Check for updates

Format of emergency department electronic health records in Europe. The European initiative and the eCREAM proposal

Guido Bertolini[1], Giulia Irene Ghilardi[1], Chiara Pandolfini[1]*,
Alessandro Bacchiega[2], Felice Catania[3], Bernardo Magnini[4],
Peter Mitro[5], George Notas[6], Gregor Prosen[7], Pankaj Sharma[8],
Jacek Górka[9] and Giovanni Nattino[1]

[1]Department of Medical Epidemiology, Istituto di Ricerche Farmacologiche Mario Negri IRCCS, Milan,
Italy, [2]Digital Health Innovation Lab, Centre for Digital Health and Wellbeing, Fondazione Bruno
Kessler, Trento, Italy, [3]Astir s.r.l., Milan, Italy, [4]Center for Augmented Intelligence, Fondazione Bruno
Kessler, Trento, Italy, [5]1st Cardiology Clinic, P.J. Safarik University, Košice, Slovakia, [6]Department of
Emergency Medicine, University of Crete School of Medicine and 7th Health Region of Crete,
Heraklion, Greece, [7]Center for Emergency Medicine, Univerzitetni Klinični Center, Maribor, Slovenia,
[8]Department of Research and Development, Ashford and St Peter's Hospitals NHS Foundation Trust,
Chertsey, United Kingdom, [9]Center for Intensive Care and Perioperative Medicine, Jagiellonian
University Medical College, Krakow, Poland

The lack of cross-border patient health data exchange in Europe is an obstacle in many ways and can negatively affect patient care and health. When clinicians have incomplete information about patients traveling or residing abroad, for example, continuity of care cannot be assured, potentially leading to poorer health outcomes. The European Electronic Health Record Exchange Format (EEHRxF) is a system being established in Europe to permit the interoperability of different healthcare systems, such as electronic health records (EHRs) and medical devices, so that they can share data to support patient care and research. The system is currently being introduced for electronic prescriptions and dispensations, patient summaries, which are part of the larger collection of health data known as the electronic health record, laboratory results and medical imaging studies and their reports, and hospital discharge reports. In emergency medicine, where research is challenging due to time and resource constraints, the EHR should no longer be seen solely as a tool to support clinical practice; it is also a source of valuable information to fuel research and improve patient care. The use of data for research, one of the stated secondary goals of the EEHRxF, thus becomes paramount here and deserves to be properly developed. It is in this context that the eCREAM (enabling Clinical Research in Emergency and Acute care Medicine through automated data extraction) project, a 5-year Horizon Europe project, was established. eCREAM will develop a system to exploit EHRs to enable research and improve decision-making, resource allocation and patient outcomes. It will address this target in two ways. First, by creating a new EHR that simultaneously meets clinical and research needs, collecting reliable, structured data that facilitate the clinical process and are readily usable for research purposes. Second, by developing an advanced natural language processing tool tailored to the specific needs of emergency medicine to automatically extract accurate, structured data from the free texts contained in EHRs. The project's innovative approach addresses current challenges in data extraction and utilization and sets a new standard for emergency medicine

in Europe in the digital age. This article provides a general overview of the eCREAM project.

Introduction

Background

The lack of cross-border patient health data exchange in Europe is an obstacle in many ways which, importantly, can negatively affect patient care and health. When doctors have incomplete information about patients traveling or residing abroad, for example, they cannot guarantee continuity of care, potentially leading to poorer health outcomes. The European Commission is working to promote health data exchange and access across the EU with its European Health Data Space (EHDS) regulation (1). Ensuring this exchange is the European Electronic Health Record Exchange Format (EEHRxF) (https://digitalhealtheurope.eu/glossary/european-electronic-health-record-exchange-format/). The EEHRxF is a system designed and set up to permit the sharing of health data both within and across country borders. Achieving this standard format entails, for example, ensuring technical interoperability, guaranteeing that different healthcare systems can effectively communicate and share data, and integrating the format with existing IT infrastructures without disrupting workflows. The interoperability of different healthcare systems, such as electronic health records (EHRs) and medical devices, will permit these systems to share data to support patient care, including cross-border electronic prescription and dispensation and access to patient summaries, laboratory results, medical imaging, and hospital discharge reports. The EHDS regulation also aims to create a strong legal framework for the re-use of health data for research, innovation, and public health purposes. In this context, the EEHRxF will enhance cross-country healthcare research by enabling seamless and secure access to standardized EHRs across member states, facilitating data sharing and collaboration among researchers.

Emergency medicine stands to benefit significantly from the EEHRxF, as this initiative addresses two of the most important challenges of the specialty. First, in emergency situations, clinicians often encounter patients with limited or no prior medical history available on-site. Immediate access to patient data would speed medical decisions, reduce errors, and avoid unnecessary repetitions of tests. Second, conducting clinical research in emergency medicine presents unique challenges and opportunities. The high volume of patients and chronic staff shortages (2–4) make traditional data collection methods impractical. However, the availability of robust data is crucial for enabling evidence-based practices tailored to the emergency department (ED) setting (5, 6). In this context, the EHR should no longer be seen solely as a tool to support clinical practice; it is also a source of valuable information to fuel research and improve patient care.

BOX 1 The eCREAM project's participants.

eCREAM involves 11 partners representing eight countries: France, Greece, Italy, Poland, Slovakia, Slovenia, Switzerland, and the UK. The consortium is coordinated by the Istituto di Ricerche Farmacologiche Mario Negri IRCCS in Milan, Italy, and includes five additional scientific/technical partners (Astir srl, Milan, Italy; Centre Hospitalier Universitaire Vaudois, Lausanne, Switzerland; European Clinical Research Infrastructure Network, Paris, France; Fondazione Bruno Kessler, Trento, Italy; Orobix life srl, Bergamo, Italy) and five clinical partners (Ashford and St Peter's Hospitals NHS Foundation Trust, London, UK; 7th Health Region of Crete, Heraklion, Greece; Univerzita Pavla Jozefa Safarika V Kosiciach, Kosice, Slovakia; Univerzitetni Klinicni Center Maribor, Maribor, Slovenia; Uniwersytet Jagiellonski, Krakow, Poland).

It is in this context that the eCREAM (enabling Clinical Research in Emergency and Acute care Medicine through automated data extraction) project started. eCREAM is a 5-year Horizon Europe program involving eight countries (Box 1) and launched in 2002. The project seeks to develop a system to exploit EHRs to enable research and improve decision-making, resource allocation, and, most importantly, patient outcomes. The project will address this target in two ways. First, a new EHR that simultaneously meets clinical and research needs is being created. This EHR is designed to collect reliable, structured data that facilitate the clinical process and are readily usable for research purposes. Second, eCREAM is developing an advanced natural language processing (NLP) tool, tailored to the specific needs of emergency medicine, that will automatically extract accurate, structured data from the free texts contained in existing EHRs. Important patient information is often embedded in unstructured free-text notes and thus poses significant extraction challenges. Together, these two strategies will permit the data to be automatically extracted from heterogeneous patient files, mainly EHRs, used in clinical practice, and to meet the different conditions required for their use in clinical research (7). This article provides a global perspective on the eCREAM project and the studies related to it, without going into detail on the methodological aspects.

Methods and analysis

The eCREAM project

The new EHR, tailored to EDs and designed by eCREAM, will create the conditions for enabling research in emergency medicine sustainable over the long term. This new EHR will be implemented in the centers that wish to adopt it and will be designed to prioritize the collection of robust,

trustworthy, research-oriented data without increasing the time staff dedicate to inputting patient data in the emergency setting. Clinicians and nurses based in the different participating countries collaborate on the EHR's design, providing information and opinions based on the different needs and practical issues in each country. The system's design will be tightly structured to meet clinical research standards while being user-friendly for everyday ED operations. It will provide real-time data quality validations, ensuring the accuracy and reliability of the information collected.

In parallel with developing a new EHR, the eCREAM project foresees at least three studies related to creating a data extraction system to exploit EHRs (see Figure 1). The first study, NLP-DeVal (8), establishes the foundational technology for the project, developing and validating a robust language model capable of extracting relevant clinical information from free text in EHRs in multiple languages. The other two studies are use cases that represent real-life situations relevant to the ED and will serve to test the entire data extraction system. The first, Use Case 1 (9), will test the validity of the data extracted from the EHRs to address the issue of ED hospitalization rate, a key factor contributing to overcrowding. The second, Use Case 2 (10), will develop a real-time dashboard to be integrated into the EHR that will allow staff to monitor healthcare indicators dynamically and interactively, enabling real-time monitoring and decision-making.

NLP-DeVal study

Natural language processing (NLP) offers promising solutions. The underlying large language models (LLMs), capable of interpreting and extracting information from free text have shown remarkable performance in various applications (11, 12). Still, their general performance in specialized fields such as the medical domain, especially in non-English languages, remains suboptimal due to their training on general data from the Internet (13).

The NLP study aims to bridge this research gap by developing a specialized language model, eCREAM-LM, tailored to the linguistic characteristics of medical texts in EDs. This model will automate data extraction, enabling large-scale clinical research without the need for labor-intensive data collection on the part of healthcare personnel. The language model will be trained to address the six eCREAM languages: English, Greek, Italian, Polish, Slovakian, and Slovenian. The NLP-DeVal study is essential as it sets the foundation for the project's subsequent phases.

Design

The NLP-DeVal study is an observational, multicenter, retrospective study including 35 centers across seven countries. It involves a comprehensive approach to developing and validating the eCREAM-LM using a combination of medical literature and clinical notes.

The study is carried out in different, sequential, iterative steps: training the model on large datasets of medical texts and non-annotated clinical notes, followed by fine-tuning and validation using annotated clinical notes.

Specifically, the development of eCREAM-LM involves training different open-source large language models on billions of medical texts from the scientific literature and millions of non-annotated clinical notes from the EHRs of participating centers. The fine-tuning of eCREAM-LM will be performed using 5,000 annotated notes, which will be used to specifically train the model to extract data to fill in the project's virtual case report form (vCRF). Annotation refers to the procedure by which trained emergency physicians will read the anonymized clinical notes received from the centers and will fill in the eCREAM vCRF, which contains the set of variables identified as necessary for carrying out the project's Use Case 1 study, i.e., estimating the probability of hospitalization from the ED. The annotated data will be used to train eCREAM-LM to learn to do the same job as the expert physician.

The clinical notes provided by the centers will fully protect patient personal data since they will first undergo an anonymization process (also carried out as part of the project) that involves removing personal identifiers and context-specific information from the text. This will ensure that individual patients cannot be re-identified, thereby permitting the study to adhere to the highest data protection standards.

Given that producing annotated datasets for fine-tuning in a new language is resource-intensive and time-consuming, and publicly available datasets are not available, an automated approach based on LLMs to extend the E3C (European Clinical Cases Corpus) (14), a freely available multilingual corpus of semantically annotated clinical narratives, to languages other than English and Italian, was also developed. This system will be used to train the eCREAM-LM in the medical domain in addition to the annotated ED data and is described in an article by Magnini et al. in this special issue.

Finally, the model's performance will be validated by comparing its ability to fill in the eCREAM vCRF against the annotations made by the trained physicians. Concordance between the model and the experts will be measured. The final validation phase includes analyzing 3,000 additional annotated notes to ensure the model's reliability and accuracy in real-world applications.

This phased approach ensures the model's robustness and accuracy in extracting relevant clinical data from EHRs.

Selection of subjects

The study population comprises adult patients who visited the participating hospitals between 1 January 2021 and 31 December 2023. Each participating center will provide clinical notes from these patients' EHRs.

Data analysis

The data analysis for the NLP-DeVal study involves rigorous validation of the eCREAM-LM against expert annotations. Cohen's κ will be used to assess their agreement on the extracted vCRF because a simple accuracy measure may be misleading, as some agreements could occur randomly. Cohen's κ takes random

FIGURE 1
The eCREAM project and its studies.
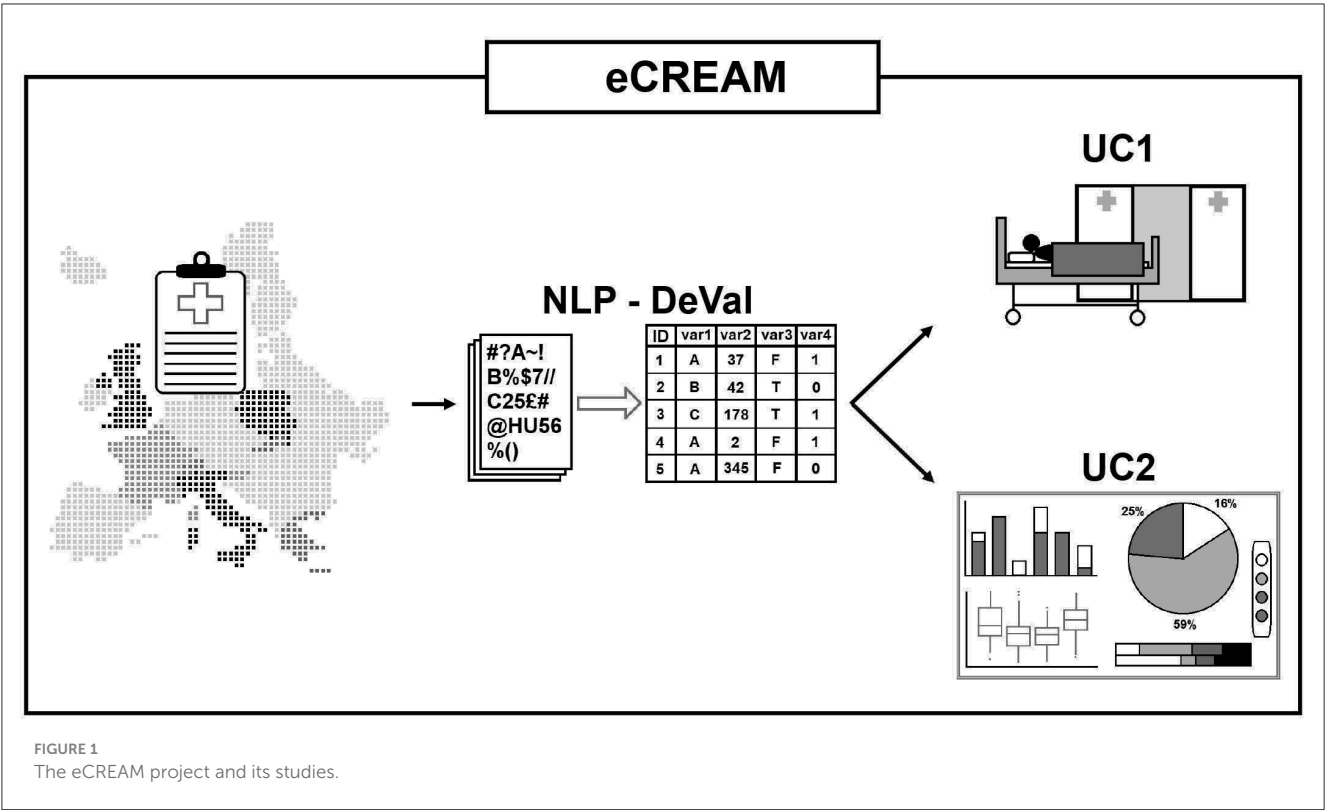
occurrences of agreement into account, correcting for the expected agreement by chance.

Use case 1 study: ED propensity to hospitalization

The Use Case 1 (UC1) study, like UC2, aims to test the soundness of the advanced natural language processing system developed to exploit EHR data for research. To test the validity of the data extracted from EHRs, UC1 will address the issue of hospitalization rate, which is tied to the negative ED overcrowding phenomenon, by developing statistical models to estimate the propensity of an ED to hospitalize patients. The hospitalization rate is the primary outcome and is a binary variable (patient admission to hospital vs. discharge). The predictor variables include clinical, demographic, physiological, and contextual factors (e.g., vital signs, comorbidities, presenting symptoms, triage category, and lab/imaging results). Because ED patients are highly heterogeneous and because UC1 is aimed at demonstrating the practical utility of the data extracted through the NLP in clinical decision-making rather than establishing conclusive results on the propensity to hospitalize from the ED, the analysis will focus on two subgroups of patients, i.e., those arriving at EDs with dyspnea or following transient loss of consciousness (TLoC). The models will be used to rigorously compare the participating centers in terms of hospitalization rate, taking into account the differences in the case mix of the different centers, enabling proper comparisons across centers and thus permitting improved decision-making and reduced unnecessary hospital admissions. A secondary objective of

UC1 is to study the association between the adjusted hospitalization rate and 30-day mortality.

Design

This observational, multicenter, retrospective study will last 24 months and involve 30 EDs across Italy, Poland, Slovakia, Slovenia, Switzerland, and the UK. Data on all patients who attended any of the participating EDs between January 2021 and December 2023 will be collected. There will be three types of data collection: data on the organizational structure of each center, transient organizational data (namely, the eCREAM restricted data collection, see following paragraph) for all patients attending the EDs, and detailed patient characteristics data (the eCREAM extended data collection) for those presenting with dyspnea or TLoC. The latter two are the standardized data collection protocols. The first type of data will be collected through questionnaires, while the latter two types will be automatically extracted from the EHRs using the IT platform created in the eCREAM project. The eCREAM platform will convert the data collected from the different countries' EDs and subject them to a semantic transformation process, including SNOMED CT, where needed, in coherence with the project's adopted codifications and ontology, in order to enable robust modeling and enhance the reproducibility of the methods. Details on the design and analyses of the UC1 study are provided in a separate article by Rubini et al. in this special issue.

Selection of subjects

The study population includes all patients who attended the participating EDs during the study period. The restricted data

collection, which will involve all ED patients, will consist of limited information and will be primarily aimed at estimating the level of ED crowding, which is expected to affect hospitalization decisions remarkably. Conversely, the extended data collection will focus on the two subgroups of interest for UC1, adult patients presenting with dyspnea and those presenting because they experienced TLoC, and will be identified from the triage section of the EHRs and final diagnoses, limiting the analysis to epilepsy or syncope-related TLoC cases.

Data analysis

The primary analysis involves developing two multivariable statistical models to estimate the probability of hospitalization for dyspnea or TLoC patients. These models will account for patient-specific, transitory-specific, hospital-specific, and societal-specific factors. The expected hospitalization rate for each ED will be calculated by summing the probabilities for all patients. For each ED, the expected rate will be compared to the one factually observed, calculating the ratio of the two quantities, resulting in the standardized hospitalization rate (SHR). The SHR will be used as an adjusted indicator to measure the individual ED's propensity to hospitalize patients. For the secondary objective, a multivariable logistic regression model will assess the association between adjusted admission rates and 30-day survival, adjusting for other prognostic factors. Each ED will receive a personalized report detailing the SHR and additional indicators to evaluate hospitalization decisions, enabling them to decide and implement measures to improve future decisions, improving patient care over time. At least 3,150 patients with dyspnea and 2,250 patients with TLoC will be recruited in each participating center. This sample size will be sufficient to develop a predictive model taking into account 20 predictors, compute the ED-specific indicators on a meaningful group of patients, and evaluate the association between propensity to hospitalization and 30-day survival (see study protocol for details).

Use case 2: quality of care dashboards

The Use Case 2 (UC2) study aims to test the soundness of the NLP system designed to automatically extract EHR data. UC2 will design real-time dashboards that will be used to enhance EHRs. The eCREAM platform, which will serve to calculate the indicators to display on the dashboards, will receive anonymous data from the EDs and from other sources in a one-directional data flow, and will produce aggregate data. Interoperability with other systems will be made possible by mapping ED data to FHIR standards. The dashboards will provide a comprehensive overview of patient flow, crowding levels, waiting times, and other key operational metrics. Possible individual indicators could be patient flow and congestion indicators, time to first provider (nurse/physician), boarding times for admitted patients waiting for inpatient beds, and triage category distribution. The dashboards, which will have customizable filters for selecting, for example, specific patient groups, will permit healthcare staff, administrators, and other stakeholders to monitor the situation in their ED, facilitating informed decision-making and

increasing ED efficiency. UC2's secondary objective is to evaluate the dashboards' usefulness, identify their strengths and areas for improvement, and evaluate their effectiveness in supporting decision-making in EDs.

Design

The study is designed as an observational, multicenter, retrospective study that focuses on developing dashboards rather than their implementation in real-time. Notably, the developed dashboards will be evaluated on a database retrospectively extracted from the participating hospitals' information systems because establishing and maintaining a continuous data stream from the hospitals is beyond the scope of the research study. The data to feed the dashboards will include administrative and clinical information such as arrival time, priority code, and hospital admission details and will involve patients who visited the participating EDs between January 1 and June 30, 2024. Information on the EDs' organizational setup will also be collected. Expert panels representing the different user groups will guide the dashboards' development through focus groups and literature reviews. Post-development, the dashboards' usefulness will be assessed through surveys and interviews with representatives of the user groups, i.e., clinicians and nurses, as well as administrators and policymakers. Details on the design and analyses of the UC2 study are provided in a separate article by Porta et al. in this special issue.

Selection of subjects

The study population includes all patients who visited the participating EDs during the data collection period. Participating centers consist of 30 EDs from Italy, Poland, Slovakia, Slovenia, the UK, and Switzerland, excluding EDs in Greece due to the lack of EHR systems.

Data analysis

Data analysis will assess the dashboards' usefulness based on the responses of the representatives of the three types of end-users to the surveys and interviews. Descriptive analyses will be used for the questionnaire, with frequency distributions, measures of central tendency, and measures of dispersion, and Cronbach's α will be used for testing survey reliability. The Chi-square test will be used to assess associations between categorical variables and the ANOVA or the Kruskal-Wallis tests will be use to compare continuous variables across multiple groups, as appropriate. A thematic analysis will be used for the interviews. A combination of deductive and inductive analysis will be conducted and use of the NVivo software to extract themes from textual data and manage coding is planned.

DATA FAIRification

One of eCREAM's aims is to FAIRify (make data Findable, Accessible, Interoperable, and Reusable) the established databases for clinicians, researchers, health policymakers, and citizens while respecting the European and national legislations. To do so, the

project will implement a strategy for making the services and databases created to address the two use cases FAIR (15) and increase their usability by a variety of users outside the consortium (e.g., patients' advocacy groups, clinicians, policymakers, and researchers). The goal is to maximize the exploitation of the collected data for secondary research and monitoring activities.

To this end, after curation, the UC1 study database will be uploaded to the Medical Informatics Platform (MIP) hosted by the Centre Hospitalier Universitaire Vaudois (CHUV) partner on the EBRAINS Research Infrastructure in Switzerland (see https://www.ebrains.eu/tools/medical-informatics-platform). The MIP is a GDPR-compliant, privacy-aware platform that allows secure, federated analysis of clinical data. Three MIP instances will be created: a public MIP for broader community access to dedicated, anonymized data, an eCREAM MIP for accredited users within the eCREAM consortium, and an extended MIP federation for partnering organizations outside the eCREAM consortium who would like to join the project's efforts. Users with different needs will, therefore, be able to access the eCREAM data for secondary research.

Discussion

The eCREAM project fits within the European EEHRxF initiative to increase EHR data access and usability. eCREAM, in particular, focuses on enabling research in emergency medicine, a medical field in which research is challenging despite its importance, thus sharing one of the secondary objectives of the European Commission's initiative, i.e., to promote the re-use of health data for research. eCREAM will design and develop a new EHR customized to support both the need for research and the needs of ED personnel. It will also enable the exploitation of data from EHRs currently in use by developing a tool to extract reliable, structured data automatically. The two use case studies will validate the system, definitively integrating artificial intelligence in the project and highlighting the transformative potential of these technologies in healthcare. The ability to process and analyze vast amounts of unstructured data from EHRs opens new avenues for research and clinical practice, significantly enhancing the quality and efficiency of clinical research and leading to evidence-based practices tailored to emergency settings. The project's design, spanning multiple languages and centers, also ensures the broad applicability and robustness of the developed tools, and these will create the conditions for making research in emergency medicine sustainable over the long term.

In addition to enabling research by exploiting EHR data, eCREAM will also impact the scientific, clinical, and public health communities by making the data it produces findable, accessible, interoperable, and reusable, based on the FAIR principles. Sharing the project's unique data will permit the use and reuse of health data to benefit patients, researchers, innovators, and regulators. The EHDS regulation is a vast initiative and although it will need to be refined (16), it will begin to pave the way for data sharing across Europe. We hope that, with this initiative as a foundation, eCREAM's achievements will break the barriers hindering research in emergency medicine.

Ethics and dissemination

Ethical considerations

The project adheres to strict ethical guidelines to ensure patient privacy and data protection. All three studies are subject to approval by the countries' ethics committees, and are currently under review by them. Clinical notes in the UC1 study will be pseudonymized before being transmitted by the participating centers, while data from the UC2 study will be transmitted in aggregate, and therefore anonymized, form. The clinical notes in the NLP-DeVal study will be stripped of any reference to the patients and contexts, and each individual patient's notes will be split into different aspects of the health record and grouped with those of other patients to ensure that the probability of patient identification is less than negligible. Additionally, an anonymization system designed and tested by eCREAM will be installed in each participating hospital to remove the possible presence of information about third parties from the notes, so that the data transmitted by the centers will be anonymized. All data made accessible to the public will be anonymized. The eCREAM project adheres to the General Data Protection Regulation (GDPR) standards. Informed consent would only be necessary for the UC1 study, but EU Regulation 2016/679 Article 9, par. 2, letter j provides that consent is not required if appropriate measures to protect the rights of the data subject are respected, and such measures will be implemented (see UC1 study protocol). As the EU Regulation allows Member States to introduce more specific provisions to ensure the lawful and fair processing of data, informed consent may or may not be required in each participating country, depending on national legislation. This will be decisive for ethics committee approval.

Dissemination

The findings from the eCREAM project will be disseminated through various channels, including academic publications, conference presentations, and stakeholder reports. A dedicated dissemination strategy will ensure that the insights and tools developed reach a wide audience, including healthcare providers, policymakers, and the broader research community.

Conclusion

The eCREAM project represents a significant step forward in employing AI to exploit EHRs and improve emergency medical care. By developing robust NLP tools and ensuring data interoperability and accessibility, in line with the European EEHRxF initiative, eCREAM aims to revolutionize clinical research in emergency medicine, providing insights to improve patient care, and operational efficiency in emergency settings.

One of the project's most valuable impacts will be to create the conditions to make such research in emergency medicine sustainable. For EDs who wish to use their current EHRs to capture structured data, eCREAM will provide its integrable data extraction system, while the project will also develop an ED-EHR expressly designed to meet even higher data quality standards for EDs who wish to adopt it. This ED-EHR will allow researchers

to address more in-depth research questions and to ensure the reliability of the results. In addition to these solutions, which are scalable and will be maintained after the end of the project, eCREAM will implement regional and national projects based on data extraction for quality-of-care monitoring. These efforts will lead to improvements in ED efficiency and patient care and will therefore be endorsed by policymakers, making eCREAM sustainable in the future.

The successful execution of this project will mark a significant advancement in the integration of technology and healthcare, setting a precedent for future research initiatives.

Author contributions

GB: Data curation, Supervision, Formal analysis, Writing – original draft, Project administration, Methodology, Validation, Investigation, Conceptualization, Funding acquisition, Writing – review & editing. GG: Project administration, Visualization, Data curation, Conceptualization, Supervision, Writing – original draft, Writing – review & editing, Methodology, Investigation. CP: Writing – review & editing, Project administration, Supervision, Writing – original draft, Methodology. AB: Software, Resources, Investigation, Methodology, Writing – review & editing, Validation. FC: Writing – review & editing, Resources, Software, Conceptualization, Methodology, Data curation, Validation, Investigation. BM: Validation, Formal analysis, Writing – review & editing, Methodology, Conceptualization, Resources, Investigation, Software. PM: Methodology, Conceptualization, Writing – review & editing. GNo: Supervision, Methodology, Conceptualization, Investigation, Writing – review & editing. GP: Conceptualization, Methodology, Writing – review & editing. PS: Conceptualization, Funding acquisition, Investigation, Methodology, Writing – review & editing. JG: Investigation, Methodology, Writing – review & editing. GNa: Writing – review & editing, Investigation, Conceptualization, Methodology, Validation, Formal analysis, Supervision, Writing – original draft.

Funding

Acknowledgments

Conflict of interest

FC was employed by Astir s.r.l.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that Gen AI was used in the creation of this manuscript. The author(s) acknowledge the use of ChatGPT4o for language editing.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. European Commission. *European Health Data Space.* (2024). Available online at: https://health.ec.europa.eu/ehealth-digital-health-and-care/european-health-data-space_en (accessed December 20, 2024).

2. Oskvarek JJ, Zocchi MS, Black BS, Celedon P, Leubitz A, Moghtaderi A, et al. Emergency department volume, severity, and crowding since the onset of the coronavirus disease 2019 pandemic. *Ann Emerg Med.* (2023) 82:650–60. doi: 10.1016/j.annemergmed.2023.07.024

3. Winters ME, Hu K, Martinez JP, Mallemat H, Brady WJ. The critical care literature 2023. *Am J Emerg Med.* (2024) 85:13–23. doi: 10.1016/j.ajem.2024.08.010

4. Vaughan L, Edwards N. Emergency hospital services are closing because of staff shortages. *BMJ.* (2023) 7:e078766. doi: 10.1136/bmj-2023-078766

5. Boaz A, Goodenough B, Hanney S, Soper B. If health organisations and staff engage in research, does healthcare improve? Strengthening the evidence base through systematic reviews *Health Res Policy Syst.* (2024) 22:113. doi: 10.1186/s12961-024-01187-7

6. Ozdemir BA, Karthikesalingam A, Sinha S, Poloniecki JD, Hinchliffe RJ, Thompson MM, et al. Research activity and the association with mortality. *PLoS ONE.* (2015) 10:e0118253. doi: 10.1371/journal.pone.0118253

7. Ercole A, Brinck V, George P, Hicks R, Huijben J, Jarrett M, et al. Guidelines for data acquisition, quality and curation for observational research designs (DAQCORD). *J Clin Transl Sci.* (2020) 4:354–9. doi: 10.1017/cts.2020.24

8. Bertolini G, Ghilardi GI, Pandolfini C, Nattino G, Lavelli A, Moretti F. Study protocol - NLP-DeVal: development and validation of a natural language processing tool to enable clinical research in emergency and acute care medicine: retrospective cohort study. *Zenodo.* (2024) doi: 10.5281/zenodo.10996542

9. Bertolini G, Banzi R, Catania F, Lavelli A, Ghilardi GI, Pandolfini C, et al. Study protocol - Propensity to hospitalize patients from the ED in European centers An observational retrospective quality-of-care study. *Zenodo.* (2024) doi: 10.5281/zenodo.10996665

10. Bertolini G, Ghilardi GI, Pandolfini C, Nattino G, Catania F, Banzi R. Study protocol - Development of a multipurpose dashboard to monitor the situation of emergency departments. *Zenodo.* (2024) doi: 10.5281/zenodo.10996699

11. Iroju OG, Olaleke JO. A systematic review of natural language processing in healthcare. *Int J Inf Technol Comput Sci.* 7:44. doi: 10.5815/ijitcs.2015.08.07

12. Zeng Z, Shi H, Wu Y, Hong Z. Survey of natural language processing techniques in bioinformatics. *Comput Math Methods Med.* (2015) 2015:674296. doi: 10.1155/2015/674296

13. AlShuweihi M, Salloum SA, Shaalan K. Biomedical corpora and natural language processing on clinical text in languages other than english: a systematic review. In: *Studies in Systems, Decision and Control* (New York, NY: Springer) (2021). p. 491–509.

14. Magnini B, Altuna B, Lavelli A, Speranza M, Zanoli R. *The E3C Project: Collection and Annotation of a Multilingual Corpus of Clinical Cases.* Torino: Accademia University Press, (2020). p. 258–64.

15. Wilkinson MD, Dumontier M, Aalbersberg IJJ, Appleton G, Axton M, Baak A, et al. The FAIR guiding principles for scientific data management and stewardship. *Sci Data.* (2016) 3:160018. doi: 10.1038/sdata.2016.18

16. Marelli L, Stevens M, Sharon T, Van Hoyweghen I, Boeckhout M, Colussi I, et al. The European health data space: too big to succeed? *Health Policy Amst Neth.* (2023) 135:104861. doi: 10.1016/j.healthpol.2023. 104861