

Core data set para la generación de datos de la vida real en esclerosis múltiple: adaptación de una iniciativa global para América Latina

Juan I. Rojas, Fernando Gracia, Tina Parciak, Ricardo Alonso, Jefferson Becker, Irene Treviño-Frenk, Marina Alonso-Serena, Diego Giunta, Patricio Abad, Edgar Carnero-Contentti, Adriana Carrá, Edgar P. Correa-Díaz, Jorge Correale, Edgardo Cristiano, José Flores, Manuel Fruns, Lorna Galleguillos, Orlando Garcea, Fernando Hamuy, Marco Lana-Peixoto, Carlos Navas, Regina Pappais-Alvarenga, Liliana Patrucco, Víctor Rivera, Silvia Tenembaum, María C. Ysraelit, Liesbet M. Peeters

Introducción. Los objetivos primarios del *core data set* son reducir la heterogeneidad y promover la armonización entre las fuentes de datos en la esclerosis múltiple (EM), reduciendo así el tiempo necesario para ejecutar esfuerzos en la recolección de datos de vida real. Recientemente, un grupo liderado por la Multiple Sclerosis Data Alliance ha desarrollado un *core data set* para la recolección de datos del mundo real en EM a nivel global. Nuestro objetivo ha sido adaptar y consensuar este conjunto de datos globales a las necesidades de América Latina para que pueda ser implementado por los registros ya desarrollados y en proceso de desarrollo en la región.

Material y métodos. Se conformó un grupo de trabajo regionalmente y se adaptó el *core data set* creado globalmente (proceso de traducción al español, incorporación de variables regionales y consenso sobre variables que se iban a utilizar). El consenso se obtuvo a través de la metodología Delphi remoto de ronda de cuestionarios y discusión a distancia de las variables del *core data set*.

Resultados. Veinticinco profesionales de América Latina llevaron adelante el proceso de adaptación entre noviembre de 2022 y julio de 2023. Se estableció un acuerdo sobre un *core data set* de nueve categorías y 45 variables, versión 2023, con la sugerencia de implementarlo en registros desarrollados o en vías de desarrollo y cohortes de EM en la región.

Conclusión. El *core data set* busca armonizar las variables recolectadas por los registros y las cohortes de EM en América Latina con el fin de facilitar dicha recolección y permitir una colaboración entre fuentes. Su implementación facilitará la recolección de datos de vida real y la colaboración en la región.

Palabras clave. América Latina. Cohortes. Datos de vida real. Epidemiología. Esclerosis múltiple. Registros.

Introducción

Los datos de la vida real se refieren a información recopilada de eventos, situaciones o actividades que ocurren en el mundo real, en contraposición a los datos generados en entornos controlados o experimentales [1-3]. Estos datos se obtienen a partir de observaciones directas, mediciones, encuestas, cohortes, registros médicos y otras fuentes similares [1-3].

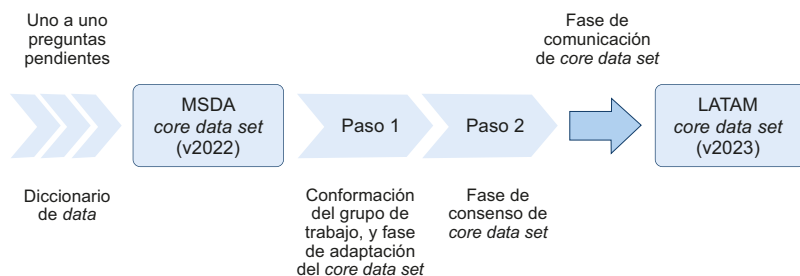
Los distintos entes regulatorios en el mundo han comenzado a valorar de manera significativa los datos de vida real, sobre todo los datos provenientes de los registros estructurados y de calidad como una fuente muy valiosa para las evaluaciones de efectividad, seguridad, riesgo-beneficio, y coste de los medicamentos durante y después de la autorización de éstos [1,2,4,5].

Si bien los aspectos metodológicos sobre el desarrollo de cohortes o registros han evolucionado significativamente en los últimos años, junto con la aparición de registros de calidad en el ámbito de la esclerosis múltiple (EM), un aspecto sin resolver aún y que complica la interoperabilidad de las fuentes (capacidad de intercambiar información entre sí) es el que concierne a la heterogeneidad en las variables de datos recolectadas en cada registro o cohortes [5-7]; específicamente, el punto de desarrollar lo que se denomina un *core data set* entre los diferentes registros existentes o set de variables comunes entre las cohortes o los registros funcionales [6]. En 2017, la Agencia Europea de Medicamentos, en su directriz sobre estudios basados en registros, promovió el uso de variables de datos comunes, esto es, una serie de variables homogéneas para usar por los distintos registros en desarrollo en EM

Centro de Esclerosis Múltiple de Buenos Aires (J.I. Rojas, E. Cristiano, L. Patrucco). Unidad de Neuroinmunología. Departamento de Neurociencia. Hospital Alemán (J.I. Rojas, E. Carnero-Contentti). Centro Universitario de Esclerosis Múltiple. Hospital J.M. Ramos Mejía (R. Alonso, O. Garcea). Facultad de Medicina. Universidad de Buenos Aires (R. Alonso, O. Garcea). Sanatorio Güemes (R. Alonso). Servicio de Clínica Médica. Hospital Italiano de Buenos Aires (M. Alonso-Serena, D. Giunta). Sección de Enfermedades Desmielinizantes. Hospital Británico (A. Carrá). Hospital Universitario Fundación Favaloro (A. Carrá). Departamento de Neurología. FLENI (J. Correale, M.C. Ysraelit). Clínica de Neuroinmunología Pediátrica. Departamento de Neurología. Hospital de Pediatría Prof. Dr. Juan P. Garrahan. Buenos Aires, Argentina (S. Tenembaum). Clínica de Esclerosis Múltiple. Servicio de Neurología. Hospital Santo Tomás. Panamá, Panamá (F. Gracia). Instituto del Cerebro Grande do Sul. Universidad Pontificia de Rio Grande do Sul. Porto Alegre (J. Becker). Centro de Investigación CIEM MS. Facultad de Medicina. Universidad Federal de Minas Gerais. Belo Horizonte (M. Lana-Peixoto). Departamento de Neurología. Universidad Federal de Rio de Janeiro. Río de Janeiro, Brasil (R. Pappais-Alvarenga). Departamento de Neurología. Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán (I. Treviño-Frenk). Instituto Nacional de Neurología y Neurocirugía. Ciudad de México, México (J. Flores). Servicio de Neurología. Hospital Metropolitano de Quito (P. Abad). Departamento



Figura 1. Pasos seguidos en el proceso de desarrollo del *core data set*.



LATAM: América Latina; MSDA: alianza de datos de esclerosis múltiple.

macéutica y agencias reguladoras, y desarrolló un *core data set* basado en ocho categorías y 42 variables (en prensa). Para garantizar la alineación entre el *core data set* y las iniciativas importantes no específicas de EM, el diccionario de datos del *core data set* contuvo códigos y términos de SNOMED que, sin duda, es un beneficio adicional para los registros [10]. El *core data set* está en vías de comunicación internacionalmente. Ese paso se tomó como el inicio y sobre eso se llevaron adelante los siguientes pasos (Fig. 1).

Nuestro objetivo ha sido adaptar y consensuar este conjunto de datos globales a las necesidades de América Latina para que pueda implementarse por los registros ya desarrollados y en proceso de desarrollo en la región.

Material y métodos

Conformación del grupo de trabajo y fase de adaptación del *core data set* al contexto de América Latina

Para esta etapa se trabajó en tres subetapas: a) etapa de conformación del grupo de trabajo; b) etapa de traducción de *core data set*; y c) etapa de incorporación de variables de acceso al *core data set* original.

- *Etapa de conformación del grupo de trabajo:* para esta etapa se conformó un grupo de trabajo metodológico encargado de llevar adelante las etapas de traducción, búsquedas, validaciones y revisiones abajo detalladas, y un grupo de trabajo de consenso. El equipo metodológico estuvo constituido por expertos en epidemiología y estadística, y en el desarrollo de consensos formales con experiencia previa en el desarrollo de este tipo de consensos en la región [11-13]. El grupo de trabajo de consenso estuvo constituido por neurólogos y profesionales médicos miembros de LACTRIMS vinculados a la atención de pacientes con EM en toda la región de América Latina. La selección de los expertos se basó en su experiencia en el seguimiento de pacientes con EM y en la generación y utilización de datos de vida real. Sólo los que acordaron participar en el proceso se incluyeron en el documento de trabajo final.
- *Etapa de traducción de core data set:* las variables creadas en el *core data set* por la MSDA se tradujeron al español en un proceso hacia adelante y hacia atrás (*backward* y *forward*). Las variables originales se tradujeron al español y luego nuevamente se tradujeron al inglés original

de Neurología. Hospital Carlos Andrade Marín. Universidad Central del Ecuador. Quito, Ecuador (E.P. Correa-Díaz). Clínica Las Condes (M. Fruns). Universidad del Desarrollo (L. Galleguillos). Clínica Alemana. Santiago de Chile, Chile (L. Galleguillos). Departamento de Neurología. Hospital IMT. Asunción, Paraguay (F. Hamuy). Clínica Enfermedad Desmielinizante. Clínica Universitaria Colombia. Bogotá, Colombia (C. Navas). Universidad de Medicina Baylor. Houston, Texas, EE. UU. (V. Rivera). Instituto de Investigación Biomédica e Instituto de Ciencia de Datos. Universidad Hasselt. Diepenbeek, Bélgica (T. Parciak, L.M. Peeters).

E-mail:

rojasjuanignacio@gmail.com

Financiación:

El estudio se llevó a cabo con una subvención irrestricta de LACTRIMS.

Aceptado tras revisión externa:

05.02.24.

Conflicto de intereses:

Deseamos confirmar que no existen conflictos de interés conocidos asociados con esta publicación y que la financiación LACTRIMS no influyó en el contenido del proyecto desarrollado.

Cómo citar este artículo:

Rojas JI, Gracia F, Parciak T, Alonso R, Becker J, Treviño-Frenk I, et al. *Core data set* para la generación de datos de la vida real en esclerosis múltiple: adaptación de una iniciativa global para América Latina. *Rev Neurol* 2024; 78: 185-97. doi: 10.33588/rn.7807.2023326.

© 2024 Revista de Neurología

con el objetivo de poder armonizar las variables y la forma en que se recolectan esas variables [6]. Sin embargo, esto no se implementó colectivamente. Todo lo anterior ha llevado a una alta heterogeneidad entre los registros en EM en lo que se refiere al contenido y, sobre todo, a la representación semántica y sintáctica de los datos que actualmente se recolectan [6,7]. Esta heterogeneidad complica la colaboración entre las fuentes, especialmente a gran escala.

Sin dudas, las partes interesadas (entes regulatorios, profesionales de salud, pacientes, pagadores de salud, asociaciones de pacientes, etc.) ven el valor de los datos del mundo real y reconocen la necesidad de generar variables de datos comunes entre los registros y cohortes en EM [6-8]. Por lo tanto, varias iniciativas en el campo de la EM ya han enfrentado el desafío de acordar un conjunto de variables homogéneas y armonizadas que se implementen entre los registros con el objetivo de que, al implementarse, la forma de recolección de datos sea homogénea, y así permita el intercambio y la complementación de la información entre las fuentes [7,8].

Recientemente, un grupo liderado por la Multiple Sclerosis Data Alliance (MSDA) ha desarrollado un *core data set* para la recolección de datos del mundo real en EM a nivel global [9]. La MSDA es una colaboración global de múltiples partes interesadas que tiene como objetivo, entre otros, desarrollar herramientas para reducir el nivel de heterogeneidad entre diferentes fuentes de datos de EM del mundo real [9]. Durante 2022, la alianza desarrolló un grupo de trabajo que constaba de 18 expertos y líderes de opinión clave en el campo de datos de vida real en EM, incluidos médicos, custodios de datos y líderes de registros cohortes de EM, industria far-

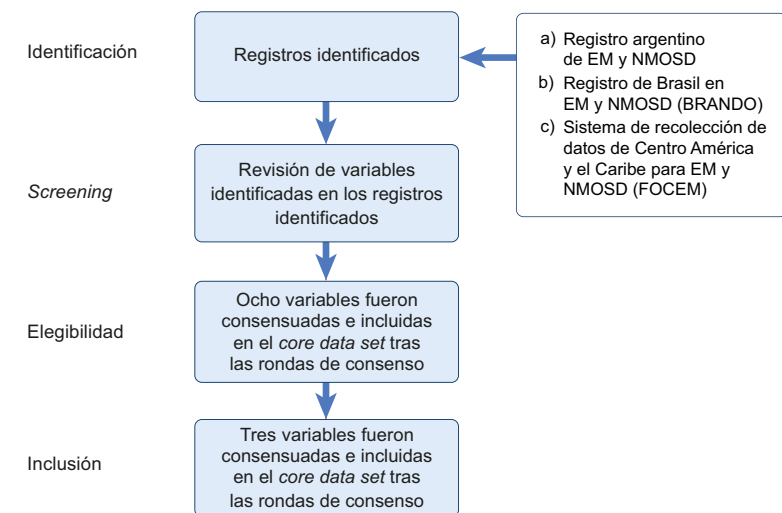
Tabla I. Síntomas considerados en el *core data set* adaptado.

Categoría	Síntoma	Categoría	Síntoma
Psicosocial	Ansiedad	Tracto urinario	Urgencia miccional
	Ataque de pánico		Trastorno esfinteriano
	Depresión		Disinergia
	Estrés		Defecación involuntaria
Dolor	Cefalea	Intestino	Estreñimiento
	Dolores articulares	Miembros superiores	Lateralización
	Neuralgia trigeminal		Incoordinación
Queja de memoria	Temblores		
Cognición	Déficit atencional	Espasticidad	Debilidad
	Afasia		Parálisis
	Trastorno del equilibrio		Rigidez
Equilibrio	Mareos	Miembros inferiores	Calambres
	Vértigo		Debilidad
	Entumecimiento		Parálisis
Sensibilidad	Parestesias	Dificultad para hablar	Caída del miembro
	Disestesias		Trastorno de la marcha
	Trastornos termoalgésicos		Dificultad para tragar
Fatiga	Fatiga	Disfunción sexual	
	Visión borrosa		
	Visión doble		
Visión	Ceguera parcial		

por nativos que dominaban el idioma inglés que viven en Argentina, los cuales se consolidaron en una versión inicial en español. Esa versión se revisó en una prueba donde se evaluó la redacción y el nivel de comprensión en ocho entrevistas realizadas por el equipo de métodos.

- *Etapa de incorporación de variables de acceso al core data set original:* (Tabla I) para esta etapa se siguió el proceso original creado por la MSDA para la incorporación de variables al *core data set*, pero aplicado a la categoría de variables de

acceso y enfocado en registros y sistemas de recolección de datos de América Latina. Se hizo: a) revisión de la bibliografía de registros en curso en América Latina; b) evaluación de los metadatos del catálogo de la MSDA (<https://msda.emif-catalogue.eu/c/msda/dbEdit/a09ae56c598ca875f7517ac5b4b6d20/95/5>), alojado en la plataforma de la web que se lanzó en 2019 y que permite a los usuarios finales buscar perfiles de metadatos de datos del mundo real de cohortes de EM enfocados en América Latina; y c) bús-

Figura 2. Resultados de la estrategia de búsqueda de variables de acceso.

EM: esclerosis múltiple; NMOSD: trastorno del espectro de la neuromielitis óptica.

acuerdo general sobre cada variable de ser parte del *data set* (<https://forms.gle/u9TmfVjrJTbKdiWE7>), así como se evaluaron las observaciones generadas. Cada variable se analizó posteriormente de forma global entre todas las respuestas. Se estableció definir la presencia de acuerdo cuando el 70% de los encuestados respondía de forma homogénea; definir la incertidumbre cuando entre el 40 y el 70% respondían de forma homogénea; y definir el no acuerdo si menos del 40% respondía de forma homogénea. Los principales métodos estadísticos utilizados fueron medidas de tendencia central y dispersión: media, mediana, moda, máximo, mínimo y desviación estándar.

Fase de comunicación del *core data set*

Una vez formado el grupo metodológico y el grupo de expertos, llevado adelante el proceso de adaptación del *core data set* y teniendo el documento local, se procedió a circularlo entre el grupo de trabajo para evaluar el consenso sobre cada una de las variables del *data set* y determinar cuál quedaba en él.

queda en Google Scholar sobre registros en América Latina. Una vez realizada la búsqueda, se enfocó en datos de acceso a la salud del '*core data element* de los hallazgos'. La estrategia de búsqueda para llegar a ese aspecto fue ('*core data set*' OR '*minimum data set*' OR '*minimal data set*' AND '*Latin America*' AND '*multiple sclerosis*') y, en segundo lugar, una estrategia más amplia ('*real-world data*' AND '*multiple sclerosis*' AND '*Latin America*').

Fase de consenso del *core data set*

Una vez identificadas las variables de acceso recolectadas por los registros, como cobertura de salud, tiempo a la cobertura, nombre de tratamientos, etc., se registraron las variables y la forma de recolección de las variables de la categoría en el proceso de adaptación del *core data set*. Para concretar el consenso se siguió la metodología Delphi remoto de ronda de cuestionarios y discusión a distancia de las variables del *core data set* (escenarios) [12-14]. Operativamente se realizó una primera, una segunda y una tercera rondas (de ser necesario) de cuestionario vía virtual con los expertos en los que cada involucrado revisó el *data set* adaptado. Se consultó sobre el grado de acuerdo de cada experto. Se recibió de cada experto la respuesta sobre cada variable y comentarios (de existir), y se vio el grado de

Resultados

Un total de 23 neurólogos y dos metodólogos, que recibieron el protocolo y el proceso de implementación, aceptaron participar en el proyecto. Estos participantes provenían de Argentina (10), Brasil (tres), Colombia (uno), Costa Rica (uno), Chile (dos), Ecuador (dos), El Salvador (uno), Guatemala (uno), México (tres), Paraguay (uno), Panamá (uno), Perú (uno), República Dominicana (uno), Uruguay (uno) y Venezuela (uno). El desarrollo del proyecto tuvo lugar entre noviembre de 2022 y julio de 2023, y contó con la colaboración metodológica y logística de la MSDA. La estrategia de búsqueda utilizada identificó las variables de acceso que se incluyeron en el conjunto de datos básicos original creado previamente por la MSDA y que fueron traducidas en el proceso previo desarrollado (Fig. 2). Durante el proceso de consenso se llevaron a cabo dos rondas virtuales para que cada profesional evaluara las afirmaciones generadas. Una vez alcanzado el consenso en las afirmaciones, se llevó a cabo una tercera ronda virtual para generar comentarios y realizar una revisión final. Las diversas discusiones del grupo de trabajo condujeron al acuerdo sobre nueve categorías y 45 variables para el *core data set*, versión 2023 para América Latina. Después de acordar las categorías, variables y valores, se finalizó el diccionario de datos del conjunto central de datos.

Además, la categoría ‘acceso a la salud’ incorporada en el conjunto de datos básicos adaptado para América Latina consensuó la inclusión de tres variables, las cuales se detallan en las tablas I y II.

Discusión

El objetivo primario del *core data set* es el de reducir la heterogeneidad y promover la armonización entre las fuentes de datos en EM, reduciendo así el tiempo necesario para ejecutar esfuerzos en la recolección de datos de vida real y optimizar la colaboración global a gran escala [5-7].

En el presente trabajo logramos adaptar el *core data set* creado globalmente (en prensa) al idioma español, identificamos las variables de acceso que tan importantes son en nuestro medio y que tanta relevancia tienen para ser homogeneizadas por los sistemas de recolección de datos, y en tercer lugar llegamos a un consenso con los principales involucrados en la generación y utilización de datos de vida real en EM en la región (Tabla III). Esto podría garantizar que el *core data set* se implemente en los registros en vías de desarrollo y desarrollados con él a fin de armonizar la recolección de datos que la facilite y contribuya a la colaboración regional a gran escala [5-7].

Es importante mencionar que hay algunas variables que *a priori* son relevantes para la enfermedad, como neurofilamentos, tomografía de coherencia óptica (OCT) o algunas de resonancia magnética, seguridad o resultados comunicados por pacientes, así como algunas que sí se incluyeron en el *core data set* original, pero que no alcanzaron el consenso suficiente en este proceso para incluirlas aquí. Algunas variables muy específicas en algunas ocasiones deben considerarse aspectos más relacionados con aspectos de investigación más profundos, que, en muchas ocasiones, no hacen el seguimiento rutinario de los pacientes con EM. Esto no quita que cada registro luego implemente las variables que considere pertinentes, pero es importante tener el concepto de mantener el *data set* lo más simple posible.

Esta adaptación también prevé una revisión periódica del conjunto de datos básicos.

Como se menciona en el desarrollo del *core data set* global (en prensa), una fortaleza es la alineación entre el *core data set* y los códigos y términos de SNOMED CT, lo que puede ser un beneficio adicional para los registros. SNOMED CT, como terminología sanitaria multilingüe internacionalmente validada, se utiliza ampliamente en todo el mun-

Tabla II. Comorbilidades consideradas en el *core data set* adaptado.

	Subcategoría
Abuso de alcohol	Espondilitis anquilosante
	Diabetes de tipo I
	Enfermedad inflamatoria intestinal
	Miastenia grave
	Psoriasis
	Artritis reumatoide
	Lupus eritematoso sistémico
Autoinmune distinta de esclerosis múltiple	Tiroiditis
	Mama
	Digestivo
	Tiroideo
	De vulva
	Pulmón
	Piel
Cáncer	Hematológico
	Diabetes de tipo 2
	Abuso de drogas
	Epilepsia
	Cataratas
	Glaucoma
	Degeneración macular
Fibromialgia	Depresión
	Psicosis
	Ansiedad
	Accidente cerebrovascular
	Hiperlipidemia
	Hipertensión
	Cardiovascular
Cardiovascular	Isquemia cardíaca
	Enfermedad vascular periférica
	Arritmia cardíaca
	Insuficiencia cardíaca congestiva
	Respiratoria
	Asma
	Enfermedad pulmonar

Tabla III. Core data set adaptado y consensuado.

Categoría	Recolectado por ^a	Nombre de la variable	Nombre variable	Colección tiempo	Formato	SNOMED definición	SNOMED código	Valores de la variable	SNOMED término	SNOMED código
Categoría 1. Demografía										
A		Fecha de nacimiento	fecha_nac	PV	aaaa-mm-dd	Fecha de nacimiento	184099003			
A		Sexo (biológico)	sexo	PV	radio/select	Sexo biológico	734000001	Mujer	Mujer (hallazgo)	248152002
								Hombre	Hombre (hallazgo)	248153007
A		País de residencia	residencia	PV, VS	lista	País de residencia	416647007	ISO 3166-1 (alpha-2)	-	-
A		Ciudad de residencia	ciudad	PV, VS	lista	Ciudad de residencia				
A		Raza	raza	PV	lista	Raza observable	103579009	Blanco, negro, afro-americano, indo-americano, nativo de Alaska, asiático, nativo de Hawái u otras islas del Pacífico, desconocido, no seguro	Desconocido (valor calificativo)	261665006
A		Etnia	etnia	PV	lista	Grupo étnico identificable clínicamente	397731000	Hispánico o latino, no hispánico o latino, desconocido, no seguro		
A		Nivel educativo	educación	PV	lista	Nivel educacional alcanzado	105421008	ISCED 0 = educación infantil; ISCED 1 = educación primaria; ISCED 2 = educación secundaria inicial; ISCED 3 = educación secundaria superior; ISCED 4 = educación secundaria no terciaria; ISCED 5 = educación terciaria incompleta; ISCED 6 = educación terciaria completa; ISCED 7 = educación universitaria; ISCED 8 = educación doctoral o post universitaria		

Tabla III. Core data set adaptado y consensuado (cont.).

Categoría	Recolectado por ^a	Nombre de la variable	Nombre variable	Colección tiempo	Formato	SNOMED definición	SNOMED código	Valores de la variable	SNOMED término	SNOMED código
A		Estado laboral	empleo	PV, VS	lista	Estado laboral en el momento de ser evaluado	224362002	Trabaja		224363007
								Trabajo completo		160903007
								Trabajo parcial		160904001
								Estudiante		413327003
								Retirado		105493001
								Retirado por problema médico		160898008
								Sin trabajo		73438004
								Vive con sus padres		224457004 / 700149001
A		Tabaquismo	fumador	PV, VS	radio/select	Tabaquismo en el momento de ser evaluado		Nunca fumo		266919005
								Fumador actual		77176002
								Fumó previamente		8517006
								Desconoce		
Categoría 2. Información sobre la historia de la enfermedad										
A		Fecha del diagnóstico	fecha_diagnostico	PV	aaaa-mm-dd	Fecha del diagnóstico	432213005			
A		Fecha de la primera recaída (inicio de la enfermedad)	fecha_inicio_enf	PV	aaaa-mm-dd	Fecha del comienzo de la enfermedad	298059007			
A		Curso de la EM	curso_EM	PV, VS	lista	Fenotipo de EM	24700007	Síndrome radiológico aislado		16415361000119105
								Síndrome clínico aislado		445967004
								EM recaída remisión		426373005
								EM secundaria progresiva		425500002
								EM primaria progresiva		428700003

Tabla III. Core data set adaptado y consensuado (cont.).

Categoría	Recolectado por ^a	Nombre de la variable	Nombre variable	Colección tiempo	Formato	SNOMED definición	SNOMED código	Valores de la variable	SNOMED término	SNOMED código
Categoría 3. Información sobre el estado de la enfermedad										
A		Fecha de evaluación	fecha_evolución	PV, VS	aaaa-mm-dd					
N		Estado de la EM (neurólogo)	ms_status_clin	PV, VS	lista	Estado de la enfermedad desde la perspectiva del neurólogo		No activa y sin progresión; activa (RM activa o recaídas) y con progresión; activa pero sin progresión; no activa pero con progresión		
A		Síntomas actuales	sintomas_actuales	PV, VS, IR	casillas/ lista			(ver hoja 'Síntomas')		
N		<i>Expanded disability status scale</i>	edss_score	PV, VS, IR	flotante	<i>Kurtzke multiple sclerosis rating scale</i> (escala de valoración)	273554001	0, 1.0, 1.5... , 10		
N		<i>25 Foot Walk Velocity Test</i>	25fwt	PV, VS	flotante			(tiempo en segundos)		
N		<i>9-Hole Peg Test</i>	9hpt	PV, VS, IR	flotante	Puntuación en el <i>9-Hole Peg Test</i> (entidad observable)	446602000	(tiempo en segundos, mano dominante)		263557007
N		<i>Symbol Digit Modalities Test</i>	sdmt	PV, VS	int	Puntuación en el <i>Symbol Digit Modalities Test</i> (entidad observable)	718387005	(número de respuestas correctas/sustituciones)		
Categoría 4. Información sobre recaídas										
A		Tuvo recaídas relapse		IR	booleano	Tuvo recaídas en el lapso de tiempo evaluado	230372003	Sí; no; no seguro		
A		Fecha de la recaída	fecha_recaída	IR	aaaa-mm-dd					
A		Usó corticoides en la recaída	corticoides_IR_recaidas		booleano	Tratamiento con corticosteroides y/o derivados de los corticosteroides (procedimiento)	788751009	Sí; no; desconozco		
Categoría 5. Investigaciones paraclínicas										
A		Resonancia magnética hecha	RMI	VS	booleano	Resonancia magnética (procedimiento)	113091000	Sí; no	s. a.	

Tabla III. Core data set adaptado y consensuado (cont.).

Categoría	Recolectado por ^a	Nombre de la variable	Nombre variable	Colección tiempo	Formato	SNOMED definición	SNOMED código	Valores de la variable	SNOMED término	SNOMED código
A		Región de la resonancia magnética realizada	RMI_region	VS	radio/select			Cerebro		816077007
								Medula cervical		241646009
								Medula dorsal		241647000
								Medula completa		24164500800%
A		Fecha de la resonancia magnética	fecha_mri	VS	aaaa-mm-dd					
N		Número de lesiones GAD+	mri_gd_les	VS	número entero			Número entero o desconocido		
N		Número de nuevas lesiones en FLAIR o T ₂	mri_new_les	VS	número entero			Número entero o desconocido		
N		Bandas oligoclonales presentes en el diagnóstico	BO_diagnostico	PV	booleano	Identificación de bandas oligoclonales en el líquido cefalorraquídeo y no en el suero en el momento del diagnóstico	113073005	Presentes; ausentes; no realizado; desconocido		
Categoría 6. Comorbilidades e infecciones oportunistas										
A		Comorbilidades	comorb	PV, VS	booleano	Comorbilidades	398192003	Sí; no; desconoce		
A		Tipo de comorbilidad	tipo_comorbilidad	PV, VS	lista			(comorbilidades)		49601007
								Cardiovascular		
								Respiratoria		50043002
								Gastrointestinal		119292006
								Psiquiátrica		74732009
								Metabólica		75934005
								Cáncer		86049000
								Musculoesquelética		928000
Autoinmune distinta a EM		85828009								
		comorb list y subcategoria_solapa comorbilities		ver solapa					74964007	

Tabla III. Core data set adaptado y consensuado (cont.).

Categoría	Recolectado por ^a	Nombre de la variable	Nombre variable	Colección tiempo	Formato	SNOMED definición	SNOMED código	Valores de la variable	SNOMED término	SNOMED código
A		Infección oportunista	infecc_opo	VS	booleano	Infección oportunista	61274003	Sí; no; desconoce		
A		Tipo de infección oportunista	tipo_infecc_opo	VS	lista			<i>LMP; HSV; VZV; CMV; pneumocystis; candidiasis; criptococosis; listeriosis; HHV8; HHV6, otra; desconocida</i>		
A		Fecha de infección oportunista	fecha_opo_VS_infecc		aaaa-mm-dd					
A		Resultado de la infección oportunista	opp_infec_out	VS	lista			En curso; recuperado con secuelas; recuperado sin secuelas; muerte		
Categoría 7. Tratamientos modificadores de la enfermedad										
A		¿En tratamiento con DMT?	dmt_status	PV, VS	select/radio			Sí; no; paciente naive de tratamiento; desconoce		844585000
A		Tipo de DMT	dmt_type	PV, VS	lista			Alemtuzumab	Alemtuzumab (fármaco)	129472003
								Cladribina	Cladribina (fármaco)	386916009
								Azatioprina	Azatioprina (fármaco)	372574004
								Otro	Campo libre para pacientes en ensayos clínicos o recibiendo otros tratamientos	
33	A	Fecha de inicio del tratamiento con DMT	dmt_start	PV, VS	aaaa-mm-dd		413946009			
34	A	Fecha del fin del DMT	dmt_stop	PV, VS	aaaa-mm-dd	Fecha de interrupción de tratamiento	413947000			
35	A	Motivo de fin de tratamiento con DMT	dmt_stop_reas	PV, VS	list	Razón (atributo)	410666004	Reacción alérgica		416093006
								Anti-JCV Ab positivo / riesgo de PML		
								Desconocido		

Tabla III. Core data set adaptado y consensuado (cont.).

Categoría	Recolectado por ^a	Nombre de la variable	Nombre variable	Colección tiempo	Formato	SNOMED definición	SNOMED código	Valores de la variable	SNOMED término	SNOMED código
Categoría 8. Tratamientos no farmacológicos										
36	P	Tratamiento	nonph_treat	PV, VS	list			Terapia ocupacional		84478008
								Fisioterapia		91251008
								Psicoterapia		75516001
								Hidroterapia		68130003
								Rehabilitación		52052004
								Fonoterapia		5154007
								Otra		
37	P	Fecha de inicio	nonph_treat_start	PV, VS	aaa-mm-dd		413946009			
38	P	Fecha de interrupción	nonph_treat_stop	PV, VS	aaaa-mm-dd	Día de finalización del tratamiento (entidad observable)	413947000			
Categoría 9. Acceso al sistema de salud y al tratamiento										
	A	Seguro de salud	seguro_salud	PV	lista	Tipo de seguro o cobertura de salud		Pública; privada; sin cobertura; desconoce		
	A	Provisión del DMT	provisión_tratamiento	VS	lista	Provisión del tratamiento durante el seguimiento por parte de su seguro de salud		Sí; no; desconocido		
	A	Cambio de marca de DMT durante el seguimiento	cambio_marca_DMT	VS	lista	Cambio de marca del tratamiento por parte de su seguro de salud		Sí; no; desconocido		

EM: esclerosis múltiple; IR: incidencia de recaída; PV: primera visita; VS: visita de seguimiento. ^a Posibilidad de ser recolectada por: A = ambos; N = neurólogo; P = paciente.

do, y permite una representación armonizada y coherente del contenido clínico [10]. Otra fortaleza, sin duda, es el proceso de consenso al que se ha llegado entre los profesionales involucrados, que buscará llevar a implementar el *core data set* en los registros en curso y en desarrollo entre quienes custodian los registros en la actualidad en la región. El *core data set* global, si bien ya tiene un tiempo de

desarrollo, su comunicación se ha iniciado recientemente de forma masiva y su implementación se ha buscado en los distintos registros en desarrollo, por lo que es probable que se cuente con experiencia de su implementación simultáneamente en las distintas regiones respecto a su factibilidad, dificultad, fortalezas y limitaciones que debe superar en el corto y mediano plazos.

En esta adaptación del *core data set*, respecto de la versión global (en prensa), la principal diferencia radica en la incorporación de variables de acceso y de la cobertura de salud de los pacientes, así como la facilitación del idioma para que las variables puedan incorporarse a los registros y sistemas de recolección de datos localmente. Estos dos puntos son de capital importancia, ya que la heterogeneidad en la traducción y en la forma de recolección que existía llevaba a diferencias significativas, que a la hora de analizarlos y compartirlos hacía difícil (casi imposible) su estandarización. Por otro lado, las variables de acceso a la salud, como si es un sistema público o privado el que cubre los estudios del paciente o los tratamientos, si recibe o no el tratamiento periódicamente y en la forma indicada en una región donde ese aspecto es verdaderamente un problema, constituyen un aspecto que proveerá información que permita optimizar el cuidado de la salud de los pacientes afectados de una forma diferencial [15-17].

Así mismo, el *core data set* busca insertarse en la comunidad de profesionales y en los sistemas de recolección que se encuentran trabajando en el desarrollo de registros y sistemas a través de su comunicación científica, la educación de profesionales involucrados y el entrenamiento sobre su utilización. Es a través de estas herramientas (comunicación, educación y entrenamiento) que se buscará profundizar la implementación del *core data set* en los distintos sistemas en desarrollo y en uso en la región [9]. De esta manera, la adaptación a América Latina del *core data set* global busca guiar a otras regiones, como Asia, Pacífico y otras, en la generación de un *core data set* adaptado a sus necesidades, para implementarlo en los sistemas locales, que permita a los registros homogeneizar la recolección de datos y así transitar un camino en común que optimice el cuidado de los pacientes afectados.

El proyecto tiene algunas limitaciones que cabe destacar. La búsqueda de variables de acceso puede haber omitido sistemas de recolección de datos que tengan variables relevantes que se deban incluir para ser evaluadas en el proceso de consenso. Sin embargo, la probabilidad de que se omita en la búsqueda un registro o una base de datos de magnitud en los que se vayan a evaluar sus variables considerando el grupo de trabajo involucrado es poco probable. Otra limitación tiene que ver con el proceso de consenso, que no incluyó a profesionales de las agencias regulatorias ni a pagadores de los sistemas de salud. De todos modos, se involucraron expertos en EM y en generación de datos de vida real que limitaron la posibilidad de sesgo. Finalmente, la cantidad de profesionales involucrados en el proceso

de consenso fue relativamente baja tratándose de un proceso regional; de todos modos, se intentó preservar la calidad del proceso metodológico y sus tiempos.

Conclusión

En conclusión, el *core data set* busca armonizar las variables recolectadas por los registros y las cohortes en EM con el fin de facilitar dicha recolección y permitir una colaboración entre fuentes apropiada. La implementación del *core data set* adaptado y consensuado permitirá alcanzar esos objetivos propuestos de armonizar y facilitar la recolección, y dará paso a una colaboración apropiada entre fuentes.

Bibliografía

1. Sherman RE, Anderson SA, Dal Pan GJ, Gray GW, Gross T, Hunter NL, et al. Real-world evidence - what is it and what can it tell us? *N Engl J Med* 2016; 375: 2293-7.
2. Schneeweiss S. Real-world evidence of treatment effects: the useful and the misleading. *Clin Pharmacol Ther* 2019; 106: 43-4.
3. McDonnell GV, Bennett A, Rae-Grant A, Bever C. Quality improvement in neurology: multiple sclerosis quality measures: executive summary. *Neurology* 2016; 86: 1361.
4. Signori A, Izquierdo G, Lugaresi A, Hupperts R, Grand'Maison F, Sola P, et al. Long-term disability trajectories in primary progressive MS patients: a latent class growth analysis. *Mult Scler* 2018; 24: 642-52.
5. Glaser A, Stahmann A, Meissner T, Flachenecker P, Horakova D, Zaratini P, et al. Multiple sclerosis registries in Europe - an updated mapping survey. *Mult Scler Relat Disord* 2019; 27: 171-8.
6. Report on Multiple Sclerosis Registries - Workshop 7 July 2017. URL://efaidnbmnnnibpcajpcglclefindmkaj/https://www.ema.europa.eu/en/documents/report/report-multiple-sclerosis-registries_en.pdf. Fecha última consulta: 01.06.2023.
7. Geys L, Parciak T, Pirmani A, McBurney R, Schmidt H, Malbasa T, et al. The multiple sclerosis data alliance catalogue: enabling web-based discovery of metadata from real-world multiple sclerosis data sources. *Int J MS Care* 2021; 23: 261-8.
8. Salter A, Stahmann A, Ellenberger D, Fneish F, Rodgers WJ, Middleton R, et al. Data harmonization for collaborative research among MS registries: a case study in employment. *Mult Scler* 2021; 27: 281-9.
9. Peeters LM, Parciak T, Kalra D, Moreau Y, Kasilingam E, van Galen P, et al. Multiple sclerosis data alliance - a global multi-stakeholder collaboration to scale-up real world data research. *Mult Scler Relat Disord* 2021; 47: 102634.
10. SNOMED - 5-Step Briefing. URL: https://www.snomed.org/snomed-ct/five-step-briefing. Fecha última consulta: 07.05.2023.
11. Carnero-Contentti E, Rojas JI, Cristiano E, Marques VD, Flores-Rivera J, Lana-Peixoto M, et al. Latin American consensus recommendations for management and treatment of neuromyelitis optica spectrum disorders in clinical practice. *Mult Scler Relat Disord* 2020; 45: 102428.
12. Cristiano E, Rojas JI, Alonso R, Alvez Pinheiro A, Bacile EA, Balbuena ME, et al. Consensus recommendations on the management of multiple sclerosis patients in Argentina. *J Neurol Sci* 2020; 409: 116609.

13. Rojas JI, Carnero-Contentti E, Abad P, Aguayo A, Alonso R, Bauer J, et al, on behalf on behalf research priorities LACTRIMS working group. Research priorities in multiple sclerosis in Latin America: a multi-stakeholder call to action to improve patients care: Research priorities in MS in LATAM. *Mult Scler Relat Disord* 2021; 53: 103038.
14. Carnero-Contentti E, Rojas JI, Cristiano E, Daccach-Marques V, Flores-Rivera J, Lana-Peixoto M, et al. Latin American consensus recommendations for management and treatment of neuromyelitis optica spectrum disorders in clinical practice. *Mult Scler Relat Disord* 2020; 45: 102428.
15. Carnero-Contentti E, Pettinicchi JP, Lopez PA, Alonso R, Garcea O, Balbuena ME, et al. Access and unmet needs to multiple sclerosis care in a cohort of Argentinean patients. *Mult Scler Relat Disord* 2019; 33: 88-93.
16. Carnero-Contentti E, Giachello S, Correale J. Barriers to access and utilization of multiple sclerosis care services in a large cohort of Latin American patients. *Mult Scler* 2021; 27: 117-29.
17. Rivera VM, Macias MA. Access and barriers to MS care in Latin America. *Mult Scler J Exp Transl Clin* 2017; 3: 2055217317700668.

Core data set for real world data in multiple sclerosis: customization for latin america from a global task force recommendation

Introduction. The primary objective of the core data set is to reduce heterogeneity and promote harmonization among data sources in EM, thereby reducing the time needed to execute real life data collection efforts. Recently, a group led by the Multiple Sclerosis Data Alliance has developed a core data set for collecting real-world data on multiple sclerosis (MS) globally. Our objective was to adapt this global data set to the needs of Latin America, so that it can be implemented by the registries already developed and in the process of development in the region.

Material and methods. A working group was formed regionally, the core data set created globally was adapted (translation process into Spanish, incorporation of regional variables and consensus on variables to be used). Consensus was obtained through the remote Delphi methodology of a round of questionnaires and remote discussion of the core data set variables.

Results. A total of 25 professionals from Latin America carried out the adaptation process between November 2022 and July 2023. Agreement was established on a core data set of nine categories and 45 variables, version 2023 to suggest its implementation in developed or developing registries, and MS cohorts in the region.

Conclusion. The core data set seeks to harmonize the variables collected by registries and cohorts in MS in Latin America in order to facilitate said collection and allow collaboration between sources. Its implementation will facilitate real life data collection and collaboration in the region.

Key words. Cohorts. Latin America. Epidemiology. Multiple sclerosis. Real world data. Registries.