01 10 2024

ANNA **GÓRSKA**, EVELINA **TACCONELLI**, FULVIA **MAZZAFERRI**, PASQUALE **DE NARDO**, ELEONORA **CREMONINI**, ELISA **GENTILOTTI**

D1.1 Clinical algorithm for the diagnosis of CA-ARTI

VALUE-Dx

Table of Contents

Abbreviations 3

1. Introduction 4

1.1 Clinical algorithm methodology 4

1.2 Systematic review of the clinical algorithms aiming to diagnose CA-ARTI's aetiology 4

2 Simulations 6

2.1 Cost profiles 6

2.2 BATCH-5 results 6

2.3 BATCH-6 results 8

2.4 Clinical Algorithms' COSTs 9

2.5 Clinical Algorithm Website 10

3. Integration of PRUDENCE data 11

3.1 PRUDENCE data summary 11

3.2 PRUDENCE Clinical algorithm simulation 11

4. Bringing systematic review to life 12

4.1 Meta-analysis automation 12

4.2 ValueDx meta-analysis update 14

5. Discussion and outlook 15

Abbreviations

CA: Clinical algorithm

CRP: C-reactive protein

ED: Emergency department

LCTF: Long-term care facility

LDH: Lactate dehydrogenase

LLMS: Large Language Models

PC: Primary care

PCT: Procalcitonin

POCTs: Point-of-Care Tests

US: Lung Ultrasound

RDT: Rapid Diagnostic Tests

NAAT: nucleic acid amplification technology

NER: Named Entity Recognition

WBC: White blood count1

WP1: Work Package 1 of the Value-Dx Project

X-ray: Chest X-ray

1. Introduction

1.1 Clinical algorithm methodology

The goal of Value-DX WP1 was to systematically review all point-of-care tests (POCTs) for their effectiveness in determining the aetiology of Community-Acquired Respiratory Tract Infections (CA-ARTIs) and to develop a clinical algorithm to guide diagnostic processes. The meta-analysis provided estimates of the sensitivity and specificity of various POCTs, encompassing clinical signs and symptoms, biomarkers like CRP and PCT, imaging techniques, rapid diagnostic tests, and nucleic acid amplification technology (NAAT), that were directly used as an input to the clinical algorithm development. The systematic review and meta-analysis was published in 2022: *Diagnostic accuracy of point-of-care tests in acute community-acquired lower respiratory tract infections. A systematic review and meta-analysis, Elisa Gentilotti, Pasquale De Nardo, Eleonora Cremonini, Anna Górska, Fulvia Mazzaferri, Lorenzo Maria Canziani, Mona Mustafa Hellou, Yudith Olchowski, Itamar Poran, Mariska Leeflang, Jorge Villacian, Herman Goossens, Mical Paul, Evelina Tacconelli, Clinical Microbiology and Infection 28 (1), 13-22.*

We develop a heuristic to derive an optimal clinical algorithm (CA) for the provided parameters: a list of POCTs and their performance, their order, their cost, pathogen prevalence, and the patient-level database for testing. The performance of POCTs was derived from the Diagnostic Test Accuracy meta-analysis. Each CA was a binary decision tree, where nodes corresponded to individual POCTs, and edges represented these tests' positive or negative outcomes. The decision trees were constructed by randomly selecting subsequent POCTs, with the available tests filtered based on preceding results and clinical guidelines and weighted according to their performance. At each node, the estimated probability of bacterial or viral aetiology was calculated based on the preceding sequence of tests and outcomes. When possible, the probability of specific pathogens was also estimated. These CAs were simulated across various settings with different available POCTs. Ultimately, the CAs were tested against patient-level data, and the best-performing algorithms were selected.

Although the datasets used for testing did not contain results for all POCTs identified in the meta-analysis, they did include the final aetiology determined by multiplex tests or at a later stage of the disease. For patients with a known aetiology who encountered a node requiring a POCT result not present in the dataset, the test result was emulated through random selection, with the probability of success reflecting the test's performance and the real aetiology of the patient's condition. This method allowed for an effective imputation of the missing values or an in-silico simulation of scenarios close to real life.

1.2 Systematic review of the clinical algorithms aiming to diagnose CA-ARTI's aetiology

A systematic review of the existing clinical algorithms aiming to diagnose the aetiology of CA-ARTIs was completed. The review identified 33 studies carried out from 1986 to 2021 and identified 64 individual clinical algorithms (CAs). All but one CAs were derived from logistic regression models based on limited datasets specific to certain settings and populations. **57% of CAs did not include any validation.** CAs were typically designed to diagnose a single disease e.g., community-acquired pneumonia (CAP) or a specific condition e.g., influenza.

The most used signs and symptoms in CAAs to predict bacterial or viral etiology in CAP patients included fever, cough, rhinitis, gastrointestinal symptoms, and age, with C-reactive-protein (CRP), Lactate dehydrogenase (LDH), and sodium levels used specifically for bacterial diagnoses, while white-blood-count (WBC) was used for both etiologies (**Figure 1**). A wide range of clinical, epidemiological, and diagnostic data are included across different algorithms, with no consensus on optimal features. The high heterogeneity in study definitions, populations, settings, and outcomes prevents meta-analysis and standardized evaluation of algorithms, few of which have been externally validated or included in evidence-based recommendations, making it difficult to recommend a specific algorithm or apply them effectively in practice.

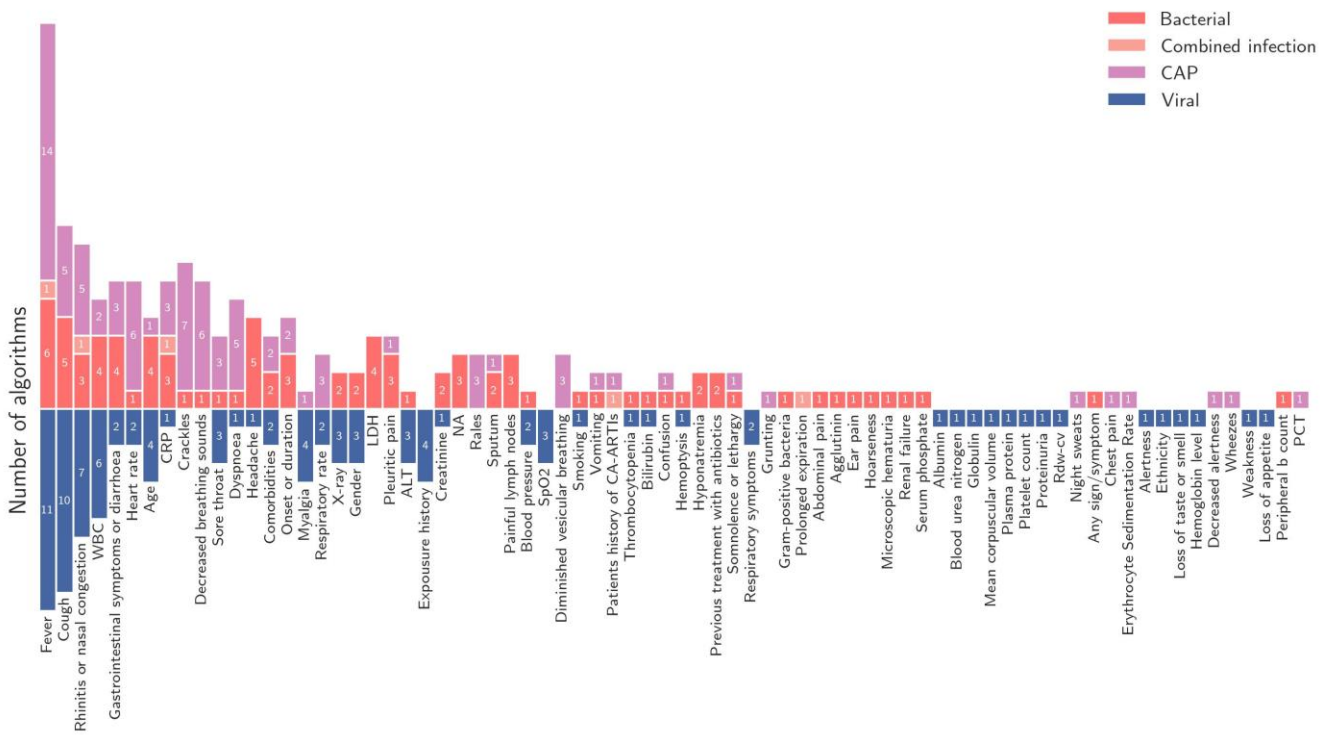


Figure 1 *Distribution of the features used within the algorithms discovered through the systematic review of the CAs. Colors correspond to the aetiology targeted by the CAs: red to bacterial, blue to viral, and pink to the diagnosis of Community-Acquired Pneumonia (CAP) without the etiological specification. Algorithms derived for infants were not shown in this plot.*

This review pointed out several key advantages of our approach to CA development allows for integrating new POCTs, utilizes robust estimates of POCT performance, and accurately reflects the step-by-step nature of the diagnostic process. Separating the shape of the CAs from the testing datasets allows for a real performance-driven design with independent validation that can be expanded and adjusted based on setting and time.

2 Simulations

2.1 Cost profiles

First, CA simulations (BATCH-5) used a stand-in cost, which assumed the following POCT order: symptoms > biomarkers > imaging > RADTs > NAATs.

Value-DX **WP5** provided updated cost estimates for the POCTs in Spain (termed **Spanish Cost Profile**), originally included in the meta-analysis and those integrated into the clinical algorithm. We assumed that visit costs remained constant, and since symptoms were documented during the visit, they were assigned a zero cost. For all RDTs, including both antibody and antigen tests, a uniform cost of 10 euros was applied. Biomarkers such as CRP and PCT were assigned a cost of 11 euros each. The cost of an X-ray was set at 80 euros. More advanced tests, including PCR and RT-PCR, were valued at 100 euros. The lung ultrasound, priced at 103 euros, emerged as the most expensive test in the set. The POCT cost profile allows for selecting CAs based on their performance and the optimal performance-to-cost ratio. Based on the **Spanish Cost Profile**, **BATCH-6** simulations were run with the POCT order: symptoms > Biomarkers, RDT > X-ray > NAATs > US.

2.2 BATCH-5 results

BATCH-5 resulted in 5,507 candidate CAs across ten test availability combinations (settings). 13,191 patients matched the inclusion criteria and were used for algorithm validation. The simulations explored 550 CAs before the rarefaction curves plateaued. For the four POCT combinations that did not use any of the biomarkers or NAATs, the simulation of trees with maximally 40% of unvisited edges failed, and those POCT combinations were deleted. For the remaining POCT combinations, the candidate CAs exhibited predictive values for both bacterial and viral aetiologies, with the bacterial AUROCs being significantly better than that of viral and often well above 50%.

Fout! Verwijzingsbron niet gevonden. **a)** presents the overall performance of the entire simulation in both the viral and bacterial AUROCs. Overall, imaging tests are the largest driver of the performance differences, with the US making for the best algorithms, X-ray the second best, and no imaging creating the worst-performing trees. While the algorithms vary significantly in their Bacterial AUROC, only minor differences were observed in their performance for the Viral aetiology. The viral performance plots exhibit a clear division: most algorithms display poor AUROC values (< 0.55), yet a few stand out in the higher-performing range.

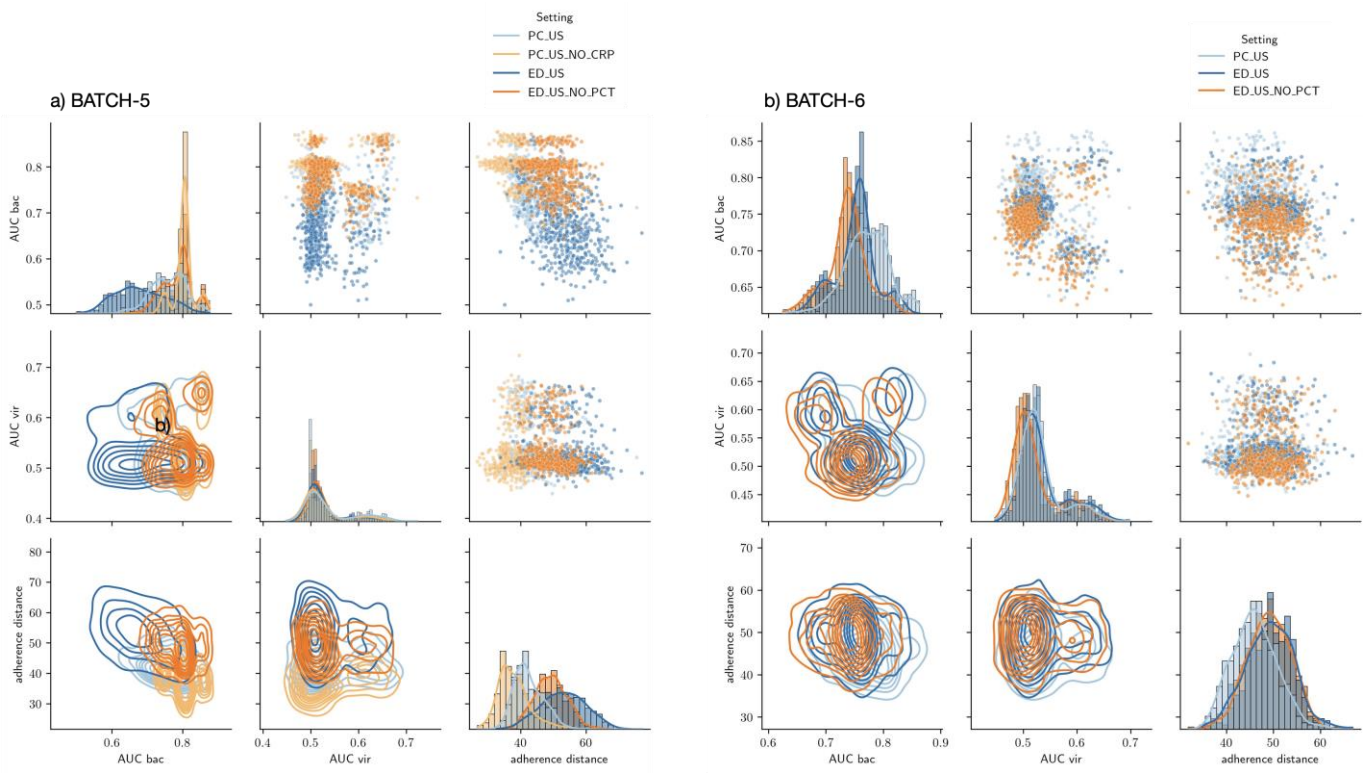


Figure 2 *2D density plots comparing the main performance parameters: AUC for Bacterial (AUC bac) and Viral (AUC vir) and adherence distance, across* ***a)*** *four simulations in BATCH-5 and* ***b)*** *three simulations in BATCH-6. Adherence distance is a sum of the absolute distance from the computed and data-derived probability of bacterial aetiology.*

Figure 2 compares the performance of the clinical algorithms represented by the 2D density plots. The split in viral AUROC is more apparent, revealing two distinct clusters. Notably, the distribution shape for viral aetiologies remains consistent regardless of the availability of POCTs. The ED_US setting explores a broader range of clinical algorithms than the other three settings, suggesting that including molecular tests allows for a wider exploration of the test space. Consequently, some of these algorithms may be better suited to specific real-life scenarios, particularly when factors such as cost are considered. Rather than analysing these distributions separately, we calculated a combined score, the average of the Bacterial and Viral AUROC scores (Bacterial+Viral AUROC)/2.

Fout! Verwijzingsbron niet gevonden. presents the results for the best trees for each setting. Sorting by the combined score did not change the selection of the best simulations. The best simulation was the PC_US_NO_CRP, i.e., the trees using the Ultrasound and no CRP; the second-best performing simulation was the ED_US_NO_PCT, using CRP, molecular tests, NAAT-tests, and Ultrasound.

Notably, the **BATCH-5** simulation took significantly longer than previous simulations, which included the WBC POCT. The simulation algorithm is designed to reject clinical algorithms with more than 40% unvisited edges. As a result, if the algorithms are implausible, the heuristic continuously rejects candidate clinical algorithms, leading to extended simulation times.

Table 1 *Best performing trees in both simulations according to the discriminatory performance. The average cost column contains the per-patient cost according to the* ***Spanish cost profile****.*

Setting abbreviation	Imaging	PCT	CRP	Molecular tests	AUROC BAC	AUROC VIR	Average cost
BATCH-5							
PC_US	Ultrasound	No	Yes	No	0.87	0.67	38.15
PC_US_NO_CRP	Ultrasound	No	No	No	0.87	0.67	35.34
ED_US_NO_PCT	Ultrasound	No	Yes	Yes	0.87	0.66	169.41
ED_US	Ultrasound	Yes	Yes	Yes	0.85	0.67	176.88
ED_NO_IMAG_NO_PCT	No	No	Yes	Yes	0.73	0.67	142.27
ED_XRAY_NO_PCT	X-Ray	No	Yes	Yes	0.72	0.63	150.68
ED_NO_IMAG	No	Yes	Yes	Yes	0.72	0.67	169.81
PC_XRAY	X-Ray	No	Yes	No	0.7	0.66	26.26
ED_XRAY	X-Ray	Yes	Yes	Yes	0.7	0.64	141.99
PC_NO_IMAG	No	No	Yes	No	0.69	0.65	3.77
BATCH-6							
PC_US	Ultrasound	No	Yes	No	0.84	0.66	32.1
ED_US	Ultrasound	Yes	Yes	Yes	0.83	0.66	175.71
ED_US_NO_PCT	Ultrasound	No	Yes	Yes	0.82	0.65	162.56
ED_NO_IMAG_NO_PCT_NO_CRP	No	No	No	Yes	0.72	0.64	148.44
ED_XRAY_NO_PCT	X-Ray	No	Yes	Yes	0.71	0.64	65.66
ED_NO_IMAG_NO_PCT	No	No	Yes	Yes	0.71	0.68	202.81
ED_NO_IMAG	No	Yes	Yes	Yes	0.69	0.65	161.21
PC_XRAY	X-Ray	No	Yes	No	0.67	0.66	25.53
PC_NO_IMAG	No	No	Yes	No	0.64	0.66	3.76

2.3 BATCH-6 results

BATCH-6 simulation ran for 550 CAs for each setting before the simulation plateaued, amounting to 4951 CAs. However, this time the four settings have failed: PC_US_NO_CRP (ultrasound, no NAATs and no CRP), ED_XRAY (molecular tests, chest x-ray), PC_XRAY_NO_CRP, PC_NO_IMAG_NO_CRP, i.e., during the time of the simulation, no CA or very few potential clinical algorithms were simulated. Otherwise, the analysis of the **BATCH-6** outcome was performed analogously to **BATCH-5** and showed that, on average, the performance was worse in **BATCH-6** than in **BATCH-5**. However, the best-performing trees are only slightly worse than those of **BATCH-6**. The shape of the space covered by the three best-performing settings differs from the analogous analysis for the **BATCH-5**.

We analysed how the probabilities of bacterial aetiologies diverge as the clinical algorithm progresses. The underlying rationale is that the distinction between bacterial and viral probabilities should become increasingly pronounced as the algorithm advances. The earlier this divergence occurs, the more effective the clinical algorithm is considered. **Figure 3** illustrates the separation between the two probability peaks, representing high and low probabilities for bacterial and viral aetiologies.

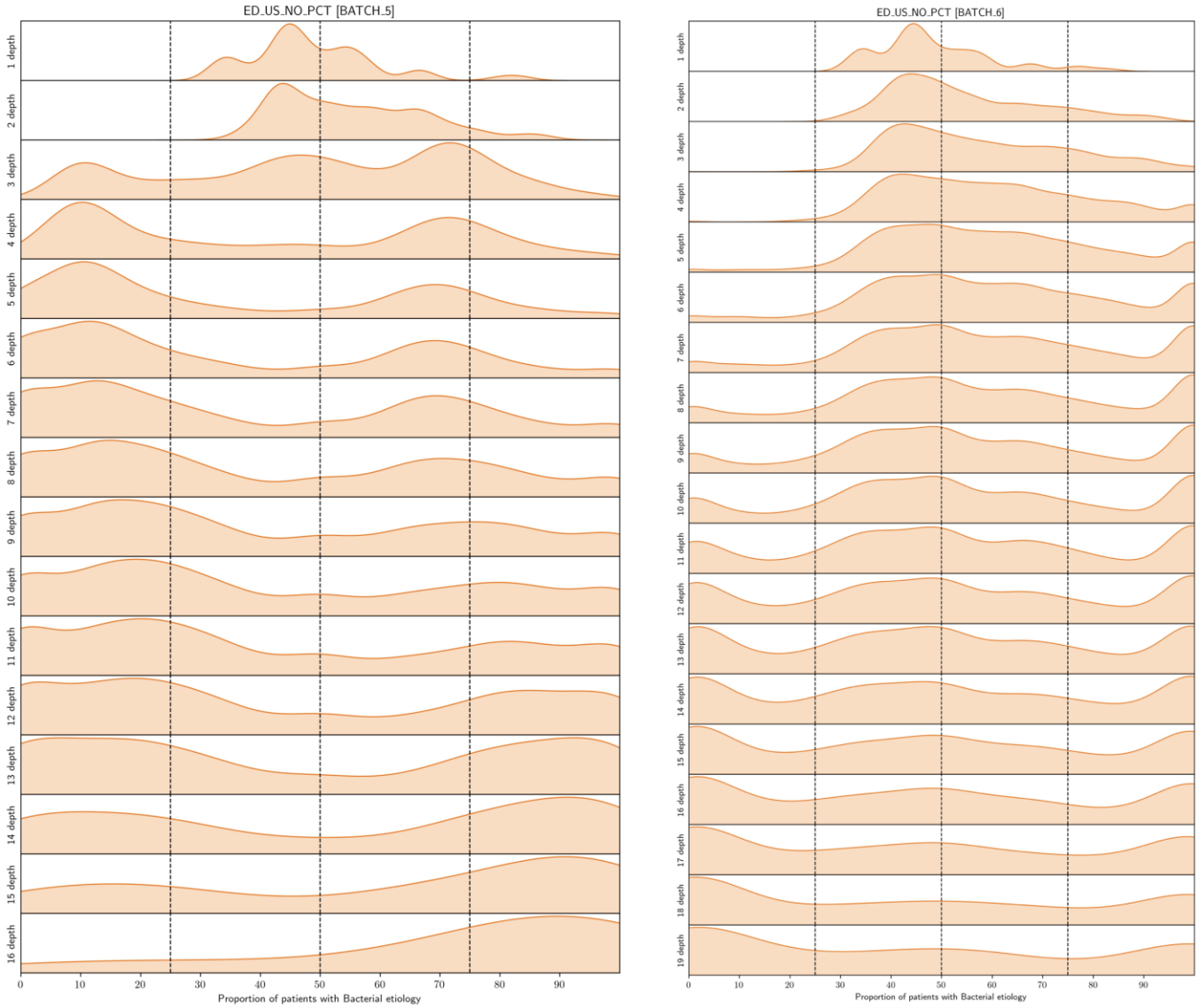


Figure 3 *Distribution of bacterial etiology based on patient-level data across all CAs within the simulation, shown alongside the clinical algorithm depth for all decision trees from both simulations. The top panels are identical, depicting the distribution of bacterial etiology probabilities for the entire population at the entry node. Subsequent panels show the distributions of the bacterial etiology at each subsequent depth level*

2.4 Clinical Algorithms' COSTs

Figure 4 illustrates that the average cost of clinical algorithms (CAs) according to the Spanish Cost Profile varies widely, ranging from 0 to as much as 500€, according to the combination of available trees, POCTs, or settings. However, no clear correlation exists between a CA's performance and cost in **BATCH-5** and **BATCH-6** simulations. Notably, a subset of algorithms demonstrates high average performance at relatively low cost. To refine the selection process, we identified the top-performing algorithms based on a new criterion: maximizing the ratio of performance, measured by average AUROC, to average cost, according to the updated cost profile.

The average cost is not the most reliable metric due to the high standard deviation, which reflects the variability in patient experiences. While some patients may receive an early diagnosis, others may require more complex and expensive diagnostic procedures. The ideal

clinical algorithm would stratify patients based on their specific needs, ensuring that everyone undergoes only the necessary tests, minimizing costs, and optimizing efficiency. **Figure 5** shows the distribution of the costs across the patients for the best-performing trees.

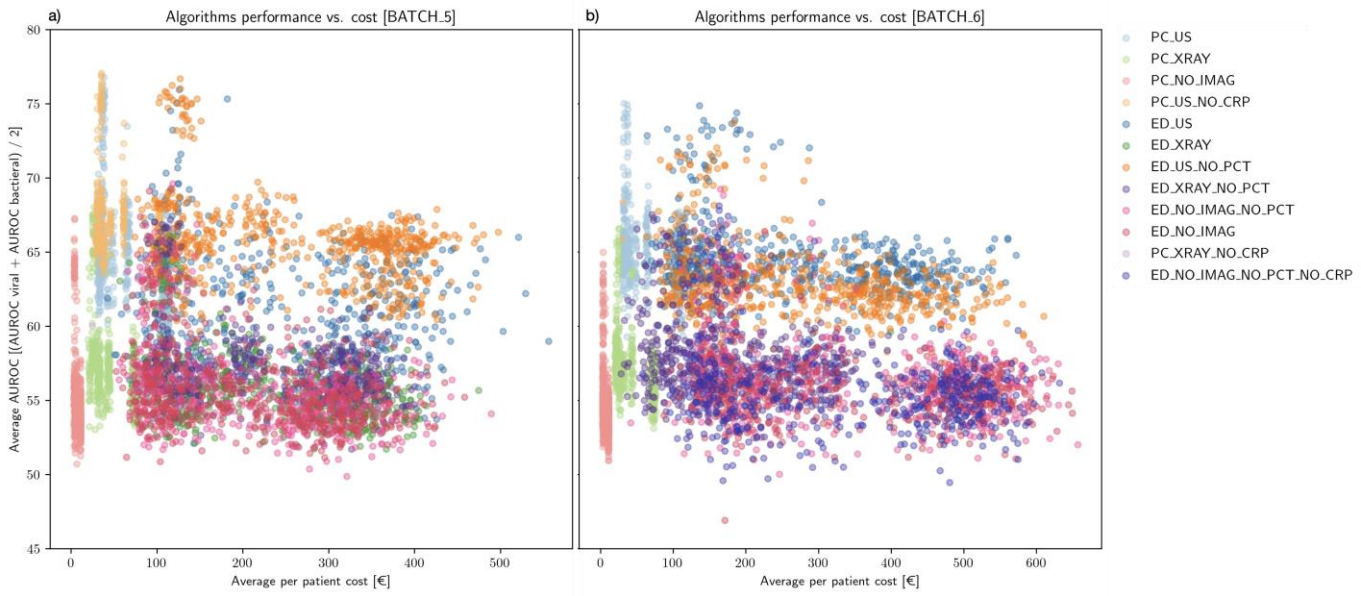


Figure 4 *Clinical algorithm's performance vs. the average cost computed for the costs according to the Spanish Cost Profile across different settings in both simulations.*

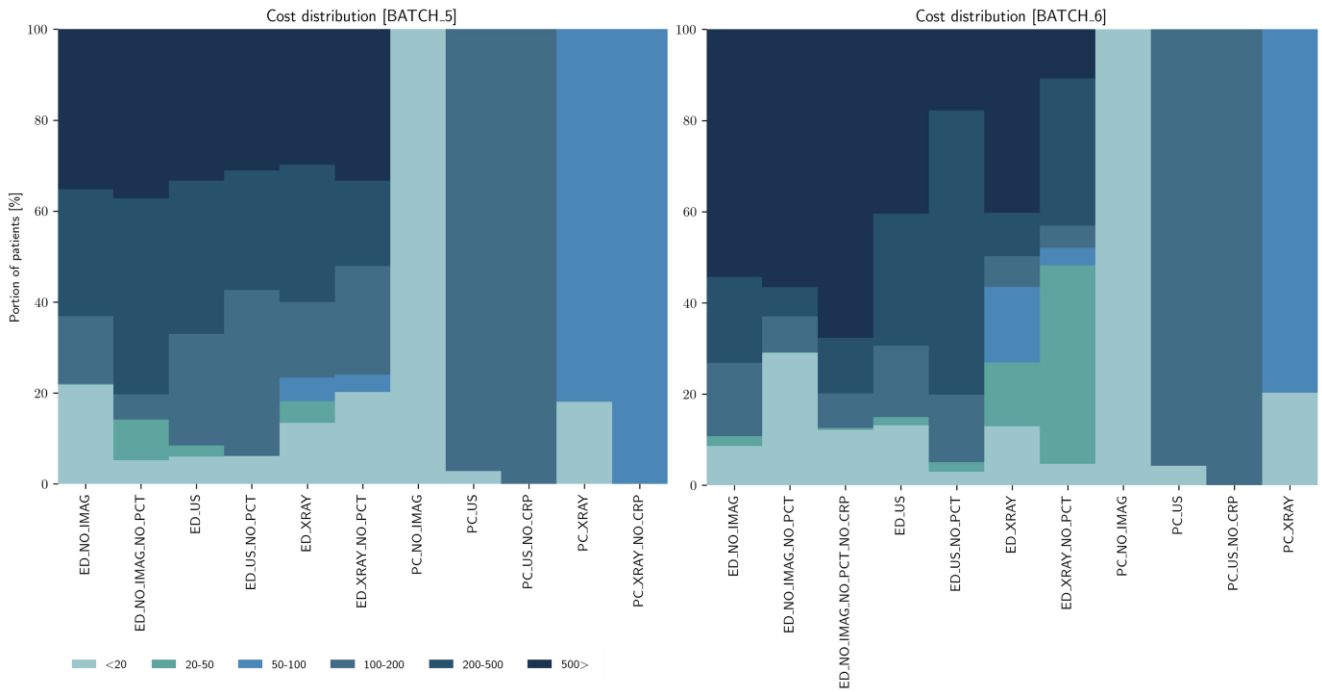


Figure 5 *Relative cost distribution among patients for the best-performing trees within each simulation in two runs.*

2.5 Clinical Algorithm Website

To deal with the complexity of the clinical algorithms, we have developed a website with interactive visualization for both the results of the meta-analysis and the clinical algorithm. The meta-analysis section is divided into five categories: signs and symptoms, biomarkers,

RADTs (antibody and antigen-based), and NAATs. Each view consists of a bubble chart of the studies' sensitivity vs. specificity, with the bubble size corresponding to the study's number of participants and the table outlining the studies linked to PubMed and QUADAs. The bubble chart and the table are interactive and connect the study-level meta-analysis results to the details of the study.

The clinical algorithm view consists of the table with the best algorithms for each combination of the POCTs (setting). The algorithm can be either visualized in its entirety or performed by providing the answers to the sequence of the POCTs.

The website is available at https://ca-valuedx.ddsp.univr.it.

3. Integration of PRUDENCE data

3.1 PRUDENCE data summary

PRUDENCE trial aimed to assess the impact of three rapid diagnostic tests—the CRP test, the Veritor Flu test, and the rapid Group A Streptococcus (GAS) test—on antibiotic prescribing practices in both primary care (PC) and long-term care facilities (LTCFs). The trial included patients of all ages presenting with cough or sore throat. Participants were asked to report whether they experienced any of nine symptoms: cough, dyspnoea, diarrhea, headache, fatigue, fever, myalgia, and sore throat at the onset of illness and over the following 28 days. The primary outcome measured was the time to recovery, with antibiotic usage also recorded.

The export PRUDENCE sub-dataset included 2,664 patients older than 16 who all survived the follow-up period. 120 patients were excluded due to a positive SARS-CoV-2 test, and 358 patients were removed as they had no recorded symptoms. The final dataset consisted of 2,186 patients, 1,392 women and 794 men, with 2,009 patients from primary care (PC) settings and 177 from long-term care facilities (LTCFs). Among the patients, 391 were tested for influenza, with 41 testing positive, while 386 were tested for GAS, resulting in 79 positive cases. Additionally, CRP values were recorded for 336 patients. The 120 120 patients with an etiological diagnosis contributed to the emulation algorithm. Adding these data enabled the comparison of the time to recovery between patients prescribed antibiotics and those not at each decision node within the clinical algorithm.

We examined whether there was a significant difference in time to recovery between patients taking antibiotics and those not taking them within subgroups of patients with positive and negative POCT outcomes (CRP, Influenza, and GAS). Overall, there was no statistical difference in recovery time between patients who took antibiotics and those who did not (5.82±5.69 days vs. 5.5±5.95 days, p=0.217). However, among patients who tested negative for Influenza, those not taking antibiotics had a shorter recovery time (antibiotics: 6.6±6.12 days vs. no antibiotics: 4.58±4.11 days, p<0.001). Conversely, in patients who tested positive for GAS, those taking antibiotics showed a significantly shorter recovery time (antibiotics: 3.5±3.54 days vs. no antibiotics: 10.22±12.58 days, p=0.001). There were no significant differences in recovery time across different thresholds of the CRP.

3.2 PRUDENCE Clinical algorithm simulation

New clinical algorithms were simulated to reflect the POCT availability in PRUDENCE, allowing the selected signs and symptoms, CRP, and Veritor Flu test, not allowing for imaging, PCT or

NAAT-based tests. However, as the PRUDENCE trial did not include patients with both the CRP and Veritor tests, another simulation was also performed to test the possible annotation of the clinical algorithms with the secondary outcomes. The simulated CAs were tested against the same patient-level databases as the previous simulations to maintain comparability.

Both simulations ran for 550 clinical algorithms and finished within minutes. **Figure 6** shows the simulation results as scatters of the viral vs bacterial performance. The CAs with the highest and most balanced performance were selected to be included on the website within the BATCH-6 section.

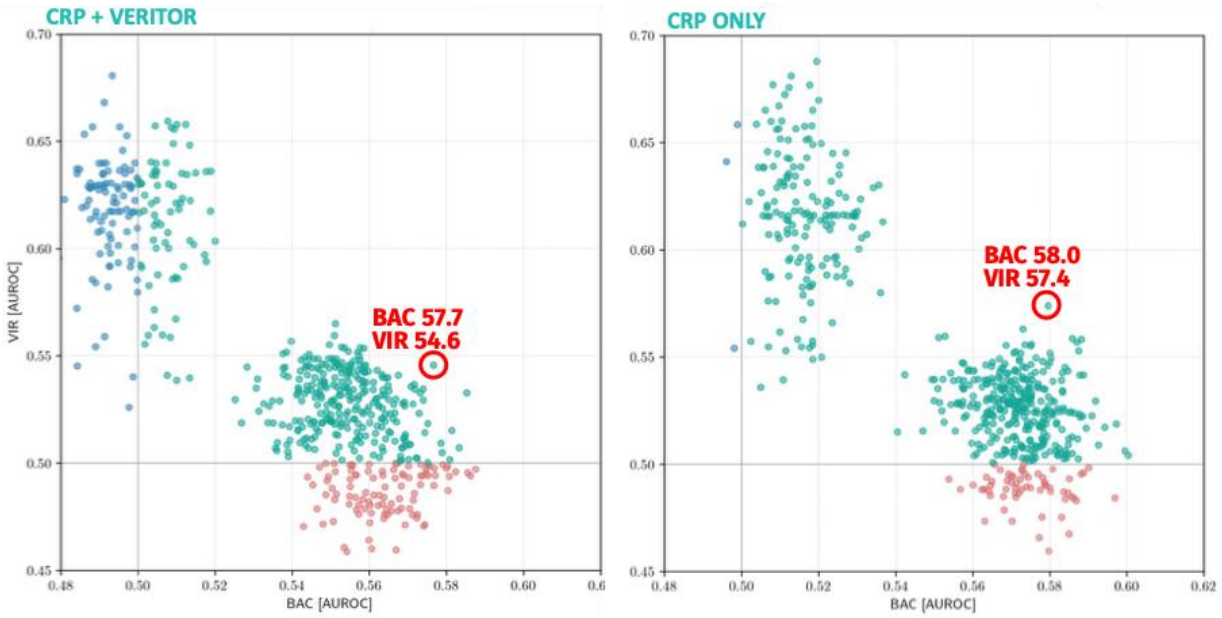


Figure 6 *Performance characteristics of the PRUDENCE clinical algorithms simulations, with the CAs selected as the best performing one denoted in red. The blue and pink points denote the clinical algorithms with insufficient discriminatory power in the bacterial and viral aetiologies, respectively. Conversely, green points correspond to the CAs with positive discriminatory power in both aetiologies. The red circle denotes the trees that can be visualized on the website.*

4. Bringing systematic review to life

4.1 Meta-analysis automation

The primary limitation of the clinical algorithm formulation and heuristic methodology stems from its strength—the scope of POCTs is inherently restricted by the breadth of their coverage in the existing literature. Consequently, important subgroups, such as children, the elderly, and specific clinical settings, were underrepresented. Also, the underlying meta-analysis must be continuously updated to ensure the algorithms remain applicable in the ever-evolving diagnostic landscape, but the systematic review and meta-analysis are highly time-consuming, with the initial effort in the ValueDX WP1 requiring nearly two years of work from three clinical researchers. Therefore, we developed a pipeline aiming to automate the entire meta-analysis process.

The inclusion/exclusion process can be approached as a supervised learning problem. We used general-purpose Large Language Models (LLMs) for text-based classification tasks with high-quality performance. Likewise, the extraction can be automated, although it is more

difficult since there might be multiple extractions from a single publication through code-level interaction with the ChatGPT. The chat can perform the higher structured tasks, e.g., "*name all POCTs mentioned in this publication abstract*". We created a custom framework for automating the two meta-analysis steps to streamline mapping the POCTs from the publications with little manual work required.

The dataset consisted of 9229 publications, with 599 accepted for the meta-analysis. Several foundation LLMs were finetuned against the classification task for inclusion in the ValueDX meta-analysis. All of these showed very good performance, reaching 95% sensitivity and specificity with 0.17 loss, using the title and the maximal possible abstract section. ***Table 2*** shows the performance of the four models fine-tuned for the screening task of the ValueDX meta-analysis.

Table 2 *The four best combinations for screening the Value-DX.*

Model	Loss	Precision	Recall
BiomedNLP-PubMedELECTRA-base-uncased-abstract	0.11	0.96	0.96
biobert-base-cased-v1.1	0.13	0.96	0.96
BiomedNLP-PubMedBERT-base-uncased-abstract-fulltext	0.17	0.95	0.95
BioM-BERT-PubMed-PMC-Large	0.23	0.94	0.94

We tested several automatic methods for extracting the values from the abstract of the publications. Each extraction column calls for a specifically tailored pipeline, and three main methods were used:

- **NER**: Named Entity Recognition with spaCy, then mapping to the manually curated and structured reference dataset. Each item in the reference consisted of the entity, synonym list, and extraction list. Therefore, a single entity could participate in several extraction variables. This pipeline was applied to each sentence in the text and subsequently merged for the abstract
- **LLM**: Supervised LLM finetuning, analogous to the screening, developed in-house using the extraction variables shared across multiple reviews or developed by other parties
- **GPT3**: OpenAI gpt-3.5-turbo-1106 (with cl100k_base encoding) via the Python API with a structured query: "You are a helpful assistant. This was a title and abstract of the scientific study [TITLE+ABSTRACT].", followed by the extraction question, e.g., "Does this study use white blood count. Answer with yes or no.". The same was used with the newer GPT model, gpt-4o (GPT4)
- **GPT3 combined**: is analogous to the GPT3 pipeline, but using the combined query – where we ask all the questions in a numerated single query, and request the answers in a numbered list

The LLM-based method was tested by splitting the dataset into training and testing. The GPT and NER were tested against the manual full-text extraction. The crucial part of the extraction, which was not automated at this stage, was the confusion matrix, describing how well the point-of-care (POCT) test was performed and the QUADAS-2-based quality criteria.

Figure 7 presents the different pipelines' performance in extracting the study's basic information directly from the abstracts for the original ValueDX meta-analysis tested against the manual extraction. The GPTs perform the best. However, there is no difference between the GPT3 and GPT4, which significantly affects the cost of this analysis. The graph also shows

a good performance of the LLM-finetuned pipeline. Its performance can be further improved by enlarging the database and extracting the population information. The publications' larger and more diverse base should enable a better prediction. Finally, we were also able to deploy the full-text-based pipeline. However, this is only the first attempt. We need to expand the pre-processing and post-processing steps to better utilize the structure of the publications.

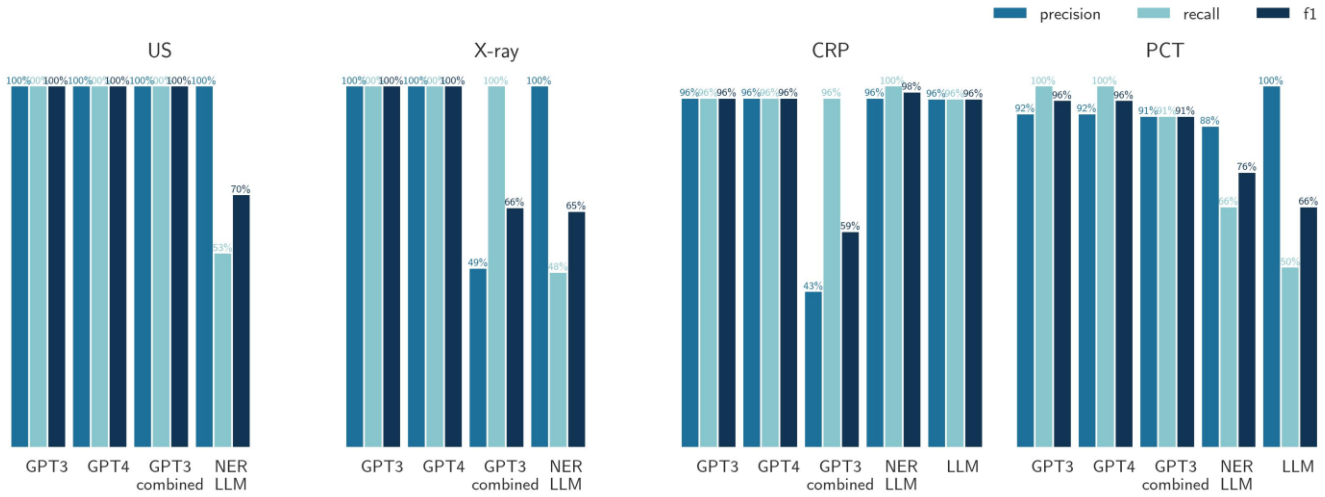


Figure 7 *Performance of the automatic extraction pipelines.*

4.2 ValueDx meta-analysis update

A new search was run on 13.08.2024, yielding 14,759 novel publications not included in the Value-DX meta-analysis, and 5,198 publications had a publication date after 30.09.2021, the last meta-analysis update. 264 publications were accepted by the finetuned PubMedELECTRA model. The model is very confident in rejecting the publications – with the majority being rejected with probability > 90%, within the accepted papers, there is more variability of the score (**Figure 8**). **Figure 910** presents the distribution of the abstract-based extraction for the accepted publications using the GPT-3 binary query.

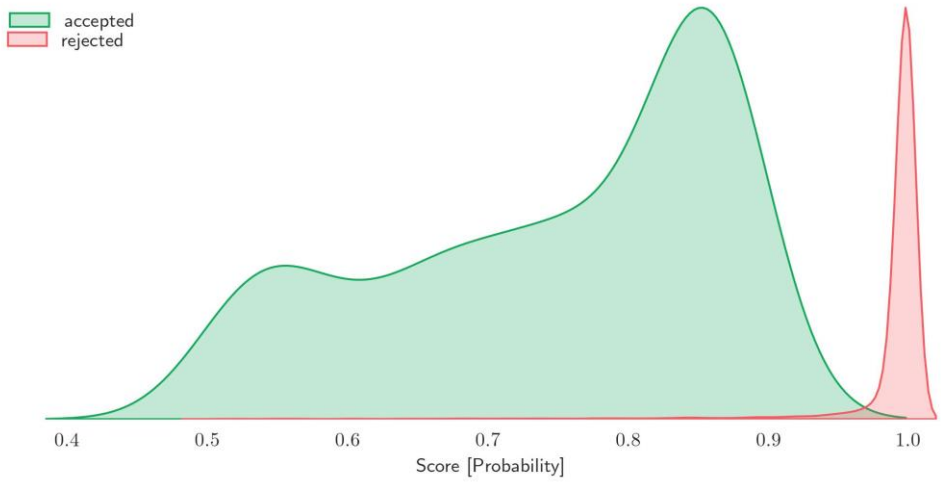


Figure 8 *Distribution of the LLM-scores for prediction of the two labels: acceptance and rejection.*

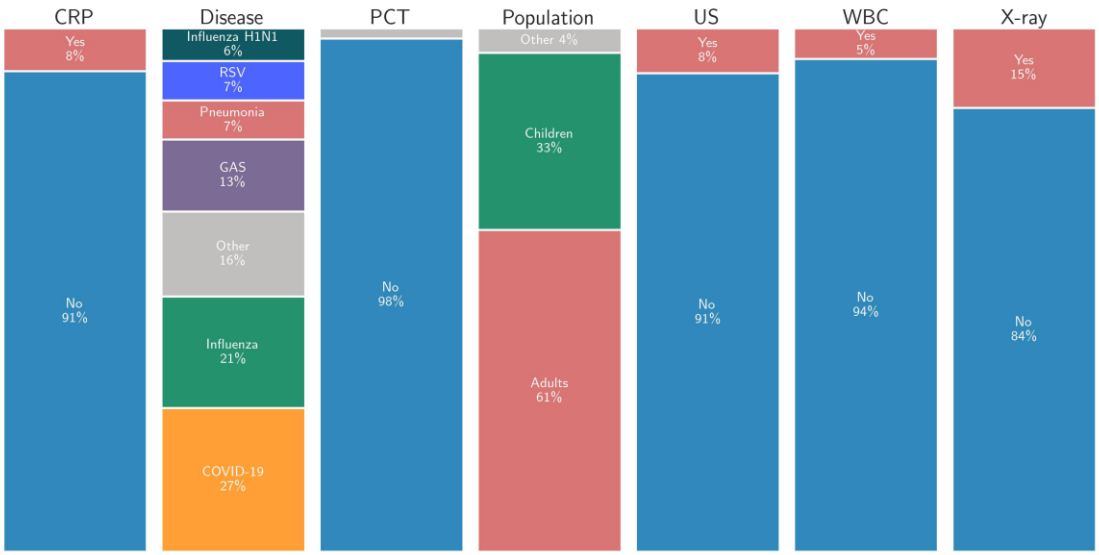


Figure 9 *Statistics of the automated extraction from the update search in August 2024.*

5. Discussion and outlook

Throughout the **VALUE-DX** project, we developed an innovative approach to clinical algorithm derivation, relying on extensive meta-analysis results rather than the traditional data-driven methods. Our findings demonstrate that the heuristics used in developing these algorithms could successfully generate solutions based on varying orders and availabilities of point-of-care tests (POCTs), each with distinct performance levels and associated costs. Among these, ultrasound consistently emerged as the most impactful POCT, significantly enhancing the algorithm's performance regardless of the test order.

Our methodology inherits several limitations of the meta-analysis, such as a lack of specificity for patient subgroups, clinical settings, and seasonal variations. These challenges could be mitigated by integrating high-quality, patient-level data. Additionally, while we assessed the algorithms' effectiveness in identifying bacterial and viral aetiologies, we did not investigate their potential impact on antibiotic prescribing practices.

The clinical algorithms and the underlying meta-analysis have been visualized on a custom, interactive website that links evidence directly to the algorithms for the first time.

The optimal algorithm might be ever-changing, adapting to the local conditions, including the patient population, pathogen provenance, POCT performance, and availability. This formulation remains highly versatile as it can support different kinds of annotations to inform on the secondary outcomes at the nodes (e.g., cost, time to recover as shown, but also potentially comorbidities), to answer different clinical questions – thus enabling further development of the specific outcomes' formulation, and further optimization of the selection of the algorithms to pursue other performance outcomes potentially.

Sustained performance of the clinical algorithms depends on the accuracy of input data, including POCT performance across subgroups, pathogen prevalence, and the representativeness of patient datasets. We developed an automated methodology for systematic reviews and meta-analyses to address this. Our flexible, complex methods for automating both screening and data extraction significantly reduced the need for manual

work. The toolkit created within the VALUE-DX project is being further refined and applied in other scientific contexts to aid future meta-analyses and ensure the long-term sustainability of these processes.

Overall, the innovations and processes developed through this project can be generalized as a pipeline for knowledge discovery (systematic review), synthesis (meta-analysis and clinical algorithm development), and delivery (interactive visualization).

Project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement No 820755. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA and bioMérieux SA, Janssen Pharmaceutica NV, Abbott, Bio-Rad Laboratories, BD Switzerland Sàrl, and The Wellcome Trust Limited.


University
of Antwerp


BIOMÉRIEUX


wellcome


UNIVERSITY OF
OXFORD


UMC Utrecht


Penta
Child Health Research


UNIVERSITÀ
di VERONA


THE UNIVERSITY
of EDINBURGH


University Medical Center Groningen

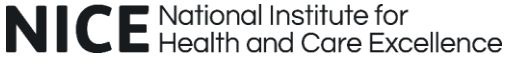

RAMBAM
Health Care Campus


UNIVERSIDAD
DE LA RIOJA


BOSTON
UNIVERSITY


BIOASTER
MICROBIOLOGY TECHNOLOGY INSTITUTE


FIND


NICE National Institute for
Health and Care Excellence


Gesundheit Österreich
GmbH


ESCMID
EUROPEAN SOCIETY
OF CLINICAL MICROBIOLOGY
AND INFECTIOUS DISEASES


ERS
EUROPEAN
RESPIRATORY
SOCIETY
every breath counts


Berry Consultants
Statistical Innovation


ZonMw


janssen


Abbott


BIO-RAD


BD

www.imi.europa.eu www.value-dx.eu