Janssen Pharmaceutica NV*

Executive Summary for Retrospective Observational Studies Using Secondary Data

Invasive E. coli Disease (IED) Characterisation in France within the "Programme de médicalisation des systèmes d'information" (PMSI): The EXCEED Study

A Feasibility Study in a French Hospital Discharge Database and Proof of Concept

Delphine Quélard[1], Jeroen Geurtsen[2], Antoine C. El Khoury[3], Maureen P. Neary[3], Katherine Theiss-Nyland[4], Franck Bruyere[5], Djillali Annane[6], Jérôme Fernandes[7], Fanny Raguideau[8], Benjamin Grenier[8]

[1]Janssen Cilag, Issy-les-Moulineaux, France; [2]Bacterial Vaccines Discovery and Early Development, Janssen Vaccines & Prevention B.V., Leiden, Netherlands; [3]Janssen Global Services, LLC, Raritan, New Jersey, United States; [4]University of Oxford, Oxford, United Kingdom; [5]Tours University Hospital, Tours, France; [6]Raymond-Poincaré Hospital (AP-HP, University of Versailles SQY, Paris Saclay University), Garches, France; [7]Côte Basque Hospital, Bayonne, France; [8]HEVA, Lyon, France

*Janssen Pharmaceutica NV is a public limited liability company incorporated and existing under Belgian Law, having its registered office at 2340 Beerse, Turhoutseweg 30, Belgium.

Title/acronym	Invasive *E. coli* Disease (IED) Characterisation in France within the PMSI: The EXCEED Study
Context and rationale for the public interest	Invasive *Escherichia coli* disease (IED) can be defined as a bacterial infection with acute systemic consequences, including bacteraemia and sepsis. Classification is based on clinical criteria and microbiological confirmation by the isolation and identification of *E. coli* from blood or other normally sterile body sites as well as from urine in patients with urosepsis and no other identifiable source of infection.[1] The burden of disease and high prevalence of antimicrobial resistance (AMR) make IED a major medical challenge; while, there is a critical unmet need for effective prevention and treatment.[2] In fact, *E. coli* is responsible for the high rates of bacteraemia in Europe, the United Kingdom and the United States.[3,4] *E. coli* is also a leading cause of community-onset bacteraemia and a prominent pathogen in community-acquired sepsis in adults in the United States.[5-7] AMR poses a major challenge in the treatment of IED. IED management relies on the timely use of effective antibiotics; however, antibiotic-resistant *E. coli* infections are becoming increasingly prevalent, leading to increased hospitalisation, mortality rates and healthcare costs.[5,8] This study, using the French hospital discharge database, "Programme de médicalisation des systèmes d'information" (PMSI), aimed to provide epidemiological, clinical and economic evidence for the burden of IED in France, and assess the feasibility of extracting such data given that there are very limited data for patients with IED available today in France.
Objectives	**Primary objective:** ● To estimate the annual incidence of hospitalisations associated with IED in France from 2015 to 2021 using PMSI data. **Secondary objectives:** ● To describe the demographic and clinical characteristics of patients with IED and their annual healthcare consumption from 2015 to 2021. ● To describe associated annual healthcare costs for patients with IED from 2015 to 2021. ● To estimate in-hospital all-cause death rates of patients with IED in France from 2015 to 2021. **Exploratory objective:** To estimate IED incidence in France among patients with hospital stays for 'invasive disease without data for a coded pathogen' from 2015 to 2021.
Methods	**Study design and study periods**
	This was a retrospective observational study performed using the hospital discharge database (PMSI) in France. Individuals ≥ 18 years of age were included in this study at the time of case identification (either IED or 'invasive disease without a coded pathogen') from 1 January 2015 to 30 November 2021. The index date was the first day of the first hospital stay for an invasive disease during the study period. Any hospitalisation for invasive disease that occurred within 28 days of the index date was considered related to the index invasive disease. After that time, any new invasive disease was considered as a new event but was not accounted for within the incidence rates. Individuals were followed from the index date until either the end of the study period (31 December 2021) or in-hospital death, whichever occurred first.

Incident patients were defined as patients with no IED during the 2-year fixed pre-index date period.

Data sources

This study was performed using data from the French nationwide hospital discharge database (PMSI).

Study populations

Study population	Description
Population 1	Infectious invasive disease due to *E. coli* OR Infectious invasive disease, unspecified AND Extraintestinal infection due to *E. coli* AND Absence of non–*E. coli* infection
Population 1A	Alternative Population 1 with revised diagnosis codes for infectious invasive disease, unspecified
Population 1B	Alternative Population 1 with revised diagnosis codes for infectious invasive disease, unspecified OR Organ dysfunction requiring intensive care unit admission AND Extraintestinal infection due to *E. coli* AND Absence of non–*E. coli* infection
Population 2	Infectious invasive disease, unspecified AND Absence of extraintestinal infection due to *E. coli* AND Absence of non–*E. coli* infection

Each population was defined by a sensitive and a specific algorithm: the sensitive algorithm used a broader set of codes; the specific algorithm used a narrower set of codes. The sensitive and specific sets of codes varied for the 'infectious invasive disease, unspecified' part of each algorithm.

To address challenges in the identification of hospitalisations for patients with IED, 2 sets of International Classification of Diseases, 10th Revision (ICD-10) codes were used to provide a lower estimate and an upper estimate of hospitalisation rates:

- A sensitive phenotype was defined using a sensitive algorithm.
- A specific phenotype was defined using a specific algorithm.

ICD-10 codes incorporated in the sensitive phenotype but not in the specific phenotype included pneumonia, cholecystitis, pyelonephritis, and fever codes; this is to reflect that a proportion of pyelonephritis cases coincide with bacteraemia and most likely qualify as IED.

Study population size

	The PMSI is a population-based data source that includes the entirety of hospitalisations in the French population, and the primary study objective was descriptive. Therefore, no power or sample size calculations were required.
	Statistical analyses
	Descriptive statistics for patient demographics, clinical characteristics and healthcare consumption were computed. **Primary objective:** ● To estimate the number of incident individuals (new patients) and the crude incidence rates among the French population over the inclusion period, by calendar year, at the national level. **Secondary objectives:** ● To describe demographic and clinical characteristics ● To describe healthcare consumption for each patient in Population 1 for the index stay, by calendar year, at the national level. ● To describe healthcare consumption–associated costs for each patient in Population 1, which were valued in 2022 and described from the National Health Insurance perspective (total costs and distribution of costs by type of resource use, by calendar year, at the national level). ● To estimate in-hospital mortality by calculating the number of deaths (any cause of death) and the mortality rate over the inclusion period, by calendar year, at the national level, for Population 1. **Exploratory objectives:** ● To calculate the number of incident individuals (new patients) with 'invasive disease without any coded pathogen' and incidence rates among the French population over the inclusion period, by calendar year, at the national level. To compute an adjusted incidence of IED based on 'invasive disease without any coded pathogen,' applying the proportion of IED among invasive diseases found in the review from Bonten et al. 2021 and Abbara et al. 2022.[9,10]
Results	Between 1 January 2015 and 30 November 2021, study Population 1 included 331,313 patients using the sensitive version of the algorithm defining IED (mean [standard deviation, SD] follow-up duration of 2.9 [2.1] years), and among them, 99,982 patients were included using the specific version of the algorithm defining IED (mean [SD] follow-up duration of 2.5 [2.1] years). Study Population 2 included 2,255,765 patients using the sensitive version of the algorithm defining IED (mean [SD] follow-up duration of 3.0 [2.2] years), and among them, 897,082 patients were included using the specific version of the algorithm defining IED (mean [SD] follow-up duration of 2.7 [2.2] years). In hospitalised patients with IED (Population 1), annual IED rates ranged from 77.0 (95% confidence interval [CI], 76.2–77.7) to 96.7 (95% CI, 95.8–97.5) per 100,000 individuals with the sensitive algorithm, and ranged from 24.1 (95% CI, 23.7–24.5) to 29.2 (95% CI, 28.8–29.7) per 100,000 individuals with the specific algorithm. When using the sensitive algorithm for Population 1 and stratifying results by age group, IED rates ranged from 32.9 (95% CI, 31.8–34.1) per 100,000 individuals in the 18–29 years age group to 533.9 (95% CI, 524.2–543.7) per 100,000 individuals in the 85+ years age group. IED rates

	were 201.1 (95% CI, 199.0–203.3) in the 60+ years age group and 238.9 (95% CI, 236.3–241.6) in the 65+ years age group. When using the specific algorithm for Population 1 and stratifying results by age group, IED rates ranged from 4.1 (95% CI, 3.7–4.5) per 100,000 individuals in the 18–29 years age group to 147.9 (95% CI, 142.8–153.2) per 100,000 individuals in the 85+ years age group. Using this algorithm, incidence rates were 64.8 (95% CI, 63.6–66.0) in the 60+ years age group and 76.1 (95% CI, 74.6–77.6) in the 65+ years age group. In Population 1A, 64,626 patients were included in the sensitive version of the algorithm, and among them 32,834 patients were included in the specific version of the algorithm. In Population 1B, 70,096 patients were included in the sensitive version of the algorithm, and among them 38,672 patients were included in the specific version of the algorithm. For Populations 1A and 1B, annual IED rates ranged from 15.5 (95% CI, 15.1–15.8) to 20.1 (95% CI, 19.7–20.5) per 100,000 individuals with the sensitive algorithm, and ranged from 6.7 (95% CI, 6.5–6.9) to 11.6 (95% CI, 11.3–11.9) per 100,000 individuals with the specific algorithm. Assuming the proportion of sepsis cases attributable to *E. coli* is 27% (as published by Bonten et al. 2021), the mean incidence rate of IED among Population 2 (without a coded pathogen) was calculated as 66.5 (95% CI, 65.8–67.2) with the specific algorithm and 168.5 (95% CI, 167.3–169.6) per 100,000 individuals with the sensitive algorithm. Using the 19% proportion of bacteraemia cases attributable to *E. coli* (as published by Abbara et al. 2022), the mean incidence rate was estimated as 47.9 (95% CI, 47.3–48.5) with the specific algorithm and 121.1 (95% CI, 120.2–122.0) per 100,000 individuals with the sensitive algorithm.
Discussion	This study was performed using the French nationwide hospital discharge database, PMSI. This claims database relies on the transposition of medical and healthcare consumption coding into diagnosis-related groups corresponding to a reimbursement income for the hospital. The algorithm for case identification that was tested in this study was initially developed within a United States setting (using International Classification of Diseases, 10th Revision, Clinical Modification [ICD-10-CM] codes) and validated in the same setting. The essential difference between the ICD-10-CM and the ICD-10 is that the ICD-10-CM offers more precision compared with the French ICD-10 (5-character codes in the ICD-10-CM and 4-character codes in the French ICD-10). The fact that the French database requires the combination of 2 codes instead of 1 for the same information could have led to an underestimation of IED case identification. In addition, 2 algorithms, 1A and 1B, adapted from algorithm 1 (algorithm 4 in Fortin et al. 2022[11]), were assessed. Although the ICD-10 codes used for the transposition of the algorithms 1A and 1B were transposed from the ICD-10-CM to the French ICD-10 by a coding physician, and results were potentially subject to the same underestimation as described above, such differences in figures might also be explained by different coding practices between countries. Pandolfi et al. analysed sepsis incidence for all pathogens overall using this same database and reported a rate of 403 per 100,000 in 2019, which are the most recent figures in a French setting to date.[12] When considering the proportion of sepsis cases attributable to *E. coli* that are described by Abbara et al. (19%)[10] and Bonten et al. (27%)[9], the projected *E. coli* incidence figures using Pandolfi et al. are consistent with what was found in the current study.[13] However, the data reported here are lower than IED rates based on laboratory culture information, regardless of whether they came from the United States (Kaiser Permanente Northwest - KP) or the United Kingdom surveillance network (UK Health Security Agency's Data Capture System - DCS).[13] For instance, an IED rate of 154.9 cases per 100,000 individuals in 2016 was reported in the KP cohort and 143.6 cases per 100,000 individuals were identified from the DCS. Differences

	were even more notable in the 65+ years age group, with incidence rates of 458.4 reported in the KP cohort and 436.0 identified by the DCS.
	Strengths and Limitations
	The main strength of the study was that it used the PMSI, a population-based nationwide hospitalisation database that provides detailed real-world data for all hospitalisations occurring among individuals who are residents in France. Data were recorded prospectively and there was an average follow-up duration of 2.9 (range, 0–7.1) years. PMSI is a data source that provides an efficient way to obtain recent, comprehensive data on a nationwide basis. One of the limitations of the study is that the case identification algorithm relied only on French ICD-10 codes, without access to results from laboratory tests in the PMSI. Additionally, the causal bacteria are not always coded within the PMSI to specify the type of invasive infection. These factors may have led to an underestimation of the incidence of IED events. Opatowski et al. 2019[14] suggested reasons for the absence of microorganisms coded within the stay including a potential lack of microbiological testing, possibly related to an empirical treatment of the infections by the physician, or a potential negative result of samplings, possibly caused by the effectiveness of empirical treatment before sampling is performed. The algorithm's performance might have been compromised by divergent coding practices and by less granularity in the French ICD-10 classifications compared with the ICD-10-CM used for the United States setting. Another limitation in the interpretation of the results of our study is the absence of data outside of the hospital setting within the PMSI database. Also, data collection occurred during the COVID-19 pandemic, which may have had some effect on data results relevant to case identification and healthcare consumption.
Conclusion	This IED feasibility study (2015 to 2021) applied sensitive and specific identification algorithms to the French nationwide PMSI hospital discharge data and showed a substantial burden of hospitalised IED, with 27.4–91.2 incident hospitalisations per 100,000 individuals overall. Incidence rates notably varied across age groups with increased rates observed in older individuals; these ranged from 64.8 to 201.1 per 100,000 in patients aged 60+ years and from 76.1 to 238.9 per 100,000 in patients aged 65+ years. The comparison of incidence rates in this study with that in the published literature suggests that the sole use of French ICD-10 codes, lack of laboratory-confirmed data, and different coding practices in France may have resulted in an underestimation of IED incidence, possibly leading to a lower reported extent of medical and economic burden of IED. A validation study that links the PMSI data to a database that includes laboratory-confirmed data would allow for validation and improvement of the algorithm used and thus should obtain a more complete and reliable estimate of the true burden of IED in the French healthcare setting.

References

1. Geurtsen J, de Been M, Weerdenburg E, Zomer A, McNally A, Poolman J. Genomics and pathotypes of the many faces of *Escherichia coli*. *FEMS Microbiol Rev*. 2022;46(6):fuac031. doi:10.1093/femsre/fuac031.

2. Antimicrobial Resistance Collaborators. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. *Lancet*. 2022;399(10325):629–655. doi:10.1016/S0140-6736(21)02724-0.

3. Jauneikaite E, Honeyford K, Blandy O, et al. Bacterial genotypic and patient risk factors for adverse outcomes in *Escherichia coli* bloodstream infections: a prospective molecular epidemiological study. *J Antimicrob Chemother*. 2022;77(6):1753–1761. doi:10.1093/jac/dkac071.

4. Gagliotti C, Balode A, Baquero F, et al. Escherichia coli and Staphylococcus aureus: bad news and good news from the European Antimicrobial Resistance Surveillance Network (EARS-Net, formerly EARSS), 2002 to 2009. *Euro Surveill*. 2011;16(11):19819. doi:10.2807/ese.16.11.19819-en.

5. Poolman JT, Wacker M. Extraintestinal pathogenic *Escherichia coli*, a common human pathogen: challenges for vaccine development and progress in the field. *J Infect Dis*. 2016;213(1):6–13. doi:10.1093/infdis/jiv429.

6. Jackson LA, Benson P, Neuzil KM, Grandjean M, Marino JL. Burden of community-onset *Escherichia coli* bacteremia in seniors. *J Infect Dis*. 2005;191(9):1523–1529. doi:10.1086/429344.

7. Rhee C, Kadri SS, Dekker JP, et al. Prevalence of antibiotic-resistant pathogens in culture-proven sepsis and outcomes associated with inadequate and broad-spectrum empiric antibiotic use. *JAMA Netw Open*. 2020;3(4):e202899. doi:10.1001/jamanetworkopen.2020.2899.

8. MacKinnon MC, McEwen SA, Pearl DL, et al. Increasing incidence and antimicrobial resistance in *Escherichia coli* bloodstream infections: a multinational population-based cohort study. *Antimicrob Resist Infect Control*. 2021;10:131. doi:10.1186/s13756-021-00999-4.

9. Bonten M, Johnson JR, van den Biggelaar AHJ, et al. Epidemiology of Escherichia coli bacteremia: a systematic literature review. *Clin Infect Dis*. 2021;72(7):1211–1219. doi:10.1093/cid/ciaa210.

10. Abbara S, Guillemot D, El Oualydy S, et al. Antimicrobial resistance and mortality in hospitalized patients with bacteremia in the greater Paris area from 2016 to 2019. *Clin Epidemiol*. 2022;14:1547–1560. doi:10.2147/CLEP.S385555.

11. Fortin SP, Hernandez Pastor L, Doua J, et al. Development and performance characteristics of novel code-based algorithms to identify invasive Escherichia coli disease. *Pharmacoepidemiol Drug Saf*. 2022;31(9):983–991. doi:10.1002/pds.5505.

12. Pandolfi F, Guillemot D, Watier L, Brun-Buisson C. Trends in bacterial sepsis incidence and mortality in France between 2015 and 2019 based on National Health Data System (Système National des données de Santé (SNDS)): a retrospective observational study. *BMJ Open*. 2022;12(5):e058205. doi:10.1136/bmjopen-2021-058205.

13. Thelwall S ON, Bhattacharya A, Wasti S, Gerver S, Davies J, et al. *Annual Epidemiological Commentary: Mandatory MRSA, MSSA and E. coli bacteraemia and C. difficile infection data 2016/17*. 2017. https://webarchive.nationalarchives.gov.uk/ukgwa/20180412112946mp_/https:/assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/634675/Annual_epidemiological_commentary_2017.pdf

14. Opatowski M, Tuppin P, Cosker K, et al. Hospitalisations with infections related to antimicrobial-resistant bacteria from the French nationwide hospital discharge database, 2016. *Epidemiol Infect*. 2019;147:e144. doi:10.1017/S0950268819000402.