es debido probablemente al escaso número de variables de conexión disponibles y a su limitado poder discriminatorio.

Se han calculado las estimaciones del número de lesionados graves y muertos a partir del resultado de cada conexión en dichas comunidades autónomas y a nivel estatal. A pesar de las importantes limitaciones metodológicas de la conexión derivadas de la falta de suficientes variables discriminatorias, se han obtenido resultados plausibles tanto para estimar el número de heridos graves como de fallecidos a 30 días por accidente de tráfico.

Recomendaciones

1. Las conexiones realizadas muestran la importante limitación de no disponer de identificadores personales. Con el fin de poder realizar la conexión de forma válida y fiable sería necesario disponer de datos individuales. El acceso a datos individuales confidenciales está limitado debido a la legislación vigente. Sin embargo, algunos autores señalan que las nuevas tecnologías informáticas permiten proveer el uso de datos individuales necesarios asegurando la protección de datos y la confidencialidad de los pacientes (CODES, 1996). Se ha sugerido la encriptación de datos y la incorporación de sistemas de seguridad basados en identificación del usuario y contraseña para su acceso como método para que los científicos puedan disponer de datos individuales (Quantin, 2000). Idealmente, sería necesario disponer de nombres y apellidos, como ya se dispone en varios países. Si esto no es posible, al menos sería conveniente disponer de algún dato individual, entre los cuales se sugiere el código postal de residencia del individuo. Esta información, de hecho, ya está disponible en la base de datos hospitalaria, pero no lo está en la policial. Asimismo, sería también importante disponer de la fecha de nacimiento en la base de datos policial en lugar de la edad del individuo, pues aumentaría considerablemente el poder discriminatorio de dicha variable.
2. Si no es posible disponer de datos personales de forma rutinaria, al menos sería conveniente poder llevar a cabo una validación de la conexión con nombres y apellidos de manera que permita valorar con un mayor grado de fiabilidad el alcance de las conexiones llevadas a cabo.
3. También es de suma importancia mejorar la calidad del código E de causa externa de la base de datos hospitalaria, pues no está recogida en un elevado número de casos, y en los casos en los que sí se recoge, con frecuencia la información que en ella se incluye es inespecífica.

4. Dadas las diferencias entre las características de los registros conectados y los no conectados, los factores de conversión se deberían calcular por separado según estas características. Dado que la subnotificación de casos no es homogénea según las características del individuo y la colisión, no es correcto calcular un único factor corrector, sino que se deben calcular varios en función de dichas características.
5. Las estimaciones de los factores de conversión para conocer el número de heridos graves y fallecidos a 30 días no pueden basarse sólo en los datos de cinco comunidades autónomas. Sería necesario disponer de datos hospitalarios de más comunidades autónomas, de manera que se cubra una mayor representatividad del territorio Español. Cabría también disponer de una mayor representatividad de las lesiones de tráfico en zona urbana, lo que implica mejorar la notificación en estas vías.
6. La conexión de datos hospitalarios con datos policiales permite además poder estudiar las características de las lesiones y su gravedad en función de las características de la colisión. Las conexiones realizadas tienen por tanto un gran potencial para estudiar los factores asociados según perfiles de usuarios y colisiones.

7. REFERENCIAS

Alsop J, Langley J. Under-reporting of motor vehicle traffic crash victims in New Zealand. *Accid Anal Prev.* 2001 May;33(3):353-9.

Amoros E, Martin JL, Laumon B. Estimating non-fatal road casualties in a large French county, using the capture-recapture method. *Accid Anal Prev.* 2007 May;39(3):483-90.

Amoros E, Martin JL, Laumon B. Under-reporting of road crash casualties in France. *Accid Anal Prev.* 2006 Jul;38(4):627-35.

Aptel I, Salmi LR, Masson F, Bourdé A, Henrion G, Erny P. Road accident statistics: discrepancies between police and hospital data in a French island. *Accid Anal Prev.* 1999 Jan-Mar;31(1-2):101-8.

Broughton J, Amoros E, Bos N, Evgenikos E, Hoeglinger S, Hollo P, Perez C and J. Tecl (2008). Estimation of the actual number of road casualties in Europe from the CARE database. SafetyNet project, Task 1.5.

Chini F, Fortini M, Tuoto R, Farchi S, Giorgi Rossi P, Amato R, Borgia P. Probabilistic record linkage for the integrated surveillance of road traffic injuries when personal identifiers are lacking. Roma: instituto Nazionale di Statistica; 2008.

Cirera E, Pérez C, Grupo de trabajo de la Sociedad Española de Epidemiología sobre la medida del impacto en salud de los accidentes de tráfico en España. Estudio de viabilidad de la conexión entre los registros de la DGT y el CMBDAH. Madrid: Ministerio de Sanidad y Consumo; 2006.

Clark DE, Hahn DR. Comparison of probabilistic and deterministic record linkage in the development of a statewide trauma registry. *Proc Annu Symp Comput Appl Med Care.* 1995:397-401.

Cook LJ, Olson LM, Dean JM. Probabilistic record linkage: relationships between file sizes, identifiers and match weights. *Methods Inf Med.* 2001 Jul;40(3):196-203.

Cryer PC, Westrup S, Cook AC, Ashwell V, Bridger P, Clarke C. Investigation of bias after data linkage of hospital admissions data to police road traffic crash reports. *Inj Prev.* 2001 Sep;7(3):234-41.

Derriks HM, Mak PM. Underreporting of road traffic casualties. International Traffic Safety Data and Analysis Group; 2007.

Dirección General de Tráfico. Seguimiento de los heridos graves a 30 días. Madrid: Ministerio del Interior, Dirección General de Tráfico: 2000.

Domingo-Ferrer J, Torra V. Disclosure control methods and information loss for microdata. En: Doyle P, Lane JJ, Theeuwes JJM, Zayatz LM, editors. Confidentiality, disclosure, and data access: theory and practical applications for statistical agencies. Elsevier; 2001.

Domingo-Ferrer J, Torra V. Distance-based and probabilistic record linkage for re-identification of records with categorical variables. En: 5e Congrés Català d'Intel·ligència Artificial. Butlletí de l'ACIA, Vol. 28. Associació Catalana d'Intel·ligència Artificial; 2002.

Fellegi IP, Sunter AB. A theory of record linkage. *J Am Stat Assoc* 1969;64:1183–210.

Ferrante AM, Rosman DL, Knuiman MW. The construction of a road injury database. *Accid Anal Prev*. 1993 Dec;25(6):659-65.

Gill L, Goldacre M, Simmons H, Bettley G, Griffith M. Computerised linking of medical records: methodological guidelines. *J Epidemiol Community Health*. 1993 Aug;47(4):316-9.

Gomatam S, Carter R, Ariet M, Mitchell G. An empirical comparison of record linkage procedures. *Stat Med*. 2002 May 30;21(10):1485-96.

Gonzalez RP, Cummings G, Mulekar M, Rodning CB. Increased mortality in rural vehicular trauma: identifying contributing factors through data linkage. *J Trauma*. 2006 Aug;61(2):404-9.

Grannis SJ, Overhage JM, McDonald C. Real world performance of approximate string comparators for use in patient matching. *Stud Health Technol Inform*. 2004;107(Pt 1):43-7.

Jarvis SN, Lowe PJ, Avery A, Levene S, Cormack RM. Children are not goldfish--mark/recapture techniques and their application to injury data. *Inj Prev*. 2000 Mar;6(1):46-50.

Johnson SW, Walker J. The crash outcome data evaluation system (CODES). Report No. DOT HS 808 338. Washington: National Highway Traffic Safety Administration; 1996.

Karlson TA, Quade C, Florey M. Nonfatal motor vehicle crash injuries: Wisconsin's experience with linked data systems. *Wis Med J*. 1996 May;95(5):301-4.

Keigan M, Broughton J, Tunbridge RJ. Linkage of STATS19 and Scottish hospital in-patient data — analyses for 1980–1995. TRL Report 420. Crowthorne: Transport Research Laboratory; 1999.

Lai CH, Huang WS, Chang KK, Jeng MC, Doong JL. Using data linkage to generate 30-day crash-fatality adjustment factors for Taiwan. *Accid Anal Prev*. 2006 Jul;38(4):696-702.

Lopez DG, Rosman DL, Jelinek GA, Wilkes GJ, Sprivulis PC. Complementing police road-crash records with trauma registry data--an initial evaluation. *Accid Anal Prev*. 2000 Nov;32(6):771-7.

Morrison A, Stone DH. Capture-recapture: a useful methodological tool for counting traffic related injuries? *Inj Prev*. 2000 Dec;6(4):299-304.

National Association of Governor's Highway Safety Representatives. So you want to link your state data. Washington: National Highway Traffic Safety Administration; 1996.

National Highway Traffic Safety Administration. The crash outcome data evaluation system (CODES). Springfield: NHTSA; 1996.

Newcombe HB. Handbook of Record Linkage: methods for health and statistical studies, administration and business. New York City: Oxford University Press 1988.

NHTSA (2009). The Crash Outcome Data Evaluation System (CODES) and applications to improve traffic safety decision making; <http://www-nrd.nhtsa.dot.gov/Pubs/811181.PDF>

Organisation for Economic Co-operation and Development. IRTAD Special Report: The availability of hospitalised road user data in OCDE member countries. OECD; 2003.

Pérez C, Cirera E, Plasència A, Grupo de trabajo de la Sociedad Española de Epidemiología sobre la medida del impacto en salud de los accidentes de tráfico en España. Estudio de la mortalidad a 30 días por accidentes de tráfico (EMAT-30). Madrid: Ministerio de Sanidad y Consumo; 2004.

Perez K, Cirera E, Borrell C, Plasencia A, on behalf of the work group of the Spanish Society of Epidemiology for the Measuring of the Impact on Health of Road Traffic Accidents in Spain. Motor vehicle crash fatalities at 30 days in Spain. *Gac Sanit*. 2006;20(2):108-15.

Petridou ET, Yannis G, Terzidis A, Dessypris N, Germeni E, Evgenikos P, Tselenti N, Chaziris A, Skalkidis I. Linking emergency medical department and road traffic police casualty data: a tool in assessing the burden of injuries in less resourced countries. *Traffic Inj Prev*. 2009 Mar;10(1):37-43.

Quantin C, Allaert F, Dusserre L. Anonymous statistical methods versus cryptographic methods in epidemiology. *Int J Med Inform*. 2000 Nov 1;60(2):177-183.

Rosman DL, Knuiiman MW. A comparison of hospital and police road injury data. *Accid Anal Prev*. 1994 Apr;26(2):215-22. Erratum in: *Accid Anal Prev* 1995 Feb;27(1):141.

Rosman DL. The feasibility of linking hospital and police road crash casualty records without names. *Accid Anal Prev*. 1996 Mar;28(2):271-4.

SafetyNet. Building the European Road Safety Observatory. Deliverable 1.15. Final report on task 1.5. SafetyNet; 2008.

Simpson HF. Comparison of hospital and police casualty data: a national study. TRL Report 173. Crowthorne: Transport Research Laboratory; 1996.

Tercero F, Andersson R. Measuring transport injuries in a developing country: an application of the capture-recapture method. *Accid Anal Prev.* 2004 Jan;36(1):13-20.

Van HT, Singhasivanon P, Kaewkungwal J, Suriyawongpaisal P, Khai LH. Estimation of non-fatal road traffic injuries in Thai Nguyen, Vietnam using capture-recapture method. *Southeast Asian J Trop Med Public Health.* 2006 Mar;37(2):405-11.

Zhu VJ, Overhage MJ, Egg J, Downs SM, Grannis SJ. An empiric modification to the probabilistic record linkage algorithm using frequency-based weight scaling. *J Am Med Inform Assoc.* 2009 Sep-Oct;16(5):738-45.