

Acción Matemática contra el Coronavirus. Protocolo de Actuación



COMITÉ ESPAÑOL DE MATEMÁTICAS

Investigador principal: Cristina Fernández Pérez (Jefa del Servicio de Medicina Preventiva y Salud Pública del Hospital Clínico San Carlos)

Investigadores colaboradores:

Alfonso Gordaliza Ramos (Presidente del CEMat) Grupo PEM (UVa UNICAN)

Ricardo Cao Abad, ITAMTI, Presidente del comité de expertos de la iniciativa "Acción Matemática contra el Coronavirus" del Comité Español de Matemáticas (CEMat)

Begoña Vitoriano (UCM, IMI Data Science Club y programa DEC-HUMLOG). Miembro del Comité de Expertos CEMat

María del Carmen Pardo (UCM)

Emilio Carrizosa (US, Presidente del IMUS)

Guadalupe Gómez (UPC y UVic)

Alejandra Cabaña (UAB)

Fermín Mallor (UPNA). Miembro del Comité de Expertos CEMat

Carmen Armero (UV)

INDICE

1.	INTRODUCCIÓN Y JUSTIFICACIÓN	3
2.	HIPÓTESIS Y OBJETIVOS	5
3.	METODOLOGÍA	6
4.	METODOLOGÍA DEL ANÁLISIS Y TRATAMIENTO DE LA INFORMACIÓN...7	
5.	CONFIDENCIALIDAD Y PROTECCIÓN DE DATOS.....	8
6.	INFORMACIÓN ADICIONAL	10

1. INTRODUCCIÓN Y JUSTIFICACIÓN

La infección por Covid19 se ha convertido en la crisis sanitaria más grave de la historia reciente. El Comité Español de Matemáticas, CEMat, está promoviendo la iniciativa *Acción Matemática contra el Coronavirus* para poner a disposición de las autoridades nuestra capacidad de análisis y modelización por si fuera útil para analizar, comprender y actuar ante la emergencia que estamos sufriendo con la pandemia COVID-19.

Tras escuchar a la comunidad investigadora española que aglutina CEMat, el comité de expertos matemáticos/estadísticos creado para esta iniciativa ha acordado trasladar a las autoridades competentes las siguientes recomendaciones.

1. Se considera importante disponer de forma directa, de las autoridades competentes, al más alto nivel, la información básica y personas de contacto en relación con los problemas más urgentes e importantes ocasionados por la epidemia, cuya complejidad aconseje un abordaje con ayuda del análisis y la modelización matemática y estadística.
2. Se considera conveniente hacer llegar a las autoridades competentes las propuestas concretas de la comunidad matemática sobre problemas, aparentemente relevantes, en los que tenemos capacidad y experiencia científica. Tras un análisis muy preliminar, se considera que la tipología de problemas que se podrían abordar comprende:
 - 2.1. La evolución y propagación de la epidemia, a nivel global y por CCAA, e incluso a nivel local en el caso de las grandes ciudades, tanto a nivel de casos confirmados, hospitalizados y en UCI, como en la proporción poblacional de portadores del virus o de individuos con presencia de anticuerpos.
 - 2.2. El efecto que tendrían en el desarrollo de la epidemia los cambios en las medidas de confinamiento y de distanciamiento social.
 - 2.3. La predicción anticipada del número de contagios, fallecidos, ingresados en hospitales, pacientes en UCI y proporción poblacional de portadores, con un horizonte de predicción de unos pocos días o incluso de semanas, a nivel de todo el Estado, de CCAA, de provincia e incluso para municipios muy poblados o con especial relevancia.
 - 2.4. La predicción de la evolución de cada paciente, a partir de la información relevante que se conozca del mismo, al objeto de poder anticiparse a su necesidad de hospitalización o ingreso en UCI, entre otros aspectos.
 - 2.5. La estimación de los tiempos relevantes en la evolución de los pacientes a partir de datos desagregados: tiempos de incubación, tiempo desde la infección hasta el cese de la enfermedad (o hasta el fallecimiento), tiempo durante el cual se puede contagiar el virus después de haberse dado de alta al paciente. Estos parámetros son fundamentales para nutrir los modelos necesarios para los puntos 3.1 a 3.3.
 - 2.6. El reparto de bienes escasos en esta situación de emergencia (mascarillas, EPIs, tests de detección, personal, etc.) entre los agentes implicados (CCAA, ciudades, hospitales, sectores de la población, etc.), persiguiendo diversos objetivos (reducir los fallecimientos, atajar la dispersión de la epidemia, limitar las diferencias de trato hacia los agentes, garantizar la rotación del personal, etc.). Esto incluye la optimización, en general, de los recursos

- materiales y humanos disponibles, tanto de tipo sanitario como, en su caso, de otros bienes y servicios de primera necesidad.
- 2.7. La colaboración con otros investigadores (biotecnólogos, virólogos, etc.) en la formulación y el desarrollo de modelos matemáticos y estadísticos para el descubrimiento vacunas y fármacos efectivos contra la enfermedad.
 - 2.8. El estudio in-silico de medidas de control de la enfermedad COVID19 no farmacéuticas que puedan colaborar al diseño de políticas de salud pública.
3. Tras la consulta a la comunidad matemática, se constata que nuestros investigadores necesitan disponer de los datos e información que se citan a continuación:
- 3.1. Series de datos diarios de la epidemia, completas y desagregadas por CCAA y por rangos de edad y sexo. Deberían contener el número de casos confirmados, de fallecidos, de pacientes curados, de enfermos hospitalizados, el número de ellos en la UCI, etc. Para analizar también el posible efecto de variables meteorológicas (temperatura, humedad, radiación solar, etc.) en la evolución de la epidemia, será necesario disponer de las curvas diarias de dichas variables facilitadas desde AEMET.
 - 3.2. Microdatos de los casos diagnosticados, incluyendo información de interés para cada paciente: tiempos relevantes (fechas de contacto con pacientes infectados, de primeros síntomas, de hospitalización, de alta, de fallecimiento), covariables relevantes (edad, sexo, ubicación geográfica de domicilio y lugar de trabajo, patologías previas, existencia de contacto previo con casos confirmados o con casos sospechosos, curvas diarias de temperatura, presión arterial del paciente, etc.).
 - 3.3. Microdatos de encuestas que se puedan estar haciendo actualmente por parte del Ministerio de Sanidad, del INE o de las CCAA en relación con la pandemia.
 - 3.4. Datos sobre recursos disponibles en el sistema de salud. Especialmente, de equipos o infraestructuras que se consideren críticos y que puedan resultar escasos. Datos de uso de los mismos, tanto en los días anteriores al inicio de la epidemia como en la actualidad y en las últimas semanas. Sería necesario conocer los problemas principales de reparto de estos bienes escasos a los que se enfrentan los mandos de gestión de la crisis y qué objetivos priorizan en sus decisiones.
 - 3.5. Registros de movilidad de la población (ficheros Big Data), incluyendo también los que está recabando el INE con la colaboración de algunas compañías de telefonía móvil.

En este contexto, contactadas ya las autoridades sanitarias del Ministerio de Sanidad, se constata las dificultades que tienen para proporcionar la información, y muy en particular, de los microdatos de los casos, así como la gestión descentralizada que lleva a que esos microdatos se encuentren solo en la fuente.

Por otra parte, con la intención de colaborar en la gestión directa de los recursos humanos y materiales, se ve necesario disponer de la información de los centros que proporcionan la asistencia sanitaria directamente. Experiencias existentes en algunas comunidades autónomas muestran que la colaboración en la gestión directa de los centros supone un avance en la gestión de los recursos permitiendo anticiparse a situaciones que podrían ser previstas.

Pero los modelos matemáticos que se pueden desarrollar para que logren su fin de ser útiles para la toma de decisiones tienen que estar basados y validados en información real y precisa.

2. HIPÓTESIS Y OBJETIVOS

El objetivo general del proyecto es colaborar con las autoridades sanitarias encargadas de gestionar esta emergencia en todos los niveles en que se considere de interés por éstas desde el entorno matemático.

Los objetivos específicos son los que se plantean en el punto 2 de la introducción, si bien algunos de ellos se encuentran ya iniciados y se pueden concretar en:

1. Proveer de una mejor predicción de la evolución y propagación de la epidemia a nivel nacional y por CCAA, mediante la construcción de un meta-predicador que facilite a las autoridades información del comportamiento a corto plazo de variables de gran interés en la expansión del virus: no es ajeno a la comunidad científica que los modelos no son únicos y que en España existen diversos grupos desarrollando modelos de evolución y predicción de la propagación de la epidemia. El objetivo del meta-predicador es aglutinar los resultados de los grupos y proveer de una estimación junto con un margen de error a partir de estas predicciones individuales. En la actualidad ya está a disposición de la comunidad científica y de la ciudadanía, en general, una web que recoge las curvas de evolución de la epidemia, en toda España y para cada Comunidad Autónoma: <https://covid19.citic.udc.es/>
2. Proveer de estimaciones y predicciones, y en su caso también un meta-predicador, para la evolución y predicción de la propagación y posible inmunización de la población a medio plazo, con el fin de asesorar sobre las medidas para salir de la crisis hacia un estado parecido a la normalidad.
3. Proveer de herramientas para la predicción de la evolución de cada paciente, a partir de la información relevante que se conozca del mismo, al objeto de poder anticiparse a su necesidad de hospitalización o ingreso en UCI, entre otros aspectos.
4. Proveer de herramientas de gestión de recursos humanos y materiales a los centros sanitarios, entre las que ya están listas para funcionar o en funcionamiento una aplicación web para los turnos de personal sanitario ante una situación extremadamente cambiante de plantillas (bajas, nuevas contrataciones, altas), y un modelo de simulación para simular la ocupación de las UCIs y plantas a corto-medio plazo por centros.

Estas son las actividades ya comenzadas e identificadas, sin detrimento de que entre los múltiples grupos de investigación que colaboran y se han ofrecido para este trabajo se estén desarrollando otras o puedan comenzarse en un futuro inmediato, en especial si se puede contar con información precisa.

Una característica de todas estas iniciativas es la necesidad de contar con información fiable. La hipótesis de partida en el mes de febrero y primera mitad de marzo era que la información pública que se tenía era fiable, cuando la realidad ha mostrado que estaba lejos de serlo. Así se muestra como primer paso necesario acudir a información menos agregada, más precisa y más cercana a donde se encuentran los pacientes.

Los estudios matemáticos que se plantean no requieren del conocimiento profundo de los pacientes, pero necesitan modelizar tiempos y probabilidades a partir de información fiable, tanto para valores agregados como individuales para las herramientas de gestión hospitalaria.

En particular, se precisa información para ajustar las distribuciones de probabilidad del tiempo que un paciente con COVID-19 grave va a estar ingresado en un hospital, el tiempo que estará en planta previo a ingreso en UCI si es el caso, la probabilidad del paso a la UCI y el tiempo de ingreso en UCI, el tiempo en su caso en planta después del paso por la UCI, las tasas de fallecimiento en planta y UCI. Estas estimaciones deben hacerse de forma global, así como de forma condicional teniendo en cuenta algunas características clínicas y sociodemográficas de cada paciente para evitar errores en la medida de lo posible.

3. METODOLOGÍA

Los objetivos planteados referidos a la evolución y predicción de la epidemia serán abordados mediante modelos dinámicos de ecuaciones diferenciales, modelos dinámicos estocásticos, modelos estadísticos. En general, están bien representados en la comunidad científica participante las distintas aproximaciones existentes para estos problemas. Sin embargo, la tarea del Comité se centra en facilitar información precisa y hacer el meta-predicador a partir de los distintos resultados.

Las predicciones de los grupos individuales se llevarán a cabo por investigadores voluntarios en el ámbito de la comunidad Matemática / Estadística / Científica de Datos a los que se ha hecho llegar el llamamiento de cooperación mediante diferentes medios de comunicación, redes sociales y el portal del CEMat con la página específica <http://matematicas.uclm.es/ceamat/covid19/>, que centralizará la información aportada por miembros de la comunidad matemática española relativos a la modelización del fenómeno de expansión del virus. A partir de la información recibida y la discusión entre la comunidad matemática, el Comité de Expertos creado a para este fin expreso elaborará una predicción a partir de las anteriores con sus correspondientes medidas de error, y la trasladará a las autoridades y, si se considera oportuno, se hará pública en la web del CEMat.

Respecto a las predicciones de más largo plazo, y que necesariamente tienen que incluir los efectos de las medidas tomadas hasta el momento, pero también de movilidad, etc. para predecir el efecto que tendrá la progresiva salida de la situación actual de confinamiento, se identificarán los modelos capaces de representar este efecto con el fin de hacer las predicciones. Necesariamente este modelo debe incluir un número realista de población afectada, y hasta cierto punto inmunizada, lo que pasa por seleccionar los modelos que mejor puedan predecir estos valores y adaptarse a condiciones cambiantes de movilidad, lo que puede incluir también el muestreo aleatorio para hacer estimaciones y para ajustar los modelos. Además del muestreo aleatorio que implica la realización de tests de forma extensiva, pues es preciso detectar los casos asintomáticos, otras fuentes de información serán también requeridas, desde la disponible en los centros de salud, como la que pueda obtenerse por encuestas y redes sociales. En este sentido, referido al muestreo aleatorio, ya se han recibido ofertas también para el diseño del muestreo, e incluso para el trabajo de campo.

En relación con herramientas para la predicción de la evolución de cada paciente, se desarrollarán modelos de clasificación supervisada. La literatura científica aporta numerosas herramientas dentro de la Ciencia de Datos, desde las del análisis de datos desde el punto de vista de la Estadístico hasta las herramientas más actuales de Aprendizaje Automático. Para la aplicación de todas estas herramientas, los equipos de investigación disponen del software necesario de código abierto y suficientemente contrastado. Para ello es preciso contar con información clínica de muestras de pacientes amplias para poder crear submuestras de entrenamiento y submuestras test de pacientes de esta infección que ya hayan recorrido todo el trayecto dentro del sistema sanitario para ajustar los distintos modelos.

Por último, referido a la gestión de recursos, hay dos iniciativas en marcha. Una es un modelo de optimización en una aplicación web comercial de turnos de personal en otro sector, y que se está adaptando para centros hospitalarios, en el corto plazo, para hacer replanificaciones. La empresa cede

el uso de la plataforma y los servidores, a la que accederían directamente los hospitales para que puedan cargar sus plantillas y poder reprogramar los turnos con facilidad ante cambios imprevistos de éstas. El personal puede instalar la aplicación en su teléfono para ver sus turnos. No hay tratamiento de datos por nadie ajeno a la entidad, salvo la grabación de la información en el servidor, y si se desea, se ofrece también la ayuda para cargar la información en la aplicación. La otra, que ya se está usando directamente en la Comunidad Foral de Navarra, se centra en la modelización mediante simulación de eventos discretos de un hospital, de modo que dada la situación en un momento dado de pacientes en el hospital, modelizados los tiempos de permanencia en planta y UCI y las respectivas tasas de paso de unas a otras así como de salida de ambas, así como la afluencia de pacientes a urgencias para los siguientes días, se puedan obtener predicciones de ocupación de planta y de UCI, así como de altas y de fallecimientos. Estas predicciones pueden ayudar no solo en la gestión con los pacientes e incluso al traslado de estos a otros hospitales, sino también en la previsión de recursos materiales y humanos requeridos para su atención.

Como puede verse, todos los modelos precisan del ajuste de patrones de probabilidad a partir de la información (datos) disponible. A continuación, se describe la información requerida que es básica para los objetivos del proyecto, así como la metodología para su tratamiento y análisis.

4. METODOLOGÍA DEL ANÁLISIS Y TRATAMIENTO DE LA INFORMACIÓN

Una de las primeras variables de interés para lograr los objetivos planteados en la sección anterior es el tiempo de estancia en hospital, desde que el paciente ingresa por COVID-19 hasta que es dado de alta o bien hasta que fallece. En muchos casos, este análisis se enfrenta al reto de que para muchos de los pacientes no se sabe el tiempo exacto de estancia porque aún están ingresados al finalizar el estudio, o simplemente por pérdida de seguimiento. El mismo problema se encuentra al estudiar el tiempo de estancia en UCI. Se trata por tanto de observaciones con un alto porcentaje de censura. La distribución de estos tiempos puede estimarse usando los estimadores límite-producto incondicional (Kaplan and Meier, 1958) y condicional (Beran, 1981), usando un selector del parámetro ventana de tipo validación cruzada (Geerdens et al, 2017). Pero incluso aunque hubiera suficientes observaciones ya finalizadas correspondientes a provincias donde la propagación de la epidemia se inició antes, es un reto dar distribuciones de probabilidad para estos tiempos, cuando no se sabe con exactitud los factores que influyen en los mismos.

En el estudio del tiempo de estancia en planta hasta que entra en la UCI también puede darse un alto porcentaje de censura, puesto que muchos pacientes estarán todavía en planta, sin entrar en UCI, al fin del estudio, o también por pérdida de seguimiento. Para algunos pacientes que estuvieron ingresados en planta se sabe que nunca entraron en la UCI porque han sido de dada alta o han fallecido antes. Pero en este análisis nos encontramos con el reto de cómo tener en cuenta que, de los pacientes todavía ingresados en planta algunos necesitarán UCI y otros no, sin conocer los factores que influyen en ello o sin conocer las particularidades del paciente. Una alternativa para tratar con esta situación es el uso de modelos de supervivencia más recientes que lo tienen en cuenta, llamados modelos de curación. Entre ellos se encuentran modelos incondicionales como el estimador de Laska and Meisner (1992), o de forma condicional, teniendo en cuenta las distintas características del paciente, como los estimadores de López-Cheda et al (2017a, b) y Safari et al (2020).

Por lo tanto, del tratamiento y análisis de los datos se pretenden obtener estimaciones puntuales y distribuciones de probabilidad para las principales variables de interés de los

pacientes dentro del sistema hospitalario, tanto globales como condicionales por las características de los pacientes, como edad, patologías, etc.

Para ello se necesita poder disponer de la siguiente información de los pacientes que han pasado por hospital:

- La **fecha** de distintos eventos de interés: positivo en COVID19, ingreso en hospital, ingreso en UCI, salida/exitus de UCI y alta/exitus de planta. En caso de no conocer las fechas exactas, se necesitaría el **número de días** hospitalizado, número de días en UCI y estado al finalizar el estudio (ingresado en planta, ingresado en UCI, alta o exitus). En cualquier caso es imprescindible conocer la fecha a la que están consolidados los datos (fecha de última actualización de los mismos).
- **Información clínica potencialmente relevante:** edad, sexo, patologías previas, serología, pruebas de diagnóstico, y toda cuanta información adicional esté disponible y sea susceptible de suministrarse dentro de la estricta observación de los protocolos de protección de datos.
- **Información sociodemográfica:** área sanitaria.

Sería deseable disponer de esta información a lo largo del tiempo para poder desarrollar el estudio longitudinal, entre otras para poder validar la modelización de la información, sobre todo ante la posibilidad que se ha comentado de datos altamente censurados, pero también por la diferente afluencia de pacientes en urgencias por COVID, así como para validar la modelización. Sin embargo, en lugares donde la epidemia se ha propagado antes, es posible que con la información de todos los pacientes de COVID que han pasado por el hospital pueda desarrollarse un modelo completo.

5. CONFIDENCIALIDAD Y PROTECCIÓN DE DATOS

Los datos en los que se basa este proyecto, si son tratados directamente de su fuente primaria (hospitales, centros de salud, etc.) son considerados por la legislación vigente en materia de protección de datos personales como categorías especiales de datos personales («datos sensibles») que requieren un tratamiento especial. Por este motivo se analiza la situación del tratamiento de datos para minimizar o eliminar el riesgo sobre el tratamiento de datos y establecer las medidas adecuadas.

La principal legislación aplicable en el tratamiento de datos es:

- Reglamento (UE) 2016/679 del Parlamento y del Consejo, de 27 de abril de 2016, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46 / CE (Reglamento general de protección de datos - RGPD) (DOUE 05/04/2016)
- Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales.
- Ley 12/1989, de 9 de mayo, de la Función Estadística Pública, así como la normativa autonómica aplicable en cada Comunidad.

El tratamiento de datos personales sensible está expresamente prohibido según establece el artículo 9.1 del RGPD con una serie de excepciones que incluirán el motivos cómo el interés público y los intereses vitales del interesado que recogen los artículos 9.c y 9 g y por ejemplo, los considerandos 46 “...“como por ejemplo cuando el tratamiento es necesario para fines humanitarios, incluido el control de epidemias y su propagación,... ” o el 52 “supervisión y alerta sanitaria, la prevención o control de enfermedades transmisibles y otras amenazas graves para la salud”, incluyendo explícitamente el tratamiento estadístico.

Esta excepción no es aplicable al CEMat en tanto que no ha recogido ni es responsable de los datos ni tiene la habilitación jurídica ni tiene la designación de una autoridad pública para llevar a cabo el proyecto.

No obstante, para el estudio estadístico no es necesaria la información identificativa de los pacientes o afectados del Coronavirus, sino que basta con la información desagregada del colectivo.

En este último supuesto, datos desagregados que **no permitan la identificación de personas**, el tratamiento de la información quedaría excluido del ámbito de la legislación en materia de protección, que define “dato personal” en su artículo 4 como:

1) «datos personales»: toda información sobre una **persona física identificada o identificable («el interesado»)**; se considerará persona física identificable toda persona cuya identidad pueda determinarse, directa o indirectamente, en particular mediante un identificador, como por ejemplo un nombre, un número de identificación, datos de localización, un identificador en línea o uno o varios elementos propios de la identidad física, fisiológica, genética, psíquica, económica, cultural o social de dicha persona;

Así pues, a fin de garantizar poder llevar a cabo el proyecto dentro de sus objetivos y sin vulnerar la legislación vigente en materia de protección de datos el Comité de Dirección del CEMat establece que en la solicitud y entrega de datos susceptibles de formar parte del estudio no deberá aparecer ningún tipo de dato que permita la identificación – directa o indirecta – de las personas sobre las que hace referencia.

Para este fin se establece el siguiente protocolo de seguridad de los datos:

1. Procedimiento de recogida

El CEMat determinará el investigador interlocutor con el centro médico, hospital o institución que recopilará los datos necesarios para la investigación, a fin de que estos investigadores tengan conocimiento de las medidas de seguridad establecidas en este protocolo.

El centro médico, hospital o institución suministrará la información y datos que puedan ser significativos para el estudio y que en ningún caso incluirán los siguientes datos que pudieran identificar personas:

- Nombre y apellidos
- DNI, NIE, pasaporte, etc.
- Dirección completa (fuera de la población o barrio)

Los datos de carácter personal serán codificados para evitar su asociación directa con los participantes, y por tanto tratados de forma confidencial en todo momento de acuerdo al Reglamento general de protección de datos (RGPD) (Reglamento (EU) 2016/679) y lo dispuesto en la Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales.

La información a suministrar se delimitará a aquella definida por el CEMat.

No se aceptará información que incluya datos personales, siendo comunicado al remitente para su revisión y corrección y siendo destruida / borrada cualquier copia o fichero que esté en posesión del CEMat de forma automática.

2. Tratamiento de la información

El CEMat centralizará la recogida de la información recogida.

Como medida adicional de seguridad se revisará y analizará toda la información para garantizar que no incluya datos personales. En caso de detectarse datos personales se borrará el fichero informático y se informará a la entidad para que remita la información según los estándares definidos.

La depuración de datos se llevará a cabo por personal designado por el CEMat. El tratamiento de la información será llevado a cabo por los investigadores voluntarios que previamente se hayan identificado y registrado en el proyecto.

CEMat, como depositaria y custodia de los datos del proyecto, aplicará las medidas de seguridad necesarias para garantizar la seguridad, confidencialidad e integridad de la información que pueda ser considerada confidencial.

6. INFORMACIÓN ADICIONAL

EQUIPO INVESTIGADOR: Investigadores en el ámbito de la Estadística y las Matemáticas de diversas universidades y organismos de investigación españoles que participan en el programa “Acción Matemática contra el Coronavirus” promovido por el Centro Español de Matemáticas (CEMat).

REFERENCIAS

Beran R (1981) Nonparametric regression with randomly censored survival data. Tech. rep., University of California, Berkeley

Laska EM, Meisner MJ (1992) Nonparametric estimation and testing in a cure model. *Biometrics* 48(4):1223–1234

López-Cheda A, Cao R, Jácome A, Van Keilegom I (2017a) Nonparametric incidence estimation and bootstrap bandwidth selection in mixture cure models. *Computational Statistics & Data Analysis* 105:144–165.

López-Cheda A, Jácome A, Cao R (2017b) Nonparametric latency estimation for mixture cure models. *TEST* 26(2):353–376

Geerdens, C. Acar, E.F. Janssen, P. (2018). Conditional copula models for right-censored clustered event time data. *Biostatistics* 19(2) 247–262.

Safari, W., López-de-Ullibarri, I. Jácome, A. (2020). A product limit estimator of the conditional survival function when cure status is partially known. *Submitted*.