EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

Use of real-world data in non-interventional studies to generate real-world evidence

Draft Reflection Paper

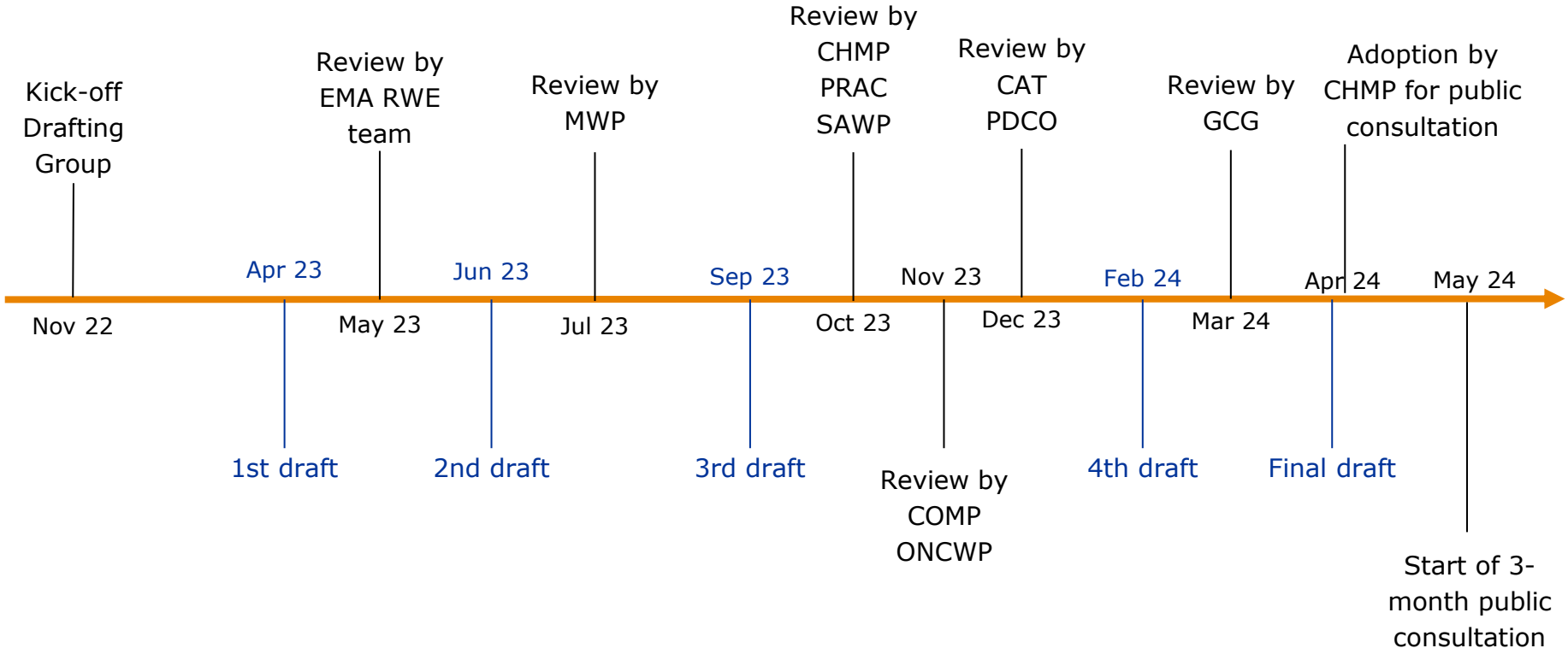
Joint HMA/EMA Big Data Steering Group workshop on RWE methods
14 June 2024

Presented by: Xavier Kurz, on behalf of the MWP

Drafting group: Olaf Klungel (MWP), Carla Torre (MWP, CHMP),
Stine Hasling Mogensen (MWP), Xavier Kurz (ESEC) and Juan Jose Abellan (EMA)

An agency of the European Union

Development history



Classified as internal/staff & contractors by the European Medicines Agency

Introduction

Use of real-world data in non-interventional studies to generate real-world evidence

Definitions:

- **RWD**: data that describe patient characteristics (including treatment utilisation and outcomes) in routine clinical practice
- **RWE**: evidence derived from the analysis of RWD
- **NIS** (definition as per EU legislation): a clinical study that does not fulfil any of the conditions defining a clinical trial (CT) in Article 2.2(2) of Regulation (EU) No 536/2014 on clinical trials on medicinal products for human use

  (a) the assignment of the subject to a particular therapeutic strategy is decided in advance and does not fall within normal clinical practice of the Member State concerned;

  (b) the decision to prescribe the investigational medicinal products is taken together with the decision to include the subject in the clinical study; and

  (c) diagnostic or monitoring procedures in addition to normal clinical practice are applied to the subjects

Classified as internal/staff & contractors by the European Medicines Agency

Introduction (2)

- Clinical trials (CTs) are the main source of evidence to evaluate the benefit-risk profile of medicines in marketing authorisation procedures
- Non-interventional studies (NIS) often used to support safety assessment but less often to support efficacy due to methodological limitations.
- Electronic healthcare data and registries provide new opportunities to use RWD and generate RWE derived from the clinical practice and gain knowledge.
- Examples of use of RWD in NIS to generate RWE supporting regulatory assessment:
  - To understand the clinical context (e.g. disease epidemiology, standards of care, clinical practice)
  - To support the feasibility assessment and the planning of non-interventional post-authorisation safety, efficacy and drug utilisation studies (e.g. impact on inclusion/exclusion on sample size)
  - To investigate associations and impact, for example to investigate the association between treatment exposure and either efficacy or safety outcomes, and monitor the implementation and the effectiveness of risk minimisation measures.

Classified as internal/staff & contractors by the European Medicines Agency

Objectives of the Reflection Paper

To discuss methodological aspects of NIS using RWD to generate RWE for regulatory purposes

- to highlight potential limitations
- to reflect on how some of these limitations could be overcome or mitigated to increase the reliability of the evidence

Classified as internal/staff & contractors by the European Medicines Agency

Content

- Scope
- Legal and regulatory requirements for use of RWD and RWE
- Study design
  - General considerations
  - Feasibility assessment
  - Studies with descriptive objectives
  - Studies with causal objectives
  - Bias and confounding
  - Effect modification
- Governance and Transparency
- Data quality
  - Reliability
  - Relevance
  - Multi-database studies
  - Data linkage
  - Data quality frameworks
- Statistical analyses
  - Hypothesis testing, estimation and precision
  - Time-dependent analyses
  - Stratified analyses
  - Sensitivity analyses
  - Missing data
  - Heterogeneity

Classified as internal/staff & contractors by the European Medicines Agency

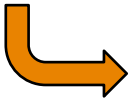
Scope

- Main scope: design, conduct and analysis of NIS using RWD to generate RWE for regulatory purpose. Use of RWD in the context of CTs is out of scope, e.g. external control groups, data source for patient recruitments, etc.
- General methodological guidelines already exist for NIS. The reflection papers focuses on principles considered critical for the conduct and assessment of NIS using RWD for regulatory purposes
- RWD can be collected through primary data collection and secondary use of data collected for another purpose. Attention to be given to the influence of primary purpose of data collection.
- Difference made between studies with descriptive objectives and studies with causal objectives given different implications for the study design.

Classified as internal/staff & contractors by the European Medicines Agency

Legal and regulatory requirements for use of RWD and RWE

- Art. 2 of Regulation (EU) No 536/2014: defines a *clinical study* and classifies clinical studies into either CT or NIS - RWD may be used in both CT and NIS to generate RWE
- There is no overall regulatory requirement to use RWD
- The regulatory assessment requires that the evidence provided is valid and reliable to support a specific regulatory objective; **on a case-by-case basis, use of RWD may be desirable or may be required to fill knowledge gaps for the specific regulatory objective**



the relevance of using RWE and the suitability of RWD for a specific application is to be discussed with regulators at an early stage of product or protocol development

- NI PASS to follow applicable legislation and GVP Module VIII; NI PAES to follow scientific guidance on PAES
- Guidance on registry-based studies applicable to RWD and RWE

Classified as internal/staff & contractors by the European Medicines Agency

Study design

1. *General considerations*

- Choice of study design to be primarily driven by the need to obtain reliable evidence for the research question
- Applicable methodological standards to be applied (ref. ENCePP Guide and Checklist for Study Protocol)

2. *Feasibility assessment*

- **Recommended as a basis of early discussions with regulators.** It may include, as applicable:
  - Evaluation of fitness-for-purpose of the candidate data source(s) (e.g. data quality, availability of data elements, sample size rationale)
  - Feasibility to implement the proposed study design, incl. choice of study populations, exposure(s), outcome(s), statistical parameters
  - Reference to feasibility of target trial emulation approach if applicable
  - Conclusion with a discussion of the relevance of the RWD sources and the study design to generate the required evidence, with proposal of different options if applicable.

Classified as internal/staff & contractors by the European Medicines Agency

Study design (2)

3. *Studies with **descriptive objectives***

- May have different goals. Their design and analysis plan may therefore vary and should follow the best practice for the research question.
- Results may be influenced by the setting in which they are observed. If this is the case, and this is relevant for the research question, the conditions that may influence the results of the study should be addressed, incl.:
  - setting (primary care, hospital setting, etc.) and type of the available data (prescription, claims, dispensing, …)
  - healthcare system of the country(-ies) where the data are collected, which may influence the availability and accessibility of the data
  - environment in which the data are collected, e.g. diagnostic criteria, prescribing practices, prescribing formularies, coding practices or reimbursement policies
  - specificities of the data sources (e.g. terminologies, CDM, data quality).

Classified as internal/staff & contractors by the European Medicines Agency

Study design (3)

4. *Studies with **causal objectives***

- Challenge is to deal with the risk of selection bias, information bias and confounding
- Target trial emulation framework should be considered as a strategy to formalise the design and analysis:
  - developed for NIS but with similarities with CTs, facilitating the evaluation of the contribution of the study to complement the evidence from CTs
  - helps the investigators to consider potential bias and adequate methods to address them
  - high level of transparency on the study design, the assumptions needed to emulate the trial and the definition of causal effects, facilitating evaluation and replicability
  - mitigation of bias, such as the prevalent user bias and the immortal time bias
  - estimand framework should be considered for the design of the hypothetical target trial.

Performs explicitely what regulators often do implicitely, i.e. a comparison with the ideal situation where a RCT would have been feasible.

Classified as internal/staff & contractors by the European Medicines Agency

TABLE 2. Specification and Emulation of a Target Trial During Pregnancy Using a Healthcare Administrative Database. As an example, the table describes 4 trials of COVID-19 booster vaccination with different outcomes: periconceptional trial, early pregnancy trial, late pregnancy trial and any-trimester trial.

Protocol Component	Target Trial	Emulation
Eligibility Criteria	• Enrollment period: January to December 2021 • Pregnant: Gestation under 12 weeks for periconceptional trial, under 20 for early trial, over 20 for late trial, and unrestricted for any-trimester trial. • Aged 18–50 years • No active SARS-CoV-2 infection (past infection allowed) • Primary vaccination schedule completed at least 6 months ago • Enrolled in the healthcare system of interest (e.g., insurance with prescription benefits or electronic health records) at least 12 months	Same. Eligibility criteria are identified via codes in the database
Treatment Strategies	1) Vaccine booster at enrolment 2) No booster during pregnancy	Same. Vaccination, including brand and date, is identified based on pharmacy dispensations and procedure codes
Assignment Procedures	Individuals are randomly assigned to one of the two vaccination strategies and are aware of the strategy to which they have been assigned.	Individuals assigned to each vaccination strategy are assumed to be comparable conditional on baseline covariates: gestational week, calendar month, age, month, region, chronic conditions, health care utilization, prior COVID-19, *etc.*
Follow-up Period	• Starts at vaccine assignment • Ends at the occurrence of an outcome of interest, 140 days after LMP (for spontaneous abortion) or 90 days after birth (for other outcomes), pregnancy loss, death, or loss to follow-up (disenrollment from insurance), whichever occurs earliest	• Starts at vaccine administration • Same except for loss to follow-up. Because pregnancy status is often ascertained by the end-of-pregnancy outcome, which forces a "complete case" approach
Outcome	Safety: • Periconceptional trial: a major congenital malformation • Early pregnancy trial: Spontaneous abortion • Late pregnancy trial: Stillbirth, preterm birth, microcephaly, gestational diabetes, preeclampsia, preterm delivery, labor induction, Cesarean section, postpartum hemorrhage, maternal death, small for gestational age, need for NICU admission, and neonatal death. Effectiveness: • Any-trimester trial: Maternal or infant COVID-19 with onset postvaccination	Same. Diagnoses are identified with algorithms based on combinations of codes in the database (this may be the same approach used in the target trial, which is pragmatic by definition)
Causal Contrasts of Interest	Intention-to-treat effect Per-protocol effect	Observational analog of per-protocol effect

Study design (4)

5. *Bias and confounding*

- *Selection bias*
  - Adequate definition and justification of inclusion/exclusion criteria (incl. codes and algorithms), taking into account possible missclassification – *consider wide inclusion criteria with restriction at analysis stage*
  -  Implicit selection criteria resulting from the method used to identify the study population and define exposure categories, e.g. health seeking behaviors leading to comparison of different populations
  - Comparison between the cohort's key characteristics and those of the target population using e.g. published research or national statistics if concern about generalisability of the association
  - Depending on research question, selection bias may be introduced with prevalent drug users. New user design prefered.
  - If external or historical cohort used for comparison: decision to be justified and the likelihood of bias and confounding to be discussed with special attention to potential differences in selection criteria.

Classified as internal/staff & contractors by the European Medicines Agency

Study design (5)

5. Bias and confounding

- *Information bias*
  - Pathway of data collection and/or extraction should be identified (diagnosis, coding, recording, data transformation, summarizing and analysis) and the different steps should be verified in order to evaluate if the data source(s) contains enough details on exposures and outcomes to correctly classify the patients.
  - Any previous validation study to be identified and evaluated; new validation study may be proposed
  - Avoid assumptions about direction of misclassification
    - impact depends on study objectives
    - more important to identify, minimise and measure if possible potential misclassification of exposures, outcomes and relevant covariates at the design stage
  - Define dates and account for timing of study entry, start of treatment and outcome occurrence as misclassification may lead to time-related bias.

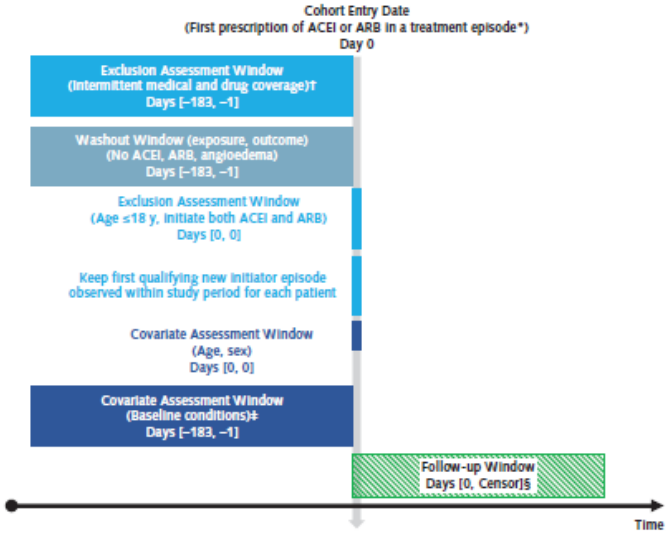
Classified as internal/staff & contractors by the European Medicines Agency

Study design (6)

5. *Bias and confounding*

- *Time-related bias*
  - Relevant time periods to be identified at the design stage in order to plan data collection or extraction of important dates and changes of patient status over time – graphical representations recommended

Figure 3. Exposure-based cohort entry where the cohort entry date is selected after application of exclusion criteria.



Graphical representations in reference to cohort entry date for different study designs presented in:

Schneeweiss et al., *Ann Intern Med. 2019;19;170(6):398-406*

Study design (7)

5. Bias and confounding

- *Confounding*
  - Potential confounders to be identified at design stage to plan data collection and analysis
  - Relevant epidemiological methods to be applied
  - Use of active comparator(s) is recommended in studies with causal objectives; more appropriate with new user design
  - Use of negative/positive control exposure/outcome may reveal residual confounding and may help in the interpretation and appraisal of results by revealing residual confounding
    - choice depends on research question and available data

Classified as internal/staff & contractors by the European Medicines Agency

Study design (8)

6. Effect modification

Need to address effect modification at the design phase in order to:

- collect/extract relevant data
- plan the data analysis

Classified as internal/staff & contractors by the European Medicines Agency

Governance and transparency

Governance

- Governance of the RWD sources used in a study to be made available in order to understand any restrictions related to the conditions of access, availability and publication of data.
- Principles of the ENCePP Code of Conduct to be applied (ADVANCE CoC for vaccines)

Transparency

MAAs/MAHs are encouraged to:

- register the NIS in the HMA-EMA Catalogue of studies together with the study protocol and study report (see guidance of GVP Module VIII on PASS and the Scientific guidance on PAES);
- register the data sources used in the NIS in the HMA-EMA Catalogue of real-world data sources; if data source is already registered, information to be updated if last update performed >12 months;
- make publicly available the codes used for the creation of the analytical data set and the programming code for the statistical analyses.

Classified as internal/staff & contractors by the European Medicines Agency

HMA-EMA Catalogues of RWD sources and studies

EUROPEAN MEDICINES AGENCY

Public **metadata repositories** that describe **RWD sources** and **studies that utilise such data to generate RWE.** They help pharmaceutical companies, researchers and regulators **identify** and **utilise RWD data** in investigating the use, safety and effectiveness of medicines.

Launched February 2024, https://catalogues.ema.europa.eu/

- One step closer to data-driven medicines regulation!
- Enhance **discoverability** of data sources and studies facilitating **collaboration and research**
- **Link RWD sources to studies** conducted which can support **study design**, **protocol evaluation**, and **results interpretation**
- **'FAIR'** data principles supported
- Promote **transparency** in observational research
- Ongoing activities on data **interoperability** and **integration** with other catalogues (e.g. EHDS, EHDEN)
- Advanced **user-friendly platform**

Help us foster the discoverability of RWD sources!

Add information on your data source directly to the Catalogues or send us an email at

me opa.eu

Benefits of including your data source in the Catalogue

- Increased identification and utilisation of your data source which can promote collaborations and studies
- Increased visibility to regulators, researchers and pharmaceutical companies
- Respond to the DARWIN EU Open Call to become a DARWIN data partner via the Catalogues

Classified as internal/staff & contractors by the European Medicines Agency

Data quality

Reliability

- Property of the data irrespective of the use in any study: do data represent the intended underlying medical concepts and are they trustworthy and credible ?
- Dimensions of reliability to be evaluated and documented by using a data quality framework
- Standard data quality management applied to the data source to be documented.

Relevance

- Study-specific: are key data elements (exposure, outcomes, covariates) available, is the size of the population adequate, is the population representative of the target population for the study objective, is the study design applied to the data source appropriate to answer the research question.
- Can be documented with the HARPER protocol template.

Classified as internal/staff & contractors by the European Medicines Agency

Data quality (2)

Multi-database studies

- Multiplicity of databases and associated increase in sample size should not reduce the quality requirements
- Information on the reliability and relevance of each of the databases should be presented.
- Heterogeneity between data sources to be addressed

Data linkage

- Protocol should describe:
  - the data elements used to link the data;
  - the linkage methodology, incl. how the performance of the matching is measured if this is a probabilistic matching, and an evaluation of the impact of imperfect matching.

Classified as internal/staff & contractors by the European Medicines Agency

Data quality (3)

Data Quality Frameworks

- Several data quality frameworks provide a set of characteristics determining the fitness-for-use of data
- Use of a data quality framework appropriate to the data source
- MAAs/MAHs to develop expertise for the implementation, analysis and interpretation of a DQF
- The HMA-EMA Data Quality Framework for EU medicines should be followed.

Classified as internal/staff & contractors by the European Medicines Agency

Statistical analyses

- Estimation and hypothesis testing
  - Analyses of NIS should focus on estimation and confidence intervals; statistically significant results may not be clinically relevant in large healthcare databases
  - Interpretation of study results to be based on estimate, clinical relevance and appraisal of study design, analysis and bias

- Time-dependent analyses
  - should be planned when appropriate to the research question in the cohort studies where events occur at different time points to account for time-dependent variables

Classified as internal/staff & contractors by the European Medicines Agency

Statistical analyses

- Stratified analyses
  - may fulfil different objectives and should be pre-specified
  - may be planned as sensitivity analyses
- Sensitivity analyses
  - to be discussed with regulators at an early stage and pre-specified in SAP
  - ICH E9(R1) Addendum on estimands to be considered
- Missing data:
  - assumptions behind any missing data and the appropriateness of the method chosen to handle it in the analysis to be justified
  - any imputation model(s) to be specified
  - ICH E9(R1) Addendum to be followed

Classified as internal/staff & contractors by the European Medicines Agency

Statistical analyses (2)

- Heterogeneity
  - to be anticipated in the study protocol or SAP and the management to be presented
  - relevance of evidence synthesis and proposed methods to be discussed

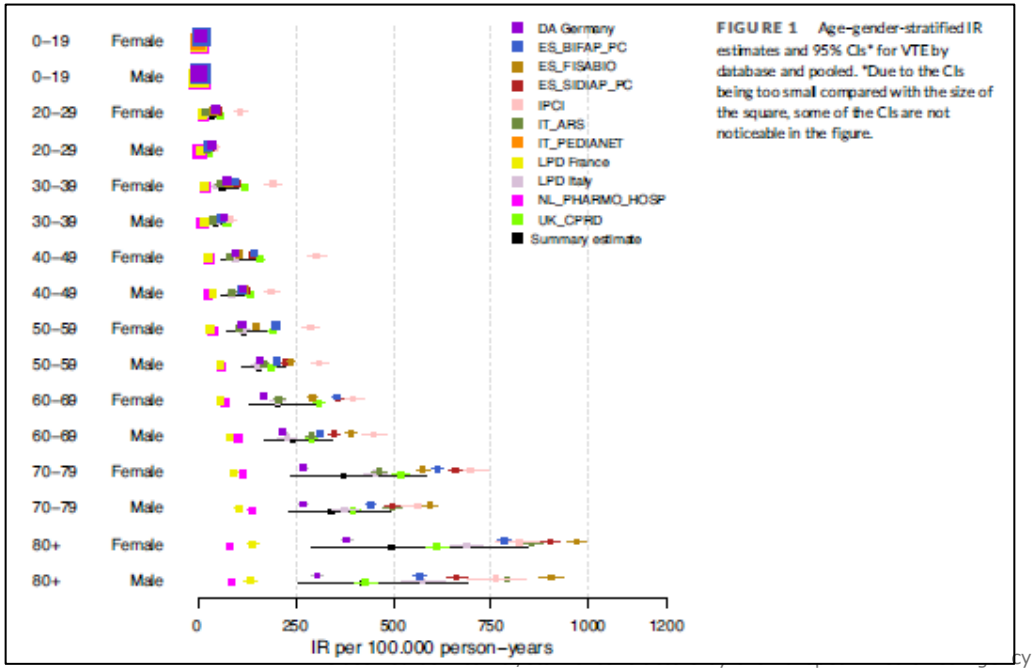


FIGURE 1 Age–gender-stratified IR estimates and 95% CIs* for VTE by database and pooled. *Due to the CIs being too small compared with the size of the square, some of the CIs are not noticeable in the figure.

Russek et al. Pharmacoepidemiol Drug Saf. 2023 ; 32(9):1032-48

Thank you !

Official address Domenico Scarlattilaan 6 ● 1083 HS Amsterdam ● The Netherlands
Telephone +31 (0)88 781 6000
Send us a question Go to www.ema.europa.eu/contact

Follow us on **@EMA_News**

Classified as internal/staff & contractors by the European Medicines Agency