Universidad Zaragoza
1542

Tesis Doctoral

Medición y comparación de resultados de servicios sanitarios: Métodos de ajuste de riesgos por carga de enfermedad para la medición y comparación de resultados de servicios sanitarios

Autor

Francisco Ramón Estupiñán Romero

Director/es

Enrique Bernal Delgado

Facultad de Medicina

2024

Tesis Doctoral

Medición y comparación de resultados de servicios sanitarios

Métodos de ajuste de riesgos por carga de enfermedad ***para la medición y comparación de resultados de servicios sanitarios***

Autor

Francisco Ramón Estupiñán Romero

Director/es

Bernal Delgado, Enrique

UNIVERSIDAD DE ZARAGOZA

Microbiología, Medicina Preventiva y Salud Pública

2024

Agradecimientos

Gracias al Dr. Enrique Bernal Delgado por su confianza, su paciencia y su insistencia en conseguir que me doctorara pese a mi determinación por agotar todos los plazos.

Gracias a la Dra. Mª José Rabanaque por su amabilidad y su comprensión.

Gracias a mis colegas y compañeros investigadores, coautores de las publicaciones presentadas en el presente trabajo. Estas publicaciones son fruto del esfuerzo multidisciplinario, del trabajo en equipo y de la predisposición para valorar, discutir y elaborar nuevas formas para la utilización de los datos sanitarios en la mejora de la calidad de la prestación de servicios sanitarios y la evaluación de los resultados en salud de políticas, servicios o tecnologías sanitarias.

Gracias a mi familia y amigos por su apoyo incondicional, y gracias a todos aquellos que, de un modo u otro, me han ayudado a llegar hasta aquí.

Tesis por compendio de publicaciones

La presente Tesis Doctoral está compuesta por un compendio de artículos previamente publicados en revistas científicas indexadas.

A continuación, se enuncian las referencias bibliográficas de los artículos que constituyen el cuerpo de la presente Tesis.

Artículo 1

Bernal-Delgado, E., Estupiñán-Romero, F. A data infrastructure for the assessment of health care performance: lessons from the BRIDGE-health project. Arch Public Health 76, 6 (2018). https://doi.org/10.1186/s13690-017-0245-1

Artículo 2

Comendeiro-Maaløe, M., Estupiñán-Romero, F., Thygesen, L. C., Mateus, C., Merlo, J., Bernal-Delgado, E., & ECHO consortium (2020). Acknowledging the role of patient heterogeneity in hospital outcome reporting: Mortality after acute myocardial infarction in five European countries. PloS one, 15(2), e0228425. https://doi.org/10.1371/journal.pone.0228425

Artículo 3

J. González-García, C. Tellería-Orriols, F. Estupiñán-Romero and E. Bernal-Delgado, "Construction of Empirical Care Pathways Process Models From Multiple Real-World Datasets," in IEEE Journal of Biomedical and Health Informatics, vol. 24, no. 9, pp. 2671-2680, Sept. 2020, doi: 10.1109/JBHI.2020.2971146

Artículo 4

Estupiñán-Romero, F., Pinilla Dominguez, J., Bernal-Delgado, E., & Atlas VPM consortium (2023). Differences in acute ischaemic stroke in-hospital mortality across referral stroke hospitals in Spain: a retrospective, longitudinal observational study. BMJ open, 13(6), e068183. https://doi.org/10.1136/bmjopen-2022-068183

Índice general

Contenido

Agradecimientos 5

Tesis por compendio de publicaciones 6

Artículo 1 6

Artículo 2 6

Artículo 3 6

Artículo 4 6

Índice general 8

Resumen 11

Antecedentes y Objetivos 11

Metodología 12

Discusión 13

Conclusiones 13

Introducción 16

Presentación de los artículos que comprende la tesis 19

Artículo 1 19

Artículo 2 20

Artículo 3 22

Artículo 4 23

Objetivos de investigación 25

Objetivo general 25

Objetivos específicos 25

Metodología 27

Aspectos metodológicos comunes 27

Reutilización de datos de salud de vida real 27

Alcance poblacional 27

Aproximación federada al análisis de datos 27

Diseño de Estudio 28

Población de Estudio 28

Unidad de análisis 29

Indicadores 29

Ajuste de riesgo por carga de enfermedad 31

Uso de modelos jerárquicos en la comparación por proveedor 32
Uso de técnicas y métodos de minería de procesos 32
Metodología propia del Artículo 1 33
Diseño 33
Fuente de información 33
Particularidades del análisis 34
Metodología propia del Artículo 2 34
Diseño 35
Fuente de información 35
Particularidades del análisis 35
Metodología propia el Artículo 3 38
Diseño 38
Fuente de información 38
Particularidades del análisis 38
Metodología propia del Artículo 4 40
Diseño 40
Fuentes de información 40
Particularidades del análisis 41
Consideraciones éticas, financiación y conflicto de interés 43
Consideraciones éticas 43
Financiación 43
Conflicto de interés 44
Discusión 45
Fortalezas y debilidades de los resultados obtenidos 45
Aportaciones del doctorando 47
Medición y comparación de resultados de servicios sanitarios 47
Implicaciones de los resultados y líneas de investigación futuras 49
Conclusiones 51
Referencias 52
PUBLICACIONES 64
Anexo I 65
Factor de Impacto (FI) y Áreas temáticas (AT) correspondientes a las publicaciones incluidas 65
Artículo 1 JCR **FI: 1.774 (Q3)** AT: **PUBLIC, ENVIRONMENTAL & OCCUPATIONAL HEALTH** 65

Artículo 2 JCR **FI: 3.24 (Q2)** AT: **MULTIDISCIPLINARY SCIENCES** 65
Artículo 3 JCR **FI: 5.772 (Q1)** AT: **MEDICAL INFORMATICS** 65
Artículo 4 JCR **FI: 2.9 (Q2)** AT: **MEDICINE, GENERAL & INTERNAL** 65
Anexo II: Publicaciones 66

Resumen

Antecedentes y Objetivos

La monitorización de los resultados en salud a lo largo del tiempo y la comparación entre proveedores fijando umbrales de efectividad comparada (e.g., comparación con los centros con mejores resultados de salud como centros de referencia) permite generar un sistema de retroalimentación positiva que informa de forma continua a los profesionales sanitarios y gestores sobre su procesos de toma de decisiones clínicas y la efectividad de sus actuaciones orientándolos en tiempo real en la aplicación de acciones de mejora más precisas.

Un aspecto adicional de este nuevo paradigma de monitorización del desempeño de los servicios sanitarios e información para la gestión de los mismos se basa en la capacidad de los sistemas analíticos de ofrecer resultados relevantes en el momento de la toma de decisiones con los datos disponibles, lo que obliga al refresco continuo de los datos o a la disposición de medidas de predicción basadas, bien en la extensión de las tendencias centrales mediante estimación ponderada a partir de los modelos habituales o a través de métodos de simulación que incorporen la información histórica como conocimiento a priori de las circunstancias del sistema. Esta forma de proceder configura lo que se ha dado en denominar 'Sistema Sanitario Inteligente' (Learning Health System). Esta necesidad de predecir las tendencias de desempeño de los servicios sanitarios con un cierto margen de confianza, en ocasiones proyectándose mucho en el futuro, obliga a incrementar la validez y eficiencia de los modelos de ajuste de riesgo para asegurar la comparación entre comparables y mejorar la precisión de las medidas, al tiempo que establece la necesidad de controlar el potencial sobre-ajuste del modelo.

La presente Tesis Doctoral, presenta un compendio de publicaciones que abordan aspectos metodológicos clave en la medición y comparación de resultados de servicios sanitarios.
Específicamente, se centra en 3 áreas principales:

a) aborda los desafíos de interoperabilidad y calidad de datos que plantea la creación de una infraestructura de datos sanitarios federada, diseñada para evaluar servicios de salud a nivel europeo;
b) subraya la importancia de analizar la interacción entre pacientes y proveedores sanitarios en la evaluación del rendimiento de los servicios. Esto incluye discernir qué proporción de la variabilidad en los resultados se debe a los proveedores y qué parte responde a la interacción con grupos específicos de pacientes, y
c) examina cómo se estructura y mantiene esa relación durante la atención médica (procesos asistenciales), su impacto en términos de continuidad de cuidados y su evolución a lo largo del tiempo cuando se introduce una nueva tecnología sanitaria efectiva.

Metodología

Los artículos incluidos en esta Tesis Doctoral se construyen como una suerte de test de validación de métodos y técnicas para la medición de resultados de los servicios sanitarios ejemplificados a través de la implementación de aproximaciones metodológicas diversas a casos de uso de evaluación comparativa de servicios y Sistemas de Salud, discutiendo sobre su oportunidad, alcance y utilidad para generar conocimiento actuable en la toma de decisiones en salud.

Los artículos incluidos se enfocan en los pasos necesarios para construir una infraestructura de datos de salud para investigación, y los métodos y técnicas para asegurar la calidad de los datos y la precisión de las medidas, para luego avanzar recomendaciones para la implementación de aproximaciones federadas de análisis a partir de la resolución de los desafíos de interoperabilidad entre los Sistema Sanitarios Europeos; para luego centrarse específicamente en validar la utilidad de tres aproximaciones metodológicas diversas y complementarias en la medición de resultados: a) el uso de modelos generalizados lineales mixtos (jerárquicos) con intercepto aleatorio y pendiente aleatoria para la caracterización de la interacción entre grupos de pacientes específicos y proveedores, b) el uso de la minería de procesos para el descubrimiento y la evaluación de los procesos asistenciales compuestos a partir de las trayectorias de interacción de los pacientes con los servicios sanitarios, y c) el uso de modelos generalizados aditivos mixtos (jerárquicos) para la caracterización de la evolución

temporal de los Sistemas de Salud como parte de la medición de resultados de los servicios sanitarios.

Discusión

En esta tesis se destacan las diferencias en variación y concordancia de las medidas de resultados a nivel hospitalario al considerar distintos modelos de evaluación. Los estudios incluidos generan conocimiento relevante para la toma de decisiones en la gestión de servicios sanitarios, sin embargo, una crítica común al estudio de estas variaciones es que, aunque pueden discriminar diferencias entre proveedores y establecer estándares de referencia, no se diagnostican las circunstancias que producen estos resultados. Los métodos y técnicas implementadas en los estudios incluidos demuestran que estas diferencias se justifican en parte por la adaptación a grupos específicos de pacientes, características del entorno y la adopción de nuevos medios y tecnologías, por lo que las perspectivas de análisis propuestas son relevantes en el ajuste de riesgos para la monitorización y evaluación comparativa de los resultados de los servicios sanitarios.

Conclusiones

En el trabajo de investigación desarrollado, que ha dado lugar a la Tesis Doctoral que se presenta, se ha podido observar la necesidad incorporar nuevos métodos y técnicas de análisis a nuestro instrumental para la evaluación de los servicios sanitarios a partir de la reutilización de datos de salud de vida real para avanzar en el desarrollo de un Sistema Sanitario Inteligente, en el que los decisores dispongan de información precisa, fiable y actuable a partir de la cual implantar medidas o políticas para la mejora de la asistencia sanitaria, la reducción de las variaciones injustificadas de la práctica médica y, en un extremo, la mejora de salud de la población.

Para ello es necesario resolver el acceso a los datos de salud a partir de la construcción de infraestructuras tecnológicas sostenidas por una gobernanza fuerte centrada en el aseguramiento de la seguridad, la confidencialidad y la calidad de los datos sanitarios y en su explotación transparente, trazable y reproducible para la evaluación de los servicios sanitarios, y considerar en esta evaluación la heterogeneidad de los pacientes atendidos, incluyendo términos de interacción entre grupos específicos de pacientes y los servicios,

explorando la organización de los procesos asistenciales para adecuarse a la mejor evidencia disponible, a las necesidades de la población atendida, pero también a las restricciones de entorno y a la cultura de los servicios; y, por último, atendiendo a la dinámica de la evolución de todos estos elementos a lo largo de los años, especialmente en aquellos procesos de alta complejidad que requieren en su atención de la coordinación e integración de cuidados entre proveedores y niveles asistenciales.

Introducción

Una parte importante de la investigación en servicios de salud se centra en la medición del rendimiento del sistema sanitario a partir del análisis y la comparación de los resultados en salud obtenidos por los proveedores de servicios a los distintos niveles de asistencia sanitaria.

Una de las claves en las que se fundamenta este tipo de análisis es el uso secundario de datos sanitarios, procedentes de fuentes de datos administrativos generados por sistemas de los servicios asistenciales, también conocidos como datos de vida real, en contraposición a los datos recopilados a raíz de un ensayo clínico o en un entorno experimental.

La explotación de estas fuentes de datos sanitarios permite aproximarse globalmente al sistema sanitario a través de la caracterización de su actividad y sus resultados, contando además con información sobre características y situación de los pacientes atendidos, sus diagnósticos y la mayor parte de las intervenciones realizadas durante su estancia hospitalaria.

La integración de distintas fuentes de datos administrativos de varios países europeos se ha revelado útil para evaluar las variaciones injustificadas de los resultados de la asistencia sanitaria.
En este trabajo se describen los procedimientos utilizados para crear una infraestructura de datos (por ejemplo, acceso e intercambio de datos, definición del acervo común mínimo de datos necesarios y desarrollo del modelo de datos lógicos relacionales) y, los métodos para producir mediciones fiables del rendimiento de la asistencia sanitaria (por ejemplo, normalización de ontologías y análisis de garantía de calidad). Por otro lado, la precisión y fiabilidad de los resultados de explotación de estos datos depende de la calidad de los mismos, que pueden presentar problemas de completitud, incorrecta codificación o mala clasificación de los registros, que es necesario evaluar para informar como potenciales limitaciones a las conclusiones de estos análisis.
El estudio de las variaciones de la práctica médica como aproximación a la medición de las diferencias sistemáticas de comportamiento entre proveedores configura un marco de referencia para el análisis de la efectividad de un servicio sanitario o un proveedor de servicios

sanitarios en base a la observación ponderada de sus resultados de salud a lo largo del tiempo. Este tipo de análisis tiene también en consideración factores asociados a estas variaciones, como: a) la demanda (i.e., características epidemiológicas, carga de enfermedad, condiciones socioeconómicas, estados de salud percibidos y preferencias personales, etc.); b) la oferta asistencial (i.e., disponibilidad de personal, técnicas y recursos sanitarios para la aplicación de ciertas intervenciones, diferencias en los modelos organizacionales, etc.); o c) la interacción de ambas cuestiones con intangibles como escuelas de pensamiento médico o usos y costumbres institucionales, o protocolos asistenciales organizados en procesos o vías clínicas, más o menos establecidos en base a distintos niveles de evidencia.

La comprobación de variaciones sistemáticas en la atención de una condición específica y la medición de diferencias injustificadas en los resultados de salud entre proveedores (e.g., incidencia de mortalidad intrahospitalaria para una condición) requiere de la exclusión o minimización de las causas razonables de variación y otros potenciales factores de confusión. Es en este punto donde demuestran su utilidad los métodos de ajuste de riesgos que permiten modelizar estos factores a nivel de paciente y de su interacción con los proveedores sanitarios, para llegar a ofrecer una perspectiva sobre la varianza explicada por cada factor y configurar un marco de comparación entre comparables.

En este sentido, existen en la literatura multitud de alternativas para la modelización del riesgo, centradas mayoritariamente en la estimación de la mortalidad intrahospitalaria y reingresos como proxy de resultados en salud fácilmente medibles a partir de estas fuentes de datos.

El uso de métodos de ajuste de riesgos basados en modelos multinivel con la posibilidad de introducción de interacciones dentro de los distintos niveles y el análisis segmentado de resultados según características específicas de subgrupos de pacientes, o la expansión del uso de modelos bayesianos basados en la simulación de eventos discretos o interacción de ecuaciones estructurales ha permitido la expansión de las capacidades interpretativas y predictivas de estos análisis dando como resultado una mayor fiabilidad y precisión de la técnica para llegar a explicar los resultados y prescribir acciones de mejora basadas en la evidencia obtenida del análisis.

En la actualidad siguen surgiendo abordajes para el problema del ajuste de riesgos por carga de enfermedad basados en las últimas técnicas de aprendizaje estadístico no supervisado (e.g., aplicación de redes neuronales al descubrimiento de estos patrones de coocurrencia de enfermedad), debido en parte a la disponibilidad de conjuntos masivos de datos asistenciales y en parte al gran interés por el desarrollo de sistemas de clasificación de patologías basadas en reglas inferidas de forma automática con relevancia para áreas de negocio relacionadas con la gestión de la demanda sanitaria en mercados de aseguramiento libre. Esto no quita para que la aplicación de estas técnicas o métodos tenga, necesariamente, que estar justificada por su capacidad de ofrecer nuevas perspectivas sobre alguna dimensión del Sistema Sanitario que pueda informar la toma de decisiones en salud de forma efectiva.

En este sentido, el problema al que nos enfrentamos en el desarrollo de este trabajo es la falta de un consenso generalizado en la metodología para la medición de resultados en los servicios sanitarios, más allá del consenso de que los métodos tradicionales basados en listas de comorbilidades deberían estar superados.

Así pues, el objetivo de este trabajo es el de entender, aplicar y analizar de forma comparada diversos métodos y modelos de ajuste de riesgos para su aplicación en la evaluación de la efectividad y calidad de los hospitales y otros servicios sanitarios, utilizando la validación empírica de estos métodos en la investigación de casos de uso de monitorización, evaluación y comparación del rendimiento de servicios sanitarios a partir de la reutilización de datos de salud de vida real.

Presentación de los artículos que comprende la tesis

Artículo 1

Bernal-Delgado, E., Estupiñán-Romero, F. A data infrastructure for the assessment of health care performance: lessons from the BRIDGE-health project. Arch Public Health 76, 6 (2018). https://doi.org/10.1186/s13690-017-0245-1

El artículo analiza la creación y utilización de una infraestructura de datos destinada a evaluar el rendimiento de la asistencia sanitaria en varios países europeos. Esta infraestructura integra distintas fuentes de datos administrativos y ha sido decisiva para evaluar variaciones injustificadas en el rendimiento de la atención sanitaria.

Se describen los procedimientos implicados en el establecimiento de esta infraestructura de datos, que incluye el acceso y el intercambio de datos, la definición de un mínimo común de datos requeridos para resolver preguntas de investigación de medición y evaluación de servicios sanitarios y el desarrollo de un modelo de datos lógico relacional para esta infraestructura. Además, esbozan los métodos utilizados para producir mediciones fiables del rendimiento de la asistencia sanitaria, como la normalización de ontologías y el análisis de garantía de calidad.

El documento aporta ideas fundacionales sobre la aplicabilidad del modelo que se propone a una infraestructura de investigación europea llegando a sugerir que una infraestructura federada basada en la interoperabilidad semántica podría ser una forma sensata de avanzar hacia un Espacio Europeo de Datos de Salud (EEDS). En una infraestructura de este tipo, los datos individuales permanecerían en el país, y los análisis estadísticos serían distribuidos entre los centros que componen la infraestructura para generar resultados locales (i.e., nacionales o regionales) que luego puedan ser comparados o meta-analizados a nivel Europeo.

Artículo 2

Comendeiro-Maaløe, M., Estupiñán-Romero, F., Thygesen, L. C., Mateus, C., Merlo, J., Bernal-Delgado, E., & ECHO consortium (2020). Acknowledging the role of patient heterogeneity in hospital outcome reporting: Mortality after acute myocardial infarction in five European countries. PloS one, 15(2), e0228425. https://doi.org/10.1371/journal.pone.0228425

El rendimiento hospitalario, presentado como la comparación de mediciones medias, descarta que los resultados hospitalarios puedan variar según los tipos de pacientes y la interacción del hospital con estos subgrupos (i.e., pacientes con características similares). El objetivo de este trabajo es destacar la importancia de tener en cuenta la heterogeneidad de los pacientes a la hora de informar sobre el rendimiento de los hospitales a fin de poder comparar entre comparables.

Se llevó a cabo un estudio observacional sobre datos administrativos de prácticamente todos los ingresos hospitalarios de 2009 por infarto agudo de miocardio (IAM) dados de alta en Dinamarca, Portugal, Eslovenia, España y Suecia. El rendimiento hospitalario se aproximó mediante la mortalidad intrahospitalaria ajustada por riesgo. Se utilizaron modelos de regresión multinivel (MLRM) para evaluar las diferencias en el rendimiento de los hospitales, comparando las estimaciones de los modelos de intercepción aleatoria (que recogen los efectos contextuales generales de los hospitales) y los modelos de pendiente aleatoria (que recogen los efectos contextuales de los hospitales para los pacientes con y sin insuficiencia cardiaca congestiva - ICC). Se utilizó el índice Kappa (KI) ponderado para evaluar la concordancia entre las estimaciones de rendimiento.

Se encontraron diferencias entre los modelos jerárquicos considerando sólo el hospital como intercepto aleatorio y los modelos que incluían también la pendiente aleatoria derivada de la interacción de paciente y hospital a nivel del modelo.

Se analizaron 46.875 ingresos por IAM, 6.314 con ICC coexistente, dados de alta en 107 hospitales. La tasa global de mortalidad intrahospitalaria fue del 5,2%, oscilando entre el 4% de Suecia y el 6,9% de Portugal. El MLRM con pendiente aleatoria superó al modelo con sólo intercepto aleatorio, destacando un GCE mucho mayor en los pacientes con ICC [VPC = 8,34 (IC95% 4,94 a 13,03) y MOR = 1,69 (IC95% 1,62 a 2,21) frente a VPC = 3,9 (IC95% 2,4 a 5,9),

MOR de 1,42 (IC95% 1,31 a 1,54) sin ICC]. No se observó concordancia entre las estimaciones [KI = -0,02 (IC95% -0,08 a 0,04].

La diferencia de efectos contextuales generales en los pacientes con IAM con y sin ICC, junto con la falta de concordancia en las estimaciones, sugiere que es necesario tener en cuenta la heterogeneidad de los pacientes para caracterizar e informar adecuadamente sobre el rendimiento de los hospitales.

Artículo 3

J. González-García, C. Tellería-Orriols, F. Estupiñán-Romero and E. Bernal-Delgado, "Construction of Empirical Care Pathways Process Models From Multiple Real-World Datasets," in IEEE Journal of Biomedical and Health Informatics, vol. 24, no. 9, pp. 2671-2680, Sept. 2020, doi: 10.1109/JBHI.2020.2971146

Los procesos asistenciales son "planes asistenciales multidisciplinares que detallan los pasos asistenciales esenciales para pacientes con problemas clínicos específicos". El presente trabajo describe cómo aplicar una metodología existente de minería de procesos para construir modelos empíricos de procesos de un proceso asistencial hiperagudo de alta complejidad (i.e., ictus). Estos modelos de proceso son una pieza única de información para la investigación en servicios sanitarios: por ejemplo, para evaluar su conformidad con el proceso teórico descrito en las guías clínicas o para evaluar el impacto del proceso en los resultados de salud.

Para ello, este trabajo se basa en el diseño e implementación de una solución que a) sintetiza el conocimiento experto sobre cómo se presta la atención sanitaria dentro y entre proveedores como un registro de actividades, y b) construye el modelo de proceso a partir de ese registro de actividades utilizando técnicas de minería de procesos. A diferencia de investigaciones anteriores basadas en capturas de datos ad hoc, el enfoque actual se basa en la vinculación de varios conjuntos de datos heterogéneos del mundo real que comparten una vinculación semántica mínima.

Este trabajo demuestra la viabilidad de construir modelos de procesos a partir de datos de vida real de vías asistenciales utilizando análisis de minería de procesos. Para ello se requiere una planificación cuidadosa de la definición de un modelo de datos y el desarrollo de un algoritmo de vinculación que garantice la coherencia temporal entre los datos disponibles de los distintos conjuntos de datos.

La solución propuesta abre la puerta a un amplio conjunto de estudios en profundidad sobre cómo se organizan los sistemas sanitarios, cómo esta organización puede afectar a los resultados en salud y cómo mejorar la organización tanto a nivel de sistema sanitario como, en último término, en la práctica clínica diaria.

Artículo 4

Estupiñán-Romero, F., Pinilla Dominguez, J., Bernal-Delgado, E., & Atlas VPM consortium (2023). Differences in acute ischaemic stroke in-hospital mortality across referral stroke hospitals in Spain: a retrospective, longitudinal observational study. BMJ open, 13(6), e068183. https://doi.org/10.1136/bmjopen-2022-068183

Este artículo continúa el trabajo anterior, evaluando las diferencias en la mortalidad intrahospitalaria por ictus isquémico agudo entre hospitales de referencia de ictus y aportar pruebas sobre la asociación de esas diferencias con la adopción a destiempo de terapias de reperfusión eficaces. Se trata de un estudio observacional longitudinal retrospectivo utilizando datos administrativos de prácticamente todos los ingresos hospitalarios desde 2003 hasta 2015. Treinta y siete hospitales de referencia de ictus del Sistema Nacional de Salud español. Los participantes fueron pacientes mayores de 18 años con un episodio hospitalario con un ingreso diagnóstico de ictus isquémico agudo en cualquier hospital de referencia de ictus (196.099 ingresos). Los indicadores de resultados principales que se evaluaron fueron (1) la variación hospitalaria en la mortalidad intrahospitalaria a los 30 días medida en términos del coeficiente de correlación intraclase (CCI); y (2) la diferencia en la mortalidad entre el hospital de tratamiento y la tendencia de utilización de terapias de reperfusión (incluida la fibrinólisis intravenosa y la trombectomía mecánica endovascular) en términos de odds ratio mediana (MOR).

La mortalidad intrahospitalaria ajustada por ictus isquémico agudo a los 30 días disminuyó durante el periodo de estudio. Las tasas de mortalidad intrahospitalaria ajustada tras un ictus isquémico agudo variaron del 6,66% al 16,01% entre hospitales. Más allá de las diferencias en las características de los pacientes, la contribución relativa del hospital de tratamiento fue mayor en el caso de los pacientes sometidos a terapias de reperfusión [ICC=0,031 (ICB 95%=0,017-0,057)] que en el caso de los que no lo hicieron [ICC=0,016 (ICB 95%=0,010-0,026)]. Utilizando el MOR, la diferencia en el riesgo de muerte fue de hasta un 46% entre el hospital con el riesgo más alto y el hospital con el riesgo más bajo en pacientes sometidos a terapia de reperfusión [MOR 1,46 (ICB 95%: 1,32-1,68)]; en pacientes no sometidos a ninguna terapia de reperfusión, el riesgo fue un 31% mayor [MOR 1,31 (ICB 95%: 1,24-1,41)].

En los hospitales de referencia de ictus del Sistema Nacional de Salud español, la mortalidad intrahospitalaria global ajustada disminuyó entre 2003 y 2015. Sin embargo, persistieron las variaciones de mortalidad entre hospitales.

Este trabajo señala la oportunidad del uso de los modelos GAMM para estimar la evolución temporal de los indicadores de rendimiento de los servicios sanitarios y caracterizar aquellos fenómenos relacionados con modificaciones organizativas como la adopción de una nueva tecnología efectiva.

Objetivos de investigación

La línea de investigación se corresponde con la medición y comparación de resultados de servicios de salud a partir del uso secundario de datos de salud de vida real, enmarcada dentro de la Investigación en Políticas y Servicios de Salud.

Esta Tesis se organiza en ese sentido en torno a un Objetivo General que se desarrolla y responde distintos Objetivos Específicos que han sido abordados en los distintos trabajos que conforman la línea de estudio.

Objetivo general

Entender, aplicar y analizar de forma comparada diversos métodos y modelos de ajuste de riesgos por carga de enfermedad para su aplicación en la evaluación de la efectividad y calidad de los hospitales y otros servicios sanitarios, utilizando la validación empírica de estos métodos utilizando datos de salud de vida real.

Objetivos específicos

1. Describir los procedimientos requeridos para establecer una infraestructura de datos sanitarios para investigación en servicios de salud (i.e., acceso a datos, distribución de código de análisis, definición de conjunto mínimo requerido de datos, desarrollo de modelo común de datos, etc.) y los métodos para producir mediciones fiables del rendimiento de los servicios sanitarios (i.e., estandarización de los procesos, normalización semántica mediante ontologías y diseño e implementación de los análisis de calidad).
2. Validar la utilidad de considerar la interacción entre los proveedores sanitarios y grupos de pacientes en el ajuste de riesgos en la medición y comparación de resultados de salud por proveedor (i.e., hospital).
3. Desarrollar y probar los métodos de minería de procesos para descubrir empíricamente procesos asistenciales complejos en los que existe continuidad de cuidados de modo que se haga posible evaluar la porción de efectividad asociada a la organización de los cuidados más allá del proveedor de asistencia sanitaria.

4. Desarrollar y validad un modelo de medición y comparación de resultados de salud entre proveedores sanitarios que considere los cambios en la práctica médica a lo largo del tiempo derivados, por ejemplo, de la introducción o adopción de tecnologías sanitarias efectivas en un proceso de alta complejidad.

Los objetivos específicos se abordan respectivamente en cada uno de los trabajos que componen esta Tesis Doctoral.

Metodología

La línea de investigación de la presente Tesis Doctoral ha desarrollado y validado varias metodologías basada en la reutilización de datos de salud de vida real en el desarrollo de estudios de investigación - también llamados casos de uso- en los que se evalúan distintos aspectos del rendimiento de los servicios sanitarios (i.e., indicadores de proceso, o de resultados en salud), bien desde una perspectiva poblacional o desde la perspectiva de la comparación entre proveedores sanitarios (i.e., hospitales). En este apartado se abordarán en primer lugar los aspectos metodológicos comunes de la línea de investigación y, posteriormente, los métodos implementados en cada uno de los trabajos.

Aspectos metodológicos comunes

Reutilización de datos de salud de vida real

Todos los estudios incluidos en la presente Tesis Doctoral se basan en el uso secundario o reutilización de datos de salud a nivel individual y pseudonimizados, generados y recopilados por los sistemas de información de salud de los Sistemas de Salud a partir de la interacción de la población (demanda) con los servicios sanitarios (oferta). En general, las principales fuentes de información consultadas para la realización de los mismos son de datos sanitarios clínicos y administrativos estructurados. En función del estudio, estos datos pueden verse complementados con otras fuentes de información sociodemográficas a nivel poblacional o datos de recursos o características de los proveedores sanitarios - principalmente hospitales.

Alcance poblacional

El alcance de todos los trabajos se extiende a toda la población de referencia cubierta por el Sistema de Salud de los países o regiones participantes en cada estudio a nivel de la unidad de análisis correspondiente, en algún caso país, región, área de salud o proveedor sanitario.

Aproximación federada al análisis de datos

En general, los trabajos presentados sientan las bases de una aproximación federada al análisis de datos. Esta aproximación implica, que en estudios comparativos en los que

participan distintas regiones o sistemas sanitarios, los datos de salud - de especial protección- no son transferidos desde los orígenes de datos en donde son custodiados por los Sistemas Sanitarios o las instituciones responsables de su gestión en cada país y región, sino que se distribuye el código (programas) de análisis estadístico basados en un modelo común de datos que recoge los requerimientos de los datos de entrada para producir resultados parciales locales a nivel de cada participante. Esto implica que a la hora de comparar sea necesario asegurar no sólo la reproducibilidad de los análisis sino también su consistencia, ya que la comparación requiere de la recolección y meta-análisis de estos resultados parciales locales de todos los participantes.

La aproximación federada se basa en el principio de 'visita del dato' o, dicho de otro modo, en el hecho de que el programa de análisis visita el dato en lugar de transferir todos los datos a un repositorio común.

Diseño de Estudio

Todos los estudios tienen un diseño observacional y, por lo tanto, retrospectivo. Salvo para el artículo 2, el periodo de estudio abarca múltiples años independientemente de que el planteamiento del estudio sea longitudinal o transversal de corte anual.

Población de Estudio

Los artículos 1 y 2 comparten la mayor parte de la población de estudio, construyéndose con base poblacional abarcando a toda la población con algún episodio de hospitalización entre 2003 y 2009 en 5 países europeos (Dinamarca, Eslovenia, España, Inglaterra y Portugal, con la incorporación de Suecia en lugar de Inglaterra en el artículo 2).

El artículo 3 se basa en la población de Aragón para construir una cohorte de pacientes que ingresan en urgencias hospitalarias o en hospital con un diagnóstico de ictus o con la sospecha de un ictus isquémico.

Por último, el artículo 4 considera como población de estudio todos los episodios de hospitalización por ictus isquémico financiados públicamente en el Sistema Nacional de Salud

en aquellos hospitales que actúan como centros de referencia para la atención al ictus isquémico (i.e., hospitales con capacidad de realización de trombectomía mecánica).

Unidad de análisis

Los estudios se enfocan principalmente en dos tipos de análisis en base a la unidad de análisis que se considera relevante para la comparación. Por un lado, tenemos a) estudios geográficos, con base poblacional, cuya unidad de análisis es una demarcación geográfica administrativa del Sistema Sanitario, generalmente referida a área de residencia del paciente, al nivel de desagregación que resulte relevante para la toma de decisiones sobre asistencia sanitaria; y, por otro, b) comparación entre proveedores asistenciales (i.e., hospitales), donde la unidad de análisis es el centro sanitario (proveedor) que sirve a un área de salud. En este último caso, aunque la comparación se realiza entre centros, los estudios tratan de capturar toda la población atendida independientemente de su origen geográfico o área de residencia.

Artículo 1. Se presenta una infraestructura de datos integrada por varios orígenes de información con el propósito general de realizar estudios de monitorización y evaluación del rendimiento de los Sistemas Sanitarios, calculándose varios indicadores tanto a nivel geográfico como a nivel de proveedor para 5 países europeos.

Artículo 2 y 4. En ambos casos el proveedor asistencial (hospital) es la unidad de análisis para la evaluación y comparación de la variación de resultados en salud, aunque en enfermedades y contextos diferenciales.

Artículo 3. En este estudio, la unidad de análisis relevante en la comparación es la región como unidad de toma de decisiones a nivel de Sistema Sanitario en el entorno de España, donde las competencias en Sanidad están transferidas a las comunidades autónomas.

Indicadores

Los artículos presentados en esta tesis utilizan una gran diversidad de indicadores de proceso y de resultados en salud en dependencia de los objetivos de cada estudio, pero con la

vocación de ofrecer una perspectiva única sobre el funcionamiento del Sistema Sanitario que permita su monitorización y evaluación por comparación.

En general, todos los indicadores usados han sido validados anteriormente en alguno de los proyectos del Atlas de Variaciones de la Práctica Médica (Atlas VPM) a nivel nacional o del proyecto European Collaboration for Hospital Optimization (ECHO-Health). Estos pueden ser indicadores de proceso, que facilitan la medición del funcionamiento de un proceso como, por ejemplo, la tasa estandarizada de cirugía de reemplazamiento total o parcial de cadera en pacientes añosos con fractura de cadera o indicadores de resultado, como por ejemplo la tasa estandarizada de muerte por infarto de miocardio, muerte por ictus isquémico. Por otro lado, estos indicadores pueden estar calculados como tasas indirectamente estandarizadas a nivel poblacional por sexo y grupo quinquenal de edad considerando la población de referencia del área geográfica de residencia, independientemente de donde es atendido el paciente, o estimarse como una razón de mortalidad estandarizada para cada hospital, representando la tasa media de muerte (reingreso, u otro resultado relevante) por una condición.

El artículo 1 presenta y discute la construcción y medición de una multitud de indicadores tanto de proceso como de resultado a partir de los datos integrados en la infraestructura de datos para investigación. Algunos de los indicadores presentados miden, por ejemplo, los ingresos hospitalarios no programados por infarto de miocardio diferenciando entre aquellos con ST-elevado y aquellos sin, los ingresos hospitalarios no programados en pacientes diabéticos con complicaciones agudas, los ingresos hospitalarios por fallo cardíaco congestivo en adultos mayores de 40 años, o la mortalidad intrahospitalaria por cualquiera de las condiciones anteriores. Un listado extenso de los indicadores producidos en este trabajo puede ser consultado en el material suplementario del artículo o en la dirección web [ECHO-Health indicators].

El artículo 2 se centra exclusivamente en el estudio de la mortalidad intrahospitalaria después de ingreso por infarto agudo de miocardio como indicador de resultado de la provisión de cuidados sanitarios para una condición urgente y de alta incidencia. Evaluando el impacto de tener en consideración en el modelo la existencia de riesgos basales de muerte distintos en función de si el paciente tiene como antecedente una cardiopatía crónica congestiva.

El artículo 3 mide la mortalidad por ictus, centrándose más específicamente en la mortalidad por ictus isquémico, aunque en este caso, los indicadores de proceso ganan relevancia al estudiarse el proceso empírico de asistencia integrada entre urgencias y atención hospitalaria a la sospecha de ictus - también conocido como 'Código Ictus' en España. En este estudio se tienen en cuenta otros indicadores como parte de la evaluación de este proceso como el tiempo puerta-aguja, desde la admisión en urgencias hasta la administración de un fármaco fibrinolítico para comparar la distribución de este tiempo en los pacientes con sospecha de ictus isquémico atendidos en el conjunto de hospitales de Aragón y elicitar diferencias entre ellos y respecto a la ventana de oportunidad de tratamiento (ventana de indicación de fibrinólisis).

Finalmente, el artículo 4 mide también la mortalidad intrahospitalaria en pacientes que ingresan con un diagnóstico de ictus isquémico, pero diferenciando entre aquellos que reciben y aquellos que no reciben ninguna terapia de reperfusión para evaluar la variación entre hospitales en ambos grupos de pacientes y a lo largo de los años considerando la paulatina adopción de las técnicas de reperfusión que nos son comunes en la actualidad.

Ajuste de riesgo por carga de enfermedad

En todos los artículos los indicadores se estiman a partir de modelos de ajuste de riesgo en base a características sociodemográficas de los pacientes y de carga de enfermedad. En todos los casos el ajuste de riesgos por carga de enfermedad se hace identificando para cada indicador el subconjunto de comorbilidades que pueda resultar relevante. En general, se utiliza como referencia el conjunto de comorbilidades propuesto por Elixhauser, introduciendo cada comorbilidad como una variable binaria independiente sin incluir ningún índice o medida sintética de carga de enfermedad. El subconjunto de variables relevantes se justifica en cada caso a los efectos de reducir la variabilidad observada debido al riesgo basal diferencial de los pacientes analizados a los efectos de mejorar la señal de la variación no justificada asociada a aspectos derivados de la oferta sanitaria, es decir, del proveedor asistencial.

En algunos estudios, en concreto, en los artículos 1 y 4 se introduce también el tiempo (i.e., año, mes, día) con objeto de capturar y ajustar estacionalidades (o componentes periódicos) de la variación a lo largo del periodo de estudio.

Uso de modelos jerárquicos en la comparación por proveedor

Los artículos 1, 3 y 4 refieren comparaciones por proveedor que se establecen a partir de la estimación de razones de mortalidad estandarizadas a partir de modelos jerárquicos, en los que la muerte intrahospitalaria por una condición determinada se estima considerando las variables sociodemográficas y de carga de enfermedad a nivel de paciente como efectos fijos, mientras que se introduce el proveedor (hospital) como efecto aleatorio permitiendo la comparación indirecta respecto a la media del conjunto de proveedores estudiados o respecto a un estándar de referencia (*benchmark*) que puede establecerse en base a un percentil deseado en función de la lógica del indicador (i.e., a mayor muerte peor rendimiento/calidad de la asistencia, a menor infección de herida quirúrgica o sepsis post-operatoria mejor rendimiento/calidad/seguridad de la asistencia sanitaria).

Uso de técnicas y métodos de minería de procesos

El artículo 3 se fundamenta en el uso de técnicas y métodos de minería de procesos como el descubrimiento empírico de los procesos asistenciales a partir de un histórico (log) de eventos construido a partir de la secuencia cronológica de actividades clínico-administrativas durante la provisión de cuidados sanitarios. La minería de procesos utiliza las secuencias de actividades asociadas a instancias asistenciales de interacción de los pacientes con una condición clínica determinada con los proveedores sanitarios en los distintos niveles asistenciales desde la demanda de servicios sanitarios hasta la resolución del proceso para crear un mapa del proceso en base al agregado de trayectorias de los pacientes atendidos. Este mapa del proceso asistencial, similar a las vías clínicas tradicionalmente establecidas para asegurar la continuidad de la asistencia permite, por ejemplo, comparar la realidad de la práctica médica con los protocolos asistenciales o las guías de práctica clínica, explorar la continuidad asistencial entre diversos dispositivos médicos o comparar las trazas más frecuentes entre Sistemas Sanitarios ofreciendo una perspectiva enriquecida a la organización de los servicios (oferta) a las necesidades de los pacientes (demanda). El uso de

este tipo de técnicas permite identificar puntos de acción en los cuales la toma de decisiones puede verse influida por la disponibilidad de información precisa y oportuna sobre aquello que funciona (comparadores de referencia).

Metodología propia del Artículo 1

El objetivo de este estudio fue establecer recomendaciones para la creación de una infraestructura de datos de salud de vida real para la investigación en servicios y políticas de salud; y sobre la implementación de métodos para producir mediciones fiables del rendimiento de los servicios sanitarios.

Diseño

Este artículo no se trata de un estudio per se sino de una revisión crítica de la construcción de una infraestructura de datos y de las medidas de evaluación de la calidad de los datos de dicha infraestructura para el aseguramiento de la precisión en la monitorización y comparación entre servicios sanitarios de varios países europeos.

Fuente de información

La infraestructura de datos a la que se hace referencia en este artículo surge del proyecto ECHO e incluye varios orígenes de datos [altas hospitalarias, información demográfica a nivel geográfico, datos socioeconómicos a nivel geográfico, características de la oferta a nivel hospitalario y vectores geográficos que representan áreas geográficas de interés (por ejemplo, zonas de captación de hospitales, autoridades sanitarias, regiones, NUTS)] con vistas a, partiendo de epígrafes individuales, monitorizar y evaluar el funcionamiento de los Sistema de Salud tanto a nivel geográfico como hospitalario, teniendo en cuenta su evolución a lo largo del tiempo (por ejemplo, procesos de fusión de hospitales).

Así, el modelo lógico de datos, se concibió como una base de datos relacional, construida sobre tres entidades principales (episodios, hospitales y áreas geográficas) y sus respectivos atributos.

Particularidades del análisis

Los análisis presentados en el artículo 1 se corresponden con análisis de calidad siguiendo el marco metodológico de análisis de la calidad de estadísticas oficiales propuesto por Eurostat. El modelo de datos de la infraestructura comentada posee tres elementos críticos:

a) dado que los episodios almacenan información individual a nivel de paciente que en realidad está integrada tanto en un hospital de tratamiento como en áreas geográficas jerárquicamente constituidas (asistencia sanitaria en regiones, regiones en países), la vinculación entre archivos y catálogos sigue un esquema 1 a 1 (cuando la vinculación se limita a atributos basados en episodios) o un esquema 1 a N (cuando los episodios están vinculados a un hospital o a un área). Los identificadores unívocos a nivel de episodio garantizan la trazabilidad y robustez de esta vinculación; b) dada la riqueza de la información contenida en cada episodio, dependiente principalmente del gran número de diagnósticos y procedimientos registrados en cada registro (~20 a 50 variables cada uno), es necesario aumentar la eficiencia computacional. Con este fin, los diagnósticos y las variables de procedimiento se dividieron en dos catálogos diferentes que contenían todos los diagnósticos o procedimientos existentes, en su máximo nivel de precisión; y c) en lugar de producir un único resultado, el modelo de datos implica la producción de tres archivos de salida independientes que contienen indicadores de proceso e indicadores de resultados medidos a nivel geográfico, indicadores de resultados medidos a nivel de hospital (proveedor) y variables intermedias ad hoc y modificadores para garantizar el control de los factores de confusión (por ejemplo, comorbilidades de Elixhauser).

Los análisis que se muestran en el artículo se corresponden con la evaluación de la coherencia, la preservación de identidad referencial de las entidades contenidas, la cardinalidad de los datos y las herencias de sus atributos, y la consistencia semántica de los indicadores entre unidades de análisis y a lo largo del tiempo.

Metodología propia del Artículo 2

El objetivo de este estudio fue el destacar la importancia de tener en cuenta la heterogeneidad de los pacientes al informar sobre el rendimiento de los hospitales, a partir de la introducción una noción del riesgo basal de distintos grupos de pacientes como función de su heterogeneidad.

Diseño

Estudio observacional transversal que utiliza datos administrativos que representan prácticamente todos los ingresos hospitalarios por infarto agudo de miocardio en pacientes de 40 a 80 años, tratados en 434 hospitales de 5 países europeos (Dinamarca, Portugal, Eslovenia, España y Suecia) en 2009, con un total de 73.812 episodios potenciales dados de alta. Los hospitales con menos de 250 episodios de infarto agudo de miocardio en 2009 (umbral discrecional) fueron excluidos de la muestra para reducir la heterogeneidad estructural entre hospitales y ganar robustez en las estimaciones. La muestra final incluyó 107 hospitales, con 46.875 episodios de infarto agudo de miocardio (63,5% de todos los episodios asistidos), de los cuales fallecieron el 5,2% (2.451 casos fatales). La insuficiencia cardíaca congestiva coexistió en el 13,5% de la muestra (6.314 episodios de infarto).

Fuente de información

Para cada país participante utilizó un equivalente, en términos de fuente de datos clínico-administrativa al conjunto mínimo de datos al alta hospitalaria (CMBD-AH) español incluyendo todos los episodios de hospitalización financiados de forma pública.

Particularidades del análisis

El resultado hospitalario de este estudio (es decir, la medida indirecta del rendimiento) fue el riesgo ajustado de mortalidad intrahospitalaria en pacientes con infarto agudo de miocardio que permanecieron ingresados hasta 30 días después de su ingreso; por lo tanto, los pacientes ingresados con el código de diagnóstico de ingreso 410* en los países que utilizan la clasificación internacional de enfermedades (CIE-9th-MC) (España y Portugal) e I21* e I22 * en los países que utilizan la CIE-10th-MC (Dinamarca, Eslovenia y Suecia), excluyéndose los ingresos debidos a embarazo, parto o puerperio.

Las variables independientes a nivel de paciente fueron: a) edad, categorizada como 40-49, 50-59, 60-69 y 70-80, utilizando el grupo más joven como grupo de referencia; b) sexo, utilizando el varón como grupo de referencia c) las comorbilidades del paciente, calculadas como puntuación de riesgo de Elixhauser obtenida a partir de la probabilidad de muerte prevista para cada uno de los episodios modelizados con una regresión logística de un solo

nivel, y d) la coexistencia de insuficiencia cardiaca congestiva en el episodio de infarto agudo de miocardio.

Los pacientes con insuficiencia cardíaca congestiva constituían el subgrupo de pacientes de interés, por ser más frágiles que el paciente normal y supuestamente requerir una mayor intensidad de cuidados. Se identificaron los pacientes con insuficiencia cardíaca cuando se encontraron los códigos CIE9 398.91 y 428*, y el código CIE10 I50*, en cualquier diagnóstico secundario registrado dentro del mismo episodio. Las definiciones y los códigos correspondientes se desarrollaron y validaron en el contexto del proyecto ECHO. A nivel de hospital, no se incluyeron variables específicas, salvo los efectos contextuales generales calculado en base a los efectos fijos del modelo. Por último, se incluyó una variable ficticia que identificaba el país de ingreso utilizando Suecia como referencia en las comparaciones.

Tras la estimación del riesgo basal de muerte asociado a las características del paciente y al país de residencia (modelo basal) a través de un modelo de regresión logística convencional de un solo nivel que incluía la edad, el sexo, la puntuación de riesgo de Elixhauser, la coexistencia de insuficiencia cardíaca congestiva y el país de tratamiento, se construyeron dos modelos de regresión lineal mixta para estimar el riesgo de muerte específico del hospital para los pacientes con infarto agudo de miocardio. Para ello, se siguió un proceso de dos etapas. La primera especificación del modelo de regresión lineal mixta incluyó un intercepto aleatorio para el nivel hospitalario en un modelo de regresión logística multinivel de dos niveles, de modo que cada hospital obtuvo su propio intercepto (es decir, el riesgo basal de muerte intrahospitalaria en cada centro). La segunda especificación del modelo de regresión lineal mixta, como extensión de la anterior, incluía una pendiente aleatoria, permitiendo que cada hospital variara su pendiente de riesgo según un grupo específico de pacientes (en nuestro caso, pacientes con insuficiencia cardíaca congestiva, con un mayor riesgo basal de muerte a priori). En la práctica, obtenemos una varianza hospitalaria para los pacientes sin insuficiencia cardíaca y una varianza hospitalaria diferente para los pacientes con esta condición. A partir de estos modelos se estimó el efecto contextual general como los coeficientes de partición de la varianza (CPV) y la Median Odds Ratio (MOR) para ambos modelos, comparándose estos entre ambos y midiendo su consistencia entre hospitales. Por último, para la evaluación de la concordancia (es decir, la concordancia en el rendimiento hospitalario en pacientes con y sin insuficiencia cardíaca), comparando los residuos de ambas partes aleatorias en el modelo ampliado, su intercepto y su pendiente. El nivel de

concordancia entre ambos residuales, especificando la distancia en términos de riesgo por hospital desde la media, se estudió utilizando una medida de acuerdo entre observadores para variables categóricas. Como el número de casos por país variaba sustancialmente, se estimó un índice Kappa ponderado. Según este enfoque, los hospitales de la muestra se clasificaron en tres posibles situaciones: mejor, neutra o peor que la esperada, convirtiéndose esta categorización en el objeto de la medición de la concordancia.

Metodología propia el Artículo 3

El objetivo de este artículo es demostrar cómo explotar de conjuntos de datos de salud del mundo real para el análisis de procesos asistenciales complejos en los que existe continuidad de cuidados de modo que se haga posible evaluar la porción de efectividad asociada a la organización de los cuidados más allá del proveedor de asistencia sanitaria. El objetivo concreto del estudio es describir la vía clínica (o trayectoria asistencial) de los pacientes con sospecha de ictus agudo y la vía específica de los pacientes con ictus isquémico, profundizando un poco más en el uso oportuno del tratamiento fibrinolítico, como hito importante en este proceso asistencial.

Diseño

El caso de uso seleccionado es la implementación del 'Código Ictus', el proceso asistencial de diagnóstico diferencial y tratamiento hiperagudo del ictus isquémico, en el sistema de salud de la Comunidad Autónoma de Aragón, un sistema sanitario público que comprende 9 hospitales de agudos que cubren una población de 1,3 millones de asegurados.

Fuente de información

Se utilizaron todos los datos clínicos y administrativos de los episodios de urgencias hospitalarias e ingreso hospitalario con diagnóstico de sospecha o confirmado de ictus, y los datos sociodemográficos de la base de datos de usuarios del Sistema de Salud de Aragón para todos los pacientes atendidos en el año 2017.

Particularidades del análisis

Los análisis implementados en este artículo se construyen sobre la base de la Metodología de Minería de Procesos (PM 2) mediante la que se procesaron tres conjuntos diferentes de datos sanitarios procedentes de tres sistemas de información distintos para aplicar el descubrimiento de procesos aunque la metodología descrita en el artículo puede aplicarse a cualquier otro proceso asistencial cuya actividad esté debidamente registrada en los datos rutinarios recogidos y mantenidos por los sistemas sanitarios.

La contribución de este trabajo consiste en construir los modelos empíricos de procesos basados en el minado de la actividad asistencial asociada al diagnóstico diferencial y el tratamiento del ictus agudo.

La metodología de minería de procesos tiene 6 tareas principales, divididas en 3 etapas: 1) Etapa de "Inicialización": compuesta por la Tarea 1 "Planificación" y la Tarea 2 "Extracción de datos"; 2) Etapa "Análisis": compuesta por la Tarea 3 "Procesamiento de datos", la Tarea 4 "Minería y análisis" y la Tarea 5 "Evaluación"; y 3) "Mejora y apoyo del proceso", una etapa de una sola tarea (Tarea 6).

Se utilizó un algoritmo de vinculación de episodios (Tarea 3) para determinar los casos con diagnóstico confirmados de ictus. A partir del pseudoidentificador de paciente este algoritmo vincula la secuencia de actividades y corrige potenciales errores de registro para asegurar la visualización de la continuidad de los episodios de Código Ictus. A partir de la colección de registros de actividad se genera el registro histórico de la secuencia de actividades asistenciales que se explotó mediante heurísticos de asociación para el descubrimiento y evaluación del proceso.

Metodología propia del Artículo 4

El objetivo del artículo fue evaluar las diferencias en la mortalidad intrahospitalaria por ictus isquémico agudo entre hospitales de referencia para ictus y aportar pruebas sobre la asociación de esas diferencias con la adopción a lo largo del tiempo de terapias de reperfusión efectivas.

Diseño

Se realizó un estudio observacional longitudinal retrospectivo con datos administrativos de prácticamente todos los episodios hospitalarios del Sistema Nacional de Salud entre el 1 de enero de 2003 y el 31 de diciembre de 2015.

Se seleccionaron en primer lugar aquellos episodios con un diagnóstico de ingreso de ictus isquémico agudo (662 997 episodios) para a continuación definir los hospitales de ictus de referencia como aquellos capaces de realizar trombectomía mecánica endovascular confirmados mediante la identificación de los procedimientos de trombectomía realizados en el último año de análisis (2015) y, de estos, seleccionar sólo los episodios con una estancia inferior a 30 días, restringiendo la observación de casos de muerte intrahospitalaria a aquellos con mayor probabilidad de estar asociados a la intervención médica sobre el ictus (es decir, las muertes en estancias más prolongadas tienen mayor probabilidad de estar asociadas a infecciones nosocomiales o al estado de salud subyacente del paciente). Por último, para reducir la extra heterogeneidad se restringió el estudio a los hospitales de referencia de ictus con al menos 2.000 episodios de ictus isquémico registrados a lo largo del periodo de estudio. La población final del estudio incluyó 196 099 episodios de 37 hospitales de ictus de referencia a nivel nacional.

Fuentes de información

Los episodios se extrajeron del conjunto de datos del proyecto del Atlas de variaciones de la práctica médica, Atlas VPM, que recoge prácticamente todos los episodios hospitalarios dados de alta en los hospitales públicos del Sistema Nacional de Salud (SNS). Además, las características de los hospitales se recopilaron del Catálogo Nacional de Hospitales (Ministerio de Sanidad). Atlas VPM reutiliza datos rutinarios procedentes principalmente de registros electrónicos de ingresos hospitalarios y visitas de atención primaria para la

identificación y selección de casos (p. ej., procedimientos quirúrgicos de interés, enfermedades específicas, eventos de calidad y seguridad), el análisis de atributos clínicos del paciente (p. ej., comorbilidades, cirugía concurrente), el análisis de características administrativas dignas de recoger (p. ej., datos de ingreso y alta, fecha de cirugía) y para la identificación del lugar de residencia (p. ej., atención primaria, sanitaria o área donde reside el paciente). Además, Atlas VPM implica un proceso de vinculación e intercambio con las 17 Consejerías de Sanidad de las comunidades autónomas españolas que comparten una agenda de investigación que evalúa las variaciones injustificadas en la práctica médica y traduce los resultados de la investigación en herramientas de elaboración de perfiles y evaluación comparativa destinadas a facilitar la toma de decisiones clínicas y políticas. En particular, como parte de su metodología de garantía de calidad de los datos, Atlas VPM lleva a cabo comprobaciones de calidad periódicas (una vez al año) para, por ejemplo, reducir los datos incompletos, evitar variaciones en la codificación de las horas extraordinarias, corregir incoherencias semánticas y sintácticas en las variables o reasignar los episodios de ingreso al lugar de residencia si se producen cambios en la geolocalización de las áreas administrativas o los proveedores hospitalarios.

Particularidades del análisis

En este estudio se utilizó un modelo generalizado mixto aditivo para tratar de modelar las asociaciones no lineales, dependientes del paso del tiempo, para estimar la mortalidad intrahospitalaria por ictus a los 30 días del ingreso, considerando el hospital como un efecto aleatorio, y las variables sociodemográficas, de carga de enfermedad y de tratamiento como efectos fijos. Este tipo de modelos permite sustituir los valores de algunas variables por funciones de suavizado que capturan comportamientos no lineales como, por ejemplo, el crecimiento en el volumen de terapias de reperfusión a lo largo del tiempo a partir de la disponibilidad generalizada de la técnica, del entrenamiento de los profesionales sanitarios en su uso o de la publicación de una estrategia o recomendación. Las funciones de suavizado se estimaron de forma no paramétrica aproximando medidas de dispersión a funciones planares finas, o a funciones polinómicas cúbicas cíclicas (en la covariable “meses”), seleccionando parámetros de ajuste óptimos mediante el método de validación cruzada para tener en cuenta la estacionalidad. Por último, se modelizó los efectos de los hospitales como efectos aleatorios independientes, añadiendo un parámetro de interacción entra la terapia

de reperfusión (a nivel de paciente) y el hospital de tratamiento. La modelización de los hospitales como efectos aleatorios nos permitió tener en cuenta la agrupación natural de los episodios de ictus isquémico dentro de cada hospital, midiendo así las diferencias entre ellos, mientras que la consideración de la interacción con la terapia de reperfusión permitió también modelizar las diferencias potenciales entre ambos grupos de pacientes, permitiéndonos diferenciar los efectos del hospital, en cada grupo de pacientes y a lo largo del tiempo.

Consideraciones éticas, financiación y conflicto de interés

Consideraciones éticas

No existen conflictos éticos relacionados con la realización de esta Tesis. La línea de investigación ha respetado la confidencialidad y la privacidad de la información en todos los trabajos. Todos los análisis utilizaron datos a nivel individual pseudonimizados, cuyo acceso se concedió a partir de la aprobación de un proyecto de investigación por los Comités de Ética de la Investigación correspondientes (i.e., CEICA en Aragón) dentro de su alcance para cumplir con los objetivos específicos de investigación reseñados en cada proyecto. En el caso del trabajo descrito en el artículo 1, el acceso a los datos requirió la firma de convenios internacionales bilaterales de transferencia y tratamiento/procesamiento de datos entre los países participantes y el Instituto Aragonés de Ciencias de la Salud dentro del alcance de los proyectos ECHO-Health, BRIDGE-Health y de la Acción Conjunta Europea InfAct-JA 'Joint Action on Health Information', continuación del anterior.

Los datos de hospitalizaciones públicamente financiadas por en el Sistema Nacional de Salud del proyecto Atlas VPM han sido cedidos por las comunidades autónomas participantes en el consorcio Atlas VPM mediante convenios, resoluciones o acuerdos con el Instituto Aragonés de Ciencias de la Salud para investigación.

Todos los investigadores participantes en estos estudios están sujetos a la Política de Seguridad de la Información a nivel de sistema del Instituto Aragonés de Ciencias de la Salud (IACS) y a acuerdos de confidencialidad y privacidad de los datos con sus instituciones.

Financiación

La línea de investigación dentro de la que se engloba la presente Tesis Doctoral ha sido financiada desde diversos ámbitos. En primer lugar, el trabajo reflejado en los artículos 1 y 2 se enmarca dentro del proyecto europeo "BRIDGE-Health - BRidging Information and Data Generation for Evidence-based Health policy and research" ('664691' in the European Union's Health Programme 2014-2020), en el que se continúan los trabajos iniciados en el proyecto europeo ECHO-Health (Grant Agreement number 242189 from FP7-HEALTH) en los que el doctorando participó como investigador. El trabajo reseñado en el artículo 3 fue parcialmente financiado por el proyecto ProCOMP (LMP 183_18) del Gobierno de Aragón, a través del Programa Operativo del Fondo Europeo de Desarrollo Regional (FEDER), "Construyendo

Europa desde Aragón"; y por el Proyecto ICTUSnet (SOE2/P1/E0623) del Programa Interreg SUDOE que apoya el desarrollo de las regiones del suroeste de Europa mediante la financiación de proyectos transnacionales a través del FEDER en los cuales el doctorando participó como investigador y coordinador de tarea. En último lugar, el trabajo del artículo 4 se enmarca dentro de los trabajos científicos desarrollados por el consorcio Atlas VPM siendo parcialmente financiado por el proyecto CONCEPT-STROKE - 'Effectiveness and efficiency of acute ischemic stroke care pathways in five Spanish Regions' (PI19/00154); enmarcado dentro de la participación del grupo de investigación en las redes REDISSEC (Red de Investigación en Servicios de Salud en Enfermedades Crónicas (RD16/0001/0007)); y su continuación en la red RICAPPS (Red de Investigación en cronicidad, atención primaria y promoción de la salud (RD21/0016/0023)) y en colaboración con el Departamento de Métodos Cuantitativos en Economía y Gestión de la Universidad de Las Palmas de Gran Canaria (ULPGC).

Por último, el doctorando es miembro del "Grupo de Ciencia de Datos para la Investigación en Políticas y Servicios de Salud" del Instituto Aragonés de Ciencias de la Salud (IACS) desde el año 2014. Este grupo está reconocido como grupo consolidado del Gobierno de Aragón y forma parte de la Red de Investigación en Cronicidad, Atención Primaria y Promoción de la Salud (RICAPPS) del Instituto de Salud Carlos III (ISCIII).

Conflicto de interés

Ni el doctorando ni el director de la Tesis presentaron conflictos de interés a la hora de desarrollar la línea de investigación. Tampoco el resto de autores de los trabajos presentados en la Tesis Doctoral tuvieron ningún conflicto de interés, tal y como se explica en los distintos artículos.

Discusión

Fortalezas y debilidades de los resultados obtenidos

En los artículos incluidos se pueden observar las diferencias en términos de variación y concordancia en las medidas de resultados a nivel de hospital al incluir estas consideraciones en los distintos modelos de evaluación. También queda patente la capacidad de estos estudios de generar conocimiento relevante para la toma de decisiones en la gestión de los servicios sanitarios.

Una crítica general que se hace a la comparación de la actuación de los servicios sanitarios a partir del estudio de la variación atribuible al proveedor (oferta) es que aunque es capaz de discriminar las diferencias entre proveedores en términos de resultados y, de este modo, establecer estándares de referencia de resultados deseables o seleccionar proveedores excelentes en su funcionamiento habitual, no puede ofrecer un diagnóstico preciso de las circunstancias que producen este comportamiento. El avance en el modelado de la interacción entre paciente y sistema más allá de los ajustes por carga de enfermedad o de la caracterización simple de los recursos asistenciales de los proveedores comparados para profundizar en la organización de los cuidados y su continuidad entre proveedores y niveles asistenciales y la evolución de la práctica clínica a lo largo del tiempo con la incorporación de nuevas recomendaciones y tecnologías sanitarias configura el marco efectivo para la implantación de un modelo de mejora continua de la calidad de los servicios sanitarios a partir de la reutilización de los datos de salud generados por el propio sistema acercándose al concepto de 'Sistema Sanitario Inteligente' (Learning Health System).

Los métodos y técnicas implementadas en los estudios incluidos en la tesis nos ofrecen resultados que demuestran que estas perspectivas de análisis son relevantes para la monitorización y la evaluación comparativa de los resultados de los servicios sanitarios. Es decir, existen diferencias significativas y sistemáticas en el desempeño de los proveedores que se justifican en la adaptación de su respuesta o funcionamiento frente a grupos específicos de pacientes, respecto a características de entorno que se traducen en cambios organizativos o respecto a la adopción de nuevos medios y tecnologías en el entorno sanitario.

Sin embargo, pese a la consistencia en los resultados, cada estudio presenta sus particularidades tanto a nivel de alcance como a nivel de las recomendaciones que se derivan de los desafíos para la implementación de cada metodología y sus potenciales aplicaciones.

Los métodos y técnicas utilizados para la evaluación de resultados de los servicios sanitarios presentados en esta tesis son completamente dependientes de la disponibilidad, granularidad y calidad de los datos lo que supone una limitación evidente respecto a aquello que se puede observar como del grado de consistencia o de detalle con el que podemos medirlo.

A pesar de la riqueza de datos de salud derivada de la digitalización de los Sistemas Sanitarios, el registro, diseño, estructura y utilización de los sistemas de información sanitaria está completamente determinado por el uso primario de los mismos para la gestión clínica y sanitaria y la planificación de la provisión de servicios y no pensando en su explotación a los efectos de monitorización y control de la calidad o incluso para investigación. A esto se une la heterogeneidad de la disponibilidad de datos y del acceso a los mismos en toda Europa que obedece al crecimiento orgánico -en función de necesidad- de estos sistemas de información, y a la ordenación de los procedimientos de acceso a datos de salud a nivel individual derivado de la interpretación de las normativas europeas y nacionales guiada por el grado de aversión al riesgo de cada institución. Con esto y con todo, existe la posibilidad de encontrar un conjunto mínimo común de datos que permita la evaluación robusta del rendimiento de los servicios sanitarios en unidades de análisis significativas tanto a nivel geográfico como por proveedores.

En términos prácticos este punto de partida requiere del abordaje y resolución de los problemas de interoperabilidad entre los distintos sistemas para asegurar la capacidad de reutilización de los datos sanitarios para investigación. Una forma de abordar estos problemas de interoperabilidad es caracterizándolos y categorizándolos según el marco europeo de interoperabilidad en sus niveles de interacción entre sistemas, a nivel legal, organizativo, semántico y sintáctico y tecnológico. A nivel legal, los Sistemas Sanitarios deben reconocerse mutuamente y establecer un interés común en la colaboración, la monitorización y la evaluación comparada de los resultados de sus servicios de salud. También requieren, a este

nivel, cumplir con los requerimientos legales en términos de aseguramiento de la confidencialidad, privacidad y seguridad en la gestión de datos que imponen tanto la normativa europea como sus marcos regulatorios nacionales. En este sentido, la aproximación federada a los análisis de datos de salud viene a facilitar el respeto de todos los agentes implicados a sus normas y políticas de seguridad de la información habilitándolos para participar en cualquier proyecto basado en el uso secundario de datos de salud. A nivel organizativo, existe la necesidad de establecer una sistemática de trabajo entre las instituciones o sistemas que participen en estos trabajos, diferenciando roles y responsabilidades entre los participantes respecto a la gestión y análisis de sus datos, y a la interpretación de los resultados de las comparaciones entre servicios. A efectos prácticos esto supone una inversión importante en establecer una gobernanza fuerte de las redes de investigación en evaluación de servicios a partir de datos de vida real y/o de las infraestructuras de datos sanitarios para investigación. Esto también requiere invertir en la capacitación de los investigadores en la implementación de estos métodos y prácticas de análisis para facilitar su diseminación y la traslación de los resultados al Sistema Sanitario. A nivel sintáctico (estructura) y semántico, la interoperabilidad de datos requiere el diseño y la generación de modelos comunes de datos que sostengan el desarrollo de indicadores de rendimiento que aborden los diferentes ámbitos de la evaluación de los Sistemas Sanitarios -utilización, equidad, calidad, seguridad, efectividad y eficiencia- además del establecimiento de rutinas transparentes y trazables de aseguramiento y gestión de la calidad de los datos. Finalmente, a nivel tecnológico se requiere la capacidad de comunicación de los datos o los resultados de los análisis entre los sistemas de información y la capacidad técnica y tecnológica de implementación y ejecución de los métodos de análisis expuestos de forma reproducible.

Aportaciones del doctorando

Medición y comparación de resultados de servicios sanitarios

La línea de investigación de esta Tesis Doctoral muestra que, en la medición y comparación de resultados de servicios sanitarios deben entrar en consideración no sólo los aspectos relacionados con el ajuste de la carga de enfermedad de los pacientes (o de la población cubierta por estos servicios), sino también relacionados con a) la interacción entre

proveedores y grupos específicos de pacientes con riesgos basales diferenciales (adaptación de la oferta a la demanda), b) la configuración de los procesos asistenciales como secuencias de interacción entre pacientes y los servicios sanitarios (i.e., vías clínicas y procesos administrativos), y c) la dinámica de estas interacciones a lo largo del tiempo modelando la evolución de las circunstancias y los medios de provisión de asistencia sanitaria (i.e., modificaciones estructurales, adopción o desinversión en tecnologías sanitarias, y/o cambios en la indicación o en las guías de intervención y tratamiento frente a una patología - incluida la introducción de políticas, estrategias o programas de salud).

En los estudios presentados se implementan y validan alternativas técnicas y metodológicas para facilitar esta consideración de forma incremental hasta conseguir adoptar una perspectiva de evaluación holística del Sistema Sanitario a partir de la comparación de resultados de sus servicios en la atención de pacientes con condiciones específicas. El alcance de la aplicación de los métodos validados es generalizable a cualquier otra condición susceptible de asistencia sanitaria, pudiendo ser adaptada para el estudio evaluativo de cualquier tipo de Sistema, o para la comparación internacional entre Sistemas Sanitarios.

A nivel particular, el trabajo descrito en el artículo 1 muestra que es posible establecer una infraestructura de datos sanitarios para investigación a nivel europeo a partir de la cual poder evaluar y comparar el funcionamiento de los servicios sanitarios; y apunta a la necesidad de desarrollar infraestructuras en red basadas en aproximaciones federadas de análisis que cumplan con los requerimientos legales de confidencialidad, privacidad y seguridad en la gestión de los datos. Finalmente, ofreciendo recomendaciones para la implementación de este tipo de infraestructuras federadas de análisis de datos de salud. Por otro lado, los artículos 2, 3 y 4 muestran la implementación en diferentes casos de uso de los métodos y técnicas exploradas en esta línea de investigación. En concreto, el artículo 2 se centra en las diferencias atribuibles al hospital en la mortalidad intrahospitalaria en pacientes ingresados después de un infarto agudo de miocardio; mientras que los artículos 3 y 4 se enfocan en las diferencias atribuibles al hospital en la mortalidad intrahospitalaria en pacientes ingresados después de un ictus pero caracterizando esta atribución a diferencias en el procesos asistencial (i.e., trayectoria del paciente) y en la evolución de la provisión de los servicios recogiendo las dinámicas de cambio dentro del Sistema de Salud.

Implicaciones de los resultados y líneas de investigación futuras

La validación de los métodos y técnicas de análisis descritas en estos trabajos abren la puerta a poder informar de forma sistemática la toma de decisiones en la gestión de los servicios sanitarios desde nuevas perspectivas destacando puntos potenciales de actuación para la mejora de la calidad de los cuidados.

Los resultados aportados en esta Tesis Doctoral se suman a la evidencia ya existente en la medición de las variaciones en la efectividad, calidad y seguridad de los cuidados sanitarios a nivel de proveedor y en la evaluación comparativa de los servicios sanitarios.

El alcance de estos trabajos deja aún algunas cuestiones que deben ser abordadas. En primer lugar, cuestiones relativas al diseño, organización y gestión e implementación de estudios basados en el análisis federado de datos como a) la selección y adecuación de los métodos de análisis estadístico a la fragmentación de la información en múltiples sitios (y el acceso parcial a la misma) y sus implicaciones en términos de precisión y consistencia en las medidas, b) la capacitación de los participantes en estos estudios, c) las oportunidades de contextualización para la interpretación de los resultados de comparación (o monitorización) en los entornos y circunstancias de cada sistema sanitario, etc. En segundo lugar, la introducción de la perspectiva de estudio de la interacción entre paciente y proveedor a través de la exploración y la explotación de los procesos asistenciales abre la puerta a un mundo de posibilidades para la identificación de los puntos de acción dentro de los propios servicios y a nivel de sistema de la evaluación del mismo desde la perspectiva de continuidad de cuidados a lo largo de toda la vida de la población servida por ese sistema. Por primera vez, la digitalización, la integración mediante la resolución de los desafíos de interoperabilidad y el análisis estadístico de la información de salud nos habilita a los investigadores en políticas y servicios sanitarios a seguir la experiencia vital de salud y enfermedad de una población "*desde la cuna a la tumba*", y en el transcurso de esa experiencia vital evaluar y amplificar el papel que juegan las políticas y servicios sanitarios como determinantes de salud de la población. En tercer lugar, esta perspectiva temporal nos obliga necesariamente a integrar las dinámicas de cambios en la provisión de los servicios sanitarios tanto a nivel estructural como a nivel funcional con la adopción de nuevas prácticas médicas basadas en cambios culturales, organizativos o en la introducción de nuevas tecnologías o en la desinversión en tecnologías obsoletas. Por último,

la incorporación rutinaria al instrumental metodológico en la evaluación comparativa de estos métodos y técnicas de análisis nos capacita para dar el siguiente paso en la medición del impacto de los servicios sanitarios en la salud de la población.

Los resultados del trabajo realizado en la línea de investigación descrita en esta Tesis Doctoral se han implementado y han continuado desarrollándose en el contexto de otros proyectos de investigación evaluativa nacionales, como AtlasVPM o los proyectos coordinados CONCEPT, e internacionales, como los proyectos europeos PHIRI, EHDS2 pilot, BY-COVID, y en los estudios de investigación de la colaboración ICCONIC.

Finalmente, es de destacar que las recomendaciones para la implementación de infraestructuras de datos para investigación y la aproximación federada a la evaluación comparativa de los servicios sanitarios a partir del uso secundario de datos de salud han servido de base a la participación del Instituto Aragonés de Ciencias de la Salud en el diseño del Espacio Europeo de Datos de Salud (EHDS) a través de su participación en las Acciones Conjuntas Europeas InfAct-JA y TEHDAS-JA, y al liderazgo en la configuración de una agenda estratégica de uso secundario de datos para la investigación en salud en el proyecto HealthyCloud.

Conclusiones

1. La medición y comparación de resultados de los servicios sanitarios se basa en la reutilización de los datos de salud recogidos por los Sistemas de Salud. En el contexto europeo, el acceso a estos datos requiere el despliegue de una infraestructura federada que integre todas las fuentes de información sanitaria y de salud de la población de cada país y que resuelva los desafíos de interoperabilidad legal, organizativa, semántica y tecnológica mediante el establecimiento de modelos comunes de datos, procesos de gobernanza centrados en el aseguramiento de la calidad de los datos y la distribución de programas de código abierto que permitan ejecutar de una forma transparente, trazable y reproducible los análisis estadísticos que se requieran para cada estudio.
2. En la evaluación comparativa de los servicios sanitarios por proveedor debería ser un requisito considerar la heterogeneidad de los pacientes más allá del ajuste por carga de enfermedad incluyendo términos de interacción entre grupos específicos de pacientes y el servicio (proveedor de servicios).
3. La aplicación de la minería de procesos permite evaluar y comparar cómo repercute la organización de los cuidados sanitarios, incluida su continuidad entre proveedores y niveles asistenciales, en los resultados sanitarios y, por tanto, ofrecer recomendaciones precisas sobre cómo mejorar la calidad de la atención sanitaria para mejorar los resultados en salud.
4. Es imprescindible considerar la propia evolución del Sistema Sanitario a lo largo del tiempo subsecuente a cambios estructurales, organizativos, culturales o cambios en la práctica médica asociados a la adopción de tecnologías sanitarias en la evaluación de los resultados de los servicios sanitarios. Especialmente en aquellos procesos de alta complejidad que requieran de la coordinación o integración de cuidados entre proveedores y niveles asistenciales.

Referencias

Artículo 1

1. European Commission. Communication on effective, accessible and resilient health systems. COM (2014) 215 final. Available at: https://ec.europa.eu/health//sites/health/files/systems_performance_assessment/docs/com2014_215_final_en.pdf [Accessed 3 Aug 2017].

2. Institute for Clinical Evaluative Sciences (ICES) Available from: http://www.ices.on.ca/Publications/Atlases-and-Reports [Accessed 3 Aug 2017].

3. Agency for Health Research and Quality (ARQH). 2010 Available at: http://www.qualityindicators.ahrq.gov [Accessed 3 Aug 2017].

4. NHS Atlas of Variations in Healthcare. Available at: https://www.england.nhs.uk/rightcare/products/atlas/ [Accessed 3 Aug 2017].

5. Dutch National Atlas of Public Health. Available at: https://www.volksgezondheidenzorg.info/onderwerp/atlas-vzinfo/inleiding#!node-kaarten [Accessed 3 Aug 2017].

6. New Zealand Atlas of Healthcare Variation. Available at: http://www.hqsc.govt.nz/our-programmes/health-quality-evaluation/projects/atlas-ofhealthcare-variation/ [Accessed 3 Aug 2017].

7. OECD/EU (2016), Health at a glance: Europe 2016: state of health in the EU cycle, OECD Publishing, Paris. http://dx.doi.org/10.1787/9789264265592-en [Accessed 3 Aug 2017].

8. Reviews of National Healthcare Quality Available at: http://www.oecd.org/els/health-systems/health-care-quality-reviews.htm [Accessed 3 Aug 2017].

9. Expert Group on Health Systems Performance Assessment (HSPA): Available at: https://ec.europa.eu/health/systems_performance_assessment/policy/expert_group_en [Accessed 3 Aug 2017].

10. E. Bernal-Delgado, on behalf of ECHO Consortium; T. Christiansen, on behalf of ECHO Consortium; K. Bloor, on behalf of ECHO Consortium; C. Mateus, on behalf of ECHO Consortium; A.M. Yazbeck, on behalf of ECHO Consortium, J. Munck, on behalf of ECHO Consortium, J. Bremner, on behalf of ECHO Consortium; ECHO: health care performance assessment in several European health systems. Eur J Pub Health 2015; 25(suppl_1): 3-7. doi: https://doi.org/10.1093/eurpub/cku219

11. Thygesen CT, Baixauli-Pérez C, Librero-López J, Martínez-Lizaga N, Ridao-López M, Bernal-Delgado E, on behalf of the ECHO Consortium. Comparing variation across European countries:

building geographical areas to provide sounder estimates. Eur J Pub Health. 25(suppl_1):8–14. https://doi.org/101093/eurpub/cku229

12. Bergdahl M, Ehling M, Elvers E, Földesi E, Körner T, Kron A, et al. Handbook on data quality assessment methods and tools. Wiesbaden: European Comission; 2007.

Artículo 2

1. Institute for Clinical Evaluative Sciences (ICES). Canada. http://www.ices.on.ca/Publications/Atlasesand-Reports (10 January 2019, date last accessed)

2. Atlas of Variations. Public Health England. https://fingertips.phe.org.uk/profile/atlas-of-variation (10 January 2019, date last accessed)

3. Agency for Health Research and Quality (ARQH). US Department of Health & Human Services, USA

4. Health Quality and Safety Commission, New Zealand. http://www.hqsc.govt.nz/our-programmes/health-quality-evaluation/projects/atlas-of-healthcare-variation/ (10 January 2019, date last accessed)

5. National Institute for Public Health and the Environment, The Netherlands. https://www.volksgezondheidenzorg.info/onderwerp/atlas-vzinfo/inleiding (10 January 2019, date last accessed)

6. The Atlas of Variation in the Spanish National Health Service, Instituto Aragonés de Ciencias de la Salud, IIS Aragón, Spain Available at: www.atlasvpm.org (10 January 2019, date last accessed)

7. OECD Health Care Quality Indicators Project, Paris http://www.oecd.org/els/health-systems/healthcare-quality-indicators.htm (10 January 2019, date last accessed)

8. The Dartmouth Atlas of Healthcare, The Dartmouth Institute, USA. Available at: http://www.dartmouthatlas.org (10 January 2019, date last accessed)

9. Ohlsson H, Librero J, Sundquist J, Sundquist K, Merlo J: Performance evaluations and league tables: do they capture variation between organizational units? An analysis of 5 Swedish pharmacological performance indicators. Med Care 2011, 49(3):327–331. https://doi.org/10.1097/MLR.0b013e31820325c5 PMID: 21263360

10. European project on healthcare outcomes, performance and efficiency. http://www.eurohope.info/ (10 January 2019, date last accessed)

11. Bernal-Delgado E, Christiansen T, Bloor K, et al. on behalf of ECHO Consortium ECHO: Health care performance assessment in several European countries. European Journal of Public Health, Volume 25, Issue suppl_1, 1 February 2015, Pages 3–7, https://doi.org/10.1093/eurpub/cku219

12. Merlo J, Ostergren PO, Broms K, Bjorck-Linne A, Liedholm H. Survival after initial hospitalisation for heart failure: a multilevel analysis of patients in Swedish acute care hospitals. J Epidemiol Community Health. 2001; 55(5):323–9. https://doi.org/10.1136/jech.55.5.323 PMID: 11297650

13. Merlo J, Gerdtham UG, Eckerlund I, et al. Hospital level of care and neonatal mortality in low- and highrisk deliveries: reassessing the question in Sweden by multilevel analysis. Med Care. 2005; 43(11):1092–100. https://doi.org/10.1097/01.mlr.0000182484.14608.b9 PMID: 16224302

14. Ghith N, Frolich A, Merlo J. The role of the clinical departments for understanding patient heterogeneity in one-year mortality after a diagnosis of heart failure: A multilevel analysis of individual heterogeneity for profiling provider outcomes. PLoS One. 2017; 12(12):e0189050. https://doi.org/10.1371/journal.pone.0189050 PMID: 29211785

15. Ghith N, Wagner P, Frolich A, Merlo J. Short Term Survival after Admission for Heart Failure in Sweden: Applying Multilevel Analyses of Discriminatory Accuracy to Evaluate Institutional Performance. PLoS One. 2016; 11(2):e0148187. https://doi.org/10.1371/journal.pone.0148187 PMID: 26840122

16. Ash AS, Fienberg SE, Louis TA, Normand SLT, Stukel TA, Utts J. Statistical issues in assessing hospital performance. Commissioned by the Committee of Presidents of Statistical Societies. The COPSSCMS White Paper Committee: 2011

17. Mohammed MA, Manktelow BN, Hofer TP. Comparison of four methods for deriving hospital standardised mortality ratios from a single hierarchical logistic regression model Stat Methods Med Res. 2016 Apr; 25(2):706–15.

18. Goldstein H, Spiegelhalter D. League Tables and Their Limitations: Statistical Issues in Comparisons of Institutional Performance. Journal of the Royal Statistical Society, SocA. 1996; 159:385–443.

19. Austin PC, Tu JV, Alter DA. Comparing hierarchical modelling with traditional logistic regression analysis among patients hospitalized with acute myocardial infarction: should we be analyzing cardiovascular outcomes data differently? Am Heart J. 2003; 145(1):27–35. https://doi.org/10.1067/mhj.2003.23 PMID: 12514651

20. Marang-van de Mheen PJ, Shojania KG. Simpson's paradox: how performance measurement can fail even with perfect risk adjustment. BMJ Qual Saf 2014; 23:701–705. https://doi.org/10.1136/bmjqs-2014-003358 PMID: 25118292

21. Sanagou M, Wolfe R, Forbes A, Reid CM. Hospital-level associations with 30-day patient mortality after cardiac surgery: a tutorial on the application and interpretation of marginal and multilevel logistic regression BMC Medical Research Methodology 2012, 12:28 https://doi.org/10.1186/1471-2288-12-12

22. Pouw ME, Peelen LM, Lingsma HF, et al. (2013) Hospital Standardized Mortality Ratio: Consequences of Adjusting Hospital Mortality with Indirect Standardization. PLoS ONE 8(4): e59160. https://doi.org/10.1371/journal.pone.0059160 PMID: 23593133

23. Merlo J, Yang M, Chaix B, Lynch J, Rastam L. A brief conceptual tutorial on multilevel analysis in social epidemiology: investigating contextual phenomena in different groups of people. J Epidemiol Community Health 2005, 59(9):729–736 https://doi.org/10.1136/jech.2004.023929 PMID: 16100308

24. Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data. Med Care 1998; 36:8–27 https://doi.org/10.1097/00005650-199801000-00004 PMID: 9431328

25. Elixhauser and Quan's ICD-9-CM and ICD-10 Coding Algorithms for Elixhauser Comorbidity Index. Med Care 2005 Nov; 43(11): 1130–9. https://doi.org/10.1097/01.mlr.0000182534.19832.83 PMID:16224307

26. European Collaboration for Healthcare Optimization (ECHO) www.echo-health.eu. Zaragoza (Spain): Instituto Aragonés de Ciencias de la Salud—Instituto Investigación Sanitaria Aragón; c2011. Estupiñán F, Baixauli C, Bernal-Delgado E on behalf of the ECHO consortium. Handbook on methodology: ECHO information system quality report; 2014 Apr 27 [accessed: 10 January 2019]; Available from: http://www.echo-health.eu/handbook/infrastructure.html

27. Goldstein H, Browne W, Rasbash J. Partitioning variation in generalised linear multilevel models. Understanding Statistics 2002, 1:223–232.

28. Merlo J, Chaix B, Ohlsson H, et al. A brief conceptual tutorial of multilevel analysis in social epidemiology: using measures of clustering in multilevel logistic regression to investigate contextual phenomena. J Epidemiol Community Health 2006, 60(4):290–297. https://doi.org/10.1136/jech.2004.029454 PMID:16537344

29. Larsen K, Merlo J. Appropriate assessment of neighbourhood effects on individual health: integrating random and fixed effects in multilevel logistic regressions. Am J Epidemiol 2005, 161(1):81–88. https://doi.org/10.1093/aje/kwi017 PMID: 15615918

30. Larsen K, Petersen JH, Budtz-Jorgensen E, Endahl L. Interpreting parameters in the logistic regression model with random effects. Biometrics 2000, 56(3):900–914.

31. Browne WJ MCMC Estimating in MLwiN v2.29. Center for Multilevel Modelling, University of Bristol 2013.

32. Leckie G, Charlton C. runmlwin—A Program to Run the MLwiN Multilevel Modelling Software from within Stata. Journal of Statistical Software, (2013); 52 (11):1–40]

33. Cohen J. A coefficient of agreement for nominal scales". Educational and Psychological Measurement 1960; 20 (1): 37–46. https://doi.org/10.1177/001316446002000104

34. https://www.socialstyrelsen.se/register/halsodataregister/patientregistret/inenglish

35. Ludvigsson JF, Andersson E, Ekbom A, et al. External review and validation of the Swedish national inpatient register. BMC Public Health. 2011; 11:450. https://doi.org/10.1186/1471-2458-11-450 PMID: 21658213

36. Merlo J, Chaix B, Yang M, Lynch J, Rastam L. A brief conceptual tutorial of multilevel analysis in social epidemiology: linking the statistical concept of clustering to the idea of contextual phenomenon. J Epidemiol Community Health 2005, 59(6):443–44 https://doi.org/10.1136/jech.2004.023473 PMID: 15911637

37. Merlo J. Multilevel analytical approaches in social epidemiology: measures of health variation compared with traditional measures of association. J Epidemiol Community Health 2003, 57(8):550–552. https://doi.org/10.1136/jech.57.8.550 PMID: 12883048

38. Merlo J. Invited commentary: a multilevel analysis of the individual heterogeneity-a fundamental critique of the current probabilistic risk factor epidemiology. American Journal of Epidemiology 2014, 180(2):208–212. https://doi.org/10.1093/aje/kwu108 PMID: 24925064

39. Merlo J, Ohlsson H, Lynch KF, Chaix B, Subramanian SV. Individual and collective bodies: using measures of variance and association in contextual epidemiology. J Epidemiol Community Health 2009, 63(12):1043–1048. https://doi.org/10.1136/jech.2009.088310 PMID: 19666637

Artículo 3

1. H. Campbell, R. Hotchkiss, N. Bradshaw, and M. Porteous, "Integrated care pathways.,"BMJ (Clinical research ed.), vol. 316, no. 7125pp. 133–7, Jan. 1998, ISSN: 0959-8138. [Online]. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2665398.

2. T. Rotter, L. Kinsman, E. L. James, A. Machotta, H. Gothe, J. WillisP. Snow, and J. Kugler, "Clinical pathways: effects on professional practice, patient outcomes, length of stay and hospital costs," Cochrane Database of Systematic Reviews, no. 3, Mar. 2010, ISSN: 14651858. DOI: 10.1002/14651858.CD006632.pub2.

3. D. J. Zand, K. M. Brown, U. Lichter-Konecki, J. K. Campbell, V. Salehi, and J. M. Chamberlain, "Effectiveness of a Clinical Pathway for the Emergency Treatment of Patients With Inborn Errors of Metabolism," PEDIATRICS, vol. 122, no. 6, pp. 1191–1195, Dec. 2008, ISSN: 0031-4005. DOI: 10.1542/peds.2008-0205.

4. J. Lenkowicz, R. Gatta, C. Masciocchi, C. Casà, F. Cellini, A. Damiani, N. Dinapoli, and V. Valentini, "Assessing the conformity to clinical guidelines in oncology: An example for the multidisciplinary management of locally advanced colorectal cancer treatment Jacopo, "Management Decision, vol. 56, no. 10, pp. 2172–2186, Oct. 2018, ISSN: 0025-1747. DOI: 10.1108/MD-09-2017-0906.

5. A. Partington, M. Wynn, S. Suriadi, C. Ouyang, and J. Karnon, "Pro-cess Mining for Clinical Processes: A Comparative Analysis of Four Australian Hospitals,"ACM Transactions on Management Information Systems, vol. 5, no. 4, pp. 1–18, Jan. 2015, ISSN: 2158656X. DOI: 10.1145/2629446.
6. W. M. P. van der Aalst, Process Mining: Data Science in Action, Second Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2016, ISBN: 978-3-662-49851-4. DOI: 10.1007/978-3-662-49851-4.
7. L. P. Garrison, P. J. Neumann, P. Erickson, D. Marshall, and C. D. Mullins, "Using Real-World Data for Coverage and Payment Decisions: The ISPOR Real-World Data Task Force Report, "Value in Health, vol. 10, no. 5, pp. 326–335, Sep. 2007, ISSN: 10983015. DOI: 10.1111/j.1524-4733.2007.00186.x.
8. M. L. van Eck, X. Lu, S. J. J. Leemans, and W. M. P. van der Aalst, "PM²: A Process Mining Project Methodology," in CAiSE 2015: Advanced Information Systems Engineering, Springer, Cham, 2015pp. 297–313. DOI: 10.1007/978-3-319-19069-3_19.
9. W. van der Aalst, T. Weijters, and L. Maruster, "Workflow mining: discovering process models from event logs," IEEE Transactions on Knowledge and Data Engineering, vol. 16, no. 9, pp. 1128–1142, Sep. 2004, ISSN: 1041-4347. DOI: 10.1109/TKDE.2004.47.
10. A. Weijters, W. M. P. Aalst van der, and A. K. Alves De Medeiros, Process mining with the HeuristicsMiner algorithm, ser. BETA publi-catie : working papers. Technische Universiteit Eindhoven, 2006, ISBN: 978-90-386-0813-6.
11. C. W. Günther and W. M. Van Der Aalst, "Fuzzy mining -Adaptive process simplification based on multi-perspective metrics," in BPM 2007: Business Process Management, vol. 4714 LNCS, Springer, Berlin, Heidelberg, 2007, pp. 328–343, ISBN: 9783540751823. DOI: 10.1007/978-3-540-75183-0_24.
12. S. J. J. Leemans, D. Fahland, and W. M. P. van der Aalst, "Discovering Block-Structured Process Models from Event Logs -A Constructive Approach," in PETRI NETS 2013: Application and Theory of Petri Nets and Concurrency, Springer, Berlin, Heidelberg, 2013, pp. 311–329. DOI: 10.1007/978-3-642-38697-8_17.
13. J. Munoz-Gama, Conformance Checking and Diagnosis in Process Mining, ser. Lecture Notes in Business Information Processing. Cham: Springer International Publishing, 2016, vol. 270, ISBN: 978-3-319-49450-0. DOI: 10.1007/978-3-319-49451-7.
14. A. Rozinat and W. van der Aalst, "Conformance checking of processes based on monitoring real behaviour," Information Systems, vol. 33, no. 1pp. 64–95, Mar. 2008, ISSN: 03064379. DOI: 10.1016/j.is.2007.07.001.
15. B. F. van Dongen, A. K. A. de Medeiros, H. M. W. VerbeekA. J. M. M. Weijters, and W. M. P. van der Aalst, "The ProM Framework: A New Era in Process Mining Tool Support,"in ICATPN 2005:

Applications and Theory of Petri Nets 2005, Springer, Berlin, Heidelberg, 2005, pp. 444–454, ISBN: 978-3-540-26301-2, 978-3-540-31559-9. DOI: 10.1007/11494744_25.

16. C. W. Günther and A. Rozinat, "Disco: Discover Your Processes," in BPM 2012: Proceedings of the Demonstration Track of the 10th International Conference on Business Process Management, CEUR Workshop Proceedings (CEUR-WS.org), 2012, pp. 40–44.

17. G. Janssenswillen, B. Depaire, M. Swennen, M. Jans, and K. Vanhoof, "bupaR: Enabling reproducible business process analysis," Knowledge-Based Systems, vol. 163, pp. 927–930, Jan. 2019, ISSN: 09507051. DOI: 10.1016/j.knosys.2018.10.018.

18. W. van der Aalst, "Process mining: discovering and improving Spaghetti and Lasagna processes," in CIDM 2011: IEEE Symposium on Computational Intelligence and Data Mining, IEEE, Apr. 2011pp. 1–7, ISBN: 978-1-4244-9927-4. DOI: 10.1109/CIDM.2011.6129461.

19. L. Maruster and R. Jorna, "From Data to Knowledge: A Method for Modeling Hospital Logistic Processes," IEEE Transactions on Information Technology in Biomedicine, vol. 9, no. 2, pp. 248–255, Jun. 2005, ISSN: 1089-7771. DOI: 10.1109/TITB.2005.847194.

20. W. Yang and Q. Su, "Process mining for clinical pathway: Literature review and future directions," in ICSSSM 2014: 11th International Conference on Service Systems and Service Management, IEEE, Jun. 2014, pp. 1–5, ISBN: 978-1-4799-3134-7. DOI: 10.1109/ICSSSM.2014.6943412.

21. E. Rojas, J. Munoz-Gama, M. Sepúlveda, and D. Capurro, "Process mining in healthcare: A literature review," Journal of Biomedical Informatics, vol. 61, pp. 224–236, Jun. 2016, ISSN: 1532-0464. DOI: 10.1016/J.JBI.2016.04.007.

22. E. Batista and A. Solanas, "Process Mining in Healthcare: A Systematic Review" in IISA 2018: 9th International Conference on Information, Intelligence, Systems and Applications, IEEE, Jul. 2018pp. 1–6, ISBN: 978-1-5386-8161-9. DOI: 10.1109/IISA.2018.8633608.

23. Á. Rebuge and D. R. Ferreira, "Business process analysis in healthcare environments: A methodology based on process mining," Information Systems, vol. 37, no. 2, pp. 99–116, Apr. 2012, ISSN: 0306-4379. DOI: 10.1016/J.IS.2011.01.003.

24. S. Suriadi, R. S. Mans, M. T. Wynn, A. Partington, and J. Karnon, "Measuring Patient Flow Variations: A Cross-Organisational Process Mining Approach," in AP-BPM 2014: Asia Pacific Business Process Management, Springer, Cham, 2014, pp. 43–58. DOI: 10.1007/978-3-319-08222-6_4.

25. R. Mans, H. Schonenberg, G. Leonardi, S. Panzarasa, A. CavalliniS. Quaglini, and W. van der Aalst, "Process mining techniques: an application to stroke care.,"Studies in health technology and informatics, vol. 136, no. eHealth Beyond the Horizon – Get IT Therepp. 573–8, 2008, ISSN: 0926-9630.

26. M. Binder, W. Dorda, G. Duftschmid, R. Dunkl, K. A. Fröschl, W. Gall, W. Grossmann, K. Harmankaya, M. Hronsky, S. Rinderle-Ma, C. Rinner, and S. Weber, "On analyzing process compliance in skin cancer treatment: An experience report from the evidence-based medical compliance cluster (EBMC^{2})," in CAiSE 2012: Advanced Information Systems Engineering, vol. 7328 LNCS, Springer, Berlin, Heidelberg, 2012, pp. 398–413, ISBN: 9783642310942. DOI: 10.1007/978-3-642-31095-9_26.

27. I. Litchfield, C. Hoye, D. Shukla, R. Backman, A. Turner, M. Lee, and P. Weber, "Can process mining automatically describe care pathways of patients with long-term conditions in UK primary care? A study protocol," BMJ Open, vol. 8, no. 12, Dec. 2018. DOI: 10.1136/bmjopen-2017-019947.

28. O. Alberti González, J. C. Aragüés Bravo, M. Bestué Cardiel, I. Campello Morer, M. J. Chopo Alcubilla, C. García Arguedas, M. J. Gimeno Peribáñez, A. Gómez Peligros, B. Gros Bañeres, L. Jarauta Lahoz, M. J. Lafuente González, J. Marta Moreno, G. Martínez Boro-bio, E. Muñoz Farjas, M. Palacín Larroy, C. Pérez Lázaro, M. Sampériz Murillo, and M. Serrano Ponz, Plan de Atención al Ictus en Aragón. Actualización 2019-2022, M. Bestué Cardiel, J. Marta Moreno, and G. Martínez Borobio, Eds. Gobierno de Aragón. Dirección General de Asistencia Sanitaria, 2018, ISBN: 978-84-09-08415-9. [Online]. Available: https://www.aragon.es/estaticos/GobiernoAragon/Departamentos/SanidadBienestarSocialFamilia/Sanidad/Documentos/Programa%7B%5C_%7DIctus%7B%5C_%7Dactualizacion2019.pdf.

29. P. Conde-Espejo, Evaluación de la eficiencia de modelos organizativos para el abordaje del ictus (unidades de ictus). Madrid: Consejería de Sanidad -D. G. de Planificación, Investigación y Formación, 2013. [Online]. Available: http://www.comunidad.madrid/publicacion/1354388439231.

30. K. Chodorow and M. Dirolf, MongoDB: the definitive guide. O'Reilly Media, 2010, p. 193, ISBN: 9781449381561. [Online]. Available: http://cds.cern.ch/record/1359920.

31. E. S. Chen and I. N. Sarkar, "Mining the Electronic Health Record for Disease Knowledge," Biomedical Literature Mining, vol. 1159pp. 269–286, 2014. DOI: 10.1007/978-1-4939-0709-0_15.

32. T. Delespierre, P. Denormandie, A. Bar-Hen, and L. Josseran, "Empirical advances with text mining of electronic health records," BMC Medical Informatics and Decision Making, vol. 17, no. 1, p. 127, Dec. 2017, ISSN: 1472-6947. DOI: 10.1186/s12911-017-0519-0.

33. E. V. Epure, P. Martin-Rodilla, C. Hug, R. Deneckere, and C. Salinesi, "Automatic process model discovery from textual methodologies," in Proceedings of 2015 IEEE 9th International Conference on Research Challenges in Information Science (RCIS), L. P. G.-P. C. Rolland C. Anagnostopoulos D., Ed., vol. 2015-June, IEEE Computer Society, May 2015, pp. 19–30, ISBN: 978-1-4673-6630-4. DOI: 10.1109/RCIS. 2015.7128860.

34. W. van der Aalst, A. Adriansyah, and B. van Dongen, "Replaying history on process models for conformance checking and performance analysis,"Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, vol. 2, no. 2, pp. 182–192, Mar. 2012, ISSN: 19424787. DOI: 10.1002/widm.1045.

35. S. Montani, G. Leonardi, S. Quaglini, A. Cavallini, and G. Micieli, "Mining and Retrieving Medical Processes to Assess the Quality of Care," in International Conference on Case-Based Reasoning, Saratoga Springs, NY, USA: Springer, Berlin, Heidelberg, 2013, pp. 233–240. DOI: 10.1007/978-3-642-39056-2_17.

Artículo 4

1. Holodinsky, J.K., Yu, A.Y.X., Assis, Z.A. et al. History, Evolution, and Importance of Emergency Endovascular Treatment of Acute Ischemic Stroke. Curr Neurol Neurosci Rep 16, 42 (2016). https://doi.org/10.1007/s11910-016-0646-5

2. National Institute of Neurological Disorders and Stroke rt-PA Stroke Study Group (1995). Tissue plasminogen activator for acute ischemic stroke. The New England journal of medicine, 333(24), 1581–1587. https://doi.org/10.1056/NEJM199512143332401

3. Hacke, W., Kaste, M., Fieschi, C., Toni, D., Lesaffre, E., von Kummer, R., Boysen, G., Bluhmki, E., Höxter, G., & Mahagne, M. H. (1995). Intravenous thrombolysis with recombinant tissue plasminogen activator for acute hemispheric stroke. The European Cooperative Acute Stroke Study (ECASS). JAMA, 274(13), 1017–1025.

4. Tsang, A.C., Yang, I., Orru', E., Nguyen, Q.A., Pamatmat, R.V., Medhi, G., Wan, Y., & Huang, S. (2020). Overview of endovascular thrombectomy accessibility gap for acute ischemic stroke in Asia: A multi-national survey. International Journal of Stroke, 15, 516 - 520.

5. Allen, M., Pearn, K., Ford, G.A., White, P., Rudd, A.G., McMeekin, P., Stein, K., & James, M.A. (2021). National implementation of reperfusion for acute ischaemic stroke in England: How should services be configured? A modelling study. European Stroke Journal, 7, 28 - 40.

6. Alvarez Sabín, J., Alonso de Leciñana, M., Gállego, J., Gil-Peralta, A., Casado, I., Castillo, J., Díez Tejedor, E., Gil, A., Jiménez, C., Lago, A., Martínez-Vila, E., Ortega, A., Rebollo, M., Rubio, F., & Grupo de Estudio de las Enfermedades Cerebrovasculares de la Sociedad Española de Neurología (2006). Plan de atención sanitaria al ictus [Plan for stroke healthcare delivery]. Neurologia (Barcelona, Spain), 21(10), 717–726.

7. Ministerio de Sanidad y Política Social. Estrategia en Ictus del Sistema Nacional de Salud. Estrategia aprobada por el Consejo Interterritorial del Sistema Nacional de Salud el 26 de noviembre de 2008. https://www.mscbs.gob.es/organizacion/sns/planCalidadSNS/docs/EstrategiaIctusSNS.pdf

8. Díez-Tejedor, E., & Fuentes, B. (2011). Stroke care in Spain. What do we have? What do we need? Neurología (English Edition), 26(8), 445–448. https://doi.org/10.1016/j.nrleng.2011.01.011
9. Ministerio de Sanidad, Servicios Sociales e Igualdad. Evaluación de la Estrategia en Ictus del Sistema Nacional de Salud. 2013. https://saludcantabria.es/uploads/pdf/noticias/Punto%2011-%20Evaluacion%20Estrategia%20en%20Ictus%20del%20SNS%202013.pdf
10. A. Cayuela, L. Cayuela, M.J. Ortega Belmonte, S. Rodríguez-Domínguez, I. Escudero-Martínez, A. González, ¿Ha dejado de disminuir la mortalidad por enfermedades cerebrovasculares en España?, Neurología, 2019. ISSN 0213-4853, https://doi.org/10.1016/j.nrl.2019.06.001.
11. M. Alonso de Leciñana, A. Morales, M. Martínez-Zabaleta, Ó. Ayo-Martín, L. Lizán, M. Castellanos. Características de las unidades de ictus y equipos de ictus en España en el año 2018. Proyecto Pre2Ictus. Neurología. 2020. ISSN 0213-4853. https://doi.org/10.1016/j.nrl.2020.06.012.
12. AHRQ QI™ Version 4.5, Inpatient Quality Indicators #17, Technical Specifications, Acute Stroke Mortality Rate https://www.qualityindicators.ahrq.gov
13. Tebé C1, Abilleira S, Ridao M, Espallargues M, Salas T, Bernal-Delgado E y Atlas VPM*. Atlas de Variaciones en el manejo de la Enfermedad Cerebrovascular Isquémica. Atlas de Variaciones de la Práctica Médica en el Sistema Nacional de Salud (Atlas VPM). Volumen 5. Número 1. Diciembre 2003. ISSN:2529-9255
14. Elixhauser and Quan's ICD-9-CM and ICD-10 Coding Algorithms for Elixhauser Comorbidity Index. Med Care 2005 Nov; 43(11): 1130–9. pmid:16224307
15. Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data. Med Care 1998; 36:8–27 pmid:9431328
16. Ministerio de Sanidad. Sistema de Información de Atención Especializada (SIAE). Registro de Actividad de Atención Especializada. Conjunto Mínimo de Datos al Alta Hospitalaria. España. 2015.
17. Ministerio de Sanidad, Consumo y Bienestar Social. Estadísticas e información Sanitaria. Catálogo Nacional de Hospitales 2015. Actualizado a 31 de diciembre de 2014. NIPO en línea: 731-19-039-5 Disponible en: https://www.mscbs.gob.es/ciudadanos/prestaciones/centrosServiciosSNS/hospitales/docs/CNH2015.pdf
18. Angulo-Pueyo, E., Comendeiro-Maaløe, M., Estupiñán-Romero, F. et al. Atlas VPM: two decades informing on unwarranted variations in health care in Spain. Res Health Serv Reg 1, 5 (2022). https://doi.org/10.1007/s43999-022-00005-3
19. OECD (2011), "In-hospital mortality following stroke", in Health at a Glance 2011: OECD Indicators, OECD Publishing, Paris. DOI: https://doi.org/10.1787/health_glance-2011-43-en

20. Lin, X., & Zhang, D. (1999). Inference in generalized additive mixed models by using smoothing splines. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 61(2), 381–400. doi:10.1111/1467-9868.00183.
21. SNIJDERS, T. A. B., & BOSKER, R. J. (1994). Modelled Variance in Two-Level Models. Sociological Methods & Research, 22(3), 342–363. https://doi.org/10.1177/0049124194022003004
22. Merlo, J., Chaix, B., Ohlsson, H., Beckman, A., Johnell, K., Hjerpe, P., Råstam, L., & Larsen, K. (2006). A brief conceptual tutorial of multilevel analysis in social epidemiology: using measures of clustering in multilevel logistic regression to investigate contextual phenomena. Journal of epidemiology and community health, 60(4), 290–297. https://doi.org/10.1136/jech.2004.029454
23. R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.r-project.org/
24. Simon Wood (2013) mgcv: Mixed GAM Computation Vehicle with Automatic Smoothness Estimation. R package version 1.1.8 https://cran.r-project.org/package=mgcv
25. Kim, J., Thayabaranathan, T., Donnan, G. A., Howard, G., Howard, V. J.,Rothwell, P. M. Thrift, A. G.,et al. (2020). Global Stroke Statistics 2019. International Journal of Stroke, 15(8), 819–838. doi:10.1177/1747493020909545
26. Lopez, A., Murray, C. The global burden of disease, 1990–2020. Nat Med 4, 1241–1243 (1998). https://doi.org/10.1038/3218
27. He, Q., Wu, C., Luo, H., Wang, Z.-Y., Ma, X.-Q., Zhao, Y.-F, et al. (2014). Trends in In-Hospital Mortality among Patients with Stroke in China. PLoS ONE, 9(3), e92763. doi:10.1371/journal.pone.0092763XXX
28. Heuschmann, P. U., Berger, K., Misselwitz, B., Hermanek, P., Leffmann, C., Adelmann, M., et al. (2003). Frequency of Thrombolytic Therapy in Patients With Acute Ischemic Stroke and the Risk of In-Hospital Mortality. Stroke, 34(5), 1106–1112. doi:10.1161/01.str.0000065198.80347.c5
29. Seghieri, C., Mimmi, S., Lenzi, J. et al. 30-day in-hospital mortality after acute myocardial infarction in Tuscany (Italy): An observational study using hospital discharge data. BMC Med Res Methodol 12, 170 (2012). https://doi.org/10.1186/1471-2288-12-170
30. Tebé C, Abilleira S, Ridao M, Espallargues M, Salas T, Bernal-Delgado E y Atlas VPM. Atlas de Variaciones en el manejo de la Enfermedad Cerebrovascular Isquémica. Atlas de Variaciones de la Práctica Médica. Diciembre 2013. Volumen 5. Número 1. ISSN:1699-4663. 2013. https://www.atlasvpm.org/wp-content/uploads/2019/06/Atlasnumero9_2.6MB.pdf
31. J.C. López Fernández, J. Masjuan Vallejo, J. Arenillas Lara, M. Blanco González, et al. Analysis of stroke care resources in Spain in 2012: Have we benefitted from the Spanish Health System's stroke

care strategy?. Neurología (English Edition), Volume 29, Issue 7, 2014, Pages 387-396, ISSN 2173-5808, https://doi.org/10.1016/j.nrleng.2013.06.009.

Universidad Zaragoza

TESIS DOCTORAL

Programa de Doctorado en Medicina

PUBLICACIONES

“Medición y comparación de resultados de servicios sanitarios”

Francisco Ramón Estupiñán Romero

Anexo I

Factor de Impacto (FI) y Áreas temáticas (AT) correspondientes a las publicaciones incluidas

Artículo 1 JCR **FI: 1.774 (Q3)** AT: **PUBLIC, ENVIRONMENTAL & OCCUPATIONAL HEALTH**

Bernal-Delgado, E., Estupiñán-Romero, F. A data infrastructure for the assessment of health care performance: lessons from the BRIDGE-health project. Arch Public Health 76, 6 (2018). https://doi.org/10.1186/s13690-017-0245-1

Artículo 2 JCR **FI: 3.24 (Q2)** AT: **MULTIDISCIPLINARY SCIENCES**

Comendeiro-Maaløe, M., Estupiñán-Romero, F., Thygesen, L. C., Mateus, C., Merlo, J., Bernal-Delgado, E., & ECHO consortium (2020). Acknowledging the role of patient heterogeneity in hospital outcome reporting: Mortality after acute myocardial infarction in five European countries. PloS one, 15(2), e0228425. https://doi.org/10.1371/journal.pone.0228425

Artículo 3 JCR **FI: 5.772 (Q1)** AT: **MEDICAL INFORMATICS**

J. González-García, C. Tellería-Orriols, F. Estupiñán-Romero and E. Bernal-Delgado, "Construction of Empirical Care Pathways Process Models From Multiple Real-World Datasets," in IEEE Journal of Biomedical and Health Informatics, vol. 24, no. 9, pp. 2671-2680, Sept. 2020, doi: 10.1109/JBHI.2020.2971146

Artículo 4 JCR **FI: 2.9 (Q2)** AT: **MEDICINE, GENERAL & INTERNAL**

Estupiñán-Romero, F., Pinilla Dominguez, J., Bernal-Delgado, E., & Atlas VPM consortium (2023). Differences in acute ischaemic stroke in-hospital mortality across referral stroke hospitals in Spain: a retrospective, longitudinal observational study. BMJ open, 13(6), e068183. https://doi.org/10.1136/bmjopen-2022-068183

Anexo II: Publicaciones

Bernal-Delgado and Estupiñán-Romero *Archives of Public Health* (2018) 76:6
DOI 10.1186/s13690-017-0245-1

Archives of Public Health

RESEARCH Open Access

A data infrastructure for the assessment of health care performance: lessons from the BRIDGE-health project

Enrique Bernal-Delgado[1,2,3*] and Francisco Estupiñán-Romero[1,2]

Abstract

The integration of different administrative data sources from a number of European countries has been shown useful in the assessment of unwarranted variations in health care performance. This essay describes the procedures used to set up a data infrastructure (e.g., data access and exchange, definition of the minimum common wealth of data required, and the development of the relational logic data model) and, the methods to produce trustworthy healthcare performance measurements (e.g., ontologies standardisation and quality assurance analysis). The paper ends providing some hints on how to use these lessons in an eventual European infrastructure on public health research and monitoring. Although the relational data infrastructure developed has been proven accurate, effective to compare health system performance across different countries, and efficient enough to deal with hundred of millions of episodes, the logic data model might not be responsive if the European infrastructure aims at including electronic health records and carrying out multi-cohort multi-intervention comparative effectiveness research. The deployment of a distributed infrastructure based on semantic interoperability, where individual data remain in-country and open-access scripts for data management and analysis travel around the hubs composing the infrastructure, might be a sensible way forward.

Background

The current EU healthcare agenda is built upon three pillars: strengthening healthcare effectiveness, increasing accessibility and improving resilience. The agenda bestows a critical role, among other strategies, to the assessment of health systems performance and to the routinely use of existing health information systems [1].

There are some countrywide examples where national health institutions have implemented actions meant to use health information systems in the evaluation of health systems performance [2–6]. Although less frequent, there are also some pre-eminent international efforts on cross-country comparisons. Notably, the OECD is regularly producing the Health at a Glance report [7] or numerous outlets from its Healthcare Quality Indicators project [8], and, the European Commission has set up an expert group on health systems performance assessment (HSPA) whose agenda is led for the exchange of HSPA experiences, the definition of HSPA priority areas and the support to national policy-makers on HSPA methods [9]. Lastly, different EU research programs have fostered the development of research initiatives aiming the cross-country analysis of health systems performance. [An extensive review of those research projects can be found at http://www.euroreach.net/compendium]. A commonality between these initiatives is the use (reuse) of routinely collected data, in particular, administrative data.

One of those projects has been ECHO (European Collaboration of Healthcare Optimization) an international effort to access and link administrative health data sources from several European countries with a view to set the basis for cross-country health systems performance assessment. ECHO accessed and reused individual-level data from hospital admissions and, demographic, socioeconomic and supply information to analyse and report on a number of health system performance (HSP) dimensions (e.g., utilisation of low-value procedures, equity of access to effective care, or quality and efficiency), at meaningful levels of analysis (either hospitals or geographic healthcare

* Correspondence: ebernal.iacs@aragon.es
[1]Institute for Health Sciences in Aragon (IACS), Zaragoza, Spain
[2]BRIDGEHealth Project, Brussels, Belgium
Full list of author information is available at the end of the article

areas) [10]. (See more details at www.echo-health.eu). Later on, integrated within the context of the BRIDGE-Health project (www.bridge-health.eu), ECHO methods and achievements have been revisited with the aim to contribute to the design and development of a sustainable European infrastructure on public health research and monitoring.

This paper provides a description of the challenges faced to build a data infrastructure aiming HSPA, and some thoughts on whether this model is suitable for a European research infrastructure.

Challenges on building a health information system based on routine data on health

Any project using routine data to provide sound international HSP research and monitoring has to face a number of challenges; thus: a) defining the minimum common dataset required to assess HSP dimensions and indicators; b) analysing the data origins, as well as the linkage mechanisms and developing the logic data model that will allow the production of comparable performance indicators; c) getting access to original data sources, curated and maintained by data authorities under a predefined legal frame; d) transforming raw data formats and categories into a common standard; e) building extensive catalogues (i.e. dictionaries) aimed to allocate data to units of analysis while considering over time modifications; f) building a common language (i.e., semantic interoperability) from different ontologies (e.g., different diagnoses and procedures classification systems); g) releasing resulting datasets that allow HSP analyses and reporting; and h) analysing the quality of those resulting datasets and, accordingly, decide on the accuracy and reliability of HSP results. In the following paragraphs, some of those critical challenges are discussed taking as a reference the works done in ECHO within BRIDGEHealth (EwB).

Definition of the minimum common dataset (MCD)

A project on HSPA aims at reporting relevant dimensions of HSP at meaningful levels of analysis. In the case of international comparisons, once performance dimensions and indicators, as well as units of analysis are decided, an HSPA project has to define the minimum common dataset (MCD) required for the production of such performance indicators.

Operationally, the MCD is the set of variables that composes the so-called 'core facts' table of a data infrastructure. The 'core facts' table is used to integrate the original administrative data from any participant country into a single coherent relational database. The MCD includes: a) patient attributes (e.g., group of age, sex, diagnoses and procedures); b) episode attributes (e.g., date of admission, type of discharge, type of admission, etc.); c) geographical location (e.g., health care area of residence, health care area of treatment); and, d) hospital of treatment.

Besides, the MCD contains univocal identifiers to grant traceability and linkage across data sources and catalogues: a) univocal identifiers at the maximum level of disaggregation (i.e., episode) and, b) univocal identifiers for the units of analysis, (hospitals, health authorities, or regions). A detail of the MCD from the EwB project is provided in Additional file 1.

Data origins, linkage mechanisms and logic data model

The EwB data infrastructure has included various data origins [hospital discharges, demographic information at geographic level, socioeconomic data at geographic level, supply features at hospital level, and geographic vectors depicting geographic areas of interest (e.g., hospital-catchment areas [11], health authorities, regions, NUTS)] with a view to, departing from individual episodes, analysing HSP both, at geographic and hospital-level, accounting for over time evolution (e.g., hospital merging processes).

So, the EwB logic data model, conceived as a relational database, has been built upon three main entities (episodes, hospitals, and geographic areas) and their respective attributes. The critical attributes for each of these entities are described along various catalogues; so, dictionaries or ontologies containing codes for diagnoses and procedures, hospital names, locations and evolution, name for the geographic areas and resident population.

EwB data model owns three critical elements: a) as episodes store individual patient-level information that is actually embedded into both a hospital of treatment and geographic areas hierarchically constituted (health care into regions, regions into countries) linkage across files and catalogues follows either a 1-to-1 scheme (when linkage is limited to episode-based attributes) or a 1-to-N scheme (when episodes are linked to a hospital or an area). The aforementioned MCD univocal identifiers grant traceability and robustness for this linkage; b) given the rich information contained in each episode, mainly dependent on the large number of diagnoses and procedures recorded in each registry (~20 to 50 variables each), there is a need of increasing computational efficiency. For this purpose, the diagnoses and the procedure variables were split into two different catalogues containing all existing diagnoses or procedures, at their maximum level of precision; and, c) instead of yielding a single output, the data model implies the production of three separate output files containing geographic-based HSP indicators, hospital-based HSP indicators, and ad hoc intermediate variables and modifiers to grant the control of confounding (e.g., Elixhauser comorbidities); all are indexed using the

EwB univocal identifiers, reducing the volume of files, allowing storage in a distributed way, and enabling parallel access and processing.

Last but not least, as any relational data model the EwB dataset has been tested to check coherence, and whether the model preserves the identity of the entities contained, their referential integrity, the cardinality of the data and, the inheritances of their attributes [12]. A detail of the logic data model is provided in Additional file 2.

Standardizing data and assuring semantic interoperability

The standardization of data from different data origins and different countries entail a series of operations that aim the comparability of the raw data. In the use of administrative data the most frequent threats to comparability are: a) the way variables are categorized might be different across data origins; b) coding precision maybe different hampering the comparability of definitions (Fig. 1); c) not all data origins might provide full coverage for each of the variables at any unit of analysis (Fig. 2); d) time-dependent phenomena might translate into data inconsistencies over the years (Fig. 3); and, e) data origins might use different ontologies to normalize their data.

The level of complexity of the remedies will differ. The differences in the way the variables are categorized will generally require a simple transformation looking for a minimum common denominator across origins. Differences in coding precision will entail careful assessment to weight the actual impact on the construct validity of the performance indicators. Missing values or time-dependent modifications will involve a logic data model that flexibly deal with both issues; so, in the former creating 'missing' categories, at each level of analysis; in the latter, creating comprehensive catalogues linked with secondary identifiers able to account for changes (e.g., in geographic boundaries, the merge of several providers or the periodic upgrade of classification systems).

Nonetheless, the greater level of complexity lays on tackling the semantic differences across ontologies; for example, whether an episode means the same in all the data origins, or whether what a coding system flags as congestive heart failure in an origin, is a congestive heart failure in another one. Semantic interoperability is at the very core of any HSPA international comparison as HSPA culminates with the development and measurement of performance indicators using information on diagnoses and procedures frequently, from different ontologies and coding systems (In EwB, for example, ICD9th, ICD 10th, ACHI, NOMESCO and OPCS).

Addressing semantic interoperability requires indicator-specific crosswalks across coding systems. For that purpose, an extensive mapping of codes has to be developed, which implies a deep knowledge of the different ontologies and the face and empirical validation of in-country experts (coders, clinicians and potential users of HSPA results) looking for coherence, anomalous data distributions and consistency over time. Taking the case of EwB, the mapping exercise has ended up producing crosswalks for 201 indicators. [For more details on the crosswalks see: http://www.echo-health.eu/handbook/getting-indicators.html].

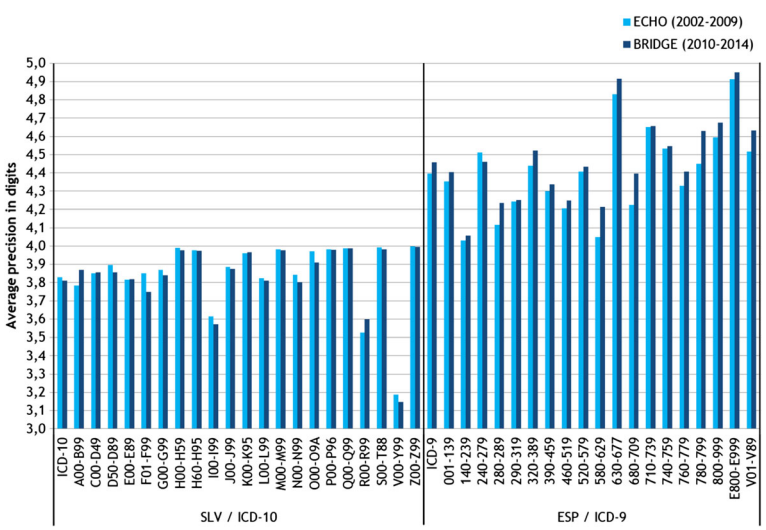


Fig. 1 Coding precision (Slovenia and Spain)NNote. Note: Coding precision refers to the level of specificity in the way an admission is coded. In the case of diagnoses or procedures, the different classification systems own different levels of specificity. For example in ICD 9th MC, diagnoses might be coded with three, four of five digits; the greatest level of specificity implies the use of 5 digits. In the figure, actual levels of specificity in Slovenia (ICD 10th) and Spain (ICD 9th) are shown. Spain shows larger coding precision in all Major Diagnostic Categories (x axis), generally improving since 2010. When building comparable indicators the level of specificity has to be equivalent

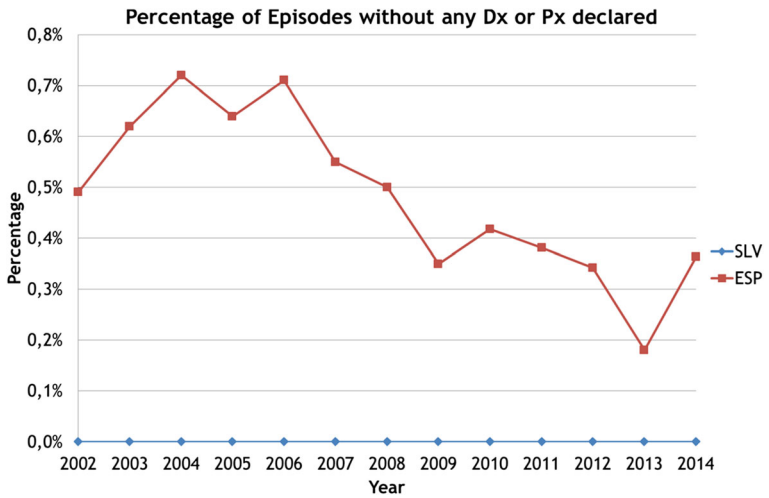


Fig. 2 Missing values in core variables. Note: The figure shows the evolution in the number of episodes with missing information on the diagnosis of admission or on the procedures performed in the episode. While Slovenia has no defaulting episodes, in Spain the number has reduced over time ranging from 0.7% to 0.4%. The toll of missing data in either the diagnosis of admission or in the procedures would impact the numerators and/or denominators of the performance rates

Assuring internal validity through data quality analysis

As any observational study, HSPA is threatened by systematic bias and confounding. The use of routine administrative data in observational studies requires further efforts to understand any potential information flaw. It is then essential a systematic analysis of the quality of the dataset upon which HSPA is developed. A concise example on in-hospital case-fatality rates in the admission for an acute myocardial infraction (AMI-IM) is provided to have a hint on how to systematically

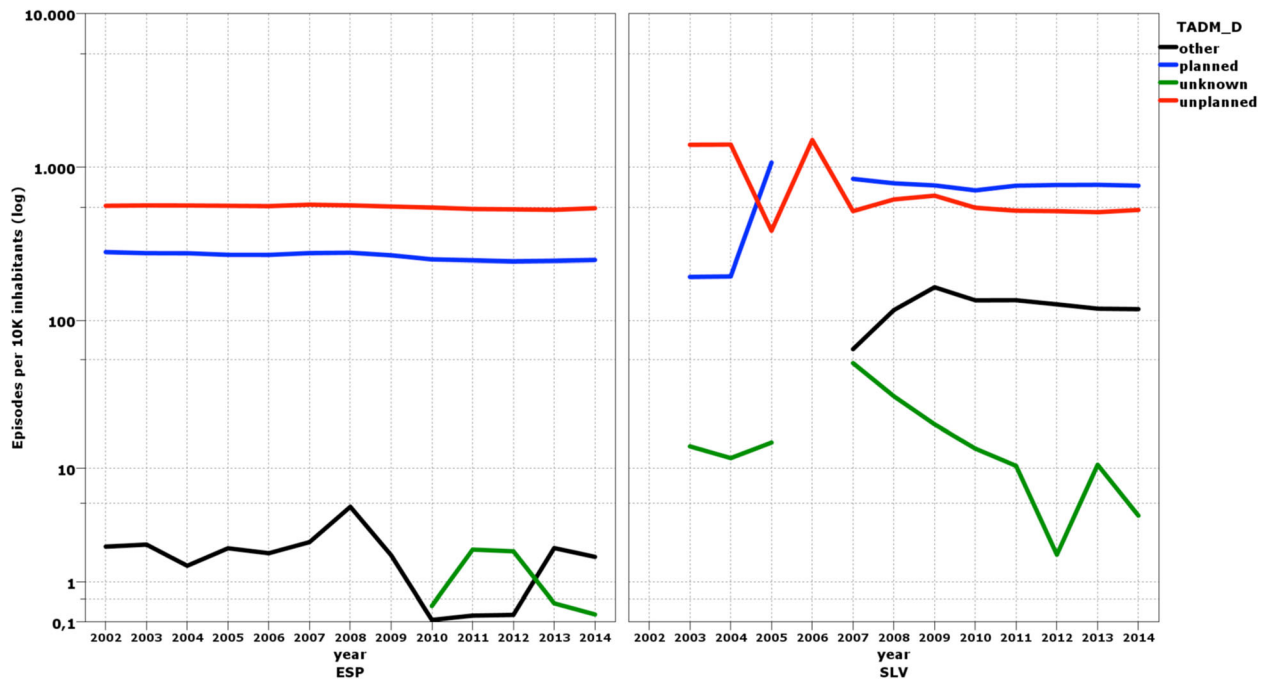


Fig. 3 Time-dependent phenomena – episodes by type of admission. Note: The figure on the right aims at highlighting the Slovene inconsistency (as compared to Spain) in the coding of planned (blue line) vs. unplanned admissions (red line), as well as disproportionally large number of 'other' types of admission. Apparently, in 2006 all admission were coded as unplanned, and since 2007 a new category 'other type of admission' starts to be used. This timedependent finding might have an impact in those performance indicators that require this variable – typically, unplanned admission are excluded when assessing surgical outcomes that might be influenced by differences in acute and severe patients' conditions

analyse information flaws in the dataset. For that purpose, data from Slovenia and Spain, a total of 65 million episodes (5 million a year in Spain and half a million a year in Slovenia, from 2002 to 2014) have been analysed and discussed hereinafter.

AMI-IM is operationally defined as the standardised in-hospital case-fatality rate for patients over 18 years old, admitted for an acute myocardial infarction. So, besides the quality analysis of case-fatalities (i.e., numerator) and patients 18 and older admitted for an AMI (i.e., denominator), as HSPA requires risk-adjustment so that hospital differences in performance, beyond differences in the underlying risks in patients, may be controlled. The quality of potential confounders has also to be analysed. (Crosswalks may be consulted at http://www.echo-health.eu/handbook/CV_AMI_MORT.html).

Looking at Fig. 4 a great stability is observed, in both countries, in the main variables concurring in this indicator; thus: a) case-fatalities steadily and smoothly decrease as expected (black line); b) AMI admissions following a similar pattern in both countries (dark blue line), suggesting that no systemic factors are affecting unevenly in the number of patients at risk; and, c) there is consistent evolution of potential confounders (AMI as STEMI or NSTEMI, the concurrence of congestive heart failure, the presence of comorbidities as diabetes with complications and hypertension with complications) with no abrupt changes as those seen in Fig. 2. In

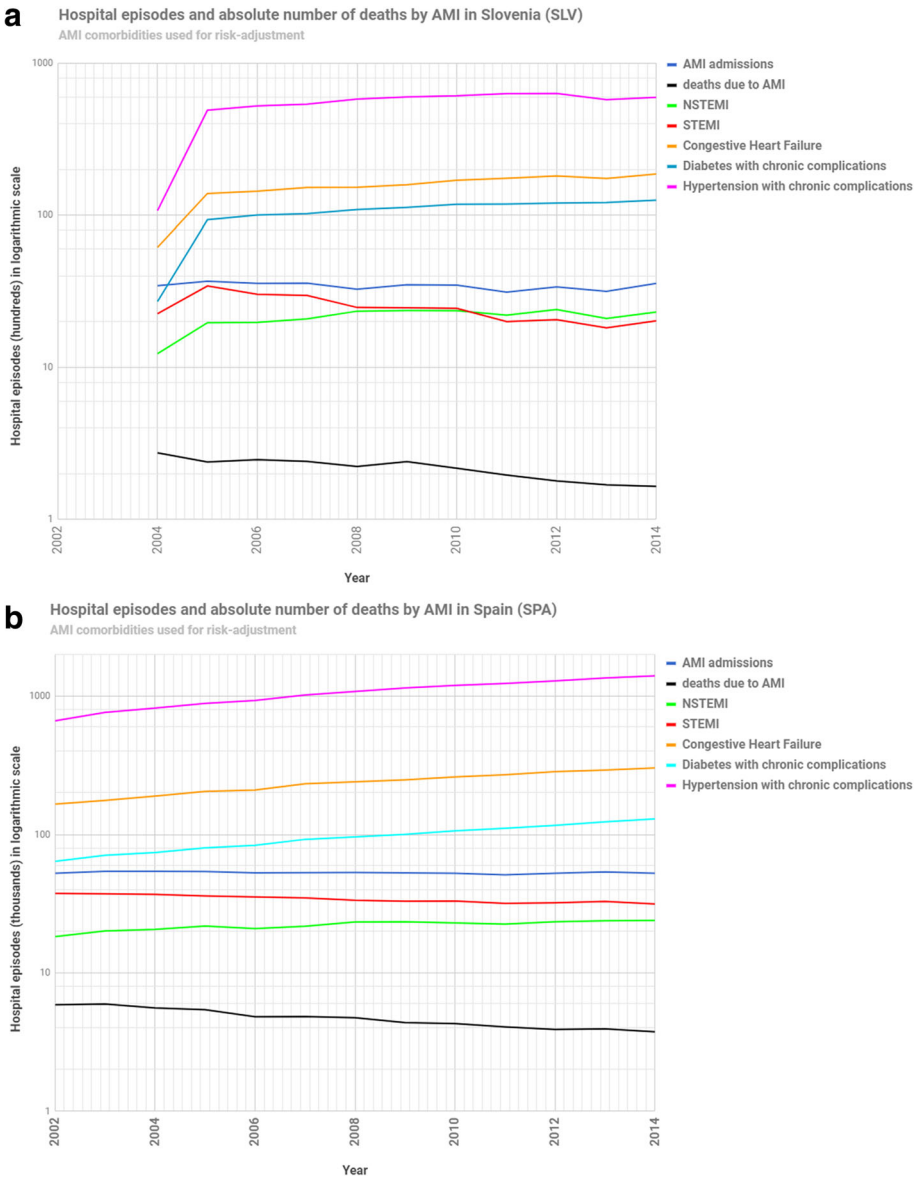


Fig. 4

general terms, HSPA comparisons with this indicator seem to be accurate. Let's look though more in depth. In the Slovene figure, stability starts at some point after 2005, so HSPA will be more reliable after that year. Most importantly, in both countries ST-elevation of myocardial infarction (STEMI) is reducing (red line) while Non-STEMI (NSTEMI) episodes (green line) are appearing more frequently –overlapping in Slovenia since 2008 and not yet converging in Spain. It is very unlikely that the AMI epidemiology is changing over the years (i.e., more NSTEMI cases in the latest years). Conversely, it is more plausible to think of a potential information bias; so, besides coding differences between Slovenia and Spain, it seems that in-country coding is very likely underestimating STEMI cases -it is well known that more severe patients that end up dying in the first 24 h (STEMI cases) have less precise information in their discharge records, end up being registered as undefined AMI. As a consequence, the use of STEMI vs. NSTEMI as potential confounders in the risk adjustment of AMI-IM, between Slovenia and Spain, has to be carefully assessed, and maybe discarded.

Developing HSPA with administrative data in an EU health information infrastructure

EwB has developed a central relational data infrastructure that stores administrative data from different data sources from various countries, with a view to carry out health systems performance research and monitoring. Might this data infrastructure be taken as a reference in the context of an eventual EU health information infrastructure? Along the following paragraphs, we reflect on those elements that should be taken into consideration in the public debate on the development of such an infrastructure at the EU level.

Some of the EwB attainments would suggest that the data infrastructure developed could be taken as a valid case study. Thus,

- EwB has proven the plausibility of creating a central repository populated with anonymised and de-identified individual information, transferred from different countries with limited administrative costs, while attaining the various legal requirements in data access, management, curation and reporting.
- As EwB builds on administrative data (i.e., data regularly collected by health or statistic authorities upon the compliance of normative provisions) it benefits of certain stability over time, irrespective of in-country health care reforms.
- Albeit the uneven richness of data across Europe, EwB has revealed that it is possible to find a minimum common dataset that eventually allows a sound comparison of health systems performance at meaningful units of analysis.
- The logic data model enables individual, hospital-specific and geographic analyses; moreover, it allows the traceability of these episodes, hospitals and geographic areas, capturing time-dependent phenomena that might alter their consistency over time, and subsequently, the reporting of performance.
- A method has been developed to assure semantic interoperability in the development of performance indicators addressing different HSP domains: utilization, equity, quality and safety, and efficiency. As aforementioned, the data model allows straightforward updates once the new ontology is available.
- All methods and techniques are transparent and publicly available with a view to assure reproducibility.

Nonetheless, a translation of the EwB model into a European Research Infrastructure would not be straightforward.

- EwB has been conceived as a demonstration project developed in a limited number of countries, a sample of convenience that, although representative of countries with different data governance schemes and data richness, might not represent the complexity of the EU28.
- EwB is confined though to the secondary use of hospital administrative data (enriched with some extra administrative data sources) aimed at specifically analysing health care performance, which may not be the only type of data sources (nor the only aim) in an eventual European Research Infrastructure. Nevertheless, the challenges addressed along this paper are not specific to hospital data, so should be quite the same for any other data source.
- In practical terms, the administrative costs for the maintenance of an expanded EwB infrastructure is unknown; on the other hand, the legal requirements for data access (and eventual transfer) will multiply, so that the exiguous governance model implemented in EwB might not suffice.
- A third element has to do with the semantic interoperability. Although the method developed to build comparable performance indicators has been shown valid, there is a need of in-country expert panels contributing in the face and empirical validation of existing or new indicators, as well as to the attentive follow up of the publication of renewed ontologies. So, the EwB governance of this task should not be the same when scaling up to more indicators and more countries.
- Last but not least, although the EwB central relational dataset has been proven accurate, effective

to compare HSP across different countries, and efficient enough to deal with hundred of millions of episodes, the logic data model might no be responsive to some future requirements. On the one hand, the data model requires individual data transfer; although this has been possible in some countries, others have found serious difficulties ending up not providing data. On the other hand, although this logic data model allows research and monitoring, it is confined to ecological or cross-sectional studies that, at most, may add a temporal perspective. According to the current developments in health systems performance beyond classical monitoring, a state-of-the-art infrastructure should aim the reuse of electronic health and medical records and conduct more complex comparative effectiveness research. Should the EU data infrastructure aim at addressing the challenge, it might be recommendable a different type of logic data model for which we propose a distributed model where individual data remain in-country and open-access scripts for data extraction, transformation, analysis and reporting travel around the hubs composing the infrastructure.

Additional files

Additional file 1: Annex 1. (PDF 85 kb)
Additional file 2: Annex 2. (PDF 260 kb)

Acknowledgments
EwB has received the precious contribution of Lau C Thygesen, at the National Institute of Public Health, University of Southern Denmark, Denmark. Mircha Poldrugovac and Tit Albrecht, at the National Institute of Public Health in Ljubljana, Slovenia; Ceu Mateus, University of Lancaster, United Kingdom; Stefan Mathis-Edenhofer and Andreas Birner at Gesundheit Österreich GmbH, Vienna, Austria; Nils Gütacker Centre for Health Economics, University of York, United Kingdom; Judit Zsuga, at University of Debrecen, Hungary; and, Pedro P. Barros, Faculdade de Economia da Universidade Nova de Lisbon, Portugal.

Authors' contributions
Both authors equally contributed in the design, drafting and discussion of the paper. Specifically, FER prepared the data quality analyses upon which the first section was developed. Both read and approved the final manuscript.

Ethics approval and consent to participate
The access and transfer of data have carefully followed the legal provisions for personal data protection and patients' rights. The episodes collected in the EwB do not contain personal data; episodes are anonymised and dissociated in origin making impossible backwards traceability. The data management was conducted in accordance with the amended Helsinki Declaration, and the International Guidelines for Ethical Review of Epidemiological Studies.

Competing interests
The authors declare that they have no competing interests.

Publisher's Note

Author details
[1]Institute for Health Sciences in Aragon (IACS), Zaragoza, Spain. [2]BRIDGEHealth Project, Brussels, Belgium. [3]Health Services and Policy Research Group - ARiHSP, Institute for Health Sciences in Aragon (IACS), San Juan Bosco 13 (CIBA building), 50009 Zaragoza, Spain.

Received: 27 September 2017 Accepted: 7 November 2017
Published online: 24 January 2018

References
1. European Commission. Communication on effective, accessible and resilient health systems. COM (2014) 215 final. Available at: https://ec.europa.eu/health//sites/health/files/systems_performance_assessment/docs/com2014_215_final_en.pdf [Accessed 3 Aug 2017].
2. Institute for Clinical Evaluative Sciences (ICES) Available from: http://www.ices.on.ca/Publications/Atlases-and-Reports [Accessed 3 Aug 2017].
3. Agency for Health Research and Quality (ARQH). 2010 Available at: http://www.qualityindicators.ahrq.gov. [Accessed 3 Aug 2017].
4. NHS Atlas of Variations in Healthcare. Available at: https://www.england.nhs.uk/rightcare/products/atlas/. [Accessed 3 Aug 2017].
5. Dutch National Atlas of Public Health. Available at: https://www.volksgezondheidenzorg.info/onderwerp/atlas-vzinfo/inleiding#!node-kaarten [Accessed 3 Aug 2017].
6. New Zealand Atlas of Healthcare Variation. Available at: http://www.hqsc.govt.nz/our-programmes/health-quality-evaluation/projects/atlas-of-healthcare-variation/[Accessed 3 Aug 2017].
7. OECD/EU (2016), Health at a glance: Europe 2016: state of health in the EU cycle, OECD Publishing, Paris. http://dx.doi.org/10.1787/9789264265592-en [Accessed 3 Aug 2017].
8. Reviews of National Healthcare Quality Available at: http://www.oecd.org/els/health-systems/health-care-quality-reviews.htm. [Accessed 3 Aug 2017].
9. Expert Group on Health Systems Performance Assessment (HSPA): Available at: https://ec.europa.eu/health/systems_performance_assessment/policy/expert_group_en [Accessed 3 Aug 2017].
10. E. Bernal-Delgado, on behalf of ECHO Consortium, T. Christiansen, on behalf of ECHO Consortium, K. Bloor, on behalf of ECHO Consortium, C. Mateus, on behalf of ECHO Consortium, A.M. Yazbeck, on behalf of ECHO Consortium, J. Munck, on behalf of ECHO Consortium, J. Bremner, on behalf of ECHO Consortium; ECHO: health care performance assessment in several European health systems. Eur J Pub Health 2015; 25(suppl_1): 3-7. doi: https://doi.org/10.1093/eurpub/cku219.
11. Thygesen CT, Baixauli-Pérez C, Librero-López J, Martínez-Lizaga N, Ridao-López M, Bernal-Delgado E, on behalf of the ECHO Consortium. Comparing variation across European countries: building geographical areas to provide sounder estimates. Eur J Pub Health. 25(suppl_1):8–14. https://doi.org/10.1093/eurpub/cku229
12. Bergdahl M, Ehling M, Elvers E, Földesi E, Körner T, Kron A, et al. Handbook on data quality assessment methods and tools. Wiesbaden: European Comission; 2007.

PLOS ONE

RESEARCH ARTICLE

Acknowledging the role of patient heterogeneity in hospital outcome reporting: Mortality after acute myocardial infarction in five European countries

Micaela Comendeiro-Maaløe[1,2], Francisco Estupiñán-Romero[1,2], Lau Caspar Thygesen[3], Céu Mateus[4], Juan Merlo[5], Enrique Bernal-Delgado[1,2]*, on behalf of the ECHO consortium[¶]

1 Health Services and Policy Research Group, Institute for Health Sciences in Aragon (IACS), Zaragoza, Spain, **2** Network for Health Services Research in Chronic Patients (REDISSEC), Madrid, Spain, **3** National Institute of Public Health, University of Southern Denmark, Copenhagen, Denmark, **4** Division of Health Research, Lancaster University, Lancaster, England, United Kingdom, **5** Unit for Social Epidemiology, Sweden & Centre for Primary Health Care Research, Region Skåne, Faculty of Medicine, Lund University Malmö, Malmö, Sweden

¶ Membership of the ECHO Consortium is listed in the Acknowledgments.
* ebernal.iacs@aragon.es

OPEN ACCESS

Citation: Comendeiro-Maaløe M, Estupiñán-Romero F, Thygesen LC, Mateus C, Merlo J, Bernal-Delgado E, et al. (2020) Acknowledging the role of patient heterogeneity in hospital outcome reporting: Mortality after acute myocardial infarction in five European countries. PLoS ONE 15(2): e0228425. https://doi.org/10.1371/journal.pone.0228425

Editor: Timir Paul, East Tennessee State University, UNITED STATES

Received: October 7, 2019

Accepted: January 14, 2020

Published: February 6, 2020

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: https://doi.org/10.1371/journal.pone.0228425

Abstract

Background

Hospital performance, presented as the comparison of average measurements, dismisses that hospital outcomes may vary across types of patients. We aim at drawing out the relevance of accounting for patient heterogeneity when reporting on hospital performance.

Methods

An observational study on administrative data from virtually all 2009 hospital admissions for Acute Myocardial Infarction (AMI) discharged in Denmark, Portugal, Slovenia, Spain, and Sweden. Hospital performance was proxied using in-hospital risk-adjusted mortality. Multilevel Regression Modelling (MLRM) was used to assess differences in hospital performance, comparing the estimates of random intercept modelling (capturing hospital general contextual effects (GCE)), and random slope modelling (capturing hospital contextual effects for patients with and without congestive heart failure -CHF). The weighted Kappa Index (KI) was used to assess the agreement between performance estimates.

Results

We analysed 46,875 admissions of AMI, 6,314 with coexistent CHF, discharged from 107 hospitals. The overall in-hospital mortality rate was 5.2%, ranging from 4% in Sweden to 6.9% in Portugal. The MLRM with random slope outperformed the model with only random intercept, highlighting a much higher GCE in CHF patients [VPC = 8.34 (CI95% 4.94 to 13.03) and MOR = 1.69 (CI95% 1.62 to 2.21) vs. VPC = 3.9 (CI95% 2.4 to 5.9), MOR of

Data Availability Statement: The access to pseudoanonimised patient-level administrative data (patient episodes: hospital admissions and discharges) supporting the manuscript findings is restricted to avoid data misuse in accordance with the System Level Security Policy of the Unit for Health Services and Policy Research at the Institute for Health Sciences in Aragon (IACS) (Available at: www.atlasvpm.org/grupo-coordinador/ARiHSPinformationsecuritypolicy) within the framework of the Spanish legal system, in particular, Regulation (EU) 2016/679, the Law 3/2018 on Personal Data Protection, the Law 14/2007 on Biomedical Research and the Law 37/2007 and Law 18/2015 on the Public Sector Information Reuse. Consequently, we have specified that only aggregated data may be accessed, upon request, throughout our data sharing agreement, by contacting the senior author of the manuscript Enrique Bernal-Delgado (ebernal.iacs@aragon.es) and/or the legal officer Ramón Launa-Garcés (rlaunag.iacs@aragon.es)

Funding: This study has been funded by the Spanish Institute for Health Carlos III through the project "The Research Network on Health Services Research on Chronic Patients (REDISSEC)", registered grant RD16/0001/0007, and co-funded by European Regional Development Fund "Investing in your future". On the other hand, the ECHO Project received funding from the 7th Framework Programme of the European Union (2010-2014) with Grant HEALTH-F3-2010-242189 and from the European Union's Health Programme (2014-2020) with Grant 664691/BRIDGE Health. Additionally, the authors Enrique Bernal-Delgado, Micaela Comendeiro-Maaløe and Francisco Estupiñán-Romero have been partially funded by the Spanish Institute for Health Carlos III through a public competitive grant (RD16/0001/0007) as research members of "The Research Network for Health Services Research on Chronic Patients (REDISSEC)". All these funding sources played no role in the study design, data collection or the analysis, decision to publish, or preparation of the manuscript. There was no additional external funding received for this study.

Competing interests: The authors have declared that no competing interests exist.

1.42 (CI95% 1.31 to 1.54) without CHF]. No agreement was observed between estimates [KI = -0,02 (CI95% -0,08 to 0.04].

Conclusions

The different GCE in AMI patients with and without CHF, along with the lack of agreement in estimates, suggests that accounting for patient heterogeneity is required to adequately characterize and report on hospital performance.

1. Introduction

The growing availability and use of administrative data are resulting in a profusion of healthcare performance assessment initiatives worldwide. Either institutionally framed or developed under the umbrella of research projects, the wealth of administrative data offers the opportunity to access larger samples of patients, covering virtually all providers in a health plan, allowing cross-country comparisons and most importantly, enabling the systematic and continuous monitoring of providers' performance. Many institutional-based [1–7] and research-oriented examples [8–15] illustrate this enormous potential. On the other hand, as performance assessment is increasingly deemed to be the basis for different value-based initiatives (e.g. benchmarking strategies, pay for performance schemes, patient choice programs, etc), decision makers are increasingly calling for trustworthy measurements and reliable reporting [16].

In this respect, analytical methods play a critical role. Once the use of ordinary (single level) regression models were shown to be inappropriate, as they circumvent the interdependence of patient outcomes within a hospital (i.e. patient risk within a hospital is more alike than patient risk from a different hospital) [9, 12, 15–19], and are at risk of the Yule-Simpson paradox [20], marginal models (Generalized Estimating Equations, GEE) or multilevel modelling (MLRM) have become increasingly popular, although their approach and interpretation are clearly different; while the use of GEE focus on the estimation of the population-averaged risk of death adjusting hospitals' heterogeneity, MLRM assumes that each hospital has their own underlying risk of an event, and this risk varies across hospitals (i.e. the probability of an event is conditional to the place where the patient is treated). Accordingly, MLRM has been suggested as a more appropriate approach when hospital-specific interpretations are needed [21].

But most importantly, variations in hospital performance are usually presented as the comparison of adjusted average measures, excluding the possibility that hospital performance may also be conditioned by patient heterogeneity, for example, determining the care responses to specific subgroups of patients [22]. One fundamental feature of MLRM in hospital performance assessment is that MLRM can drop the assumption that the underlying risk for an individual is the same for all hospitals, allowing this risk to vary at hospital level; therefore, the hospital effect also becomes a function of patient heterogeneity [23]. In practical terms, this property, which implies the inclusion of random slopes, allows the development of specific performance measurements for subgroups of patients. Therefore, the observation of better or worse performance will refer not just to the hospital outcome obtained for the regular patient but also to the hospital achievement for specific subgroups of individuals. This well-known property of MLRM has scarcely been exploited in the assessment and reporting of hospital performance.

In this paper, we use MLRM to draw out the relevance of accounting for patient heterogeneity when reporting on hospital performance, using in-hospital mortality in acute myocardial

infarction as a case study. Including a random slope for AMI patients with coexistent CHF will show the relevance of accounting for patient heterogeneity in hospital performance reporting.

2. Methods

2.1. Design, population and setting

An observational cross-sectional study utilising administrative data representing virtually all hospital admissions for AMI in patients aged from 40 to 80, treated in 434 hospitals from 5 European countries (Denmark, Portugal, Slovenia, Spain and Sweden) in 2009, totalling 73,812 potential discharged episodes. Hospitals accounting for fewer than 250 AMI episodes in 2009 (discretionary threshold) were excluded from the sample in order to reduce structural heterogeneity across hospitals and gain robustness in the estimations (Fig 1). The final sample accounted for 107 hospitals, accounting for 46,875 AMI episodes (63.5% of all assisted episodes), from which 5.2% deceased (2,451 case-fatalities). CHF coexisted in 13,5% of the sample (6,314 AMI episodes).

2.2. Endpoints

Our work comprised two consecutive endpoints; firstly, the variation in the hospital effect (i.e. GCE) when including a random slope for CHF patients in the MLRM; and, additionally, the level of agreement in hospital outcomes, contrasting both types of hospital GCE (i.e. under the assumption that the underlying risk for an individual level association is the same for all the hospitals or under the assumption that the underlying risk for CHF patients varies across hospitals).

2.3. Variables in the models

As aforementioned, the hospital outcome in this study (i.e. proxy of performance measure) was the adjusted in-hospital mortality risk in AMI patients who stayed for up to 30 days after admission; thus, inpatients with admission diagnosis code 410* in those countries using ICD-MC 9th (Spain and Portugal) and I21* and I22* in those countries using ICD 10th (Denmark, Slovenia and Sweden). Those admissions due to pregnancy, puerperium or childbirth were excluded (codes ICD-MC 9th O00*-O99* or ICD10th 630–677). [detailed in S1 Appendix]

The patient-level independent variables were: a) *age*, categorized as 40–49, 50–59, 60–69 and 70–80, using the youngest group as the reference group; b) *sex*, using male as the reference

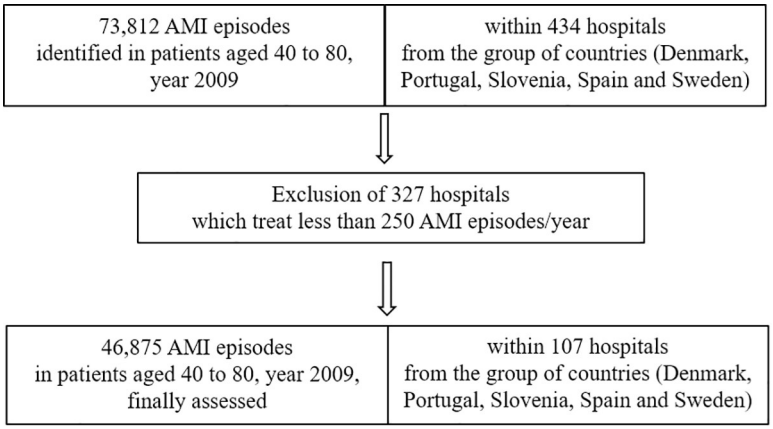


Fig 1. Study population. Flow diagram showing the episodes with an Acute myocardial infarction diagnosis according to hospital selection.

https://doi.org/10.1371/journal.pone.0228425.g001

category; c) patient comorbidities, computed as an Elixhauser risk score [24, 25], obtained from the predicted probability of death for each of the episodes modelled with a single level logistic regression; and d) the coexistence of congestive heart failure (CHF) in the episode of AMI. Patients with CHF constituted the subgroup of patients of interest, being more fragile than the regular patient and supposedly with the requirement for a higher intensity of care. A CHF was flagged when ICD9th codes 398.91 and 428*, and ICD10th code I50*, were found in any secondary diagnosis recorded within the same episode. The definitions and corresponding codes were developed and validated in the context of the ECHO project [26]. When it comes to the hospital-level, no specific variables where included except the GCE captured as a random-effect. Finally, a dummy variable identifying the country of admission was included using Sweden as a reference in the comparisons.

2.4. Analyses

Upon the estimation of the basal risk of death associated to patient features and country of residence (basal model) throughout a conventional single level logistic regression model including age, sex, the Elixhauser score of risk, the coexistence of CHF, and the country of "treatment" (see variable definitions above), two MLRM models were built to estimate the hospital-specific risk of death for patients with AMI. For that purpose, we followed the methodology described elsewhere in a two-stage process [15].

The first MLRM specification included a random intercept for the hospital level in a two-level multilevel logistic regression model, so that each hospital got its own intercept (i.e. basal risk of death) (Eq 1).

$$ln\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \gamma_{00} + \sum_{n=1}^{N}\gamma_{nj}X_{nj} + \sum_{k=1}^{K}\gamma_{kj}D_j + \gamma_{1j}Z_{ij} + \gamma_{2j}CHF_{ij} + u_{oj} + \varepsilon_{ij} \quad (1)$$

Where u_{oj}+ε_{ij} is the random effect part of the model

$$u_{oj} \sim iid\ N(0, \sigma_u^2)$$

$$\varepsilon_{ij} \sim iid\ N(0, \sigma_\varepsilon^2)$$

The second MLRM specification, as an extension of the previous one, included a random slope, allowing each hospital to vary their risk slope according to a specific group of patients (in our case, patients with CHF). In practice, we obtain a hospital variance for patients without CHF and a different hospital variance for patients with CHF (see Eq 2).

$$ln\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \gamma_{00} + \sum_{n=1}^{N}\gamma_{nj}X_{nj} + \sum_{k=1}^{K}\gamma_{kj}D_j + \gamma_{1j}Z_{ij} + \gamma_{2j}CHF_{ij} + u_{oj} + u_{2j}CHF_{ij} + \varepsilon_{ij} \quad (2)$$

Where

X_{nj} are the N variables characterising the gender and age of patients

Z_{ij} is the probability of death for a patient according to the concurrence of Elixhauser comorbidities, except CHF

$$Z_{ij} = ln\left(\frac{p_i}{1-p_i}\right) = \gamma_{00} + \sum_{k=1}^{K}\gamma_{k0}x_{ki} + \varepsilon_i$$

D_j are dichotomous variables which identify the countries where hospitals belong

$u_{oj}+u_{2j}\ CHF_{ij}+\varepsilon_{ij}$ is the random effect part of the model

$$\begin{pmatrix} u_{oj} \\ u_{2j} \end{pmatrix} \sim iid\ N\left(0, \Omega_u^2\right),\ \Omega_u = \begin{pmatrix} \sigma^2_{u0} \\ \sigma^2_{u0}\sigma^2_{u2} \end{pmatrix}$$

$$\varepsilon_{ij} \sim iid\ N(0, \sigma^2_{\varepsilon})$$

2.4.1. Estimation of the general contextual effect. The GCE was estimated for both models, the random intercept model and the extended model which adds a random slope. For both models, the hospital variance derivatives, the Variance Partition Coefficient (VPC) and the Median Odds Ratio (MOR), were also calculated.

(i) We calculated the VPC based on the latent response formulation of the model as [21, 22, 27]:

$$VPC = \frac{\sigma^2_u}{\sigma^2_u + \frac{\pi^2}{3}}$$

Where σ^2_u denotes the hospital variance, and $\frac{\pi^2}{3}$ the variance of a standard logistic distribution ($\pi = 3.1416$).

VPC is reported as a percentage that goes from 0% to 100%. If hospital differences (i.e. variance) were not relevant for understanding the individual differences in the latent propensity of death, the VPC would be 0%. That is, the hospitals would be similar to random samples taken from the whole patient population.

(ii) The median odds ratio (MOR) is an alternative interpretation of the magnitude of hospital variance [28–30]. The MOR is defined as the median value of the distribution of odds ratios (OR) obtained when randomly picking two patients with the same covariate values from two hospitals with a different underlying risk of an event of interest, and comparing the one from the hospital with the higher risk to the one from the hospital with the lower-risk. In simple terms, the MOR can be interpreted as the median increased odds of reporting the outcome if a patient is treated in another hospital with a higher risk. The MOR is calculated as:

$$\exp(\sqrt{2\sigma^2_u}\Phi^{-1}(0.75))$$

where $\Phi^{-1}(\cdot)$ represents the inverse cumulative standard normal distribution function. In the absence of hospital variation (i.e. $\sigma^2_u = 0$), the MOR is equal to 1. Theoretically, the MOR values may extend from 0 to ∞ and the higher the MOR value, the more relevant the hospital effect in terms of patient outcome. The MOR translates the hospital variance estimated on the log-odds scale to the widely used OR scale, making MOR values comparable to the individual OR covariates in the model.

For the estimation of the models, we used the Restricted Iterative Generalized Least Squares (RIGLS) method to obtain the values needed to finally run the Markov Chain Monte Carlo (MCMC) estimation method [31]. The goodness-of-fit of the models was assessed through the Bayesian Diagnostic Information Criterion (BDIC).

We performed the analyses using MLwiN run on Stata® statistical software: Release 13, College Station, TX: StataCorp LP and MlwiN version 2.35, The Centre for Multilevel Modelling, University of Bristol [32].

2.4.2. Concordance in hospital performance. Finally, for the assessment of concordance (i.e. agreement in hospital performance on patients with and without CHF), we compared the residuals from both random parts in the extended model, the intercept [u_{oj}] and the slope

$[u_{2j}]$. The level of concordance between both residuals was studied using a measurement of agreement between observers for categorical variables. As the number of cases per country varied substantially, a weighted Kappa Index was estimated [33]. The choice of this estimator depends on how commonly performance measurements are reported through funnel plots, so units of analysis are categorized as: hospitals with residuals which are statistically above the average (i.e. exhibiting a higher risk of death than expected), hospitals with residuals which are statistically below the average (i.e. exhibiting a lower risk of death than expected), and hospitals that did not differ statistically from the expected risk of death, irrespective of their actual position above or below (i.e. hospitals within the funnel boundaries). According to this approach, hospitals in the sample were classified into three possible situations: better, neutral or worse than the expected, this categorization becoming the subject of the concordance measurement. As for interpretation purposes, the higher the Kappa Index, the higher the concordance between the two estimated hospital effects, which could suggest that hospitals perform equally in the patients without CHF as in the patients with CHF. Conversely, low concordance could suggest that hospitals perform differently.

2.5. Data sources

Hospital admissions from Denmark, Portugal, Slovenia and Spain were extracted from the database consolidated and validated during the ECHO project [30]. In turn, the Swedish Patient Register [34, 35] provided the hospital data for Sweden. Both pseudonymised datasets were linked into a single database, stored, validated and analysed in a secure server set up in the premises of the Faculty of Medicine at Lund University (Malmo, Sweden), as foreseen in the access policies of the Swedish Register.

2.6. Ethics statement

This study, observational in design, used retrospective anonymized, non-identifiable and non-traceable data, and was conducted in accordance with the amended Helsinki Declaration, the International Guidelines for Ethical Review of Epidemiological Studies, and Spanish laws on data protection and patients' rights. The study implies the use of pseudonymised data, using double dissociation (i.e. in the original data source and once the data are stored in the database for analysis) which actually impedes patients' re-identification. The information supplied for the European collaboration presented the same strong characteristics of confidentiality as the other collaborating countries.

3. Results

The final sample was composed of 46,875 episodes with a primary admission diagnosis of AMI, discharged from 107 hospitals. Overall, 6,314 patients underwent a concomitant CHF. By countries, Denmark treated 4,635 of those AMI episodes in 6 hospitals (9.9% of the episodes in the sample); Portugal accounted for 6,217 from 16 hospitals (13.3% of the episodes), while Slovenia yielded 1,898 episodes in 3 of the hospitals (4.1% of the admissions analysed). In turn, Spain treated 23,043 AMI episodes in 56 hospitals (49.2% of the episodes in the sample) while Sweden dealt with 11,082 of the AMI episodes in 26 hospitals (23.6% of the episodes).

The sample had 38.2% of patients aged 70 to 80, varying across countries, with 33.2% in Denmark and 42.5% in Sweden. Overall, 26.4% of the patients were female, ranging from 23.6% in Spain to 30.6% in Sweden. The average risk score (i.e. predicted probability of death according to the Elixhauser comorbidities) for the whole sample was 5.2, ranging from 4.5 in

Table 1. Description of the study sample, per country (2009).

	DNK	PRT	SVN	ESP	SWE	TOTAL
AMI patients (n)	4,635	6,217	1,898	23,043	11,082	46,875
Age distribution						
% patients in age group 40–49	11.28%	11.57%	10.75%	12.51%	6.41%	10.75%
% patients in age group 50–59	22.37%	22.33%	26.45%	22.21%	17.08%	21.20%
% patients in age group 60–69	33.12%	28.55%	27.34%	27.65%	34.03%	29.80%
% patients in age group 70–80	33.23%	36.56%	35.46%	37.63%	42.48%	38.24%
Sex—Percentage of women	26.11%	28.57%	29.72%	23.65%	30.64%	26.44%
Risk Score (mean)	4.51	5.77	5.55	5.47	4.67	5.23
minimum	4.16	4.16	4.16	4.16	4.16	4.16
maximum	68.09	73.04	57.75	76.85	68.34	76.85
CHF patients per hospital (mean)*	78	56	79	58	57	59
minimum	38	19	35	13	28	13
maximum	114	141	117	197	139	197
Deceased (n)	206	432	84	1,286	443	2,451
In-hospital Crude Mortality Rate (per 100 patients at risk in 2009)	4.82	6.91	4.16	5.63	3.99	5.34
minimum	2.83	3.72	1.77	0.46	1.11	0.46
maximum	10.30	13.09	7.17	9.68	7.12	13.09
HOSPITALS	6	16	3	56	26	107
Episodes per hospital (mean)	772	389	633	412	426	438
minimum	505	254	283	258	250	250
maximum	951	604	1015	753	1054	1054

* CHF stands for Congestive Heart Failure

https://doi.org/10.1371/journal.pone.0228425.t001

Denmark to 5.8 in Portugal. Finally, the overall proportion of AMI patients with congestive heart failure was 59%, ranging from 56% in Portugal to 79% in Slovenia (Table 1).

Overall, 5.3 per 100 AMI patients died in hospital (2,451 cases out of 46,875) in the period of study; the crude mortality rate ranged from 0.5 to 13.1 per 100 patients at risk, for an inter-quartile interval of 1.63. By countries, Sweden, Slovenia and Denmark showed the lowest in-hospital mortality rates, 4, 4.2 and 4.8 per 100 patients at risk respectively, while Portugal showed the highest with 6.91 per 100 patients at risk, followed by Spain with an in-hospital mortality rate of 5.6 per 100 patients at risk (Table 1).

Table 2 shows the estimated adjusted-risks of death in the basal model, the basic GCE and the extended RS model. As observed in the basal model, the AMI risk of death increased with age (as compared to patients younger than 50), with the highest risk amongst the oldest (4.8 times more likely to die), the presence of comorbidities (2.1 times more likely), and the coexistence of CHF (2.8 times more likely). As compared to Sweden, patients living in Portugal were at 78% more risk of death, Denmark and Spain showing a 36% increased risk, while Slovenia barely registered a 6% increase. Being female did not increase the risk of death. Patient-level and country-level estimates were similar in both MLRM (second and third column in Table 2).

Both MLRM models confirmed the existence of a GCE; thus, beyond individuals' features, we observed an increase in the risk of death associated to the hospital of treatment. Moreover, in the specific case of the extended model with a random slope (the best model according to BDIC), the GCE was much higher in CHF patients, [VPC of 8.34 (CI95% 4.94 to 13.03) and a MOR value of 1.69 (CI95% 1.62 to 2.21)] than in those without CHF [VPC = 3.9 (CI95% 2.4 to 5.9), MOR of 1.42 (CI95% 1.31 to 1.54)].

Table 2. Factors associated to in-hospital mortality in AMI patients (2009).

	Basal model (single level)	MLRM (random intercept)	MLMR (random slope)
Gender			
Male	Ref	Ref	Ref
Female	1.09	1.08	1.08
	(0.99 1.19)	(0.98 1.18)	(0.98 1.18)
Age			
40–49	Ref	Ref	Ref
50–59	1.68	1.71	1.70
	(1.29 2.19)	(1.33 2.25)	(1.34 2.12)
60–69	2.61	2.63	2.63
	(2.03 3.35)	(2.08 3.44)	(2.10 3.20)
70–80	4.84	4.84	4.87
	(3.80 6.18)	(3.87 6.24)	(3.87 5.91)
Risk score [1]			
Healthier patients	Ref	Ref	Ref
More complex patients	2.07	2.11	2.12
	(1.90 2.27)	(1.93 2.30)	(1.93 2.33)
Congestive heart failure (CHF)			
	2.80	2.88	2.84
	(2.56 3.07)	(2.64 3.14)	(2.45 3.24)
Country			
Sweden	Ref	Ref	Ref
Denmark	1.36	1.46	1.41
	(1.14 1.61)	(1.03 2.01)	(1.00 1.96)
Portugal	1.78	1.83	1.88
	(1.55 2.05)	(1.37 2.31)	(1.43 2.42)
Slovenia	1.06	1.01	1.07
	(0.83 1.36)	(0.61 1.85)	(0.66 1.80)
Spain	1.36	1.42	1.38
	(1.21 1.52)	(1.14 1.73)	(1.10 1.66)
Hospitals			
MOR on patients w/o CHF		1.40	1.42
MOR on CHF patients			1.69
Variance			
Hospital intercept (95% credible interval)		0.12 (0.08 0.18)	0.13 (0.08 0.21)
CHF-mortality slope (95% credible interval)			0.31 (0.17 0.49)
Variance Partition Coefficient			
Patients without CHF		3.59	3.9
Patients with CHF			8.34
Goodness of fit			
Bayesian DIC		17384.88	17317.35

Models: Estimations in the basal model are obtained from single-level logistic modelling; estimations of the general contextual effect are obtained from MLRM modelling hospitals as random effect, by first just modelling a random intercept, then adding a random slope for patients with CHF. Figures represent Odds Ratios and 95% confidence intervals, except in the case of hospital estimates where MOR and confidence intervals are used. Note that, while in the random intercept model there is one value for MOR, in the model with random slope we obtain a different MOR for patients with and for patients without CHF.

https://doi.org/10.1371/journal.pone.0228425.t002

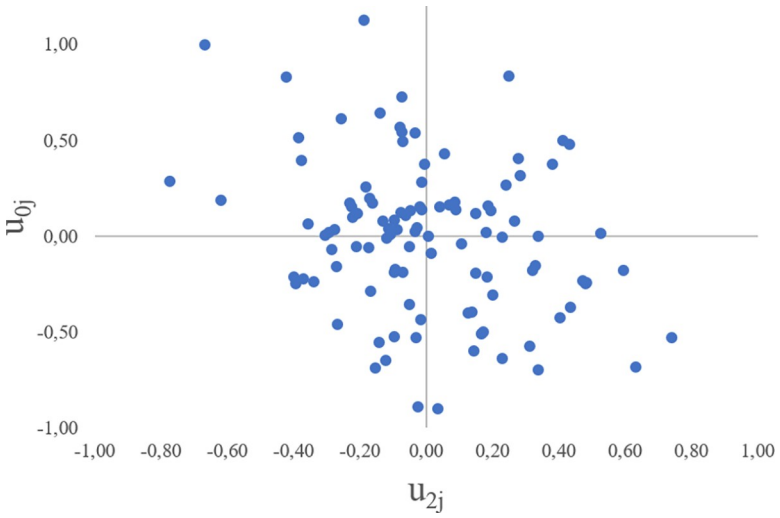


Fig 2. Comparison of the hospital effect for the patient without CHF and the hospital effect conditioned to CHF patients. Weighted Kappa Index -0,02 (CI95% -0,08 to 0.04).

https://doi.org/10.1371/journal.pone.0228425.g002

3.1. Is the hospital effect consistent between estimations?

Once was the existence of hospital variance and the observation of a better goodness-of-fit of the model with random slope examined and proved, its residuals were compared. Fig 2 represents the comparison of the hospital effect for patients without CHF [u_{oj}] and the hospital effect conditioned to the coexistence of CHF [u_{2j}].

Once hospitals were classified in accordance with their level of performance (high, moderate or low), the agreement in the classification of hospital performance was non-existent [weighted Kappa Index value of -0,02 (CI95% -0,08 to 0.04)] suggesting a distinct performance in both groups of patients.

4. Discussion

Assuming the construct validity of AMI case-fatalities as a measure of hospital quality, this performance assessment exercise, based on 46,875 hospital admissions from five countries, shows that hospital outcomes differ when it comes to specific subgroups of patients, in our case, patients with CHF. Indeed, the greater MOR for the model including a random slope (i.e. assuming an interaction term for patients with CHF) reveals a greater influence of "hospital of treatment" when it comes to the case mortality rates for CHF patients (from MOR 1.42 in patients without CHF to MOR 1.69 with CHF).

Finally, the lack of correlation between the hospital effects on the non-CHF AMI patients and AMI patients with CHF (weighted Kappa Index = -0.02), prompts the need for analysing hospital effects on regular and specific subgroups of patients.

4.1. Caveats with regard to the lack of concordance

Despite the mathematical robustness of the results in terms of goodness-of-fit of the model and precision, two questions might be challenging the lack of concordance between the hospital effect on non-CHF vs. CHF AMI patients.

We could hypothesize, for example, that systemic factors could affect the GCE estimations distinctly, if the number of CHF patients per hospital is uneven across the sample (e.g. because of biased coding practices, because more complex patients arrive at certain hospitals, or

because of differential expertise in the treatment of more fragile patients between centres). Although we have reduced this potential risk by excluding the smaller hospitals from the sample, if those phenomena are true they could have an influence on the estimations of the hospital contextual effect in the specific subgroup of patients, resulting in a higher risk of death associated to the hospital of treatment in those hospitals with more CHF patients. Fig 3 showing the potential correlation between the prevalence of high-risk patients (x axis) and the estimated risk in terms of u_{2j}(y axis) shows that this is not the case for the hospitals in the sample, ruling out this possibility.

Another point that could eventually affect the hospital contextual effect differently on non-CHF vs. CHF patients is the surviving bias in those with no concurrent CHF. Indeed, AMI patients with concomitant CHF (most of them STEMI cases) are supposed to be more likely to die within the first 24 hours. In these cases, patients might die in the emergency room. After analysing the survival curves for both groups of patients, the negligible differences observed in the first 24 hours after in-hospital admission strongly suggest that under-recording is not likely happened in our sample [S2 Appendix shows survival curves for each country]. However, as patients who died in the emergency room are not part of our dataset, we cannot discard some under-representation in those CHF patients. Whether this fact could imply any bias in the estimation is unknown.

4.2. Implication of the use of random-slope MLRM in hospital performance assessment

In contrast with single level estimations, MLRM takes into account the multilevel structure of the variance existing in the data (e.g. patients nested within hospitals), accounting for the interdependence of patient outcomes within a hospital and allowing a less biased estimation of uncertainty, providing weighted estimations of average hospital risk (i.e. shrunken residuals) and allowing a more reliable assessment of the units under study.

As compared to GEE, both MLRM and GEE assume the existence of a GCE assessing hospital performance. This contextual effect is termed "general" because it reflects the influence of the hospital context as a whole, without specifying any contextual characteristics other than the very boundaries that delimit the hospital [36]. This GCE expresses the joint effect of an array of factors like, for instance, the skills and specialization of the physicians, the available access to adequate technology as well as the quality of treatment and care in the hospital. In

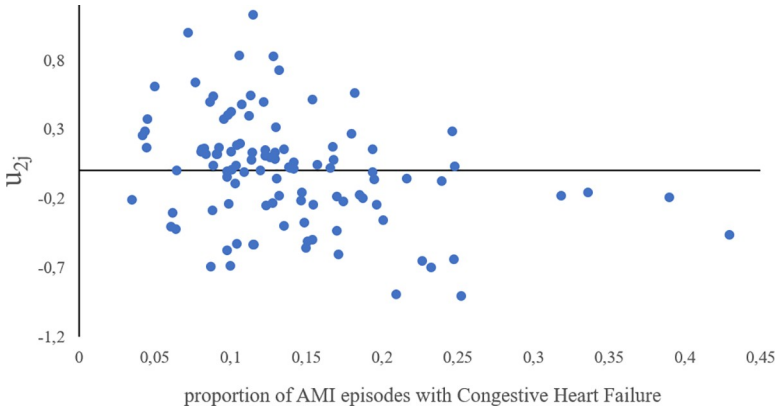


Fig 3. Correlation between the prevalence of CHF patients and the hospital effect conditioned to those patients (u2j). This graph contrasts the possibility of a higher hospital contextual effect due to the higher prevalence of CHF patients admitted in the hospitals in the sample. No positive correlation is observed.

https://doi.org/10.1371/journal.pone.0228425.g003

such a way, the hospital context may condition patient outcomes beyond individual characteristics; that is, the same patient might have a different outcome if he or she is treated in a different hospital. However, while GEE modelling takes for granted that this GCE can be quantified by measuring differences between hospital averages only, in MLRM the GCE is measured by the share of the total patient variance that is between hospital averages; that is, the MLRM does not dislocate the individual patients from the hospital averages, but rather considers that there is a distribution of individual outcomes that can be decomposed into two levels of analysis, the individual and the hospital [9, 12, 14]. Therefore, "hospital effects" (i.e. GCE) are not properly appraised by studying the differences between hospital averages alone, but by quantifying the share of the total patient heterogeneity (i.e. variance) that exists at the hospital level [37–39]. To do this, MLRM estimates the hospital variance and its derivatives partition coefficient as a measure of the hospital GCE. Thus, when studying a specific quality outcome in patients from different hospitals, the higher the hospital variance, the more relevant is the hospital context to understanding the differences between patient outcomes [12].

More importantly, unlike other methods used to analyse clustered information (i.e. patients nested within hospitals) MLRM considers individual-level associations to be hospital-specific and drops the assumption that individual level associations are the same for all the hospitals. Consequently, in an extended MLRM with RS, hospital variance, and thereby hospital GCE, becomes a function of the patients' heterogeneity. In other words, by including a RS for a specific subgroup of patients, the hospital effect is not just a function of the very boundaries of the hospital but also a function of patients' features of interest (i.e. in our case having CHF). In practice, for a dichotomous variable, we obtain a hospital variance (i.e. a GCE) for patients without CHF and a different hospital variance for patients with CHF. This becomes, beyond considerations of interpretation, the analytical advantage of MLRM as opposed to GEE modelling.

4.3. Implications for hospital performance reporting

Some authors have already suggested, while acknowledging the risk of using indirect standardization in hospital performance assessment [20, 22] or in the context of social epidemiology [23], that not considering patient heterogeneity could lead to an inappropriate assessment of performance. This paper empirically underpins the need for exploring both the hospital effect for patients with or without CHF.

Therefore, a clear message is conveyed to those interested in the public reporting of performance measures. Beyond the assumption that performance assessment using administrative data is not a firm diagnostic tool but rather an instrument for screening, reporting mechanisms, more specifically league tables or funnel plots, [9, 18] should represent hospital performance according to the results of the MLRM. If the model without a random slope prevails (which is not the case in our example), a single representation for the average patient might be enough; however, if a MLRM with random slope better explains the difference in hospital outcomes, then public reporting should represent hospital effects separately for specific subgroups of patients.

One last important implication for decision-makers is that MLRM provides a measure of the effect of size (i.e. to what extent the hospital contextual effect is relevant to the differences in health outcomes) through a number of statistics (hospital variance, variance partition coefficients, and MOR) not yielded by the popular indirect standardization methods or the GEE models. This feature makes MLRM findings more actionable than other approaches.

5. Conclusions

The hospital contextual effect in 107 hospitals from five different European countries was different in non-CHF AMI patients and AMI patients with CHF, suggesting that accounting for

patient heterogeneity should be a requirement for adequately characterising and reporting hospital performance.

MLRM is flexible enough to allow the joint analysis of both overall effects and patient-specific hospital effects, providing accurate estimations of performance as well as a measure of the actual relevance of the hospital contextual effect.

Supporting information

S1 Appendix. Description of selected codes. Inclusion and exclusion criteria and codes for episode selection.
(DOCX)

S2 Appendix. Survival curves testing differential underreporting in CHF patients.
(DOCX)

Acknowledgments

We are indebted to Mircha Poldrugovac as a member of the Slovenian team in ECHO and Natalia Martínez-Lizaga for her support in the extraction and preparation of data from the ECHO dataset for the purposes of this study, and also to the Spanish Health Authorities who granted access to the hospital datasets.

ECHO Consortium: IACS's team (Bernal-Delgado E, García-Armesto S, Martínez-Lizaga M, Comendeiro-Maaløe M, Seral-Rodríguez M, Estupiñán-Romero F, Angulo-Pueyo E, Ridao-López M, and Baixaulí C, Librero J as affiliated researchers), University of Southern Denmark's team (Christiansen T, Thygesen LC), University of Nova Lisboa's team (Mateus C, Nunes C, Joaquim I), National Institute of Public Health of Ljubljana's team (Yazbeck AM, Galsworthy M, Albreht T), UMIT's team (Munck J, Güntert B) and EHMA's team (Bremmer J, Giepmans P, Dix O).

Author Contributions

Conceptualization: Micaela Comendeiro-Maaløe, Juan Merlo, Enrique Bernal-Delgado.

Data curation: Micaela Comendeiro-Maaløe, Francisco Estupiñán-Romero, Juan Merlo.

Formal analysis: Micaela Comendeiro-Maaløe.

Investigation: Francisco Estupiñán-Romero, Lau Caspar Thygesen, Céu Mateus, Enrique Bernal-Delgado.

Methodology: Micaela Comendeiro-Maaløe, Juan Merlo, Enrique Bernal-Delgado.

Project administration: Enrique Bernal-Delgado.

Resources: Francisco Estupiñán-Romero, Lau Caspar Thygesen, Céu Mateus.

Supervision: Juan Merlo, Enrique Bernal-Delgado.

Validation: Micaela Comendeiro-Maaløe, Lau Caspar Thygesen, Céu Mateus, Juan Merlo, Enrique Bernal-Delgado.

Visualization: Micaela Comendeiro-Maaløe, Francisco Estupiñán-Romero.

Writing – original draft: Micaela Comendeiro-Maaløe, Enrique Bernal-Delgado.

Writing – review & editing: Micaela Comendeiro-Maaløe, Francisco Estupiñán-Romero, Lau Caspar Thygesen, Céu Mateus, Juan Merlo, Enrique Bernal-Delgado.

References

1. Institute for Clinical Evaluative Sciences (ICES). Canada. http://www.ices.on.ca/Publications/Atlases-and-Reports (10 January 2019, date last accessed)
2. Atlas of Variations. Public Health England. https://fingertips.phe.org.uk/profile/atlas-of-variation (10 January 2019, date last accessed)
3. Agency for Health Research and Quality (ARQH). US Department of Health & Human Services, USA
4. Health Quality and Safety Commission, New Zealand. http://www.hqsc.govt.nz/our-programmes/health-quality-evaluation/projects/atlas-of-healthcare-variation/ (10 January 2019, date last accessed)
5. National Institute for Public Health and the Environment, The Netherlands. https://www.volksgezondheidenzorg.info/onderwerp/atlas-vzinfo/inleiding (10 January 2019, date last accessed)
6. The Atlas of Variation in the Spanish National Health Service, Instituto Aragonés de Ciencias de la Salud, IIS Aragón, Spain Available at: www.atlasvpm.org (10 January 2019, date last accessed)
7. OECD Health Care Quality Indicators Project, Paris http://www.oecd.org/els/health-systems/health-care-quality-indicators.htm (10 January 2019, date last accessed)
8. The Dartmouth Atlas of Healthcare, The Dartmouth Institute, USA. Available at: http://www.dartmouthatlas.org (10 January 2019, date last accessed)
9. Ohlsson H, Librero J, Sundquist J, Sundquist K, Merlo J: Performance evaluations and league tables: do they capture variation between organizational units? An analysis of 5 Swedish pharmacological performance indicators. Med Care 2011, 49(3):327–331. https://doi.org/10.1097/MLR.0b013e31820325c5 PMID: 21263360
10. European project on healthcare outcomes, performance and efficiency. http://www.eurohope.info/ (10 January 2019, date last accessed)
11. Bernal-Delgado E, Christiansen T, Bloor K, et al. on behalf of ECHO Consortium ECHO: Health care performance assessment in several European countries. European Journal of Public Health, Volume 25, Issue suppl_1, 1 February 2015, Pages 3–7, https://doi.org/10.1093/eurpub/cku219
12. Merlo J, Ostergren PO, Broms K, Bjorck-Linne A, Liedholm H. Survival after initial hospitalisation for heart failure: a multilevel analysis of patients in Swedish acute care hospitals. J Epidemiol Community Health. 2001; 55(5):323–9. https://doi.org/10.1136/jech.55.5.323 PMID: 11297650
13. Merlo J, Gerdtham UG, Eckerlund I, et al. Hospital level of care and neonatal mortality in low- and high-risk deliveries: reassessing the question in Sweden by multilevel analysis. Med Care. 2005; 43 (11):1092–100. https://doi.org/10.1097/01.mlr.0000182484.14608.b9 PMID: 16224302
14. Ghith N, Frolich A, Merlo J. The role of the clinical departments for understanding patient heterogeneity in one-year mortality after a diagnosis of heart failure: A multilevel analysis of individual heterogeneity for profiling provider outcomes. PLoS One. 2017; 12(12):e0189050. https://doi.org/10.1371/journal.pone.0189050 PMID: 29211785
15. Ghith N, Wagner P, Frolich A, Merlo J. Short Term Survival after Admission for Heart Failure in Sweden: Applying Multilevel Analyses of Discriminatory Accuracy to Evaluate Institutional Performance. PLoS One. 2016; 11(2):e0148187. https://doi.org/10.1371/journal.pone.0148187 PMID: 26840122
16. Ash AS, Fienberg SE, Louis TA, Normand SLT, Stukel TA, Utts J. Statistical issues in assessing hospital performance. Commissioned by the Committee of Presidents of Statistical Societies. The COPSS-CMS White Paper Committee: 2011
17. Mohammed MA, Manktelow BN, Hofer TP. Comparison of four methods for deriving hospital standardised mortality ratios from a single hierarchical logistic regression model Stat Methods Med Res. 2016 Apr; 25(2):706–15.
18. Goldstein H, Spiegelhalter D. League Tables and Their Limitations: Statistical Issues in Comparisons of Institutional Performance. Journal of the Royal Statistical Society, SocA. 1996; 159:385–443.
19. Austin PC, Tu JV, Alter DA. Comparing hierarchical modeling with traditional logistic regression analysis among patients hospitalized with acute myocardial infarction: should we be analyzing cardiovascular outcomes data differently? Am Heart J. 2003; 145(1):27–35. https://doi.org/10.1067/mhj.2003.23 PMID: 12514651
20. Marang-van de Mheen PJ, Shojania KG. Simpson's paradox: how performance measurement can fail even with perfect risk adjustment. BMJ Qual Saf 2014; 23:701–705. https://doi.org/10.1136/bmjqs-2014-003358 PMID: 25118292
21. Sanagou M, Wolfe R, Forbes A, Reid CM. Hospital-level associations with 30-day patient mortality after cardiac surgery: a tutorial on the application and interpretation of marginal and multilevel logistic regression BMC Medical Research Methodology 2012, 12:28 https://doi.org/10.1186/1471-2288-12-12

22. Pouw ME, Peelen LM, Lingsma HF, et al. (2013) Hospital Standardized Mortality Ratio: Consequences of Adjusting Hospital Mortality with Indirect Standardization. PLoS ONE 8(4): e59160. https://doi.org/10.1371/journal.pone.0059160 PMID: 23593133

23. Merlo J, Yang M, Chaix B, Lynch J, Rastam L. A brief conceptual tutorial on multilevel analysis in social epidemiology: investigating contextual phenomena indifferent groups of people. J Epidemiol Community Health 2005, 59(9):729–736 https://doi.org/10.1136/jech.2004.023929 PMID: 16100308

24. Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data. *Med Care* 1998; 36:8–27 https://doi.org/10.1097/00005650-199801000-00004 PMID: 9431328

25. Elixhauser and Quan's ICD-9-CM and ICD-10 Coding Algorithms for Elixhauser Comorbidity Index. *Med Care* 2005 Nov; 43(11): 1130–9. https://doi.org/10.1097/01.mlr.0000182534.19832.83 PMID: 16224307

26. European Collaboration for Healthcare Optimization (ECHO) www.echo-health.eu. Zaragoza (Spain): Instituto Aragonés de Ciencias de la Salud—Instituto Investigación Sanitaria Aragón; c2011. Estupiñán F, Baixauli C, Bernal-Delgado E on behalf of the ECHO consortium. Handbook on methodology: ECHO information system quality report; 2014 Apr 27 [accessed: 10 January 2019]; Available from: http://www.echo-health.eu/handbook/infrastructure.html

27. Goldstein H, Browne W, Rasbash J. Partitioning variation in generalised linear multilevel models. *Understanding Statistics* 2002, 1:223–232.

28. Merlo J, Chaix B, Ohlsson H, et al. A brief conceptual tutorial of multilevel analysis in social epidemiology: using measures of clustering in multilevel logistic regression to investigate contextual phenomena. J Epidemiol Community Health 2006, 60(4):290–297. https://doi.org/10.1136/jech.2004.029454 PMID: 16537344

29. Larsen K, Merlo J. Appropriate assessment of neighbourhood effects on individual health: integrating random and fixed effects in multilevel logistic regressions. Am J Epidemiol 2005, 161(1):81–88. https://doi.org/10.1093/aje/kwi017 PMID: 15615918

30. Larsen K, Petersen JH, Budtz-Jorgensen E, Endahl L. Interpreting parameters in the logistic regression model with random effects. *Biometrics* 2000, 56(3):900–914.

31. Browne WJ MCMC Estimating in MLwiN v2.29. Center for Multilevel Modelling, University of Bristol 2013

32. Leckie G, Charlton C. runmlwin—A Program to Run the MLwiN Multilevel Modelling Software from within Stata. Journal of Statistical Software, (2013); 52 (11):1–40]

33. Cohen J. A coefficient of agreement for nominal scales". Educational and Psychological Measurement 1960; 20 (1): 37–46. https://doi.org/10.1177/001316446002000104

34. https://www.socialstyrelsen.se/register/halsodataregister/patientregistret/inenglish

35. Ludvigsson JF, Andersson E, Ekbom A, et al. External review and validation of the Swedish national inpatient register. BMC Public Health. 2011; 11:450. https://doi.org/10.1186/1471-2458-11-450 PMID: 21658213

36. Merlo J, Chaix B, Yang M, Lynch J, Rastam L. A brief conceptual tutorial of multilevel analysis in social epidemiology: linking the statistical concept of clustering to the idea of contextual phenomenon. *J Epidemiol Community Health* 2005, 59(6):443–44 https://doi.org/10.1136/jech.2004.023473 PMID: 15911637

37. Merlo J. Multilevel analytical approaches in social epidemiology: measures of health variation compared with traditional measures of association. *J Epidemiol Community Health* 2003, 57(8):550–552. https://doi.org/10.1136/jech.57.8.550 PMID: 12883048

38. Merlo J. Invited commentary: multilevel analysis of individual heterogeneity-a fundamental critique of the current probabilistic risk factor epidemiology. *American Journal of Epidemiology* 2014, 180(2):208–212. https://doi.org/10.1093/aje/kwu108 PMID: 24925064

39. Merlo J, Ohlsson H, Lynch KF, Chaix B, Subramanian SV. Individual and collective bodies: using measures of variance and association in contextual epidemiology. *J Epidemiol Community Health* 2009, 63 (12):1043–1048. https://doi.org/10.1136/jech.2009.088310 PMID: 19666637

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2020.2971146, IEEE Journal of Biomedical and Health Informatics

Construction of Empirical Care Pathways Process Models from Multiple Real-World Datasets

Juan González-García, Carlos Tellería-Orriols, Francisco Estupiñán-Romero, Enrique Bernal-Delgado

***Abstract*—Care pathways (CPWs) are "multidisciplinary care plans that detail essential care steps for patients with specific clinical problems". While CPWs impact on health or cost outcomes is vastly studied, an in-depth analysis of the real-world implementation of the CPWs is an area that still remains underexplored. The present work describes how to apply an existing process mining methodology to construct the empirical CPW process models. These processmodels are a unique piece of information for health services research: for example to evaluate their conformance against the theoretical CPW described on clinical guidelines or to evaluate the impact of the process in health outcomes.**

To this purpose, this work relies on the design and implementation of a solution that a) synthesizes the expert knowledge on how health care is delivered within and across providers as an activity log, and b) constructs the CPW process model from that activity log using process mining techniques. Unlike previous research based on ad hoc data captures, current approach is built on the linkage of various heterogeneous real-world data (RWD) sets that share a minimum semantic linkage. RWD, defined as secondary use of routinely collected data as opposite to ad hoc data extractions, is a unique source of information for the CPW analysis due to its coverage of the caregiving activities and its wide availability. The viability of the solution is demonstrated by constructing the CPW process model of Code Stroke (Acute Stroke CPW) in the Aragon region (Spain).

***Index Terms*—Real-World Data, Process Mining, Data Mining, Health System Research, Code Stroke, Comparative Effectiveness Research**

I. Introduction

CARE pathways (CPWs) (also called clinical pathways, integrated care pathways or care maps) are "multidisciplinary care plans that detail essential care steps for patients with specific clinical problems" [1]. As stated in the Cochrane review [2], which included a total of 27 studies of CPW implementation, CPWs "aim to link evidence to practice and optimize clinical outcomes whilst maximising clinical efficiency". This same review concluded that “Care pathways are associated with reduced in-hospital complications and improved documentation without negatively impacting on length of stay and hospital costs”.

While CPWs benefits on health or cost outcomes have been studied and properly demonstrated in the Cochrane review [2], an in-depth evaluation of the real-world implementation of the CPWs is currently an active research area. The study of CPW implementation in the health care has evolved from basic before/after comparison of some variables of interest of works such as [3], to a more complex analysis to answer questions such as which is the adequacy of the real-world implementation of the CPW with respect to the theoretical or normative definition of the CPW, as in [4], or what is the effect of actual CPW implementation in terms of health outcomes for the different patients that traverse the pathway, as in [5]. From the computing and data science community process mining techniques have emerged as an alternative approach to solve those questions.

Process mining is a relatively new research field in computing science that combines unsupervised and supervised data mining methods with business process analysis techniques [6]. The goal of process mining is to discover and study the structural organization of productive processes undertaken in an organisation, namely the business process models.

In the context of health services and policy research, process mining can be used to capture the CPWs’ process model applying process discovery techniques from the logs available in their information systems. This empirical CPW process model, as it comes from RWD, can be later used to answer the questions suggested previously serving as example: it can be compared with the theoretical model, described in clinical guidelines, using conformance checking techniques; the paths or traces patients followed within the CPW, may be used to enrich a survival analysis to measure the effect of the exposition to the different paths.

The purpose of this paper is to demonstrate how to exploit the use of Real-World Data (RWD) sets for CPW analysis. RWD may be defined formally as “data used for decision-making that are not collected in conventional randomized controlled trials (RCTs)” [7] or, in other words, a secondary use of the administrative routine data collected in the health care systems as opposed to specific ad hoc data extractions. In this paper, RWD sets from a large health system are used within a process mining methodology to evaluate the actual implementation of a complex CPW. The use case selected is the Code Stroke implementation, the CPW for stroke management, in the Aragón Region health system, a public health system that comprises 9 acute hospitals covering a population of 1.3 million insurees.

This paper leverages the Process Mining Methodology (PM^2) described in [8] for the Code Stroke analysis use case. Three different RWD sets coming from three different information systems were processed to apply process discovery. Please, note that the approach presented in the paper can be applied to any other CPW whose activity is properly recorded in the routine data collected and maintained by the health care systems.

J. González-García and C. Tellería-Orriols are with the Biocomputing Unit and the Health Services and Policies Research (ARiHSP) group, Institute for Health Sciences in Aragon (IACS), Aragon, Spain and Health Services Research network on Chronic Patients (REDISSEC)

F. Estupiñán-Romero and E. Bernal-Delgado are with the Health Services and Policies Research (ARiHSP) group, Institute for Health Sciences in Aragon (IACS), Aragon, Spain and Health Services Research network on Chronic Patients (REDISSEC)

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2020.2971146, IEEE Journal of Biomedical and Health Informatics

II. Related work

A. Process mining

Process mining is a relatively new research field in computing science that can be categorized under the umbrella of the data mining methods. Its main objective is to gain insight on how corporations or institutions organize their production, i.e. understand, analyse, and, at the end, improve productive processes. The extensive digitalisation in almost all industries made the case for process mining. In this field, the book "Process Mining: Data Science in Action" [6] by Wil M.P. van der Aalst may be considered the main reference.

In brief, the process mining cornerstone are the activity logs generated by different information systems in an organisation. Activity logs are files, usually in plain text, that contain a sequence of the ordered activities that took place in the productive process. Activity logs are processed to *discover* the process model by means of algorithms such as the α-algorithm [9], the Heuristics Miner [10], the Fuzzy mining [11] or the Inductive Miner [12]. Example of process models are depicted in results section.

As detailed in [6], the output of the process discovery algorithms is a map of the *control flow* perspective of the process. Depending on the extra information available in the activity logs, for example the specific timestamp when the activities took place or the person or resources involved in the activity, a broader insight can be achieved, for example a time-related perspective, a performance perspective or organisational perspective. Even it is not the main aim of the present work, the results section contains examples for control-flow, time-related and resource perspectives.

Pursuing the improvement of the production processes, conformance checking techniques and algorithms conform also an important area of process mining research. Extensively covered in the Munoz-Gama awarded PhD. thesis [13], conformance checking techniques aim to evaluate and quantify if the process models discovered follow the expected behavior and quantify. The differences between the actual and the expected behavior may suppose an opportunity for process improvement. Quantifications may be done from basic measurements of certain parameters of the process models, as the presented in the results section, to more complex ones based on topologies and graph algorithms, such as the ones collected in the Munoz-Gama PhD, or in the work by Rozinat et. al. in [14].

Almost all algorithms in process mining field are implemented as part of the ProM tool [15], the reference process mining tool developed at the Technical University of Eindhoven. Commercial tools such as Disco [16] have done an effort to approach process mining to businesses by providing a neat interface. In the present paper the tool use for process mining tasks was `bupaR` [17], a R library for process mining. It was selected due to the interaction with all the available packages the R language offers and the possibility to create a reproducible process mining analysis workflows (note that ProM and Disco are principally interactive tools).

Last but not least, the process mining community has worked in giving a systematic approach to process mining projects by designing analysis methodologies. In this aspect, the `L*` life-cycle model [18] is one of the most accepted within process mining works. This paper follows a plus modern synthesis of the *L** life-cycle model, namely the PM^2 methodology, described in [8]. The PM^2 methodology schema is depicted in Figure 1 and is described in Section III.

B. Process mining in healthcare and health services research

The use of process mining has attracted the healthcare and, specifically, the health services research community attention since its initial steps. The 2005 work of Măruşter and Jorna in [19] is one of the very first examples of process oriented analysis. In this work authors proposed an initial approach to capture the interaction between "multidisciplinary" patients, i.e. pluri-pathogical patients that require multiple specialty care professionals, with the health system. In fact, this work appeared before the dawn of major Process Mining works and algorithms.

In recent years, up to 3 reviews covered the process mining works related to healthcare: in 2015, Yang and Su analyzed 37 works in [20]; in 2016, Rojas et. al. produced the most extensive review up to date [21], including 74 works; and, finally, Batista and Solanas reviewed 55 works in 2018 [22].

Among these three reviews, it is interesting to highlight the one by Yang and Su [20], as it explicitly focuses on the application of process mining for CPW. The conclusions of Yang and Su after reviewing 37 works pointed that there is an important limitation in the usability of process mining for CPW analysis, mainly due to the lack of structure in the processes or in the clinical data itself. The review also pointed to limitations when dealing with integrated care as, in general, the data used was limited to reduced data sets.

It is in the data structure and availability context where the current work presents a notable improvement. In the vast majority of previous works (not cited here for space limitations) the data used wasn't previously modelled and usually corresponded to a single hospital information system (HIS) or, in some other cases, ad hoc data gatherings. Few works such as [23] presents a neat data model usable for a broad spectrum of analyses. Same applies for data sets, few works analysed data from multiple sources, for instance in [5] and [24] the authors gathered data from four different hospitals that share the same HIS to evaluate the pathway underwent by patients with chest pain symptoms, or in [25] where authors also gathered an ad hoc data registry in four Italian hospital to analyse stroke CPW; and very few of them linked multiple (RWD) sets, for example in [26], the authors linked the Austrian cancer registry with a treatments database.

The present work is an explicit effort in two facets: first, to define a data model able to capture the analysis requirements to exploit the use case data using process mining; second, to effectively link RWD sets routinely collected in the health system from multiple hospitals and emergency room departments, a challenging task to provide coherence to disjoint points of view. It is interesting to note that current health services research works share this fundamental vision of extensively link and use RWD sets to enable integrated care analysis, such as the study protocol for UK-wide primary care analysis detailed by Litchfield et. al. in [27].

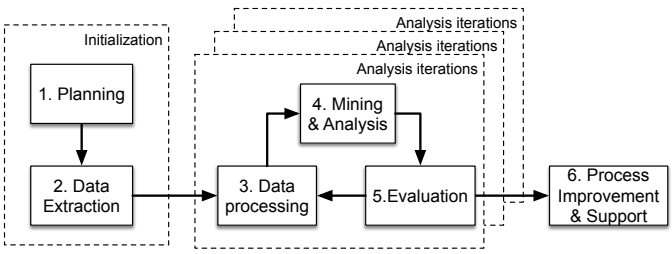


Fig. 1. PM^2 methodology schema adapted from [8]

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2020.2971146, IEEE Journal of Biomedical and Health Informatics

III. Methods

The contribution of this paper is to construct the empirical CPW process models using RWD sets by applying process mining techniques. To this purpose, the methodology used is an adaptation of the PM^2 methodology. PM^2 is a general methodology for process-mining-based projects proposed by Maikel L. van Eck in [8]. Figure 1 contains a redraw of methodology schema from the original paper. As can be seen in the Figure, the methodology has 6 major tasks, divided en 3 stages:

- "Initialization" stage: composed by Task 1 "Planning" and Task 2 "Data Extraction".
- "Analysis" stage: composed by Task 3 "Data processing", Task 4 "Mining and Analysis" and Task 5 "Evaluation".
- "Process improvement and support", a single task stage (Task 6).

In order to ease the clarity of the current section it has been organised to align with the PM^2 methodology tasks.

Note that the source code of the implemented solution, which covers mainly the "Analysis" stage tasks, is available in a GitHub repository [1]. Sample datasets are also included to permit a full reproducibility of the approach.

A. Task 1: Planning. The Code Stroke use case

The use case selected to illustrate the present work is the construction of the Code Stroke process model in the Aragón region (Spain). Code Stroke is a complex organizational intervention that aims at delivering, in time, the best treatment available for acute stroke patients. Code Stroke provides a clear decision-making algorithm, i.e. a care pathway, privileging access to defined resources and treatments when the "Code Stroke" is *activated* [28]. Stroke represents the second cause of death in general population and first in female population, being also the first case of disability, generating high social costs. Evidence [29] showed that a proper management of stroke patients may lead to a mortality risk reduction as large as 45% (Odds ratio 0.61: CI95% 0.47-0.78).

The aim of the study is to describe the overall pathway for patients with a suspicion of acute stroke and the specific pathway for patients with ischaemic stroke, deepening a bit further in the timely use of fibrinolytic treatment, as a major milestone in this care process. For this particular illustration, we retrieved the information for all suspected cases of an acute stroke episode in 2017. The episode linking algorithm of Tasks 3 determined the actual stroke cases.

This work used retrospective pseudonymised data and was conducted in accordance with the amended Helsinki Declaration, the International Guidelines for Ethical Review of Epidemiological Studies, and Spanish laws on data protection and patients' rights. The data went through a double dissociation process, i.e. in the original data source and once data is stored in the database, impeding patients re-identification.

B. Task 2: Data Extraction Task. Data model and RWD sets

1) The data model: a basic element in the data extraction task is the process model construction to identify and clarify the semantics of the data and its structure. The data model may limit the possibilities of further analyses or the comparability among different health systems once the process model has been established. To this respect, our data model was a straightforward adaptation of the data model present in the RWD sets used in this work. In different settings, the nonexistence of a clear data model or the necessity of dealing with datasets with different semantics will require an extensive pre-processing work to ensure not just the syntactic interoperability, but also the semantic interoperability.

To design the data model is important to understand an abstract structure of the CPW and its translation into the typical information systems available in the health system. In the present work, the real-world implementation of a CPW is abstracted as a set of episodes. An episode is a sequence of caregiving events related to the treatment of a given patient for an acute occurrence of a given illness, considered from an entry point (e.g., admission to an emergency room) to a final discharge (e.g. hospital discharge). The events are each individual record registered in the information system or the database of a given healthcare service (e.g., emergency room information system or a hospital information system). An event includes the series of activities that took place in the specific service for a specific episode. An activity is a timestamped clinical or administrative act that took place in a given service during the caregiving episode of an acute occurrence of a given illness (e.g., the emergency room admission or the hospital surgery).

The data model designed to capture the previous description is depicted in Figure 2 using a Unified Modeling Language (UML) schema. Three main classes represent the data that will be captured from the RWD sets: "Patient", "Urgent Care Event" and "Hospital Event". Locations and facilities are, in fact, attributes of the "Patients" and "Events" but are promoted to classes for time-related expressivity. The superclass "Event" is used to provide a single interface independently of the type of the event. Note that the two timestamps attributes of the "Event" class correspond to the minimum time (`start_time`) and maximum time (`end_time`) observed in the subclasses. In the case of "Hospital Event" the granularity of the time attributes is casted from date to timestamp by adding midnight time.

Finally the "Episode" objects are created during the Task 3, detailed in III-C, by capturing the event continuity among the CPW that took place during a patient treatment.

2) RWD sets extraction: the RWD sets used for this study were three: emergency room data warehouse (SUH BI), the hospital discharge database (CMBD), the insurance database (BDU). These three datasets are distributed across the Aragon Health System facilities with different security policies, data models, but with a common identifier available in all of them: the patient's personal identification code (indicated as `patient_id`).

In a first extraction step the data was captured into a single repository leveraged to a health data platform. This platform is a hardware/software solution to integrate the health system data sources into a single RWD Data Lake and a computing platform to exploit the the RWD Data Lake. The platform integrates the Extract, Transform and Load (ETL) processes

[1] https://github.com/IACS-Biocomputing/process-mining-RWD

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2020.2971146, IEEE Journal of Biomedical and Health Informatics

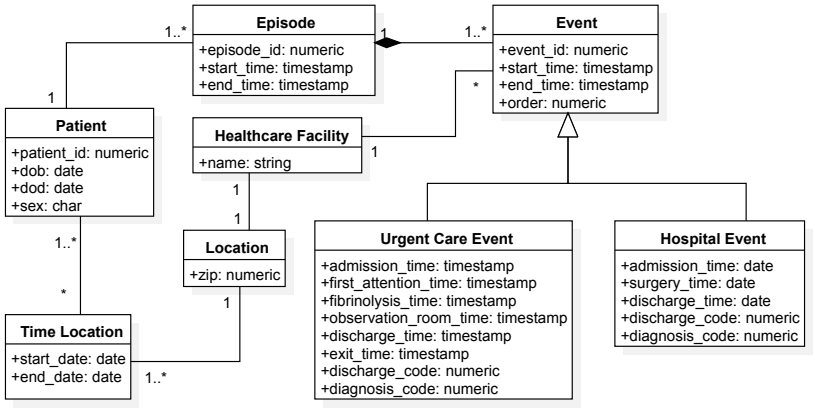


Fig. 2. UML diagram of the Code Stroke data model

that captured the data from the aforementioned and store them in tabular formats.

A second extraction step was required to gather the specific events and patients related to the Code Stroke and create the objects in the data model, creating the objects for all classes except the "Episode". This data gathering is a complex task divided in four data selections: 1) SUH BI emergency room records where `diagnosis_code` corresponds to one of the ICD-9/ICD-10 stroke codes provided by neurologists and the minimum timestamp among its activities takes place during 2017; 2) CMBD records where variable `diagnosis_code` corresponds to one of the ICD-9/ICD-10 stroke codes, where the date reflected in the variable `admission_time` occurs in 2017; 3) records from SUH BI where variable patient `id` appears into the union of patient `id` variables from records in two previous selections; 4) records from CMBD where variable `patient_id` appears into the union of `patient_id` variables from records in selections 1 and 2; 5) records from BDU where variable `patient_id` appears into the union of `patient_id` variables from records in data selection 1 and 2.

The rationale of these selections is the following: first two data selections correspond to the direct detection of suspicious stroke events in ER or hospital facilities; selections 3 and 4 are the events that may reflect side conditions of the pathway but did not contain an explicit stroke diagnosis (e.g., a patient who came to ER with problems with speech and it is confirmed as a stroke in the hospital). Data selections 1 and 3 are merged and codified as "Emergency Room Event" objects, and equivalently done with data captures 2 and 4 to generate "Hospital Event" objects. Finally, data selection 5 is transformed into "Patient" objects.

C. Task 3: Data processing Task. Episode linking algorithm and activity Log generation

The episode linking algorithm is the core algorithm of the present work. As stated previously, `patient_id` is the only common identifier across the datasets and no other information links the continuity of the Code Stroke episodes. The episode linking algorithm is in charge of determine the correctness of the episodes and construct the episode continuity not explicitly recorded in the RWD sets.

The episode linking algorithm is a Python application that runs as follows:

Single event timestamp correctness and correction procedure. Those events with incoherent timestamps are marked as incorrect, e.g. hospital surgery out of bounds of the hospitalization. Events marked as incorrect take part of the whole linking procedure to capture and *consume* the rest of the episode events that will be also considered incorrect avoiding the use of them in other episodes.

A correction procedure processes manually-inserted activity timestamps (e.g., ER Fibrinolysis activity). The correction consists on checking the coherence between year, month and day of the manually-inserted activities and the rest of activity timestamps available in the same event. If there is a discrepancy between the majority of the timestamps and the manually-inserted ones, the manually-inserted are corrected by using the most observed value. This correction may be to the lowest level of detail for ER events (i.e., days) and only for years in Hospital events.

Event bound detection and censoring. Detect the minimum and maximum times in all events and store them in dummy variables for each event (`start_time` and `end_time`). In case the event bounds are out of the study period (initially defined) are marked as left or right censored appropriately.

Global sorting. When two events from same patient occur the same **date**, the Emergency Room precedes the Hospital Event. This definition avoids the timestamp granularity mismatch between types of events. This step generates the `order` variable in the events.

Linking rules execution. Apply chronological and logical linking rules comparing pairs of ordered events of each patient. The chronological linking rules consider the types of events to compare and focus in the time sequence adding a time tolerance between the end of an event and the beginning of the following one. The tolerance depends on the type of events compared. Linking rules complement chronological linking rules ensuring that the `discharge_code` of the first event is compatible with the second event type.

Timestamp synchronization procedure. This procedure adjusts the timestamps across the different events in an episode and transforms the granularity of all timestamps to date plus

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2020.2971146, IEEE Journal of Biomedical and Health Informatics

time (hours, minutes and seconds). Those episodes where there is an overlap between Emergency Room events, the latter activity timestamp of the first event is substituted by the first activity timestamp of the following event. For those episodes that include an Emergency Room event followed by a hospital event, the admission activity timestamp of the hospital event (`admission_time`) is copied from the timestamp of the last activity of the previous Emergency Room event (`end_time`) to include the time granularity. This time part is also added to the rest of activities in the hospital event. When only hospital events take part in an episode, midday time are added to the hospital activities.

Using the use case data for 2017, the episode linking algorithm processed a total of 5802 suspect stroke episodes. Just 3265 were determined, the rest were discarded due to some of the previously commented reasons, for example 268 corresponded to right censored episodes or 319 correspond to any failed linking rules. Once finished the episode linking, the correct episodes are given a unique identifier (`episode_id`) and stored in MongoDB [30]. The storage is done in a patient-centric collection (the term used in MongoDB to refer to group of documents with similar structure and contents). Each document of the collection corresponds to a patient present and includes the stroke episodes described plus the patient socio-economic data captured from the BDU dataset. MongoDB was selected to handle stroke episodes as it is able to manage a variable number of events within an episode and a variable number of episodes for each patient easily within the structure of the patient-centric collection.

The log generation is a transformation process of the patient collection into a new MongoDB collection, in which each document contains the triplet `<episode_id, activity_id, timestamp>`.

The transformation process, currently included in the episode linking script, traverses all the patients in the patient collection. For each patient, the process takes all the correct episodes and then traverses the events that conform the episode. Each event is then decomposed on the activities it contains, and this information is stored in the activity log collection. In the specific case where an episode contains two consecutive hospital related events and the first activity discharge type indicates a discharge to a long stay hospital, the activities of the second event are specifically distinguished as long stay hospital activities. The collection is then sorted by the timestamp, resulting the orthodox organization of a classical log.

D. Task 4: Mining and Analysis. Care Pathways process model generation

The care pathways process model generation leverages process discovery techniques. The tool selected for process model is `bupaR` [17], a R language package that offers an exhaustive variety of process mining analysis techniques.

In a R script, the activity log collection is transformed to an activity log data frame (the term used in R for its data table structure). The activity log data frame is used by `bupaR` to construct the care pathway process model. `bupaR` offers then multiple ways to present and explore the resulting process model. As presented in the following section, process traces, process maps and process timelines are the most useful process model outputs for the purposes of the current work.

E. Task 5: Evaluation

This section contains the outputs of the process mining analysis that were provided to the Aragón Code Stroke Implementers Group, as part of a face-validation approach of the results with domain experts. The Code Stroke Implementers Group is conformed by neurologists, emergency room managers, hospital managers and the Code Stroke strategy directors. The face-validation obtained positive results, confirming the high plausibility of the results with experts' vision of the CPW process. It remains for further iterations of the process mining analysis the inclusion of minor suggestions regarding the merge of some the CPW traces shown below, when the empirical solution does not entail any difference in the management of the Code Stroke.

The outputs presented here are divided in two different scopes: frequency considerations and time considerations.

1) CPW frequency considerations: The two main outputs to analyse the frequency elements of the Code Stroke CPW are process traces and process map with frequency information. In terms of health services research, information on the counts of crude rates of certain activities provides useful insight to understand the intensity of use, whether the patients follow the expected paths, and allows to discover outlier behaviors.

Process traces, see Figure 3, are an account of the different sequences of activities observed in the activity log, in other words, the different types of episodes that took place in the care pathway. In this plot, the x-axis represent the sequence of the activities performed, the y-axis represents the different traces observed and the colours represent different activities. Process traces are ordered decreasingly with respect to the percentage of episodes that correspond to the given trace (the percentage in the grey box at the right side of the trace graph). In the plot of Figure 3 only those traces that cover more than 95% of the total episodes have been depicted.

The process traces depicted in Figure 3 clearly point that the Code Stroke CPW has a two major traces on its real implementation that cover nearly 50% of the most observed episodes. First trace represents those stroke episodes that start in the Emergency Room department and finish in a hospitalisation, with no other specific interventions (no ER fibrinolysis, no Hospital surgery). Second most observed trace is nearly equal to the first one with the main difference that the "observation room" activity in the ER is not present. The third most frequent trace correspond to those patients directly treated in the Hospital and, the fourth most frequent trace correspond to those patients only treated in Emergency Room.

This frequentist approach may be visually complemented with the frequency process map depicted in Figure 4. A process map is a directed graph where the nodes are the activities, whose name appears in the node label, and an edge or an arc between activity A and B indicates that activity A is directly followed by activity B. Edges contain a figure indicating the number of episodes where the specific transition was observed in the activity log being the edge thickness

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2020.2971146, IEEE Journal of Biomedical and Health Informatics

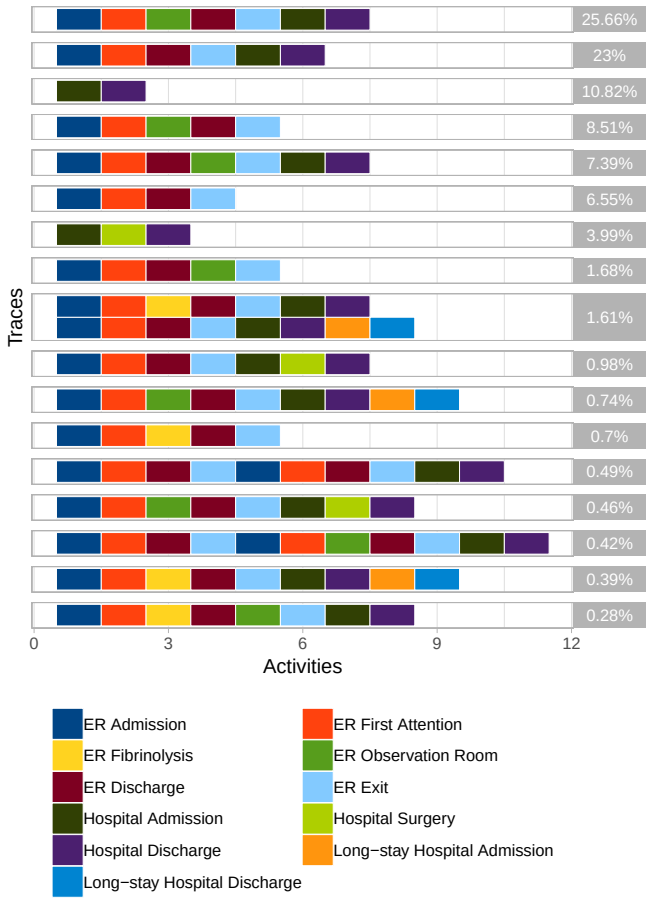


Fig. 3. Process traces covering more than 95% of total stroke episodes

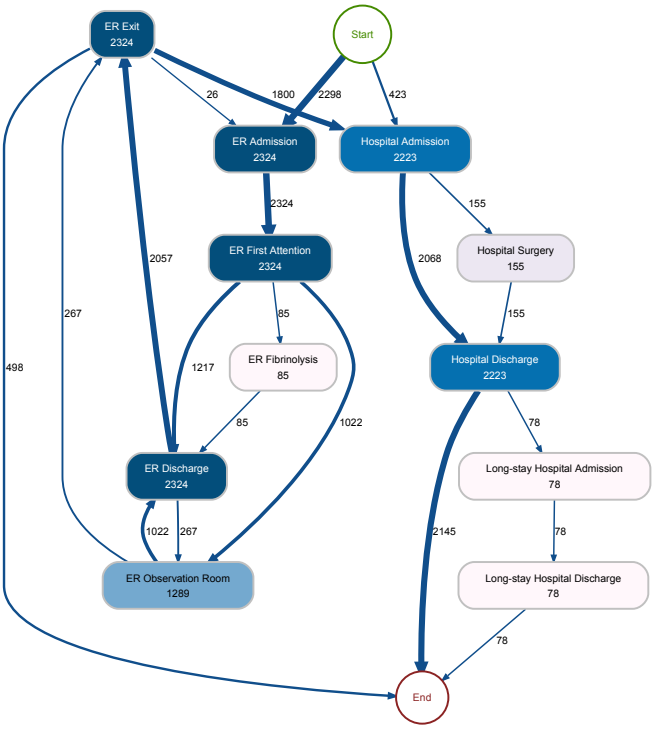


Fig. 4. Process map of episodes covering more than 95% of total episodes with transitions' frequency information

proportional to this value. The coloring of the nodes is a gradient from white to dark blue, indicating the number of episodes where the specific activity appears, a number also indicated in the node label.

Analysing this figure in depth, it is possible to obtain a finer level of granularity to understand the relationships between activities that take place in the real-world implementation of the Code Stroke pathway. Some examples are the following: the typical entry point to the Code Stroke pathway is ER Admission (2298 patients vs. 423 patients that went directly to Hospital Admission); the typical exit point is the Hospital Discharge (2145 patients vs. 576 from other activities); within the ER activities, nearly half of the patients move from ER Admission to ER Discharge (1217 patients of 2324 total), nearly the other half move to ER Observation Room (1022 patients of 2324 total), and a minority receive fibrinolysis treatment (85 patients of 2324 total); typically, ER Exit is followed of a Hospital Admission (1800 patients vs. 498 patients that leave pathway); and, typically, the patients leave the pathway after the Hospital Discharge (2145 patients vs. 78 patients that move to a Long-stay Hospital Admission).

2) CPW time considerations: when it comes to timing across the pathway, timeline plot, depicted in Figure 5, offers an overall picture of all episodes. In this timeline the x axis represents time, the y axis represents the different episodes, and the color indicate the activity type. In the timeline of Figure 5 the time expressed is the *relative* episode time, i.e. the time elapsed since the episode start[2], as a coarse view of the time distribution.

Using this representation, it is possible to distinguish the initial set of episodes that finish in ER Discharge (light blue dots), whose duration is within a day. Then, there is the vast majority of the episodes, that finish in Hospital Discharge, with a duration between 1 day and 25 days. Finally, there is also a small set of episodes, that include Long-stay Hospital activities (orange dots and darker blue dots).

Complementing this coarse analysis, the process discovery also produces a process map with time information output which gives a fine grain knowledge of Code Stroke pathway. This plot, depicted in Figure 6, is essentially the same directed graph depicted in Figure 4, but the *extra* information refers to the time dimension: the edges contain the transitions' median time observed in the activity log, and the edges the median duration of the activities, always set to zero due to the characteristics of the activity log where activities have only a single time.

Analysing the process map of Figure 6 it is possible to observe the time spent in the different transitions between activities as a straightforward approach to evaluate the time-conformance of the actual implementation of the pathway. Some interesting results observing the time information in Figure 6 are: the median time till the ER First Attention is 18 minutes; the median time to fibrinolysis, adding time from ER admission to ER first attention activities and time from ER first Attention to ER fibrinolysis, is around 1 hour (see detailed information about fibrinolysis in the following section); the

[2]It is also possible to depict the *absolute* time, i.e. wall-clock time of activities

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2020.2971146, IEEE Journal of Biomedical and Health Informatics

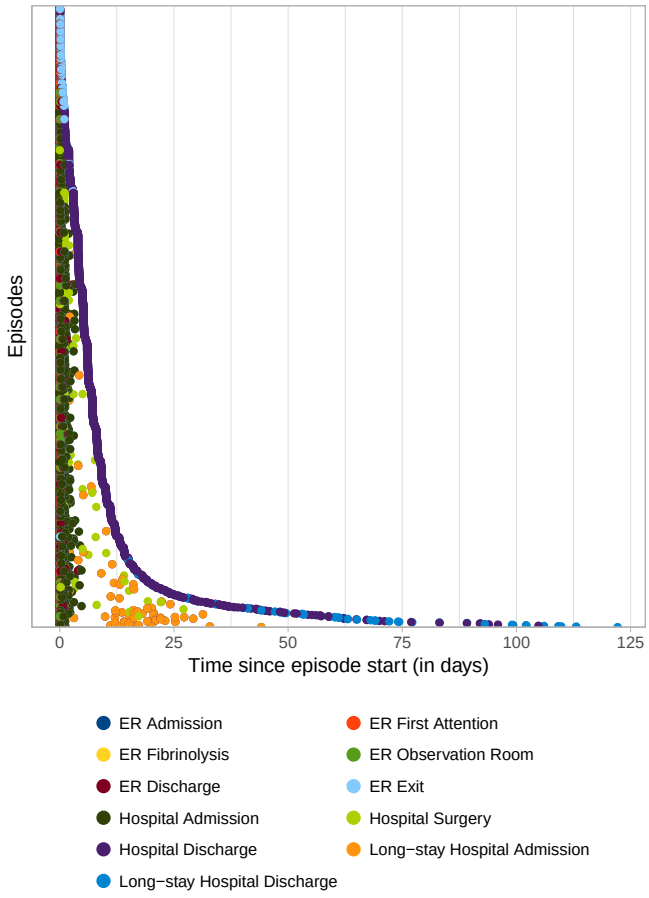


Fig. 5. Timeline of episodes covering more than 95% of total episodes

Fig. 6. Process map of episodes covering more than 95% of total episodes with transitions' median duration information

median time from ER first attention to ER observation room is around 1 hours and 48 minutes; the median time from ER discharge and actual ER exit is around 1 hour and 20 minutes; the hospital median length of stay is around 144 hours (6 days) in admissions without surgical procedure and 96 hours (4 days) when the patient underwent a surgery; finally, the longest transition observed in the pathway corresponds to stay at long-stay hospitals, whose median time is 708.5 hours (~29 days).

F. Task 6: Process improvement and support. On the fibrinolysis treatment

Although it is not the objective of the present paper to present a detailed clinical approach of the Code Stroke CPW real-world implementation, the fibrionolysis treatment analysis is a clear example to demonstrate the usefulness of the approach to evaluate the uptake of symptoms, the adequacy of the health care provided and ultimately, the impact on health outcomes. Fibrinolysis is a clinical treatment with high influence in patients' survival but requires to be provided in tight time frame since the onset of the stroke symptoms (4.5 hours since the start of the symptoms according to Code Stroke Aragon guideline [28]). Due to its importance, the treatment with fibrinolysis is captured as a specific activity.

A quick evaluation of appropriateness may be done by combining the information of those traces that include "ER Fibrinolysis activity, depicted in Figure 3 and the process map in Figure 7 . This process map has been constructed by filtering those episodes including the "ER Fibrinolysis" activity and then trimming the episodes by discarding all those activities after the "ER Fibrinolysis" (a single filtering function of the `bupaR` packages). This filter increases the legibility of the process map enabling a clear view that, in all the possible episodes that include the treatment, the median time-to-fibrinolysis is below the desired 4.5 hours (within the accounted time frame). Transposing the traces that include the "ER Fibrinolysis" activity in Figure 3 to the process map in Figure 7, i.e. following the sequence of activities indicated in the traces within the process map, the largest median time observed is around 240 minutes.

In addition, thanks to the information present in process models obtained from the input datasets, such as the first ER that treated the patient, it is possible to easily create detailed analysis plots as the one in Figure 8. This Figure exhibits the distribution of time-to-fibrinolysis grouped according the first ER of treatment, i.e. the ER facility where the patients entered to the pathway. The box-plots show that the median time-to-fibrinolysis for all hospitals is below 100 minutes. Hospitals 4, 7, 8, 11, 12 and 18 correspond to those smaller hospitals in the region where Code Stroke was recently deployed, have a median time-to-fibrinolysis between 50 and 100 minutes, being the hospitals 8 and 11 the ones with larger variability. This median time-to-fibrinolysis decreases below 50 minutes in hospitals 1, 9 and 10, the largest hospitals in the region with high experience in the Code Stroke CPW.

IV. Methodological Caveats and Limitations

Once the utility of the process model construction based on RWD sets has been documented, it is important to highlight

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2020.2971146, IEEE Journal of Biomedical and Health Informatics

Fig. 7. Process map with transitions' median duration trimmed to end at fibrinolysis activity

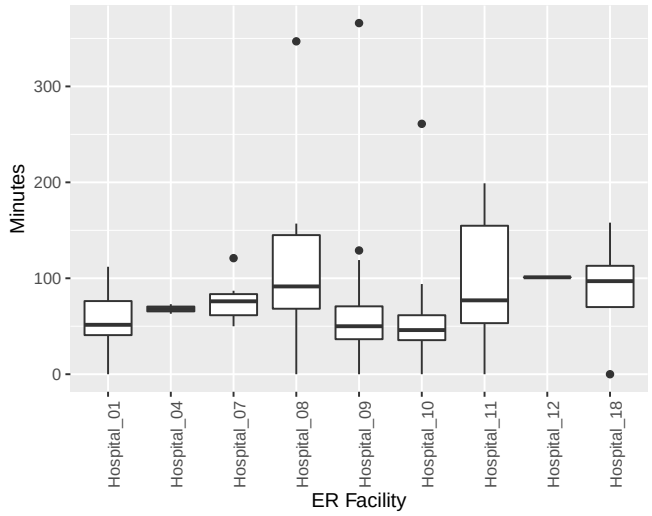


Fig. 8. Time to fibrinolysis treatment per ER facility of first contact

here a set of elements that should be considered when facing this task in other "experimental" settings.

A. Data model

It has been mentioned in the text the importance of the data model for the analysis purposes. The data model may limit the possibilities of further analyses or the comparability among different process models obtained from different providers or health systems.

B. Data availability

To the purpose of this work, three RWD sets were used. The RWD sets correspond to structured administrative data available in the Aragón health system. Undoubtedly, the addition of more RWD sets could have enriched the result, for example including the unstructured data coming from the clinical notes or reports written in natural language. Works such as [31] and [32] serve as an illustration on how to use natural language processing (NLP) approaches to capture unstructured data from electronic health records (EHR) and clinical reports, the most typical RWD sets available in health systems nowadays.

In order to asses whether not considering the inclusion of unstructured data could have implied a major limitation in the current study, an ad hoc exploration of a number of stroke discharge reports revealed two additional timestamped activities not present in the RWD sets used in this paper: functional analysis of patients and CT imaging time. However, these activities were not consistently in all reports. In any case, the inclusion of unstructured data via NLP techniques should be a key point in the future work of the current research.

C. Data quality

As in any other data analysis research, the data quality is a crucial aspect to take into account. Closely related to the data model, it is strictly necessary to gain a high expertise of data semantics, previously cited, in order to ensure a high data quality. In the present work, this expertise is reflected in multiple steps of the methodology: first, the ability to capture the information from the multiple RWD sets, detecting the sources but also translating the case definition to extract the desired data from these datasets; second, the synthesis of the rules that capture the care continuity in the episode linking algorithm, an algorithm that deals with the datasets coverage and harmonizes the semantics; and third, the evaluation of the adequacy of the resulting episodes and the process mining outputs with the expert knowledge of the acute stroke CPW.

D. Data granularity

Even being a data quality element, the data granularity should be treated on its own. This is one of the most sensitive elements when dealing with timestamped data and event logs. As detailed in the episode linking, Section III-C, the more granularity the data has, the better to link the RWD sets. The Episode linking algorithm synthesizes all the logic required to deal with the time coherence among datasets with different time granularity. However, an additional cavet should be considered when it comes to comparability: those datasets with less granularity will result in worse or impossible comparisons, for example, in the current paper the measurement of the time lagging from admission to treatment with fibrinolysis has been possible due to the availability of a high degree of granularity (hour and minute) in the timestamps of these activities.

V. Application to Health Services Research

For health services research, the construction of empirical process models from RWD sets provides a unique point-of-view to understand both how patients and health system interact and also how health systems organize in the reality to guide these interactions. Some potential uses of process discovery in health services research are the following.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2020.2971146, IEEE Journal of Biomedical and Health Informatics

A. Conformance checking

The conformance checking is the evaluation of adequacy of the actual practices observed in the empirical process model as compared to those described in the theoretical or normative CPW. Conformance checking is one of the most useful applications of the resulting process models when focusing on the organizational point-of-view.

The time-to-fibrinolysis evaluation presented in section III-F is an informal example of the conformance checking, contrasting the real-world observed times against the defined in the Code Stroke guideline [28]. This is an analysis that provides a high value when evaluating Code Stroke CPW implementation and was performed within minutes with few source code lines of the `bupaR` package.

A more sophisticated conformance checking approach will result by comparing the empirical process model discovered from the RWD sets to a manually codified model of the theoretical or normative CPW or the process model extracted from the clinical guidelines using NLP in a similar way to the work done in [33] for an archaeology manual. Independently on how the reference model is constructed, the comparison between the empirical process and the reference model may be done, using for example *event log replay* techniques [34], i.e. simulate how real-world episodes can follow the theoretical process model, or graph dissimilarity metrics, i.e. measure how different the real-world process maps are compared to theoretical process maps as described in [35].

B. Benchmarking of providers

The benchmarking is a derived result of the conformance checking introduced in the previous point, and has been hinted in the comparison of the different hospital ER facilities presented when exposing the adequacy of the treatment with fibrinolysis in the Section III-F.

Thus, having the different degrees of adequacy and conformance between the real-world implementation of the CPW and theoretical or normative CPW in the different care providers can be used to compare and contrast the organizational approach of these providers. Those best, i.e. similarity to the normative path, will serve as benchmark providers to improve the quality of the healthcare services. The study presented in [5] proposed a similar benchmark.

It is interesting to note that the work presented in this paper is being used to perform a provider benchmarking within the European project ICTUSnet [3], comparing CPW from 7 regions of 4 different European countries.

C. Comparative effectiveness research

Remains as next step of the current research the analysis of how the exposure to health system affect to health outcomes by exploiting individual and health system variables included in the RWD sets, and thus, in the inferred process model. This type of analyses aim at comparing effectiveness in two different facets.

First, measuring the effects of the traces that patients follow, comparing first health outcomes and/or costs of patients with different characteristics that follow the same trace and second comparing patients with same characteristics that follow different traces. Second, evaluating those decision points in the process model, i.e. those bifurcations in the process map. Applying decision mining techniques and using patients and health system variables, it is possible to measure which variables affect to the decision to follow one path or another and how the decision affect to health outcomes and/or costs.

[3] http://ictusnet-sudoe.eu/en/

D. Clinical translation

Although the focus of this paper is on the health services research point of view, the results of the CPW analyses could ultimately be applied in the clinical practice. Serving as an example: 1) the results of the conformance checking may lead to the reorganization of those services that do not attend to the specific clinical practice guidelines; 2) the benchmarking of providers may serve to stablish new cross-organizational policies according to the best performing providers' structures and management; 3) the results of the comparative effectiveness research can be directly applied in regular medical practice as it will unveil those practices and interventions within the CPW that result in better health outcomes; and 4) process models enriched with clinical data may be used with other data mining/machine learning techniques (e.g., random forests, neural networks) to predict adverse events within the CPW.

VI. Conclusions

It has been shown and discussed the viability of constructing care pathways' process models by linking RWD sets and using process mining analysis. To this end, there was a careful planning to define a data model and to develop a linking algorithm that guarantees the time coherence across the data available from the different datasets. The solution proposed solves the limitations pointed by W. Yang and Q. Su in [20] regarding the data availabilty. The actual source code of the solution is available in a GitHub repository, with a total guarantee of reproducibility for further experimentation.

The solution proposed has been shown to have great value for different stakeholders as it opens the door to a wide set of in-depth studies of how health systems organize, how this organisation may impact health outcomes and how to improve the organisation both at health system level and, in a final term, in the day-to-day clinical practice.

Funding

This work was partially funded by the project ProCOMP (LMP 183_18) of the Aragon Government through the Operative Programme of the European Regional Development Fund (ERDF), "Construyendo Europa desde Aragón"; and ICTUSnet Project (SOE2/P1/E0623), financed within the Interreg SUDOE Program that supports the development of south-west Europe regions by financing transnational projects through the ERDF.

References

[1] H. Campbell, R. Hotchkiss, N. Bradshaw, and M. Porteous, "Integrated care pathways.," *BMJ (Clinical research ed.)*, vol. 316, no. 7125, pp. 133–7, Jan. 1998, ISSN: 0959-8138. [Online]. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2665398.

[2] T. Rotter, L. Kinsman, E. L. James, A. Machotta, H. Gothe, J. Willis, P. Snow, and J. Kugler, "Clinical pathways: effects on professional practice, patient outcomes, length of stay and hospital costs," *Cochrane Database of Systematic Reviews*, no. 3, Mar. 2010, ISSN: 14651858. DOI: 10.1002/14651858.CD006632.pub2.

[3] D. J. Zand, K. M. Brown, U. Lichter-Konecki, J. K. Campbell, V. Salehi, and J. M. Chamberlain, "Effectiveness of a Clinical Pathway for the Emergency Treatment of Patients With Inborn Errors of Metabolism," *PEDIATRICS*, vol. 122, no. 6, pp. 1191–1195, Dec. 2008, ISSN: 0031-4005. DOI: 10.1542/peds.2008-0205.

[4] J. Lenkowicz, R. Gatta, C. Masciocchi, C. Casà, F. Cellini, A. Damiani, N. Dinapoli, and V. Valentini, "Assessing the conformity to clinical guidelines in oncology: An example for the multidisciplinary management of locally advanced colorectal cancer treatment Jacopo," *Management Decision*, vol. 56, no. 10, pp. 2172–2186, Oct. 2018, ISSN: 0025-1747. DOI: 10.1108/MD-09-2017-0906.

[5] A. Partington, M. Wynn, S. Suriadi, C. Ouyang, and J. Karnon, "Process Mining for Clinical Processes: A Comparative Analysis of Four Australian Hospitals," *ACM Transactions on Management Information Systems*, vol. 5, no. 4, pp. 1–18, Jan. 2015, ISSN: 2158656X. DOI: 10.1145/2629446.

[6] W. M. P. van der Aalst, *Process Mining: Data Sience in Action*, Second Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2016, ISBN: 978-3-662-49851-4. DOI: 10.1007/978-3-662-49851-4.

[7] L. P. Garrison, P. J. Neumann, P. Erickson, D. Marshall, and C. D. Mullins, "Using Real-World Data for Coverage and Payment Decisions: The ISPOR Real-World Data Task Force Report," *Value in Health*, vol. 10, no. 5, pp. 326–335, Sep. 2007, ISSN: 10983015. DOI: 10.1111/j.1524-4733.2007.00186.x.

[8] M. L. van Eck, X. Lu, S. J. J. Leemans, and W. M. P. van der Aalst, "PM² : A Process Mining Project Methodology," in *CAiSE 2015: Advanced Information Systems Engineering*, Springer, Cham, 2015, pp. 297–313. DOI: 10.1007/978-3-319-19069-3_19.

[9] W. van der Aalst, T. Weijters, and L. Maruster, "Workflow mining: discovering process models from event logs," *IEEE Transactions on Knowledge and Data Engineering*, vol. 16, no. 9, pp. 1128–1142, Sep. 2004, ISSN: 1041-4347. DOI: 10.1109/TKDE.2004.47.

[10] A. Weijters, W. M. P. Aalst van der, and A. K. Alves De Medeiros, *Process mining with the HeuristicsMiner algorithm*, ser. BETA publicatie : working papers. Technische Universiteit Eindhoven, 2006, ISBN: 978-90-386-0813-6.

[11] C. W. Günther and W. M. Van Der Aalst, "Fuzzy mining - Adaptive process simplification based on multi-perspective metrics," in *BPM 2007: Business Process Management*, vol. 4714 LNCS, Springer, Berlin, Heidelberg, 2007, pp. 328–343, ISBN: 9783540751823. DOI: 10.1007/978-3-540-75183-0_24.

[12] S. J. J. Leemans, D. Fahland, and W. M. P. van der Aalst, "Discovering Block-Structured Process Models from Event Logs - A Constructive Approach," in *PETRI NETS 2013: Application and Theory of Petri Nets and Concurrency*, Springer, Berlin, Heidelberg, 2013, pp. 311–329. DOI: 10.1007/978-3-642-38697-8_17.

[13] J. Munoz-Gama, *Conformance Checking and Diagnosis in Process Mining*, ser. Lecture Notes in Business Information Processing. Cham: Springer International Publishing, 2016, vol. 270, ISBN: 978-3-319-49450-0. DOI: 10.1007/978-3-319-49451-7.

[14] A. Rozinat and W. van der Aalst, "Conformance checking of processes based on monitoring real behavior," *Information Systems*, vol. 33, no. 1, pp. 64–95, Mar. 2008, ISSN: 03064379. DOI: 10.1016/j.is.2007.07.001.

[15] B. F. van Dongen, A. K. A. de Medeiros, H. M. W. Verbeek, A. J. M. M. Weijters, and W. M. P. van der Aalst, "The ProM Framework: A New Era in Process Mining Tool Support," in *ICATPN 2005: Applications and Theory of Petri Nets 2005*, Springer, Berlin, Heidelberg, 2005, pp. 444–454, ISBN: 978-3-540-26301-2, 978-3-540-31559-9. DOI: 10.1007/11494744_25.

[16] C. W. Günther and A. Rozinat, "Disco: Discover Your Processes," in *BPM 2012: Proceedings of the Demonstration Track of the 10th International Conference on Business Process Management*, CEUR Workshop Proceedings (CEUR-WS.org), 2012, pp. 40–44.

[17] G. Janssenswillen, B. Depaire, M. Swennen, M. Jans, and K. Vanhoof, "bupaR: Enabling reproducible business process analysis," *Knowledge-Based Systems*, vol. 163, pp. 927–930, Jan. 2019, ISSN: 09507051. DOI: 10.1016/j.knosys.2018.10.018.

[18] W. van der Aalst, "Process mining: discovering and improving Spaghetti and Lasagna processes," in *CIDM 2011: IEEE Symposium on Computational Intelligence and Data Mining*, IEEE, Apr. 2011, pp. 1–7, ISBN: 978-1-4244-9927-4. DOI: 10.1109/CIDM.2011.6129461.

[19] L. Maruster and R. Jorna, "From Data to Knowledge: A Method for Modeling Hospital Logistic Processes," *IEEE Transactions on Information Technology in Biomedicine*, vol. 9, no. 2, pp. 248–255, Jun. 2005, ISSN: 1089-7771. DOI: 10.1109/TITB.2005.847194.

[20] W. Yang and Q. Su, "Process mining for clinical pathway: Literature review and future directions," in *ICSSSM 2014: 11th International Conference on Service Systems and Service Management*, IEEE, Jun. 2014, pp. 1–5, ISBN: 978-1-4799-3134-7. DOI: 10.1109/ICSSSM.2014.6943412.

[21] E. Rojas, J. Munoz-Gama, M. Sepúlveda, and D. Capurro, "Process mining in healthcare: A literature review," *Journal of Biomedical Informatics*, vol. 61, pp. 224–236, Jun. 2016, ISSN: 1532-0464. DOI: 10.1016/J.JBI.2016.04.007.

[22] E. Batista and A. Solanas, "Process Mining in Healthcare: A Systematic Review," in *IISA 2018: 9th International Conference on Information, Intelligence, Systems and Applications*, IEEE, Jul. 2018, pp. 1–6, ISBN: 978-1-5386-8161-9. DOI: 10.1109/IISA.2018.8633608.

[23] Á. Rebuge and D. R. Ferreira, "Business process analysis in healthcare environments: A methodology based on process mining," *Information Systems*, vol. 37, no. 2, pp. 99–116, Apr. 2012, ISSN: 0306-4379. DOI: 10.1016/J.IS.2011.01.003.

[24] S. Suriadi, R. S. Mans, M. T. Wynn, A. Partington, and J. Karnon, "Measuring Patient Flow Variations: A Cross-Organisational Process Mining Approach," in *AP-BPM 2014: Asia Pacific Business Process Management*, Springer, Cham, 2014, pp. 43–58. DOI: 10.1007/978-3-319-08222-6_4.

[25] R. Mans, H. Schonenberg, G. Leonardi, S. Panzarasa, A. Cavallini, S. Quaglini, and W. van der Aalst, "Process mining techniques: an application to stroke care.," *Studies in health technology and informatics*, vol. 136, no. eHealth Beyond the Horizon – Get IT There, pp. 573–8, 2008, ISSN: 0926-9630.

[26] M. Binder, W. Dorda, G. Duftschmid, R. Dunkl, K. A. Fröschl, W. Gall, W. Grossmann, K. Harmankaya, M. Hronsky, S. Rinderle-Ma, C. Rinner, and S. Weber, "On analyzing process compliance in skin cancer treatment: An experience report from the evidence-based medical compliance cluster (EBMC^{2})," in *CAiSE 2012: Advanced Information Systems Engineering*, vol. 7328 LNCS, Springer, Berlin, Heidelberg, 2012, pp. 398–413, ISBN: 9783642310942. DOI: 10.1007/978-3-642-31095-9_26.

[27] I. Litchfield, C. Hoye, D. Shukla, R. Backman, A. Turner, M. Lee, and P. Weber, "Can process mining automatically describe care pathways of patients with long-term conditions in UK primary care? A study protocol," *BMJ Open*, vol. 8, no. 12, Dec. 2018. DOI: 10.1136/bmjopen-2017-019947.

[28] O. Alberti González, J. C. Aragüés Bravo, M. Bestué Cardiel, I. Campello Morer, M. J. Chopo Alcubilla, C. García Arguedas, M. J. Gimeno Peribáñez, A. Gómez Peligros, B. Gros Bañeres, L. Jarauta Lahoz, M. J. Lafuente González, J. Marta Moreno, G. Martínez Borobio, E. Muñoz Farjas, M. Palacín Larroy, C. Pérez Lázaro, M. Sampériz Murillo, and M. Serrano Ponz, *Plan de Atención al Ictus en Aragón. Actualización 2019-2022*, M. Bestué Cardiel, J. Marta Moreno, and G. Martínez Borobio, Eds. Gobierno de Aragón. Dirección General de Asistencia Sanitaria, 2018, ISBN: 978-84-09-08415-9. [Online]. Available: https://www.aragon.es/estaticos/GobiernoAragon/Departamentos/SanidadBienestarSocialFamilia/Sanidad/Documentos/Programa%7B%5C_%7DIctus%7B%5C_%7Dactualizacion2019.pdf.

[29] P. Conde-Espejo, *Evaluación de la eficiencia de modelos organizativos para el abordaje del ictus (unidades de ictus)*. Madrid: Consejería de Sanidad - D. G. de Planificación, Investigación y Formación, 2013. [Online]. Available: http://www.comunidad.madrid/publicacion/1354388439231.

[30] K. Chodorow and M. Dirolf, *MongoDB : the definitive guide*. O'Reilly Media, 2010, p. 193, ISBN: 9781449381561. [Online]. Available: http://cds.cern.ch/record/1359920.

[31] E. S. Chen and I. N. Sarkar, "Mining the Electronic Health Record for Disease Knowledge," *Biomedical Literature Mining*, vol. 1159, pp. 269–286, 2014. DOI: 10.1007/978-1-4939-0709-0_15.

[32] T. Delespierre, P. Denormandie, A. Bar-Hen, and L. Josseran, "Empirical advances with text mining of electronic health records," *BMC Medical Informatics and Decision Making*, vol. 17, no. 1, p. 127, Dec. 2017, ISSN: 1472-6947. DOI: 10.1186/s12911-017-0519-0.

[33] E. V. Epure, P. Martin-Rodilla, C. Hug, R. Deneckere, and C. Salinesi, "Automatic process model discovery from textual methodologies," in *Proceedings of 2015 IEEE 9th International Conference on Research Challenges in Information Science (RCIS)*, L. P. G.-P. C. Rolland C. Anagnostopoulos D., Ed., vol. 2015-June, IEEE Computer Society, May 2015, pp. 19–30, ISBN: 978-1-4673-6630-4. DOI: 10.1109/RCIS.2015.7128860.

[34] W. van der Aalst, A. Adriansyah, and B. van Dongen, "Replaying history on process models for conformance checking and performance analysis," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 2, no. 2, pp. 182–192, Mar. 2012, ISSN: 19424787. DOI: 10.1002/widm.1045.

[35] S. Montani, G. Leonardi, S. Quaglini, A. Cavallini, and G. Micieli, "Mining and Retrieving Medical Processes to Assess the Quality of Care," in *International Conference on Case-Based Reasoning*, Saratoga Springs, NY, USA: Springer, Berlin, Heidelberg, 2013, pp. 233–240. DOI: 10.1007/978-3-642-39056-2_17.

BMJ Open

Differences in acute ischaemic stroke in-hospital mortality across referral stroke hospitals in Spain: a retrospective, longitudinal observational study

Francisco Estupiñán-Romero ,[1] Jaime Pinilla Dominguez ,[2] Enrique Bernal-Delgado ,[1] on behalf of the Atlas VPM consortium

To cite: Estupiñán-Romero F, Pinilla Dominguez J, Bernal-Delgado E, *et al*. Differences in acute ischaemic stroke in-hospital mortality across referral stroke hospitals in Spain: a retrospective, longitudinal observational study. *BMJ Open* 2023;**13**:e068183. doi:10.1136/bmjopen-2022-068183

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2022-068183).

Received 09 September 2022
Accepted 05 June 2023

[1]Data Science for Health Services and Policy Research, Aragon Health Sciences Institute, Zaragoza, Spain
[2]Department of Quantitative Methods for Economics and Management, University of Las Palmas de Gran Canaria Faculty of Economics and Business Administration, Las Palmas de Gran Canaria, Spain

Correspondence to
Dr Francisco Estupiñán-Romero; festupinnan.iacs@aragon.es

ABSTRACT

Objective To assess differences in acute ischaemic stroke (AIS) in-hospital mortality between referral stroke hospitals and provide evidence on the association of those differences with the overtime adoption of effective reperfusion therapies.

Design Retrospective, longitudinal observational study using administrative data for virtually all hospital admissions from 2003 to 2015.

Setting Thirty-seven referral stroke hospitals in the Spanish National Health System.

Participants Patients aged 18 years and older with a hospital episode with an admission diagnosis of AIS in any referral stroke hospital (196 099 admissions).

Main endpoints (1) Hospital variation in 30-day in-hospital mortality measured in terms of the intraclass correlation coefficient (ICC); and (2) the difference in mortality between the hospital of treatment and the trend of utilisation of reperfusion therapies (including intravenous fibrinolysis and endovascular mechanical thrombectomy) in terms of median OR (MOR).

Results Adjusted 30-day AIS in-hospital mortality decreased over the study period. Adjusted in-hospital mortality after AIS rates varied from 6.66% to 16.01% between hospitals. Beyond differences in patient characteristics, the relative contribution of the hospital of treatment was higher in the case of patients undergoing reperfusion therapies (ICC=0.031 (95% Bayesian credible interval (BCI)=0.017 to 0.057)) than in the case of those who did not (ICC=0.016 (95% BCI=0.010 to 0.026)). Using the MOR, the difference in risk of death was as high as 46% between the hospital with the highest risk and the hospital with the lowest risk of patients undergoing reperfusion therapy (MOR 1.46 (95% BCI 1.32 to 1.68)); in patients not undergoing any reperfusion therapy, the risk was 31% higher (MOR 1.31 (95% BCI 1.24 to 1.41)).

Conclusions In the referral stroke hospitals of the Spanish National Health System, the overall adjusted in-hospital mortality decreased between 2003 and 2015. However, between-hospital variations in mortality persisted.

STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ This study has covered virtually all acute ischaemic stroke admissions in referral stroke hospitals in the Spanish National Health System from 2003 to 2015.
⇒ The use of a generalised additive mixed model has yielded a robust assessment of differences in mortality across hospitals as it has allowed modelling time-dependent associations relevant to this topic of study; for example, the uneven adoption of new technologies in referral stroke hospitals, changes in the organisation of stroke care or the evolving indication patterns due to overtime changes in patients' characteristics.
⇒ Although lack of adjustment for patient severity across hospitals can be a study limitation based on the secondary use of routinely collected health data, restricting the comparison to referral stroke hospitals, likely treating similar case-mixes of patients in large numbers, helps to reduce the risk of residual confounding.

INTRODUCTION

In the last decade, endovascular treatment strategies for acute ischaemic stroke (AIS) have been continuously developing, resulting in a reduction in short-term mortality and a significant increase in functional independence.[1]

However, since their introduction, effective reperfusion therapies, such as intravenous fibrinolysis or endovascular mechanical thrombectomy (EMT),[2 3] have been unevenly implemented across health systems and healthcare providers depending on adequate access to fibrinolysis drugs, lack of specialised resources (eg, stroke units, endovascular operating theatres), adequate access to urgent CT scan imaging or limited coverage of neurointerventional surgeons.[4 5]

In Spain, the first Stroke Health Care Plan was developed in 2006[6] by the Spanish Society of Neurology. However, the widespread adoption of reperfusion therapies within the Spanish National Health System (SNHS)

occurred in November 2008 after adopting the National Stroke Care Strategy.[7]

The National Stroke Care Strategy aimed to harmonise stroke care across autonomous communities (ACs), covering all the services to be provided to patients who had a stroke (primary and secondary prevention of stroke, care in the acute phase, rehabilitation and return to everyday life, as well as training and research). Notably, in acute care, the strategy incepted the Stroke Code, which is the same for all the ACs, including clinical practice guidelines and protocols, the organisation of care pathways, and the required resources for adequate coverage and provision (telestroke programmes, stroke units, neurosonology resources, diffusion and perfusion magnetic resonance and neurovascular interventions). Instead, the Stroke Code set up a hierarchy of stroke hospitals, with a smaller group acting as referral stroke hospitals for endovascular interventions.[8]

Research on the impact of the National Stroke Care Strategy has primarily focused on monitoring healthcare processes,[9] case studies,[10] narrative descriptions of clinicians' responses within care guidance and epidemiological studies on the evolution of the health determinants of stroke.[11] Still, there has yet to be an assessment of variations in hospitals' outcomes in a period that has witnessed substantial policy and organisational changes in treating AIS.

This study aimed to assess differences in AIS in-hospital mortality between referral stroke hospitals and provide evidence on the association of those differences with the overtime adoption of effective reperfusion therapies.

METHODS

Design

We conducted a retrospective, longitudinal observational study using administrative data for virtually all SNHS hospital episodes between 1 January 2003 and 31 December 2015.

Population and setting

From all the hospital admissions in the study period, we first selected those episodes with an admission diagnosis of AIS (662 997 episodes). Then, we defined *referral stroke hospitals* as those capable of performing EMT confirmed through identifying EMT procedures performed in the last year of analysis (2015).

From those episodes treated in hospitals identified as referral stroke hospitals, we kept only the episodes with a stay shorter than 30 days, restricting the observation of case fatalities to those more likely to be associated with the medical intervention on stroke (ie, deaths in more extended stays are more likely associated with nosocomial infections or the underlying health status of the patient). Finally, to reduce extra-heterogeneity, which may produce an overestimation of the intraclass coefficient, we restricted the study to those referral stroke hospitals with at least 2000 ischaemic stroke episodes recorded throughout the study period. The final study population included 196 099 episodes from 37 referral stroke hospitals (see flow chart in online supplemental figure 1).

Main endpoints

Hospital variation in 30-day in-hospital mortality measured in terms of the intraclass correlation coefficient (ICC); and (2) the difference in mortality between the hospital of treatment and the trend of utilisation of reperfusion therapies in terms of median OR (MOR).

Variables

In-hospital mortality within 30 days from admission in patients with AIS was defined following the Agency for Healthcare Research and Quality's quality indicators definition[12] (see indicator definition in online supplemental table 1), validated for the SNHS by the Atlas VPM.[13]

Eliciting the potential effect of the hospital of treatment on the risk of death, regardless of the differences in the patients attended, required risk adjustment. Patients' age, sex and the list of Elixhauser comorbidities were used in the risk adjustment following a data-driven approach, thus, retaining in the model only those comorbidities statistically associated with the outcome (significantly associated for an alpha error of 5%).[14 15] Finally, we included time variables for each episode characterising long-term structural trends and monthly seasonality and identified special admission days, such as bank holidays or weekends.

Data sources

Episodes were extracted from the Atlas VPM dataset that collects virtually all hospital episodes discharged in the public hospitals of the SNHS.[16] In addition, hospital features were collected from the National Hospital Catalogue.[17] Atlas VPM reuses routine data mainly from electronic records from hospital admissions and primary care visits for the identification and selection of cases (eg, surgical procedures of interest, specific diseases, quality and safety events), the analysis of clinical attributes of the patient (eg, comorbidities, concurrent surgery), the analysis of administrative features worth to collect (eg, admission and discharge data, date of surgery) and for the identification of the place of residence (ie, primary, healthcare or area where the patient resides). In addition, Atlas VPM implies a linkage and exchange process with the 17 Departments of Health of the Spanish regions sharing a research agenda assessing unwarranted variations in medical practice and translating research outcomes into profiling and benchmarking tools meant to facilitate clinical and policy decision-making.[18] Notably, as part of its data quality assurance methodology, Atlas VPM conducts regular (once a year) quality checks to, for example, reduce incompleteness, avoid overtime coding variations, correct semantic and syntactic inconsistencies in the variables, or reallocate admission episodes to the place of residence if there are changes in the geolocation of administrative areas or hospital providers.

Table 1 Background characteristics of the population hospitalised due to an acute ischaemic stroke

	2003	2006	2009	2012	2015
In-hospital mortality (%)*	**12.99**	**12.19**	**11.39**	**10.75**	**10.46**
Number of deaths (#)	1686	1647	1735	1809	1848
Number of admissions (#)	12 976	13 505	15 227	16 829	17 658
Reperfusion therapy (%)	0.52	2.49	7.00	12.10	15.45
Age, years (SD)	73.59 (12.17)	73.46 (12.75)	73.68 (13.04)	74.27 (13.11)	74.04 (13.44)
Gender, female (%)	46.59	46.60	46.95	47.59	46.15
Public holiday admission (%)	26.88	27.09	26.98	27.71	27.44
Congestive heart failure (%)	4.09	4.15	4.55	5.41	5.41
Pulmonary circulation disease (%)	1.03	1.21	1.60	1.77	2.07
Chronic obstructive pulmonary disease (%)	9.65	8.34	8.37	9.30	9.44
Renal failure (%)	2.89	3.47	5.62	6.83	7.89
Lymphoma (%)	0.22	0.33	0.35	0.30	0.42
Metastatic cancer (%)	0.94	1.20	1.18	1.36	1.44
Coagulopathy (%)	0.62	0.70	0.72	1.31	1.26
Fluid and electrolyte disorders (%)	1.52	1.47	1.90	2.35	2.83
Drug abuse	0.28	0.59	0.72	0.66	0.83

*Absolute and average values every 3 years to highlight their evolution.

Bias control

Our sample includes patients who were alive at the time of admission. Although sudden death is less likely to happen in ischaemic stroke, there might have been an unknown toll of patients dying before arriving at the hospital. Should this number be relevant, this could have affected the estimation of the trend in the first years of the series when emergency care needs were likely undercovered.

Another potential source of bias could be the accuracy of our data source to capture all deaths after AIS within the hospitals in our sample. We checked conditions for in-hospital mortality for ischaemic stroke in Spain, as documented by the Organization for Economic Cooperation and Development (OECD) 2011 Health at a Glance report.[19]

Since coding practice has increased over the years, it is more likely to find more secondary diagnoses per episode in the last years of the series. Elixhauser comorbidities are sensitive to coding practice, and coding practice may vary across hospitals; risk adjustment may be affected, artificially smoothing down adjusted rates in those hospitals with a more intensive coding practice. We conducted a sensitivity analysis considering independent coding variables such as age and sex in the models to assess to what extent this might have misclassified hospitals.

Finally, patients can access hospital care differently because of the time distance from their location of residence to the referral hospital, which may imply a selection bias not fully corrected by using the random hospital effect included in our models. Therefore, as the time to treatment may potentially lead to the treatment decisions and the patient's prognosis, we performed sensitivity analyses considering travelling distances from the area of residence to the nearest hospital when that information was available.

Analysis

Given the time-dependent non-linear association between our endpoint (*in-hospital mortality 30 days from admission*) and some of the predictors, and the hierarchical nature of the data, we used a generalised additive mixed model (GAMM)[20] with a logit link function to assess the influence of potential risk factors and other covariates on in-hospital mortality within 30 days of admission. In addition, we added smoothing functions for the covariates with continuous values such as age, months, and the interaction between the time trend and the 'reperfusion therapy' variable. We estimated the smoothing functions non-parametrically using a scatterplot smoother. Thin-plate splines and cyclic cubic splines (in the 'months' covariate) were used as basic functions, and optimal tuning parameters were selected using the cross-validation method to account for seasonality. Finally, we modelled the hospitals' effects as independent random effects adding an interaction parameter between reperfusion therapy (patient level) and the hospital of treatment. Modelling hospitals as random effects allowed us to account for the natural clustering of the ischaemic stroke episodes within each hospital, thus measuring differences among them while considering the interaction with reperfusion therapy enabled also modelling the potential differences between both groups of patients. ICC[21] and CIs, MOR[22] and Bayesian credible intervals (BCIs) were calculated to estimate the independent effect of the hospitals—beyond differences in the underlying risks of patients—on ischaemic stroke case fatalities. The ICC was calculated using

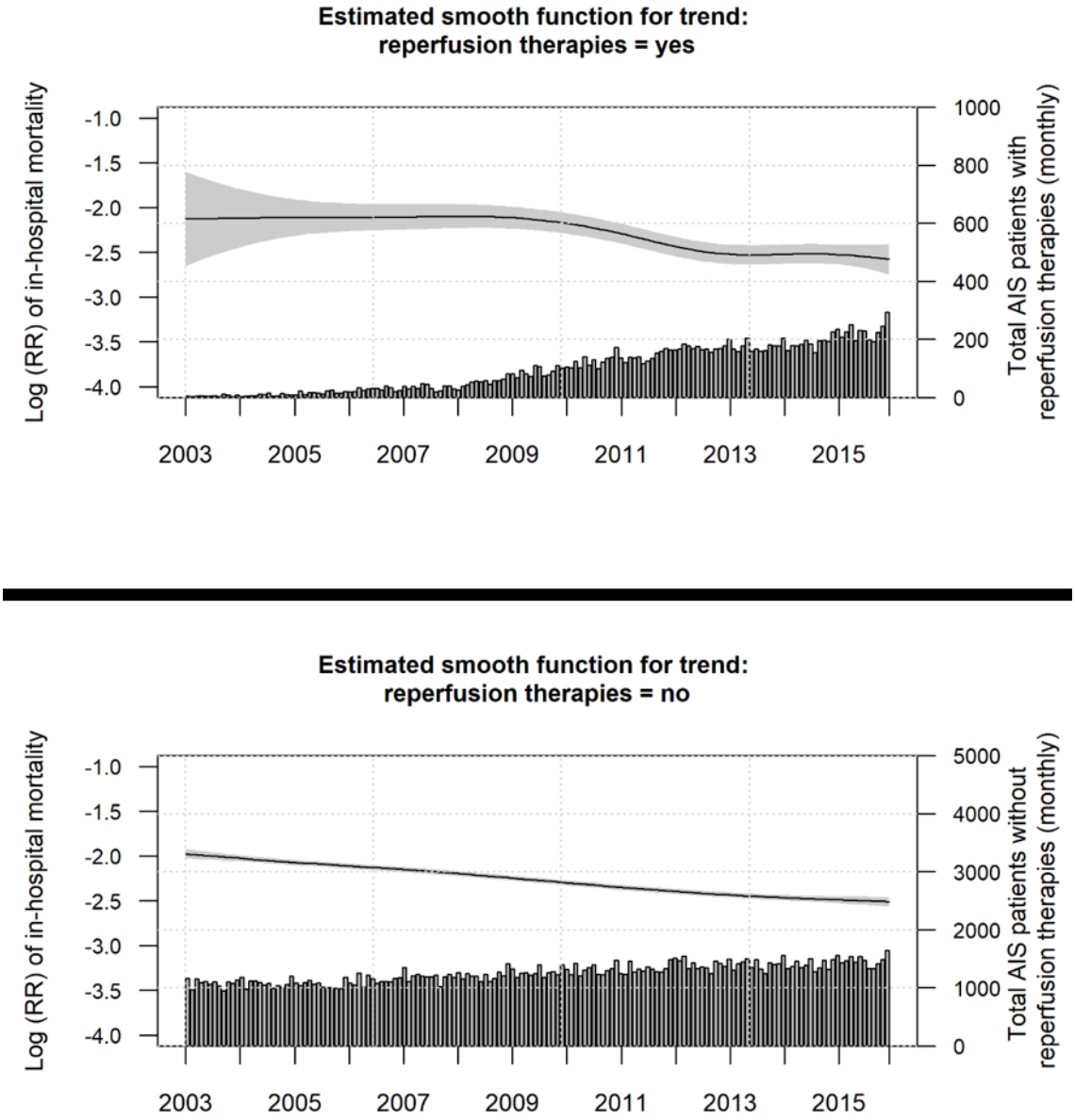


Figure 1 Fitted smooth functions for the in-hospital case fatality trend related to receiving or not reperfusion therapy. The model is fitted using thin-plate splines in the generalised additive mixed model. The main ordinate (left-side y-axis) represents the natural log of the relative risk of in-hospital case fatality due to AIS (log(RR)). The abscissa (x-axis) represents the time period variable (months). The solid line represents the relationship between log(RR) of in-hospital case fatality and the time period variable, with the shaded area representing the 95% CI. The right-side y-axis represents the evolution of the volume of patients with AIS (ie, bars) without reperfusion (above) or with reperfusion (below). AIS, acute ischaemic stroke.

the linear threshold model method, which is just a function of the area-level variance and consequently independent of the prevalence of the outcome. ICC informs of the proportion of total variation in the outcome that the hospital can explain. The MOR translates the variance attributed to the hospital into a more intuitive estimate that informs on the different relative risks for an individual if treated in a hospital with a different underlying risk of the outcome. The MOR is defined as the median value of the distribution of ORs obtained when randomly picking two patients with the same covariate values from two hospitals with a different underlying risk of an event of interest and comparing the one from the hospital with the highest risk with the one from the hospital with the lowest risk. In simple terms, the MOR can be interpreted as the median increased odds of reporting the outcome if a similar patient (ie, receiving reperfusion therapy or not) were treated in another hospital with a higher risk.

We used t-statistics to test the individual parametric coefficients' significance and evaluated the smooth terms' significance through the effective df. We tested the approximate significance of the smooth terms using X^2 statistics. Given the large numbers in this study, we performed all hypothesis tests using a type 1 error rate of 0.001. Finally, we considered the receiver operating characteristic curve for the estimation of the goodness of fit of the model. The decision on the final model presented in the results was taken according to the Akaike information

Table 2 Hospital-level random effects on case fatalities considering the differential use of reperfusion therapies

	Estimate	(95% Bayesian credible interval)	
ICC (reperfusion therapy=no)	0.0160	0.0100	0.0258
ICC (reperfusion therapy=yes)	0.0311	0.0167	0.0572
MOR (reperfusion therapy=no)	1.3089	1.2354	1.4087
MOR (reperfusion therapy=yes)	1.4589	1.3161	1.6806

ICC, intraclass correlation coefficient; MOR, median OR.

criteria. All statistical analyses were programmed using the R software,[23] including the *mgcv* package,[24] to fit the GAMM.

Patient and public involvement

None.

RESULTS

Table 1 describes the population of the study stratified for selected years. Over the period, there was a 36% increase in admissions from 12 976 to 17 658 patients with AIS. In turn, the proportion of patients who received some reperfusion therapy substantially increased from 0.52% to 15.45%. Regarding patients' profiles, the age and sex kept somewhat similar, with patients around 74 years old and 46% female, and an increased number of comorbidities registered, notably renal failure, metastatic cancers, coagulopathy and electrolytic disorders. On the contrary, the number of episodes registering lymphoma or chronic obstructive pulmonary disease (COPD) as comorbidity held over time.

The count of in-hospital mortality within 30 days of admission increased at 9.61%, from 1685 to 1848 in absolute numbers, with a decrease in in-hospital mortality rates from roughly 13% of in-hospital mortality in 2003 to 10.46% in 2015.

Patients' factors explaining differences in adjusted in-hospital mortality after AIS

Once seasonal effects were smoothed, patients who did not get any reperfusion therapy were less likely to die over the years, while those undergoing reperfusion therapy observed a steeper decrease in deaths, specifically after August 2008 (figure 1). Older patients were more likely to die within 30 days after admission. For instance, the OR for those patients in their 70s versus those in their 40s is 1.89 (95% CI 1.79 to 2.01). This is likewise the case for patients with metastatic cancer (OR=5.10, 95% CI 4.66 to 5.58), fluid and electrolyte disorders (OR 2.43; 95% CI 2.25 to 2.62), congestive heart failure (OR 2.30; 95% CI 2.12 to 2.42), pulmonary circulation disease (OR 1.69; 95% CI 1.54 to 1.85) and coagulopathy (OR 1.54; 95% CI 1.34 to 1.75). Other comorbidities also significantly associated with in-hospital mortality were renal failure (OR 1.38; 95% CI 1.30 to 1.46) and COPD (OR 1.18; 95% CI 1.12 to 1.23). Finally, patients admitted on a public holiday were more likely to die during their stay (OR 1.05; 95% CI 1.01 to 1.08). Finally, patients receiving reperfusion therapies had an increased mortality risk (OR 1.47; 95% CI 1.25 to 1.73). No statistically significant differences were found in the rest of the variables (see all statistically significant variables associated with the outcome in online supplemental table 2). All fitted smooth terms, including the interaction terms introduced in the model, were found as statistically significant (see online supplemental table 2).

Differences in adjusted mortality across referral stroke hospitals

Adjusted in-hospital mortality rates after AIS varied from 6.66% to 16.01% among hospitals. Beyond differences in patients' characteristics, the relative contribution of the hospital of treatment (table 2) was higher in the case of patients undergoing reperfusion therapies (ICC=0.031 (95% BCI=0.017 to 0.057)) than in the case of those who did not (ICC=0.016 (95% BCI=0.010 to 0.026)). Using the MOR, the difference in risk of death was as high as 46% between the hospital with the highest risk and the hospital with the lowest risk in patients undergoing reperfusion therapy (MOR 1.46 (95% BCI 1.32 to 1.68)), while in patients not undergoing any reperfusion therapy, the risk was 31% higher (MOR 1.31 (95% BCI 1.24 to 1.41)). This MOR difference between the two groups of patients was not statistically significant for a $\chi 2$ test of differences ($\chi 2>0.13$; p=0.7178). Figure 2 shows the observed variation across hospitals.

DISCUSSION

This study found a 19.5% decrease in AIS in-hospital mortality rate from 2003 to 2015, despite increasing AIS hospital admissions (36%) with more comorbidities on average.

A steep reduction in AIS in-hospital mortality was observed in 2008 (figure 1) in those patients receiving reperfusion therapy, concurrent with the widespread implementation of reperfusion therapies as part of the National Strategy on Stroke Care. However, this change in the trend was not observed in patients not receiving reperfusion therapy. Interestingly, after the decrease, in-hospital mortality rates in patients receiving any reperfusion therapy stagnated even though reperfusion therapies had been widely implemented across hospitals, potentially denoting a new phase in their implementation and utilisation.

These findings are consistent with the evolution of AIS hospital mortality rates in different international studies on the matter, such as the global stroke statistics at the national level,[25] the Global Burden of Disease Study,[26] or population-based studies in China,[27] Germany[28] or

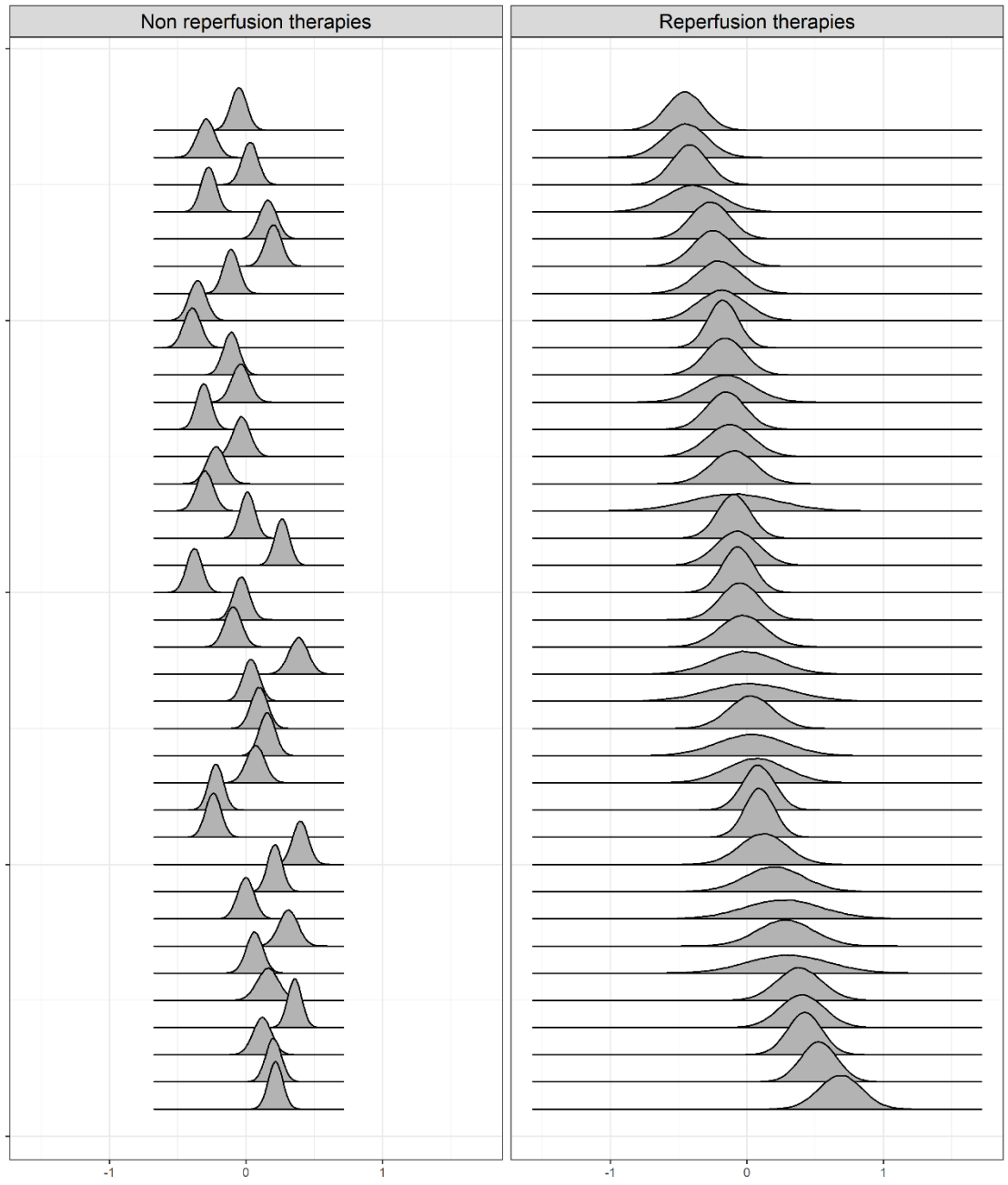


Figure 2 Hospital-level random effects: in-hospital case fatality probability density distribution with and without reperfusion therapy. The model is fitted using thin-plate splines in the generalised additive mixed model. The ordinate (y-axis) represents the referral hospitals for stroke analysed. The abscissa (x-axis) represents the normalised residuals at hospital level (or the in-hospital case fatality probability density distribution) in patients who did not undergo reperfusion therapies and in patients who received reperfusion therapies. The probability density distributions of patients who received reperfusion therapies are sorted by hospital from best to worst performance (in-hospital case fatality rate) to highlight the magnitude of the variation attributed to the hospital in those cases.

England.[29] In addition, prior studies on unwarranted variations in AIS in-hospital outcomes in Spain had also elicited differences in mortality rates as large as 4.7-fold after adjusting for age, sex and various comorbidities.[30]

Regarding our main endpoint, our study shows that referral stroke hospitals explained part of the variation in adjusted in-hospital mortality after AIS after patients' differences and time effects were adjusted in both patients undergoing reperfusion therapy and those who did not. The risk of dying was estimated to be 31% lower in the best-performing hospitals in patients not receiving reperfusion therapies and 46% lower in those patients receiving reperfusion therapies. The uneven adoption of intravenous fibrinolysis (figure 2), the early adoption of endovascular mechanical thrombectomy only in a small set of referral stroke hospitals,[31] different learning curves influencing the performance of those more complex interventions, organisational factors that affected the

implementation of the Stroke Code (for example, the existence of Stroke Units) differentially and finally, some unobserved latent factors as uneven patterns of patients' transfers from secondary hospitals to the referral stroke hospitals may explain these differences between hospitals.

Strengths and limitations

A first caveat interpreting our results is that this study focuses on referral stroke hospitals. Thus, any generalisation of the results could be referred to those hospitals accredited to provide high-profile treatments requiring 24/7 on-call specialists. Second, our data sources are limited to registering patients dying during the hospital stay. We did not have access to the patient's status before arriving at the hospital or after being discharged alive, leading to a possible underestimation of the overall mortality. In addition, different kinds of access to treatment because of geographical barriers or distance to the referral hospital, a time-sensible treatment condition, could also lead to differences in the use of reperfusion therapies and, consequently, to higher mortality. To assess the plausibility of this bias, we re-estimated AIS in-hospital mortality rates considering the travel distance (in minutes) from the centroid of the patient's area of residence to the nearest referral stroke hospital for all patients with valid information about the area of residence from 2012 to 2015. As a result, most patients receiving reperfusion therapies (92.42%) lived at a travel distance of 60 min or less, and all, irrespective of the treatment, lived at a travel distance of 180 min or less. This model provided similar results regarding the ICC, remaining the hospital of treatment effect on mortality rates (see online supplemental figure 2A).

Third, it is unlikely that we may have failed in capturing in-hospital deaths after AIS, provided that the dataset used for this study is part of the mandatory national statistics. In addition, our figures are comparable with those documented in the 2011 OECD report[19]—11.0 case fatalities per 100 patients who had a stroke in the OECD report vs 11.4 case fatalities per 100 patients who had a stroke in our sample.

Additionally, Elixhauser comorbidities in the model are subject to coding practices, which may vary across hospitals, biasing the adjustment. However, the sensitivity analysis results using just age, sex and time variables showed no differences in our estimates (see online supplemental figure 2B).

All in all, the results of our study are conditioned to the information registered in the administrative records. This issue may entail a limitation, as we may not have been able to adjust for some residual differences in patients' severity across hospitals. However, limiting the comparison to those referral stroke hospitals, likely to treat similar case-mix of patients, and the large numbers in this study, may have reduced the risk of residual confounding.

The differences among referral stroke hospitals elicited in this study point towards highly relevant opportunities to reduce unwarranted variations in medical practice and improve the provision of healthcare in conditions where effective interventions are available, for instance, by learning from best-performing hospitals.

CONCLUSION

Overall, in-hospital mortality rates within 30 days after admission decreased between 2003 and 2015, coincidentally with the widespread adoption of reperfusion therapies within referral hospitals in Spain. However, over the years, between-hospital variations in mortality persisted beyond differences in the patients treated. After its implementation, the National Strategy on Stroke Care seems to have prompted changes within the SNHS but unevenly across referral stroke hospitals covering the population in different regions.

Twitter Francisco Estupiñán-Romero @pacoramon

Acknowledgements The authors thank Natalia Martínez-Lizaga, from the Data Science for Health Policy and Services Research, at the Health Sciences Institute in Aragón, for her support in the selection and preparation of data, and to the Spanish Health Authorities participating in the Atlas VPM collaboration who granted access to the hospital datasets.

Collaborators on behalf of the Atlas VPM consortium team including Natalia Martínez-Lizaga, Manuel Ridao-López, Ester Angulo-Pueyo, Miriam Seral-Rodríguez, Ramón Launa-Garcés, Javier González-Galindo, Santiago Royo-Sierra

Contributors EB-D and FE-R worked on the conceptualisation and design. FE-R prepared the dataset. JPD implemented the analytical scripts and ran the analyses. Finally, FE-R and EB-D drafted the paper. All authors read, edited and approved the final manuscript and supplementary files.

Funding This study was partially funded by grants from CONCEPT-STROKE: Effectiveness and efficiency of acute ischemic stroke care pathways in five Spanish Regions (PI19/00154); REDISSEC: Red de Investigación en Servicios de Salud en Enfermedades Crónicas (RD16/0001/0007); and RICCAPS: Red de Investigación en cronicidad, atención primaria y promoción de la salud (RD21/0016/0023).

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not required.

Ethics approval This study, observational by design, uses retrospective anonymised non-identifiable and non-traceable data, and was conducted in accordance with the amended Helsinki Declaration, the International Guidelines for Ethical Review of Epidemiological Studies, and Spanish laws on data protection and patients' rights. The study was reviewed and approved by the Research Ethics Committee of Aragon (CEICA), which waived the need for written informed consent from the participants.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request. The data underlying this article will be shared upon reasonable request to the corresponding author.

ORCID iDs
Francisco Estupiñán-Romero http://orcid.org/0000-0002-6285-8120
Jaime Pinilla Dominguez http://orcid.org/0000-0002-7126-4236
Enrique Bernal-Delgado http://orcid.org/0000-0002-0961-3298

REFERENCES

1 Holodinsky JK, Yu AYX, Assis ZA, *et al*. History, evolution, and importance of emergency Endovascular treatment of acute ischemic stroke. *Curr Neurol Neurosci Rep* 2016;16:42. 10.1007/s11910-016-0646-5 Available: https://doi.org/10.1007/s11910-016-0646-5
2 National Institute of Neurological Disorders and Stroke rt-PA Stroke Study Group. Tissue plasminogen activator for acute ischemic stroke. *N Engl J Med* 1995;333:1581–7. 10.1056/NEJM199512143332401 Available: https://doi.org/10.1056/NEJM199512143332401
3 Hacke W, Kaste M, Fieschi C. Intravenous Thrombolysis with recombinant tissue plasminogen activator for acute Hemispheric stroke. *JAMA* 1995;274:1017.
4 Tsang ACO, Yang I-H, Orru E, *et al*. Overview of Endovascular Thrombectomy accessibility gap for acute ischemic stroke in Asia: A multi-national survey. *Int J Stroke* 2020;15:516–20.
5 Allen M, Pearn K, Ford GA, *et al*. National implementation of reperfusion for acute ischaemic stroke in England: how should services be Configured? A Modelling study. *Eur Stroke J* 2022;7:28–40.
6 Alvarez Sabín J, Alonso de Leciñana M, Gállego J, *et al*. Plan de Atención Sanitaria al Ictus [plan for stroke Healthcare delivery]. *Neurologia* 2006;21:717–26.
7 Ministerio de Sanidad y Política Social. Estrategia en Ictus del Sistema Nacional de Salud. Estrategia Aprobada Por El Consejo Interterritorial del Sistema Nacional de Salud El 26 de Noviembre de 2008. n.d. Available: https://www.mscbs.gob.es/organizacion/sns/planCalidadSNS/docs/EstrategiaIctusSNS.pdf
8 Díez-Tejedor E, Fuentes B. Stroke care in Spain. What do we have? What do we need *Neurología (English Edition)* 2011;26:445–8. 10.1016/j.nrleng.2011.01.011 Available: https://doi.org/10.1016/j.nrleng.2011.01.011
9 Ministerio de Sanidad, Servicios Sociales e Igualdad. Evaluación de la Estrategia en Ictus del Sistema Nacional de Salud. 2013. Available: https://saludcantabria.es/uploads/pdf/noticias/Punto%2011-%20Evaluacion%20Estrategia%20en%20Ictus%20del%20SNS%202013.pdf
10 Cayuela A, Cayuela L, Ortega Belmonte MJ, *et al*. Ha Dejado de Disminuir La Mortalidad Por Enfermedades Cerebrovasculares en España *Neurología* 2022;37:550–6. 10.1016/j.nrl.2019.06.001 Available: https://doi.org/10.1016/j.nrl.2019.06.001
11 Alonso de Leciñana M, Morales A, Martínez-Zabaleta M, *et al*. Características de Las Unidades de Ictus Y Equipos de Ictus en España en El Año 2018. *Neurología* 2023;38:173–80. 10.1016/j.nrl.2020.06.012 Available: https://doi.org/10.1016/j.nrl.2020.06.012
12 AHRQ QI version 4.5, inpatient quality indicators #17, technical specifications, acute stroke mortality rate. n.d. Available: www.qualityindicators.ahrq.gov
13 Tebé C, Abilleira S, Ridao M, *et al*. Atlas de Variaciones en El Manejo de la Enfermedad cerebrovascular Isquémica. *Atlas de Variaciones de La Práctica Médica En El Sistema Nacional de Salud (Atlas VPM)* 2003;5.
14 Elixhauser and Quan's ICD-9-CM and ICD-10 coding Algorithms for Elixhauser Comorbidity index. *Med Care* 2005;43:1130–9.
15 Elixhauser A, Steiner C, Harris DR, *et al*. Comorbidity measures for use with administrative data. *Med Care* 1998;36:8–27.
16 Ministerio de Sanidad. Registro de Actividad de Atención Especializada. *Conjunto Mínimo de Datos al Alta Hospitalaria España* 2015.
17 Ministerio de Sanidad,Consumo y Bienestar Social. Estadísticas E Información Sanitaria. Catálogo Nacional de Hospitales. Actualizado a 31 de Diciembre de 2014. NIPO en Línea: 731-19-039-5. 2015. Available: https://www.mscbs.gob.es/ciudadanos/prestaciones/centrosServiciosSNS/hospitales/docs/CNH2015.pdf
18 Angulo-Pueyo E, Comendeiro-Maaløe M, Estupiñán-Romero F, *et al*. Atlas VPM: two decades informing on unwarranted variations in health care in Spain. *Res Health Serv Reg* 2022;1.
19 OECD Publishing. In-hospital mortality following stroke. In: *Health at a glance 2011: OECD Indicators*. Paris: OECD Publishing, 2011.
20 Lin X, Zhang D. Inference in generalized additive mixed models by using smoothing Splines. *J R Stat Soc Series B Stat Methodol* 1999;61:381–400.
21 Snijders TAB, Bosker RJ. Modeled variance in two-level models. *Sociol Method & Res* 1994;22:342–63.
22 Merlo J, Chaix B, Ohlsson H, *et al*. A brief conceptual Tutorial of Multilevel analysis in social epidemiology: using measures of clustering in Multilevel logistic regression to investigate Contextual phenomena. *J Epidemiol Community Health* 2006;60:290–7. 10.1136/jech.2004.029454 Available: https://doi.org/10.1136/jech.2004.029454
23 R Core Team. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing, 2019. Available: https://www.R-project.org/
24 Wood S. Mgcv: mixed GAM computation vehicle with automatic smoothness estimation. R package version 1.1.8. 2013. Available: https://CRAN.R-project.org/package=mgcv
25 Kim J, Thayabaranathan T, Donnan GA, *et al*. Global stroke Statistics 2019. *Int J Stroke* 2020;15:819–38.
26 Lopez AD, Murray CC. The global burden of disease, 1990–2020. *Nat Med* 1998;4:1241–3. 10.1038/3218 Available: https://doi.org/10.1038/3218
27 He Q, Wu C, Luo H, *et al*. Trends in in-hospital mortality among patients with stroke in China. *PLoS ONE* 2014;9:e92763.
28 Heuschmann PU, Berger K, Misselwitz B, *et al*. Frequency of thrombolytic therapy in patients with acute ischemic stroke and the risk of in-hospital mortality. *Stroke* 2003;34:1106–13.
29 Seghieri C, Mimmi S, Lenzi J, *et al*. 30-day in-hospital mortality after acute myocardial infarction in Tuscany (Italy): an observational study using hospital discharge data. *BMC Med Res Methodol* 2012;12:170. 10.1186/1471-2288-12-170 Available: https://doi.org/10.1186/1471-2288-12-170
30 Tebé C, Abilleira S, Ridao M, *et al*. Atlas de Variaciones en El Manejo de la Enfermedad cerebrovascular Isquémica. *Atlas de Variaciones de La Práctica Médica* 2013. Available: https://www.atlasvpm.org/wp-content/uploads/2019/06/Atlasnumero9_2.6MB.pdf
31 López Fernández JC, Masjuan Vallejo J, Arenillas Lara J, *et al*. Analysis of stroke care resources in Spain in 2012: have we benefitted from the Spanish health system's stroke care strategy. *Neurología (English Edition)* 2014;29:387–96. 10.1016/j.nrleng.2013.06.009 Available: https://doi.org/10.1016/j.nrleng.2013.06.009

ONLINE SUPPLEMENTARY MATERIALS

Table S1. Inpatient Quality Indicator 17 (IQI 17) Acute Stroke Mortality Rate – Stratum C (Ischemic stroke) (validated definition of the SNHS by Atlas VPM)

Description: In-hospital deaths with acute ischemic stroke as a principal diagnosis for patients ages 18 years and older
Numerator: Number of deaths among cases meeting the inclusion and exclusion rules for the denominator
Denominator: Discharges, for patients ages 18 years and older, with a principal ICD-9-CM diagnosis code for acute ischemic stroke
Denominator exclusions: · transferring to another short-term hospital · MDC 14 (pregnancy, childbirth, and puerperium) · with missing discharge disposition, gender, age (age group), or principal diagnosis
Additional Denominator exclusions (decided by the authors): · with length of in-hospital stay greater than 30 days
Ischemic stroke diagnosis codes: **icd-9th code** | **icd-9th description** 43301 | BASI ART OCCL W/ INFARCT 43391 | PRECER OCCL NOS W/ INFRCT 43311 | CAROTD OCCL W/ INFRCT 43401 | CERE THROMBOSIS W/ INFRCT 43321 | VERTB ART OCCL W/ INFRCT 43411 | CERE EMBOLISM W/ INFRCT 43331 | MULT PRECER OCCL W/ INFRCT 43491 | CEREB OCCL NOS W/ INFRCT 43381 | PRECER OCCL NEC W/ INFRCT
Extract from the STATA 14 script for IQI 17 selection by Natalia Martínez-Lizaga ********** ***TPIQ17 MODELO ISQUEMICO con 436 ********** gen byte acute_isq=0 foreach var of varlist c1 { replace acute_isq=strmatch("433.01",substr(`var',1,6))+acute_isq replace acute_isq=strmatch("433.11",substr(`var',1,6))+acute_isq replace acute_isq=strmatch("433.21",substr(`var',1,6))+acute_isq replace acute_isq=strmatch("433.31",substr(`var',1,6))+acute_isq replace acute_isq=strmatch("433.81",substr(`var',1,6))+acute_isq replace acute_isq=strmatch("433.91",substr(`var',1,6))+acute_isq replace acute_isq=strmatch("434.01",substr(`var',1,6))+acute_isq replace acute_isq=strmatch("434.11",substr(`var',1,6))+acute_isq replace acute_isq=strmatch("434.91",substr(`var',1,6))+acute_isq replace acute_isq=strmatch("436",substr(`var',1,3))+acute_isq } replace acute_isq=1 if acute_isq>1 gen tpiq17_isq=0 if acute_isq==1 & (mdc!=1) & edad>17 & tipalt!=2 replace tpiq17_isq=1 if tpiq17_isq==0 & tipalt==5 ***Discharge type ==5 corresponds to death

Table S2. *Generalized additive mixed model (GAMM)*: (A) Parametric coefficients, (B) Approximate significance of the fitted smooth terms

Table S2 (A): Parametric coefficients of the GAMM.

	Estimate	**Std.Error**	**Pr(>|z|)**
Intercept	-2.41932	0.03964	<0.001
Reperfusion therapy	0.38705	0.08206	<0.001
Public holiday admission	0.04554	0.01651	0.00582
Metastatic cancer	1.62924	0.04643	<0.001
Fluid and electrolyte disorders	0.88655	0.03875	<0.001
Congestive heart failure	0.83324	0.02613	<0.001
Pulmonary circulation disease	0.52546	0.04684	<0.001
Coagulopathy	0.42883	0.06794	<0.001
Lymphoma	0.32624	0.11520	0.0046
Renal failure	0.32109	0.02817	<0.001
Chronic obstructive pulmonary disease	0.16217	0.02423	<0.001

Area under the ROC curve for the GAMM = **0.7188063**

Table S2 (B): Approximate significance of the fitted smooth terms.

	Edf	**Ref.df**	**Pr(>|z|)**
s(age)	6.146	7.098	<0.001
s(T):reperfusionT0	3.046	3.787	<0.001
s(T):reperfusionT1	4.000	4.941	<0.001
s(Nmonth)	3.395	8.000	<0.001
s(Hosp):reperfusionT0	34.429	36.000	<0.001
s(Hosp):reperfusionT1	25.228	36.000	<0.001

N.B. *Tables S1 (A) and (B) are direct outputs of the final GAMM model. Table S1 (A) shows the results for the parametric coefficients, while table S1 (B) shows the outputs for those fitted smooth terms introduced in the same model. While parametric coefficients (variables traditionally introduced as categorical or binary variables in a regression) are reported using their beta coefficients (and standard errors), fitted smooth terms can only be reported using their effective degrees of freedom for the polynomial grade used to fit their interaction with the estimated outcome (Edf | Ref.df).*

Estupiñán-Romero F, *et al. BMJ Open* 2023; 13:e068183. doi: 10.1136/bmjopen-2022-068183

Figure S1. Flowchart of the hospitalization episodes due to acute ischemic stroke selected for the study.

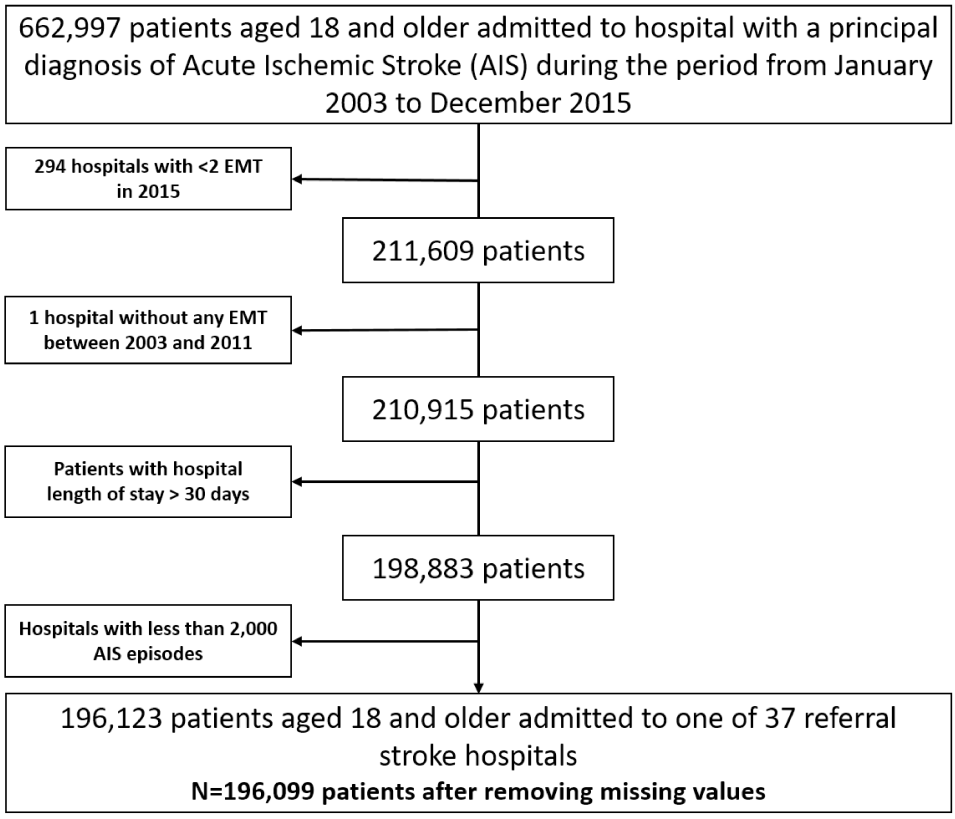


Figure S2. Sensitivity analysis: (A) Model comparison including distance; (B) Model comparison without Elixhauser comorbidity variables

Fig. S2 (A): Differences in hospital level random effects parameters between Model with traveling distance and Model without traveling distance.

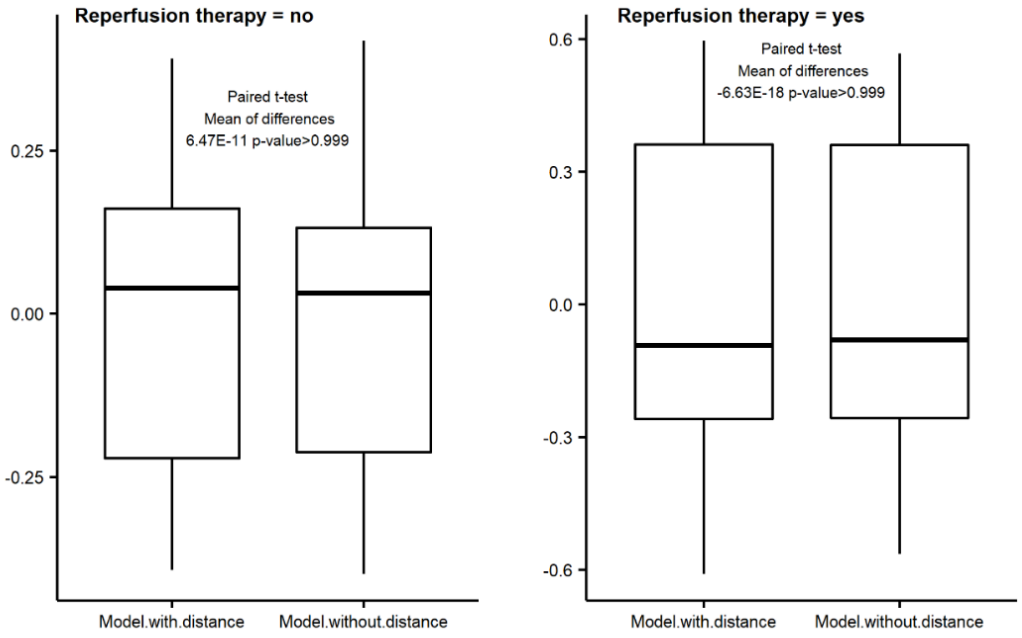
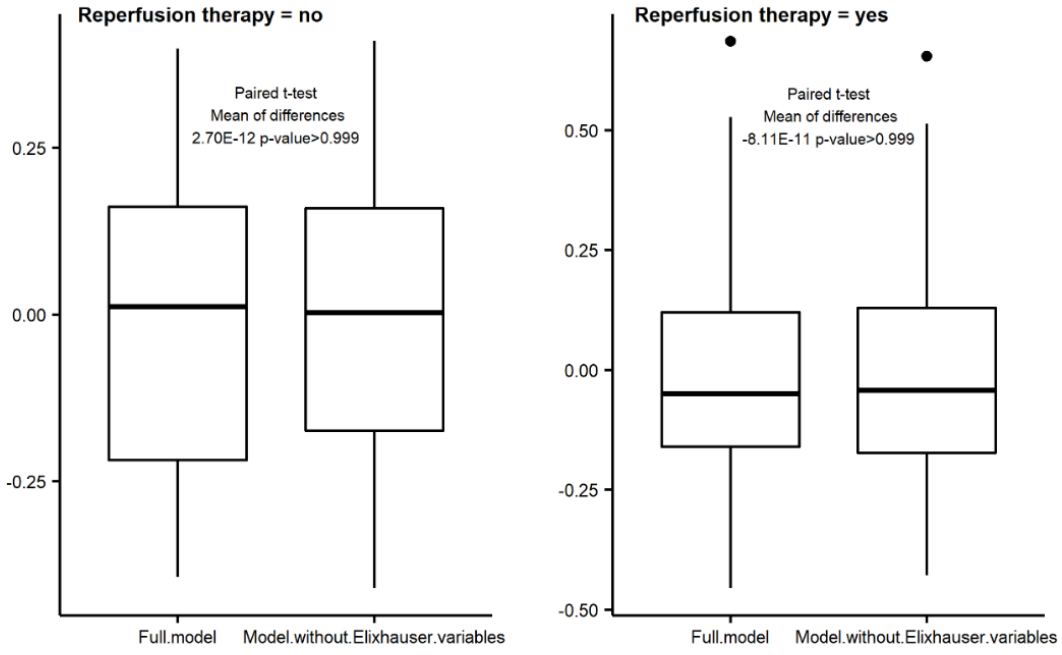


Fig. S2 (B): Differences in hospital level random effects parameters between Full Model and Model without Elixhauser comorbidity variables.



N.B. *Figure S2 offers direct comparison between the distributions of the hospital level random effects parameters. The difference is tested using a paired T-test on the means of both distributions. Figure S2 (A) compares the final model used to estimate AIS 30-day in-hospital mortality (Model.without.distance) with the same model adding a parameter considering the traveling distance from patients' location to the nearest stroke referral hospital (Model.with.distance). Figure S2 (B) compares the final model (Full.model) with the same model without introducing the Elixhauser variables (Model.without.Elixhauser.variables). The comparison of the distribution of the hospital level random effect parameters is shown segmented between those patients not receiving reperfusion therapy (left) and those receiving reperfusion therapy (right) to facilitate the interpretation of the interaction term between hospital of treatment and the trend of utilization of reperfusion therapy.*