30|09|2024

Author name: Fernando Antoñanzas

D5.3 Economic analysis of the WP4 clinical trials

VALUE-Dx

This 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

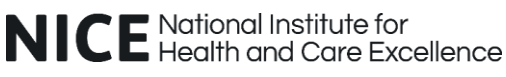

RAMBAM
Health Care Campus


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

health-economic analysis of
diagnostics strategies

D5.3 Economic analysis of the WP4 clinical trials

Lead contributor	University of La Rioja
Other contributors	Universitair Medisch Centrum Groningen, NICE, BioMérieux SA.

Document History

Version	Date	Description
V1.1	25-9-2024	**Economic analysis of the WP4 clinical trials**

CONTENTS

Abstract

Objective 1: To design a health-economic framework (HEF) to assess and demonstrate the value of diagnostics both for individual patients and for public health impact by reducing antibiotic use and subsequent antibiotic resistance among patients.
A short-run economic evaluation analysis has been elaborated for both PRUDENCE and ADEQUATE trials to assess the value of introducing point of care diagnostic tests before deciding the antibiotic prescription. In these two short-run models, an analysis of the two databases from the two separated trials has been carried out.

PRUDENCE trial did not yield statistically significant differences between the two arms (POCT and standard of care) and their numerical differences were minimum (about 1.4%); furthermore, the quality of life analysis did not show any differences between the two arms either. Hence, a cost analysis was developed for the PRUDENCE database, whose final results indicated that the healthcare system costs of treating the disease process were 121€ in the POCT arm and 112€ in the standard of care arm. Costs are 264€ in both arms when productivity costs are incorporated into the analysis. Interestingly, the quality of life analysis indicated that the utility of patients at day one of their randomization was 0.80 and that utility increased up to 0.93 at day 14, showing a plateau, as at day 28 the same value was estimated. However, as expected, the QALY difference between the two arms was small (about 0.00014) at day 14. The cost analysis was presented for each participant country using the purchase power parity conversion.

ADEQUATE trial did not show statistically significant differences between the two arms, however, the numerical differences allowed to carry out an economic evaluation analysis. Regarding the quality of life study, differences between the two arms were not found either, hence, the calculation of a cost-utility analysis was not advisable. However, a cost-effectiveness analysis was elaborated instead. The results for the primary endpoint (number of days on antibiotic consumption at day 14) indicated that the POCT is a dominant option versus the standard of care, that is to say, it cost less and it avoided antibiotic consumption and generated more days alive out of hospital. The same result holds for the 30 day period. It is important to remark that the results are subject to a high variability, in the sense that a few more days of hospitalizations or of ICU stays, due to their high cost, could change the sign of the difference of total cost between the two arms, i.e. making the POCT arm more costly. The distribution of the number of hospital and ICU stays between the two arms (fewer in the POCT arm) may be caused by a random effect of the trial and it is difficult to determine without uncertainty, that it was due to the use of the POCT.

Max ½ page

OBJECTIVE

To design a health-economic framework (HEF) to assess and demonstrate the value of diagnostics both for individual patients and for public health impact by reducing antibiotic use and subsequent antibiotic resistance among patients.

To elaborate a short-run economic evaluation analysis for both PRUDENCE and ADEQUATE trials to assess the value of introducing point of care diagnostic tests before deciding the antibiotic prescription. In these two short-run models, an analysis of the two databases from the two separated trials is carried out.

MATERIALS AND METHODS

The materials for this analysis came from the PRUDENCE and ADEQUATE trials conducted by the Value-Dx project team. Both trials collected data on antibiotics prescription and consumption, utilisation of health resources, demographic information and quality of life status of participants. Data were not combined for the economic evaluation analysis, although the methods were the same. Methods of the economic evaluation were specified in advance in the health economic analysis plan (HEAP), see ANNEX I.

Aim of the short-term economic evaluation

The primary objective of the within trial economic evaluation was to provide the short-term cost-effectiveness using individual patient level data on costs and outcomes, including antibiotic prescriptions and quality of life.

Overview of economic analysis

Resource use and effectiveness outcomes were measured for each participant in the PRUDENCE and ADEQUATE trials. Costs and effectiveness outcomes were calculated for each participant. The differences between the two arms (POCT and standard care) were calculated to give an incremental cost-effectiveness ratio. The data from the two trials were reported separately.

Trial data

The data for the economic evaluation were taken from the two trial databases. Several assumptions were made for this task (see annex I). A blinded interim dataset was used to develop the economic analysis. The final dataset contained information for all participants in the trials. Primary data from each trial were conveniently treated using an Excel spreadsheet to get aggregated intermediate results on resource utilisation and

health outcomes. These calculations provided the amount of health resources in physical units (e.g. number of medical visits, number of different diagnostic tests, number of different drugs, number of hospitalisations, etc.) as well as other resources related to productivity and school attendance of participants (e.g. number of work days lost or days out of school due to the disease process).

Cost data

For the economic evaluation, a cost associated with each resource used is required (a unit cost). As there is no comprehensive European database with the costs of health resources for all countries in the trials several articles focused on economic evaluation of health technologies were reviewed to identify the unit costs. However, it was not possible to identify the costs required for all countries. For the sake of making the calculations uniform, it was decided to use only data of the UK and Spain as extensive data collections of unit costs were available for these countries. The reference year for the analysis was 2023. Data from different years were inflated to 2023 using consumer price indexes of both countries (for Spain: https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736176802&menu=ultiDatos&idp=1254735976607, and for the UK: https://www.ons.gov.uk/economy/inflationandpriceindices). If required, the economic evaluation results could then be adapted to the other participant countries using the World Bank purchasing power parity (PPP) table for each country (https://www.worldbank.org/en/programs/icp). Other resource costs such as the cost of a working hour were obtained from the Eurostat database (https://ec.europa.eu/eurostat/data/database). Furthermore, some resources were paid on an out-of-pocket basis by patients from each country and reported in local currency. For their translation into euros the PPP table was used.
The trials were designed to obtain statistically significant aggregated data for the primary variables of the study. Therefore, individual country results were not powered to be used in the economic evaluation. Since there may be an interest in obtaining specific results of the efficiency of the point-of-care tests for each participant country, aggregated results of health outcomes and resource utilization from the whole trial were combined with the unit costs at the country level. This practice is common for the economic evaluation of drugs when data from multinational trials that provide aggregated health outcomes and resource utilization are used to populate cost-effectiveness studies for specific countries (see for instance, Vlachaki et al. (2022), Viljoen et al. (2023) and Lueza et al (2024)).

Outcome data

Both trials also collected data on quality of life as perceived by patients. The visual analogue scale (VAS) and the EuroQol 5-Dimension 5-level (EQ-5D-5L) questionnaire were administered to patients at three timepoints in the Prudence trial (at day 1 and 14 by written questionnaire, and at day 14 and 28 by phone call, i.e. at day 14 there were two responses, one in a written document and the other obtained by a phone call). In the Adequate trial, it was at day 1, 14 and 30.
For the patients belonging to those countries that have validated utilities of the EQ-5D-5L their local values are applied, otherwise, the UK utilities were used. A “t” Student test was used to test for differences in the patients´ utilities at day 1 between the two arms of the

trial. Separated “t” Student tests for paired data were also applied to test for differences in each arm between day 1, day 14 and day 28 (or 30). The results are useful to understand the recovery profile within the trial framework. A “t” Student test is also used to test for differences between the utilities of days 14 and 28 in the two arms. As the trial designed was not powered to show differences in quality of life (rather its assumption was that there should not be any health outcome difference across the two arms), likely no statistically significant differences are expected. Nevertheless, Quality adjusted life years (QALYs) are obtained during the 28 days (or 30) period and used to estimate the cost per QALY in the aggregate analysis. As the recovery is assumed to be progressive, the utility gain adopts the trapezoidal shape; hence, QALY gains are the difference between the two trapezoidal areas corresponding to the two arms ((difference in utilities x times the duration(28/365) or (30/365) divided by two)). The whole analysis is included as an annex to this report.
The economic evaluation of the PRUDENCE trial calculated the incremental cost-effectiveness (ICER) ratios for the following variables:

- avoided antibiotic prescriptions,
- avoided days on antibiotics,
- working hours lost and number of school days lost,
- QALY.

The analysis calculated the additional costs of reducing antibiotic consumption derived from the use of point-of-care tests (that is directly related to the expected reduction of the future antimicrobial resistances, the goal of this project), as well as other potential advantages for society, namely, the reduction of missed work days and days out of school. The intention to treat approach was followed to run the models.
With regard to ADEQUATE, the final version of this trial only recruited paediatric patients (and about 100 adults whose data were not used for the economic analysis given the small sample size).

The primary health outcomes used in this trial were:

- Days on antibiotic treatment at day 14 and 30
- Days alive out of hospital at day 14 and 30

Consequently, we used these two end points for the health economic analysis. They were selected to show the short term implications of the introduction of the POCT in the decision on antibiotic prescription. Nevertheless, as the duration of the trial was 30 days, we also replicated the analysis for this longer horizon to capture in a more complete way the total antibiotic consumption.

If a patient had no information regarding the end day of the hospitalisation, it was assumed by the clinical team that the patient still remained hospitalised until day 14 or 30, depending on the time horizon considered in the analysis.

Despite having administered questionnaires of quality of life to the participants, targeted for each age category (we did use the EQ-5D-Youth version for children aged 8-15, and proxy version for younger children (age 4-8), no questionnaire for those below age 4 - as stipulated by EuroQoL)as initially planned, we have acknowledged that their validity is limited for this population (the questionnaires may be completed by parents or caregivers and the results

may not be truly representative of the child's quality of life). It is acknowledged that the methods for measuring and valuing health-related quality of life (HRQoL) and generating quality-adjusted life years (QALYs) in adults pose challenges when applied to children (Thorrington (2015)). Several key challenges have been identified:

- Cognitive complexity: Children may struggle to understand and respond to the abstract and complex questions of EQ-5D due to their level of cognitive development. (Devlin, Lovett and Rowen (2021)).
- Challenges in preference assessment: Preference assessment in children can pose additional ethical and practical challenges, especially in terms of understanding abstract concepts and the ability to make informed decisions (Devlin (2022)).
- Lack of relevant dimensions for children: EQ-5D may lack specific dimensions that are relevant to children's health and well-being, such as emotional development, social functioning, and participation in school and recreational activities.

Despite these challenges, we calculated the QALYs for the ADEQUATE trial, following a similar approach as in the PRUDENCE trial. The whole analysis is included as an annex to this report. However, we did not use the quality of life results and their derived indicators of quality adjusted life years (QALYs) in the economic evaluation of this trial (See also Petrou (2022)). In consequence, the ICER used days on antibiotic and days alive out of hospital during the 14 days follow-up period, that were the primary endpoints of this trial, as well as the same analysis for the 30 days follow-up period.

Short term economic model

The model structure classified patients in two arms (POCT and standard of care), then data from the surrogate endpoints of each participant related to the health outcomes were added up in each arm, calculated its per capita average according to the intention to treat final population, and subtracted at the denominator of the ICER. Should this difference became statistically significant, a cost-effectiveness analysis would be elaborated; otherwise, according to the economic evaluation methodology, a simple cost-analysis would be carried out.

A similar approach was followed for the costs. For each patient, we added up the monetary value of all used resources (number of physical units of each resource times the corresponding unit cost), then we obtained the grand totals for both arms and subtracted them in the numerator (again, on a per capita basis), as shown in the formula.

$$ICER = \frac{\sum cost\ A - \sum cost\ B}{\sum \text{Effectiveness A} - \sum \text{Effectiveness B}}$$

The numerator of the ICER is the difference of the cost of all the resources in each arm:

Cost A = sum of costs in the POCT arm

Cost B = sum of costs in the standard of care arm

The denominator of the ICER is the difference of the endpoints in each arm:

Effectiveness A= the sum of health outcomes in the POCT arm

Effectiveness B= the sum of the health outcomes in the standard of care arm

Base-case analysis

For the base case analysis, a healthcare system perspective was used which means that only health resources costs supported by the system or by patients (when purchasing out-of-pocket drugs) were included. Spain was considered to be the reference country for the unit cost analysis of the aggregated results; also the NICE unit cost data was applied to the aggregated results, so that the UK have its separated set of the ICER results. We followed the Spanish guidelines for the economic evaluation of drugs (https://www.sanidad.gob.es/areas/farmacia/comitesAdscritos/prestacionFarmaceutica/docs/20240227_CAPF_Guia_EE_definitiva.pdf).
No discounting of either costs or health outcomes was applied as the economic analysis was conducted using a 28-day follow-up of the trial participants in PRUDENCE. Therefore, time horizon for this analysis is 28 days. The base-case cost-effectiveness analysis used avoided antibiotic prescriptions, avoided days on antibiotics, working hours lost and number of school days lost as the main outcomes.
The PRUDENCE trial used two different tests (Afinion and Veritor). The ICER gave the aggregated results for the POCT compared to the standard of care as well as the separated results for each test compared to the standard of care. To balance the number of participants in the specific POCT arms a per patient average was used. However, the economic evaluation comparing the results of each test head-to-head was not presented as it was not the purpose of the project.
The ADEQUATE trial used the BIOFIRE test; the base-case analysis gave an ICER for the aggregated results for the POCT compared to the standard of care at days 14 and 30.

Complete case analysis

An additional scenario analysis was conducted using a societal perspective, which included additional costs (namely, productivity losses and other costs supported by patients during their sickness period, such as domestic help). This analysis supplements the healthcare system perspective for the decision-making process. To avoid double counting of some costs, the ICER was consequently adapted (for instance, when calculating the ICER for the lost working days, the indirect costs of the lost productivity are not included in the numerator of the ratio). This complete analysis was applied to PRUDENCE; data on productivity for ADEQUATE were only received at a late stage and too complex to include in the current analysis.

Sensitivity analysis

A deterministic sensitivity analysis was also performed for the unit costs of the resources. In particular, we considered a plus-minus 20% variation for each one. Results of this analysis are shown in a tornado diagram. For the other costs related to the management of the POCT, we considered a cost of zero (as if they were totally integrated in the medical visit costs).

Furthermore, as aforementioned, a country analysis was also carried out considering the different value of the unit costs from the reference country (Spain) by using the purchasing power parities of the World Bank. If patients stated out-of-pocket expenses expressed in

local currencies, they were converted to euros to make the calculations homogeneous. In these cases, purchasing power parities were not used for those specific items but just directly applied to the calculations of the ICER.
As the UK has a comprehensive unit cost database, a separate analysis was performed using reported UK costs instead of using the purchasing power parity data. Despite the large number of recruited patients in the UK, the results of the health outcomes are coming from the total sample (are not specific for the UK only), as the trial was not powered to provide statistical significant data from individual countries.
A probabilistic sensitivity analysis was also elaborated.

Identification, measurement and valuation of resources

Costs of health resources used by healthcare systems during the diagnosis and therapeutic process in both trials
Data relating to resource use collected as part of the PRUDENCE and ADEQUATE trials were classified into the following categories:

- point-of-care tests,
- antibiotics,
- other prescribed drugs (inhaled, antivirals, antihistaminic, paracetamol and other NSAIDS, anticough and other),
- other supplementary tests (SARS-CoV-2, total white blood cell count, chest X-ray, etc.),
- other resources such as additional GP and paediatrician visits, out-of-hour service, accident and emergency room visits, specialist visits, additional pharmacy visits, additional costs for children day care centre, additional costs for gifts, and additional costs for other help; there were also additional cost of adverse events that included hospital stays, hospital X-ray and admission to ICU; finally, data referred to hours of lost activity were also considered.

Unit costs were applied to each resource use at the individual level to calculate their total cost of resource use over the duration of each trial. Unit costs data from the NICE database were used as the main source for the UK (https://www.england.nhs.uk/national-cost-collection/#ncc1819). For Spain, official databases for hospitalisations (Registro de Actividad de Atención Especializada -RAE-CMBD. Conjunto Mínimo Básico de Datos-Hospitalización (CMBD-H), available at https://pestadistico.inteligenciadegestion.sanidad.gob.es/publicoSNS/S) together with official healthcare prices published in regional bulletins for healthcare costs, as well as the database of the General Council of Pharmacists for the list prices of drugs (available at https://botplusweb.farmaceuticos.com/) were used.
All data were updated to 2023 prices using the consumer price index, when needed. With regard to the unit cost of hospital stays, we used the cost of hospitalisations due to respiratory diseases.
The costs of prescribed antibiotics were calculated by family of antibiotics according to a weighted basket that considered the most prescribed formats; a similar approach is followed for the costs of other prescribed drugs (number of prescriptions) and other declared drugs (number of doses) that were calculated from the cost of the active principle

dispensed in the most frequent formats, (see the Annex II).

Cost of the prescribed antibiotics

The unit costs of the prescribed antibiotics for both trial are shown in table 1. As aforementioned, the costs were obtained from the Bot-Plus database of the Spanish General Council of Pharmacists (https://botplusweb.farmaceuticos.com/) and from the NICE database of drugs (https://bnf.nice.org.uk/drug/).

Table 1. Unit costs of the prescribed antibiotics for the PRUDENCE trial (per prescription)

	UNIT COSTS	
Prescribed Antibiotics	**UK (£)**	**ES (€)**
Tetracycline	6.22	5.11
Narrow spectrum penicillin	2.39	6.61
Broad-spectrum penicillin	2.05	3.93
Co-amoxiclav	2.98	8.03
Macrolide	7.79	6.69
Quinolone	9.38	10.48
Cephalosporin	12.07	12.59
Others	8.15	10.00
Other prescribed AB for current illness	9.0	8.5
Other prescribed AB out of the trial	5.0	6.0
Inhaled	12.56	17.48
Antivirals	10.0	10.4
Antihistamines	2.29	3.51
Paracetamol and other NSAID	0.45	1.60
Anticough	3.91	1.80
Other (Lozenges, mouth washes or gargles, nose spray, ear drops and vitamins)	9.03	10.00

Source: For Spain, the costs were obtained from the Bot-Plus database of the Spanish General Council of Pharmacists (https://botplusweb.farmaceuticos.com/) and for the UK, from the NICE database of drugs (https://bnf.nice.org.uk/drug/). For the category "Other prescribed AB for current illness", we summarised the data of the column "AB_CLASS_OTH_SP" of the file "prescribing" in PRUDENCE and then calculated the average cost that was 0.53 euros/dose or 8.5 euros/package (9 pounds for the UK). For the category "Other prescribed AB out of the trial", after reviewing the database, it came out that most of the antibiotics were amoxicillin and phosphomicine. The weighted average cost was 6 euro/package (and 5 pounds in the UK).

As the classification of antibiotics is not the same for both trials, we present the unit costs for the Adequate in table 2.

Table 2. Antibiotic costs for the ADEQUATE trial

Antibiotics	UNIT COSTS	
	UK (£)	ES (€)
Amoxicillin	1.37	5.71
Amoxicillin-clavulanate	2.09	5.84
Ampicillin	24.31	3.36
Ampicillin/Sulbactam	3.36	3.06
Azithromycin	1.46	5.62
Cefaclor	7.50	11.86
Cefotaxime	21	3.43
Ceftriaxone	9.58	7.9
Cefuroxime	9.25	4.84
Erythromycin	18.20	3.84
Penicilin V	6.02	5.89
Other	8.15	10

Source: https://bnf.nice.org.uk/drug/ (for UK) and
https://botplusweb.farmaceuticos.com/FichaMUH/298045 (for Spain)

Cost of other resources

The unit costs of other resources are shown in table 3.

Table 3. Unit costs of other resources

	UNIT COSTS		Source/justification	
Other resources	**UK (£)**	**ES (€)**	**UK (*)**	**ES**
Additional GP & Paediatrician visits	33.00	47.00		(1)
Out of hours service	86.00	94.00		(1)
Accident and Emergency visits	200.00	139.00		(2)
Specialist visits	155.00	137.00		(3)
Additional pharmacy visits	5.15	n.a.		
Hospital stays	827.12	681.00		(4)
Hospital X Ray	31.00	23.00		(5)
Admission to ICU	6,834.54	5,013.00		(6)
Paid work (patient)	24.50	16.80		(7)
Paid work (patient´s caregiver)	24.50	16.80		(7)
Dependent care (patient)	24.50	16.80		(7)
Caregiver care (patient´s caregiver)	24.50	16.80		(7)

Sources: (1) Orden SAN/35/2017, de 15 de diciembre, por la que se fijan las cuantías de los Precios Públicos de los Servicios Sanitarios prestados por el Servicio Cántabro de Salud; (2) Orden SAN/35/2017, de 15 de diciembre, por la que se fijan las cuantías de los Precios Públicos de los Servicios Sanitarios prestados por el Servicio Cántabro de Salud; (3) ORDEN 727/2017, de 7 de agosto, del Consejero de Sanidad, por la que se fijan los precios públicos por la prestación de los servicios y actividades de naturaleza sanitaria de la red de centros de la Comunidad de Madrid https://www.bocm.es/boletin/CM_Orden_BOCM/2017/08/21/BOCM-20170821-1.PDF, (4) cost per day. Portal Estadístico del Ministerio de Sanidad, (5) Orden SAN/35/2017, de 15 de diciembre, por la que se fijan las cuantías de los Precios Públicos de los Servicios Sanitarios prestados por el Servicio Cántabro de Salud; (6) cost per average stay (6,7 days), Portal Estadístico del Ministerio de

Sanidad. (7) Eurostat database.
(*) The source for UK data is https://www.england.nhs.uk/national-cost-collection/#ncc1819;
The costs of other prescribed medications for viral infections regarding the Adequate trial are shown in table 4.

Table 4. Antiviral medication costs

	Unit Costs	
	UK (£)	**Spain (€)**
Oseltamivir	15.41	31.56
Lopinavir	76.85	151.76
Zanamivir	n.a.	22.90
Remdesivir	340	531.35
Ribavirin	67.08	348.43
Hydroxychloriquine	9.5	12.16

Source- https://bnf.nice.org.uk/drug/(for UK) and
https://botplusweb.farmaceuticos.com/FichaMUH/298045 (for Spain).

Costs of point-of-care tests

Manufacturers of the point-of-care tests provided information on their costs. Besides the unit costs of the tests, additional costs related to their management (described below) were also considered.
In the case of the PRUDENCE trial, the two tests used were Veritor and Afinion. The cost of the Veritor analyzer was 300-350 USD and allowed to be used for about 10,000 tests, therefore, implying a negligible fixed cost per test. According to the information provided by the manufacturer, the cost per test for COVID assays is in the range of 8-12 USD, for flu and RSV assays, in the range of 7-11 USD and for Strep, 4-6 USD. Hence, an estimation of 8€ would be an appropriate value for the cost of this test.
The Afinion costs for some participant countries where the test was directly marketed by the firm are shown in table 5. According to the provided values, we calculated the mean value that was 7.1€ and 5.6 GBP for the base-case analysis.

Given that the range of costs of Veritor are less detailed but cover the point estimate of the mean value of Afinion, we used the final aforementioned costs of Afinion for both tests.

Table 5. Unit costs of AFINION

Country	CRP test price	Explanation/justification of price
UK	4.51-6.78 GBP/test	Range depending on business model /type of contract (leasing/buying)
France	5.47-6.53 €/test	Range depending on business model /type of contract (leasing/buying)
Germany	4.9- 7€/test	Reimbursement for CRP-POC in Germany with statutory health insurance payers and health physicians association. Includes rented/sold, quality control tests and taking of blood sample.
Belgium	5.10–5.50 €/test	Range depending on business model /type of contract (leasing/buying)
Italy	8.93-12.40 €/test	Range depending on business model /type of contract (leasing/buying)
Spain	5.27-10.96 €/test	Range depending on business model /type of contract (leasing/buying)

Source: Abbott, 2023

As for the additional costs of the resources needed to manage and use the test results by the healthcare system, several recent studies such as Larson (2012), Fawsitt (2022), Boere (2022), and Holmes (2018), among others, listed the main items whose cost should be accounted for; they referred to the energy and space where the equipment is located, the associated connectivity, training time for the personnel, internal and external quality control, maintenance, test failures, time of health care professionals to obtain the sample and to operate the equipment, as well as the time added to the standard consultation when a point-of-care test is performed. From these studies, we can conclude that the most important factors related to the introduction of point-of-care tests are the time of the healthcare personnel, namely, physicians, that need more time for their consultations, and other assistants, that operate the equipment and withdraw the samples. The additional personnel time cost estimated in these studies is in the range of 7.5-15 euros. The cost of the personnel in the study by Fawsitt referred to Ireland is rather high, about 5 euros per minute of physician´s time, what contributes with 15 euros per test assuming additional 3 minutes; however, in other jurisdictions the per minute cost is about 1-2 euros. The other costs supported by the healthcare system are lower, altogether adding-up to about 3 euros per test. Thus, we can conclude that including a lump-sum for the healthcare system supported costs of 10-12 euros per test would be a fair estimate for this concept.
The OECD published in 2023 a report on AMR in which it emphasized the convenience of using POCT and indicated the costs of managing them: "The per capita cost of the intervention is estimated to vary between USD PPP 0.53-2.15 across countries. The estimated costs take into account the cost of buying the POC CRP tests, costs related to training the prescribers on the clinical guidelines related to the use of POC CRP tests and informational materials. The total intervention cost also includes some administrative expenses and expenses covering monitoring and evaluation activities at the national and

local levels. No additional costs were included in these estimates to account for any additional time that prescribers may spend to perform the POC CRP tests because the tests are assumed to be carried out during the standard period that the prescriber spends with the patient to make a diagnosis". We observe that this latter calculation omits some items accounted by other authors (energy, space, connectivity, etc.); as it is an issue subject to different approaches and as these calculations are done on a per capita basis cannot be directly applied to our analysis. However, we used an almost zero cost to show the results of an alternative scenario, that assumes a negligible impact of the other costs in the healthcare system, as it could be the case in some jurisdictions.

As a summary, table 6 states the costs of the point-of-care tests used in the economic analysis.

Table 6. Unit costs of point-of-care tests

	Unit costs		**Source**	
Tests	**UK (£)**	**ES (€)**	**UK**	**ES**
Point-of-care test	5.6	7.1	Abbott and BD	Abbott and BD
Other costs of test	0.4-12.00	0.5-12.00	Larson (2012), Fawsitt (2022), Boere (2022), and Holmes (2018), OECD(2023)	Larson (2012), Fawsitt (2022), Boere (2022), and Holmes (2018), OECD(2023)

In the case of the ADEQUATE trial, the test used was BIOFIRE, whose unit costs per test for each of the countries, as well as the average, are shown in table 7. To impute the costs of the test device to individual test, the manufacturer provided this formula:

Cost of the device per individual test = device list price per country/shelf-life years/mean number of tests in average per year per country

For both panels the manufacturer assumed a **placement model** for a **TORCH2** system, a duration of 5 **years** and an average consumption o**f 1000 tests per year**.

Table 7. Unit costs of BIOFIRE

Country	**RP2.1+ (in EUR)**	**PN+ (in EUR)**
UK	115.4	188.3
Spain	90.0	150.0
Belgium	112.7	142.2
Germany	93.0	147.0
France	99.2	138.5
Georgia	210.0	230.0
Greece	95.0	150.0
Italy	100.0	160.0
Israel	85.0	140.0

Portugal	90.0	150.0
Ireland	116.0	173.0
Hungary	93.0	133.0
Average	**108.3**	**158.5**

Source: Biomerieux, 2023

Please note that for the ADEQUATE study which was finally only addressing children , the BIOFIRE panel used was the RP2.1+

Costs of additional tests

In both trials, additional tests other than the POCT were ordered for some patients. Table 8 states these tests as well as their unit costs, together with the sources where they come from.

Table 8. Cost of additional tests

	Unit costs		Source	
Tests	**UK (£)**	**ES (€)**	**UK**	**ES**
SARS-CoV-2 test	4.00	2.94	(1)	(3)
Total white blood cell count test	9.00	23.81	(1)	(4)
Chest X-ray test	31.00	23.00	(1)	(5)
Other tests	18.00	26.05	(2)	(6)

(1) https://www.england.nhs.uk/national-cost-collection/#ncc1819;
(2) https://www.leedsth.nhs.uk/assets/71432c14fa/NIHR-2020-Investigation-and-Intervention-Tariff-1-v2.2-1.pdf#;
(3) https://www.boe.es/diario_boe/txt.php?id=BOE-A-2022-560;
(4) https://boc.cantabria.es/boces/verAnuncioAction.do?idAnuBlob=320839;
(5) BOCM»núm.198, de 21 de agosto de 2017;
(6) Other test: calculated as the weighted average cost of the "TESTS_OTHER_SPEC" column in the "addtest_20230301" sheet. Source: several Spanish official bulletins.

Finally, there were also some tests ordered in the emergency care unit whose costs are stated in table 9.

Table 9. Unit costs of tests ordered in the Emergency Department in the ADEQUATE trial

Unit cost	UK (pounds)	Spain (euros)
Haematology and biochemistry	9	33.18
Microbiology –Respiratory	18	21
Microbiology – Urine	8	32
Microbiology – Feces/Rectal Swab	18	10
Microbiology – Blood Culture	18	20.33
Microbiology – Serology (repeating form)	12	20

Source: For the UK, https://www.england.nhs.uk/national-cost-collection/#ncc1819 and https://www.leedsth.nhs.uk/assets/71432c14fa/NIHR-2020-Investigation-and-Intervention-Tariff-1-v2.2-1.pdf#. For Spain, the source is https://www.bocm.es/boletin/CM_Orden_BOCM/2017/08/21/BOCM-20170821-1.PDF,

RESULTS

Prudence trial

Clinical results from the PRUDENCE trial showed no statistical difference between the two arms in terms of the primary endpoint: antibiotic prescribing during 28 days after the index consultation. Regarding the other health outcomes considered for the economic evaluation (avoided days on antibiotics, working hours lost, and number of school days lost), as shown in table 10, no statistical differences were found at the 5% significance level. Furthermore, the numerical difference in a per capita basis is small for the endpoints (e.g. just 0.04% for the number of days on antibiotics). Hence, it is not advisable to calculate the cost-effectiveness ratios for each one of the endpoints.

Table 10. Statistical analysis of the difference in outcomes used for the economic evaluation

	POCT (%)	***Standard of care (%)***	**Total**	**Difference (POCT-Standard)**	**P-value**
Number of patients (1)	1,448	*1,191*	2,639		
Antibiotic prescription (2)	662 (45.70)	*561 (47.10)*	1,223	-1.40	0.330
No antibiotic prescription	786 (54.30)	*629 (52.90)*	1,415	1.40	0.164
Declared days on antibiotics per capita	3,349/1,448 = 2.31	*2,810/1,191 = 2.35*	6,159/2,639 = 2.33	-0.04	0.352
Working hours lost per capita	11,593.50/1,448 = 8.00	*10,150.15/1,198 = 8.52*	21,743.65/2,639 = 8.23	-0.52	0,264
Number of school days lost per capita	1,876.50/1,448 = 1.29	*1,409/1,191 = 1.18*	3,288/2,639 = 1.24	0.11	0,321

(1) It corresponds to the intention to treat results
(2) 1 patient is missing

The number of days on antibiotics are approximately 2.3 on a per capita basis or 5.0 per prescription in both arms. On average, 8 hours of work were lost by the disease process and 1.2 days of school for the participants in the trial. The dependent care hours lost by the disease process were about 0.5.
Furthermore, as it can be observed in Annex III, where the detailed QALY calculation is presented, there was not a significant difference in the quality of life and therefore in the

QALYs gained in the POCT arm versus the standard of care (the numerical difference slightly favours the standard of care arm where QALY gains were greater). Given that a full QALY equals one year of life spent in perfect health, the QALY difference at 30 days observed here would represent a difference of around one hour (on a lifetime) spent in perfect health, which can be regarded as clinically irrelevant. Table 11 summarizes this finding. Thus, a cost-utility analysis that would consider the QALY as the outcome measure is not advised either. As a consequence of these health outcomes, the economic evaluation study adopted the form of a cost analysis, that measured the cost of health care resources as well as other costs supported by society in each arm.

Table 11. Differences between usual care group and diagnostic intervention group at day 1, day 14 and day 28.

		Usual care group	**Diagnostic intervention group**	**Difference** (1)	
Day 1 (4)	**Utility**	0.80700	0.80550	-0.001493	
	N	817	1,033		
		Usual care group	**Diagnostic intervention group**	**Difference** (1)	**QALYs** (2)
Day 14 (4)	**Utility**	0.92854	0.92251	-0.00602	-0.0001440
	N	718	934		
		Usual care group	**Diagnostic intervention group**	**Difference** (1)	**QALYs** (3)
Day 28 (4)	**Utility**	0.93740	0.92923	-0.00817	-0.0001852
	N	905	1,107		

N= sample size

(1) Difference = Diagnostic intervention group utility – Usual care group utility

(2) QALYs calculate as $\frac{(-0.00602+(-0.001493))\ x\left(\frac{14}{365}\right)}{2}$

(3) QALYs calculate as $\frac{(-0.00817+(-0.001493))x\left(\frac{28}{365}\right)}{2}$

(4) The variances of samples are not equal: the two-sample t test is performed assuming unequal variances

Results of cost for each arm are presented in table 12, that summarizes the resource use in physical units as well as in monetary terms and in an aggregated basis. It is interesting to remark that the most important factors in terms of costs of the POCT arm are hospitalisations (37%), Visits and other medical resources (33%), and POCT (14%), being antibiotics 2% of total costs (obtained from the cost calculations in euros). Health care costs represent 46% of the total costs in the POCT arm. Productivity losses are 94% of total indirect costs.

Table 12. Health outcomes and cost analysis

	Standard		POCT		Afinion PCR		Veritor	
	1,191		**1,448**		**726**		**722**	
Prescribed Antibiotics	561		662		323		339	
Declared days of consumption of prescribed AB	2,810		3,349		1,586		1,763	
Lost work hours	10,150		11,594		5,882		5,712	
Lost number of school days	1,409		1,877		633		1,244	
	UK (£)	**ES (€)**	**UK (£)**	**ES (€)**	**UK (£)**	**ES (€)**	**UK (£)**	**ES (€)**
TOTAL COSTS	**402,889**	**315,036**	**480,547**	**382,210**	**248,710**	**198,170**	**231,836**	**184,040**
DIRECT COSTS	139,094	134,148	179,919	176,066	96,675	93,918	83,244	82,148
Costs per test	--	--	22,589	24,761	11,326	12,415	11,263	12,346
Antibiotic consumption	2,969	3,899	3,514	4,621	1,889	2,269	1,625	2,352
Consumption of other medications	6,137	7,583	8,159	9,870	4,868	5,826	3,292	4,044
Costs for diagnostic tests	8,492	7,969	10,359	9,313	6,631	5,818	3,729	3,494
Hospitalization costs	78,121	64,290	81,472	65,785	44,582	36,049	36,890	29,736
Costs for medical visits and others	41,564	48,797	51,582	59,722	25,945	30,264	25,637	29,458
Other additional costs for help and care	1,811	1,610	2,244	1,995	1,436	1,276	809	719
INDIRECT COSTS	263,795	180,888	300,627	206,144	152,035	104,252	148,593	101,892
Lost work productivity	248,679	170,523	284,041	194,771	144,097	98,809	139,944	95,962
Other indirect costs	15,117	10,366	16,587	11,374	7,938	5,443	8,649	5,930

Table 13 calculates the monetary value of those resources in per capita terms (as the arms contain a different number of patients); final costs are classified as healthcare costs and

indirect costs, namely, productivity related and other costs. Health care related costs are 121€ in the POCT, just about 9 euros more than in the standard of care arm. Should the cost of the management by health care professionals of the POCT were considered zero, then the cost of the standard of care cost would be less than one euro greater than in the POCT arm. If we consider also the indirect costs, then 264€ is the total per capita costs in both the POCT and standard of care arm.

Table 13. Per capita costs of direct and indirect costs.

	Standard		POCT		Afinion PCR		Veritor	
	UK (£)	**ES (€)**	**UK (£)**	**ES (€)**	**UK (£)**	**ES (€)**	**UK (£)**	**ES (€)**
TOTAL COSTS	**338.28**	**264.51**	**331.87**	**263.96**	**342.58**	**272.96**	**321.10**	**254.90**
DIRECT COSTS	**116.79**	**112.63**	**124.25**	**121.59**	**133.16**	**129.36**	**115.30**	**113.78**
Costs per test	--	--	15.60	17.10	15.60	17.10	15.60	17.10
Antibiotic consumption	2.49	3.27	2.43	3.19	2.60	3.13	2.25	3.26
Consumption of other medications	5.15	6.37	5.63	6.82	6.70	8.03	4.56	5.60
Costs for diagnostic tests	7.13	6.69	7.15	6.43	9.13	8.01	5.16	4.84
Hospitalization costs	65.59	53.98	56.27	45.43	61.41	49.65	51.09	41.19
Costs for medical visits and others	34.90	40.97	35.62	41.24	35.74	41.69	35.51	40.80
Other additional costs for help and care	1.52	1.35	1.55	1.38	1.98	1.76	1.12	1.00
INDIRECT COSTS	**221.49**	**151.88**	**207.62**	**142.36**	**209.41**	**143.60**	**205.81**	**141.12**
Lost work productivity	208.80	143.18	196.16	134.51	198.48	136.10	193.83	132.91
Other indirect costs	12.69	8.70	11.45	7.85	10.93	7.50	11.98	8.21

As it can be observed, the differences in costs between Veritor and Afinion arms are mainly driven by the differences of the hospitalisations in each arm. There is no clinical argument though to assume that these differences in hospitalizations were caused by the POCT strategy, since treatment decision were not significantly impacted by the diagnostic tests and hospital admission was quite a rare event in the study population.

Table 14 shows the per capita costs for each participant country in its local currency (except the UK whose data are already shown in sterling pounds), obtained through the application of the purchase power parity, that provides a more meaningful measure of the costs in terms of the price level and wealth of each jurisdiction. Spain is considered the benchmark for this comparison as it is one of the countries where unit cost data came

from. In this way, a wealthier country as Germany shows a total cost per patient of 317€ compared to the cost of 264€ in Spain. In other words, the societal effort involved in this type of disease process is perceived as 317€ in Germany, while exactly the same effort is valued as of 264€ in Spain, given the differences in their acquisition power. These results may be useful to understand the costs involved in each arm of the test, derived from the disease process in each country.

Table 14. Per capita costs by participant country according to the purchase power parity conversion.

Country	Unit	Purchasing Power Parities for GDP (1)	PPA_final € 2022 (US$/€)	Type costs (2)	Standard	POCT	Afinion PCR	Veritor
Belgium	Euro	0.694643	*1.200*	TC	**317.45**	**316.78**	**327.59**	**305.92**
				DC	135.18	145.93	155.25	136.55
				IC	182.27	170.86	172.34	169.37
Denmark	Danish Krone	6.153561	*10.631*	TC	**2,812.15**	**2,806.24**	**2,901.97**	**2,709.98**
				DC	1,197.46	1,292.70	1,375.31	1,209.63
				IC	1,614.69	1,513.54	1,526.65	1,500.35
France	Euro	0.674142	*1.165*	TC	**308.08**	**307.43**	**317.92**	**296.89**
				DC	131.19	141.62	150.67	132.52
				IC	176.89	165.81	167.25	164.37
Germany	Euro	0.694484	*1.200*	TC	**317.38**	**316.71**	**327.51**	**305.85**
				DC	135.14	145.89	155.22	136.52
				IC	182.23	170.82	172.30	169.33
Greece	Euro	0.509116	*0.880*	TC	**232.66**	**232.17**	**240.09**	**224.21**
				DC	99.07	106.95	113.79	100.08
				IC	133.59	125.22	126.31	124.13
Hungary	Forint	156.944391	*271.150*	TC	**71,722.91**	**71,572.09**	**74,013.63**	**69,117.04**
				DC	30,540.85	32,969.83	35,076.92	30,851.07
				IC	41,182.06	38,602.26	38,936.71	38,265.96
Ireland	Euro	0.737879	*1.275*	TC	**337.21**	**336.50**	**347.98**	**324.96**
				DC	143.59	155.01	164.92	145.05
				IC	193.62	181.49	183.06	179.91
Israel	New Israeli Sheqel	3.539881	*6.116*	TC	**1,617.71**	**1,614.31**	**1,669.38**	**1,558.93**
				DC	688.85	743.63	791.16	695.85
				IC	928.86	870.67	878.22	863.09
Italy	Euro	0.595658	*1.029*	TC	**272.21**	**271.64**	**280.91**	**262.32**
				DC	115.91	125.13	133.13	117.09
				IC	156.30	146.51	147.78	145.23
Poland	Zloty	1.787327	*3.088*	TC	**816.80**	**815.08**	**842.89**	**787.12**
				DC	347.81	375.47	399.47	351.34
				IC	468.99	439.61	443.42	435.78
Portugal	Euro	0.523337	*0.904*	TC	**239.16**	**238.66**	**246.80**	**230.47**
				DC	101.84	109.94	116.97	102.87
				IC	137.32	128.72	129.84	127.60
Spain	Euro	0.578810	*1.000*	TC	**264.51**	**263.96**	**272.96**	**254.90**
				DC	112.63	121.59	129.36	113.78
				IC	151.88	142.36	143.60	141.12
Georgia	Lari	0.964973	*1.667*	TC	**440.99**	**440.06**	**455.07**	**424.97**
				DC	187.78	202.72	215.67	189.69
				IC	253.21	237.35	239.40	235.28

(1) National currency per US dollar (Year 2022) TC = total cost; DC = direct cost; IC = indirect cost

Adequate trial

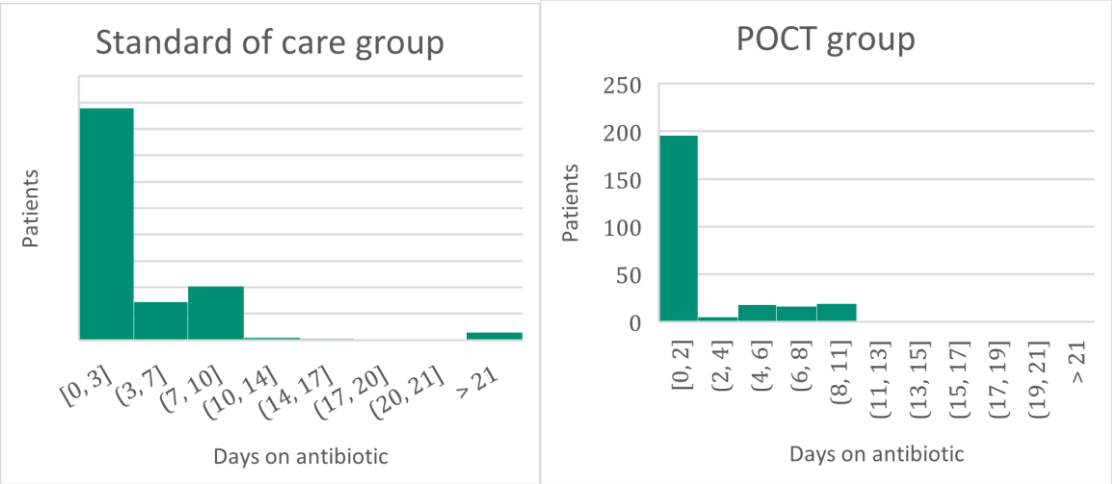
A total of 524 patients were recruited, although due to eligibility criteria data from 14 participants were disregarded for the final clinical and health economics analysis. Clinical results from the trial showed a skewed profile for the primary endpoint of Days on antibiotic treatment at day 14 (Table 15 and figure 1). The majority of patients did not consume antibiotics (medians equal to zero for days on antibiotic at day 14, and a short interquartile range (IQR)). In this situation, the standard t-Student based tests for the differences between the two arms are not recommended (the mean is not representative as about 190 patients did not consume any antibiotic). Interestingly, 6 patients in the control arm were responsible for an antibiotic consumption outside the first 14 days accounting for a total of 115 days; in the POCT arm, one patient exceeded during 24 days the antibiotic consumption outside the first 14 days. For the variable days alive out of hospital at day 14, the situation is different as most of patients were not admitted to the hospital and therefore they registered 14 days as the median and IQR. The ICER calculation is to be applied to the variable days on antibiotic treatment, accounting up to the 14th day and also up to the 30th day (as for the duration of the trial). The same is applicable to the variable days alive out of hospital at day 14, despite the small difference between the two arms (48 days of a total nearing 7400 days in each arm for the 30 days follow-up, or a difference of 0.25%, being larger in the POCT arm).

Table 15. Clinical results of the Adequate trial

Primary endpoint	**Standard of care**	**Intervention**	**Difference in mean log days (95% CI)**	**p-value from the t-test of the difference between log days**
N (randomised)	262	262	-	-
Lost to follow-up (baseline)	1	1	-	-
Lost to follow-up (day 14)	6	6	-	-
N (Full Analysis Set)	255	255	-	-
EP1: Days on antibiotic treatment at day 14 (median, IQR)	0 [0, 5]	0 [0, 1]	ne	n/a
Total days on antibiotic at day 14	637	419		
Total days on antibiotic at day 30	715	447		
EP2: Days alive out of hospital at day 14 (median, IQR)	14 [14, 14]	14 [14, 14]	ne	n/a
Total days alive out of hospital at day 14	3,368	3,401		
Total days alive out of hospital at day 30	7,415	7,463		

ne: not estimated; n/a: not available

Figure 1. Histogram of the days on antibiotic therapy by arm.



Again, as it can be observed in Annex 2, where the detailed QALY calculation is presented, there was not a significant difference in the quality of life and therefore in the QALYs gained in the POCT arm versus the standard of care (the numerical difference slightly favours the POCT arm where QALY gains were greater). The number of patients included in the quality of life analysis is different from the total number in the trial due to the fact that many patients had not responded to the EQ-5D questionnaire, especially at the later time points. Table 16 summarizes this finding. Thus, a cost-utility analysis that would consider the QALY as the outcome measure is not advised either. We also remarked the methodological concerns of this type of analysis, when applied to paediatric population, in the Methods section above. A full analysis of the quality of life is included in Annex IV.

Table 16. Differences between standard of care group and diagnostic intervention group at baseline, day 14 and day 30.

		Standard of care group	Diagnostic intervention group	Difference (1)	QALYs (2)
***Baseline*(3)**	Utility	0.80830	0.83010	+0.02179	
	N	179	170		
***Day 14*(3)**	Utility	0.96965	0.98460	+0.01494	+0.00150
	N	41	38		
***Day 30*(3)**	Utility	0.98145	0.99376	+0.01231	+0.00140
	N	33	34		

N= sample size

(1) Difference = Diagnostic intervention group utility – Control group utility

(2) QALYs calculate as $\frac{Difference\ of\ utilities\ at\ day\ 14\ or\ 30 + difference\ of\ utilities\ at\ baseline\ x\left(\frac{30}{365}\right)}{2}$

(3) The variances of samples are not equal: the two-sample t test is performed assuming unequal variances

As shown in table 17, total costs were 154,595€ and 151,657€ in the standard and POCT arm, respectively, for the 14 days horizon. Hospitalization costs represented the highest category of costs, 85% for the standard and 76% for the POCT arm. POCT contributed to 19.89% of total costs. The standard of care arm happened to have registered ICU patients that stayed in that unit for 8 days and cost about 8.5% of total; however, no patient was

admitted at the ICU in the POCT arm. Aligned to these costs are those ones calculated in sterling pounds. This study does not include the indirect costs as they were received at a late stage of the project. It is interesting to remark that the cost of antibiotic consumption (on an outpatient basis) and the costs for diagnostic tests had just a single cost result as the cost of the consumption is measured on a per package basis (that takes place at the beginning of the period, hence, costs are applied to the first 14 days) and tests at the emergency room are also ordered at the beginning of the period (other tests costs are included in the cost per hospital stay, when ordered to inpatients).

Table 17. Cost analysis for the 14 and 30 days horizon.

RESULTS	**Standard group**		**BioFire**	
SAMPLE	255		255	
Days alive out of Hospital (14days/30days)	3,368 / 7,415		3,401 / 7,463	
Days on Antibiotic Therapy (14days/30days)	637 / 715		419 / 447	
	UK (£)	**ES (€)**	**UK (£)**	**ES (€)**
TOTAL COSTS (14days/30days)	183,741 / 211,036	154,097 / 176,570	175,897 / 190,785	151,657 / 163,915
COSTS PER TEST	--	--	32.487	30.166
Antibiotic consumption	119,11	293.26	61,09	210.43
Costs for diagnostic tests	4.936	8.322	3.566	6.191
Ordinary hospitalization costs(14days/30days)	160,461 / 187,756	132,114 / 154,587	139,783 / 154,671	115,089 / 127,347
ICU costs	18,225	13,368	0	0

Given these results, for the 14 days period, as the number of days on antibiotic was smaller in the POCT than in the standard of care (218 days less) and costs also smaller (20,251€), the ICER, although calculated, must be interpreted as the corresponding to a dominant option; that is to say, the POCT is both a cost and antibiotic consumption saving option. Similar results hold for the 30 days period. In the case of the co-primary endpoint of days alive out of hospital, again, as the POCT is a cost saving option and patients remained more days alive out of hospital, it is a dominant strategy. Results are also maintained for the 30 days period. As it could have been anticipated, the longer the duration of the follow-up the more likely total costs would be higher in both arms; however, as in the standard of care were admitted more patients to the hospital, it made the difference of costs between the two arms even higher for the 30 days follow-up period, as the fixed costs of the POCT are distributed, somehow, in that longer horizon, and also because the bulk of the other medical costs took place in the immediate period after randomization.
Again, as in PRUDENCE, the uncertainty surrounding these results may be driven by the variability in the databases where unit costs were obtained. Nevertheless, this uncertainty is common to both arms and therefore no biases are expected in the final results.

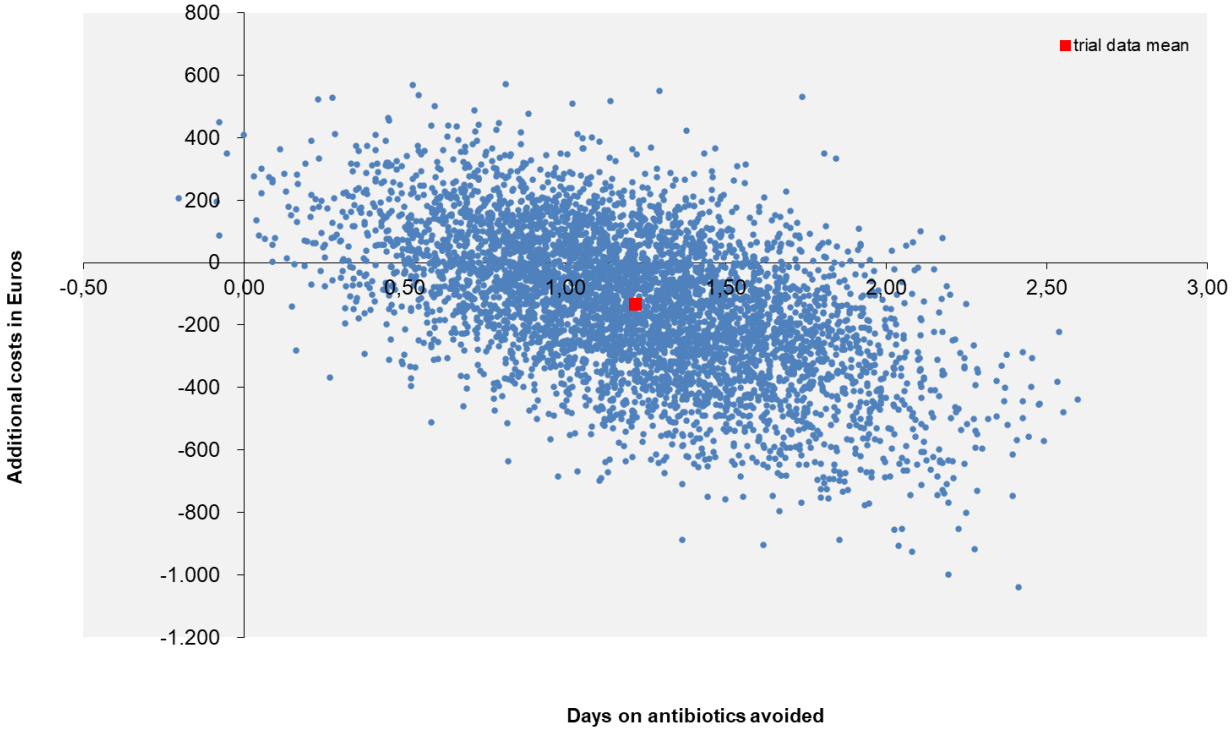
Table 18. Incremental cost-effectiveness results for the 14 and 30 days analysis.

	Biofire-Standard	
	UK	**ES**
Differential Total costs(14/30 days)	-7,844 / -20,251	-2,440 / -12,655
Differential results Days alive out of Hospital (14/30 days)	33 / 48	33 / 48
ICER with Total costs	**-238/ -422**	**-74/ -263**
Differential results Days on Antibiotic Therapy (14/30 days)	-218 / -268	-218 / -268
ICER with Total costs	**36 / 75**	**11 / 47**

A sensitivity analysis assessed the influence of different variables in the ICER, when costs varied in a plus minus 20% range; it showed that most influential variables are POCT (both its own cost and the cost of managing it), ICU and hospitalisation costs. In the case of hospitalisations, should their cost were 20% less, then the POCT was not a dominant strategy as the cost of the test offset the savings of the now less costly hospitalizations. The other variables scarcely contributed to changes in the ICER given that not all patients consumed those medical resources and because they were much cheaper; the POCT is administered to all patients in the specific arm of the trial and, consequently, it influences the final result. However, even though the POCT cost 20% more, this intervention still would show a cost saving result (the cost of other resources remaining unchanged). Interestingly, when hospitalization costs were 20% higher, due to the distribution of patients between the arms (more inpatients at the standard of care), total savings, as a difference of costs between the two arms, would become even greater. The sensitivity analysis results hold for both the 14 and 30 days period.

The bootstrap analysis (5000 replications) is also presented in figure 2; at the 30 days follow-up, when the POCT became a dominant strategy, lower costs and fewer days on antibiotic consumption per patient are obtained and represented. The positive effects of the X-axis refer to the avoided days on antibiotic consumption; the Y-axis reflects the cost-savings as they are in the negative quadrant. The 5000 replications are predominantly located in the west quadrants, which implies that there is substantial certainty that the POCT strategy will result in less days on antibiotic therapy; there is less certainty however on whether the POCT is also cost-saving given that around 30% of replications are in the north-east quadrant where POCT is more expensive than care as usual.

Figure 2. Bootstrap analysis for the variable per capita avoided average days on Antibiotic Therapy at day 30 (unit costs from Spain).



DISCUSSION

Results from the PRUDENCE trial showed no statistical differences between the two arms. A further analysis of the prescriptions, as shown in table 19, indicate that some prescriptions, namely, 177 (46+81+50) were issued despite the test indicated a negative result (GAS and CRP), and 9 prescriptions when the influenza test was positive (6+3). In total there were 186 prescriptions (177+9) that were issued according to other clinical criteria, different from the test result. This factor may be one reason explaining the lack of difference between the arms that led to apply a cost-analysis for the economic evaluation of this project. This observation and other reasons for absence of a clear effect are also discussed in the reports of the clinical trials.

Table 19. Prescriptions according to the results of the POCT

	POSITIVE TEST RESULT		**NEGATIVE TEST RESULT**	
	ANTIBIOTIC PRESC		**ANTIBIOTIC PRESC**	
	YES	**NO**	**YES**	**NO**
GAS test				
Both measurements of Veritor (seen as one trial arm)	33	0	46	118
Gas measurement of Veritor (Flu not needed outside flu season)	60	1	81	117
TOTAL	**99%**	**1%**	**35%**	**65%**
INFLUENZA test				
Both measurements of Veritor (seen as one trial arm)	3	15	74	104
Flu measurement of Veritor (Gas not needed for cough)	6	21	96	116
TOTAL	**20%**	**80%**	**44%**	**56%**
CRP test				
CRP test <5 mg/L			50	213
CRP test >200 mg/L	3	1		
CRP test 5.0 to 200 mg/L	250	153		
TOTAL	**62%**	**38%**	**19%**	**81%**

An internal analysis of the database has also been made, so that some subgroups of patients could be identified, mainly, those showing higher differences between the two arms in terms of the antibiotic prescription. It is a non-prespecified subgroup analysis and, consequently, its results must be taken just as exploratory or useful to design a new research.

Specifically, when classifying patients by their symptom severity, we found that the subgroup of moderate patients seems to have a larger difference in the number of prescriptions, as it can be observed in table 20. For instance, in this case, when we applied the per capita costs obtained from the general study (table 13, costs in euros) to a difference of 0.2, in round figures (column of "mean_ab_prescib"), of the proportion of prescriptions between the two arms across the subgroups also defined by other

characteristics (cough, influence season and Covid status), the ICER results in (121-112)/0.2= 50€ per avoided antibiotic prescription. In other words, testing patients becomes an attractive initiative in some cases, so that antibiotic prescription could be reduced at an affordable cost.

Table 20. Subgroup analysis: antibiotic prescription in the moderate subgroup of patients

	covid_status	influenza_season	cough	arm	symptom_severity	mean_ab_prescrib	alpha_ab_prescrib	beta_ab_prescrib
1	FALSE	FALSE	FALSE	base	moderate	0.7731092	92	27
2	FALSE	FALSE	FALSE	veritor_gas	moderate	0.6739130	93	45
3	FALSE	FALSE	TRUE	base	moderate	0.6525822	139	74
4	FALSE	FALSE	TRUE	crp	moderate	0.6157407	133	83
5	FALSE	TRUE	FALSE	base	moderate	0.7971014	55	14
6	FALSE	TRUE	FALSE	veritor_gasinf	moderate	0.6219512	51	31
7	FALSE	TRUE	FALSE	veritor_inf	moderate	0.0000000	0	1
8	FALSE	TRUE	TRUE	base	moderate	0.6000000	72	48
9	FALSE	TRUE	TRUE	crp	moderate	0.6696429	75	37
10	FALSE	TRUE	TRUE	veritor_gasinf	moderate	1.0000000	1	0
11	FALSE	TRUE	TRUE	veritor_inf	moderate	0.6330275	69	40
12	TRUE	TRUE	TRUE	base	moderate	0.3333333	10	20
13	TRUE	TRUE	TRUE	crp	moderate	0.3870968	12	19

As part of this project, a Point Prevalence Audit Survey (PPAS) was carried out during the period 2019-2020 by the WP4 Prudence team. This PPAS showed that when physicians prescribed antibiotics in primary care patients for similar conditions, they stated that were “sure or very sure” in about 90% of the cases. In other words, physicians doubted only in about 10% of the cases when adopting this sort of decision.

The ADEQUATE trial showed quite skewed results for the two primary endpoints, making difficult to stablish without uncertainties the statistical difference between the two arms. We have elaborated a cost analysis for the two arms and a cost-effectiveness analysis for the variable Days on antibiotic until both day 14 and 30 based on the numerical differences between the two arms, without stablishing that there was a significant statistical difference. The results indicated that the cost of avoiding a day on antibiotics by using the POCT are smaller than in the standard of care arm, being this result mainly sensitive to the cost of the POCT, ICU and hospitalisations. It is remarkable that when the follow-up period is 30 days, savings are even greater. POCT came out to be a dominant strategy in the sense that it cost less and generated fewer hospitalizations and lower antibiotic consumption. However, these results must be observed with some concerns as few patients of the sample contributed to the antibiotic consumption and even fewer to the ICU and hospitalisations costs. In other words, results are highly conditioned by random effects affecting those variables. For instance, an additional patient who would remain hospitalised until the end of the trial would contribute with about 20,000€ or 25,000 GBP to the cost of the arm where she was randomized, hence, modifying significantly the final economic results in favour or against the POCT. Assuming, for example, that an additional patient of this characteristic were hospitalised in the POCT, this strategy would not be longer dominant. Again, as aforementioned for the PRUDENCE trial, there is no clinical argument to assume that these differences in hospitalizations were caused by the POCT strategy, since treatment decision were not significantly impacted by the diagnostic tests and hospital admission was not a common event in the study population.

We have observed in the database that some hospitals showed different antibiotic prescription rates, independently of the analysed arm. In other words, the internal protocols for the diagnosis and treatments of each centre may have implications in both the POCT and standard of care arm results. Their influence is highly dependent on the number of recruited patients and the prescription of antibiotics rate for the analysed condition in that centre. The results of the ICER corresponded to aggregated results of all the participant centres; hence, the efficiency of the POCT refers to a sort of average hospital and may need local adaptations to better understand the benefits of the new diagnostic tool, both in terms of antibiotic prescription and hospitalisation costs. That is to say, the generalisation of the ICER results in not straightforward and needs additional data from different countries where variations in the protocols of use of antibiotics for this condition may be found. Some further insights on this matter can be found in D5.7 on transferability. Furthermore, as clinicians did not receive in the trial protocol any recommendation on how to use the test results to decide upon the potential prescription of antibiotic, we did not perform a deeper analysis to check if there was a sort of over-prescription in the POCT arm as we did in the PRUDENCE trial. We must remark the costs used in the ICER calculation corresponded to two countries and are average costs; hence, again, local adaptations on this ground will also be needed for the adoption of decisions about the implementation of the POCT across jurisdictions.

Finally, we acknowledge that other potential indirect outcomes of the POCT related to caregivers time and their productivity have not been included in this analysis, as we maintained it more restricted to the healthcare system area.

No relevant deviations from the Description of the Action or contingency plans took place in WP5.3

CONCLUSION

The health economic evaluation of the PRUDENCE trial adopted the form of a cost analysis, given that the primary health outcome in terms of antibiotic consumption was not statistically different between the two arms. Per capita costs of the POCT arm were 121€, just about 9 € more than in the standard of care arm. A detailed cost analysis using the purchase power parity is presented for each participant country. Quality of life results between the two arms were not statistically different either. Interestingly, the analysis of the database provided some insights of the utility changes between the day of randomization (when patients went to the medical consultation due to their disease process) and the day 14 and 28 of the follow-up period. According to the EQ-5D results, patients had a utility of about 0.8 at day 1, and 0.93 at day 14 and 30. It is the first time that the results of the quality of life of patients seeking medical care for Community Acquired Acute Respiratory Tract Infection [CA-ARTI] in Europe had been obtained.

The economic analysis of the ADEQUATE trial adopted the form of a cost-effectiveness study despite the results of the two arms did not show statistically significant differences. However, numerical differences indicated that the POCT arm had a lower number of days on antibiotic and a greater number of days alive out of the hospital. As medical costs in

the POCT were lower than in the standard of care arm, the incremental cost-effectiveness ratio showed that the POCT was a dominant strategy (cost less and generated better health outcomes measured as fewer days on antibiotic and more days alive out of hospital). Costs in the two arms are highly dependent on the number of hospital and ICU stays, that were higher in the standard of care arm, as well as on the cost of the POCT. As few patients received antibiotic therapy and even fewer patients were hospitalized, the final results in terms of the efficiency of the POCT are dependent on these parameters subject to variability across hospitals.

REFERENCES:

1. «BOE» núm. 293. de 4 de diciembre de 2014. páginas 99932 a 99959 (28 págs.) Orden DEF/2277/2014. de 28 de noviembre. por la que se establecen los precios públicos por la prestación de servicios y actividades de naturaleza sanitaria en el ámbito del Ministerio de Defensa. https://www.boe.es/buscar/doc.php?id=BOE-A-2014-12614

2. «BOE» núm. 308. de 24 de diciembre de 2021. páginas 161960 a 162106 (147 págs.) Resolución de 22 de diciembre de 2021. de la Mutualidad General de Funcionarios Civiles del Estado. por la que se publica el Concierto suscrito con entidades de seguro para el aseguramiento del acceso a la asistencia sanitaria en territorio nacional a los beneficiarios de la misma durante los años 2022. 2023 y 2024. https://www.boe.es/diario_boe/txt.php?id=BOE-A-2021-21337

Boere TM. El Alili M. van Buul LW. Hopstaken RM. et al. Cost-effectiveness and return-on-investment of C-reactive protein point-of-care testing in comparison with usual care to reduce antibiotic prescribing for lower respiratory tract infections in nursing homes: a cluster randomised trial. *BMJ Open* 2022;12:e055234. doi:10.1136/bmjopen-2021-055234

Devlin N. Valuing Child Health Isn't Child's Play. Value in Health 2022: 25 (7); 1087-1089. doi: https://doi.org/10.1016/j.jval.2022.05.009

Devlin N. Lovett R. Rowen D. Challenges in Measuring and Valuing Children's Health-Related Quality of Life. ISPOR News 2021. Vol. 7. No. 6. ISSN 2375-8678

Fawsitt CG. Lucey D. Harrington P. Jordan K. Marshall L. et al. A cost-effectiveness and budget impact analysis of C-protein point-of-care- testing to guide antibiotic prescribing for acute respiratory tract infections in primary care settings in Ireland: a decision-analytic model. *Family Practice*. 2022;29;3:389–397 https://doi.org/10.1093/fampra/cmab123

Holmes EAF. Harris SD. Hughes A. Craine N. Hughes DA. Cost-effectiveness analysis of the use of point-of-care C-reactive protein testing to reduce antibiotic prescribing in primary care. Antibiotics 2018:7;7(4):106. doi: 10.3390/antibiotics7040106

Larson B. Schnipel K. Ndibongo B. Long L et al. How to estimate the cost of point-of-care CD4 testing inprogram settings: an example using the Alere Prima (TM) analyzer in South Africa. Plos One 2012:7;4: e35444.

Lueza B. Auperin A. Rigoud C. Gross T G et al. Cost-effectiveness analysis alongside the inter B-NHL ritux 2010 trial:rituximab in children and adolescents with B cell non-

Hodgkin's lymphoma. European Journal of Health Economics 2024: 25:307-317.

OECD (2023). Embracing a One Health Framework to Fight Antimicrobial Resistance. OECD Health Policy Studies. OECD Publishing. Paris. https://doi.org/10.1787/ce44c755-en.

Petrou S. Methodological challenges surrounding QALY estimation for paediatric economic evaluation. Cost Effectiveness and Resource Allocation 2022: 20: 10

Thorrington D. Eames K. Measuring Health Utilities in Children and Adolescents: A Systematic Review of the Literature. PLoS One. 2015: 14;10(8): e0135672. doi: 10.1371/journal.pone.0135672. PMID: 26275302; PMCID: PMC4537138.

Viljoen A. Chubb B. Malkin S J P. Berry S et al. The long-term cost-effectiveness of once-weekly semaglutide 1 mg vs. delaglutide3 mg and 4.5 mg in the UK. European Journal of Helath Economics 2023: 24: 895-907.

Vlachaki I. Zinzi D. Falla E. Mantopoulos T et al. Cost-effectiveness analysis of vaborem for the treatment of carbapenem-resistant Enterobacteriaceae-Klebsiella pneumoniae carbapenemase (CRER-KPC) infections in the UK. European Journal of health Ecponomics 2022: 23:537-549.

WEB Pages

https://www.sanidad.gob.es/areas/farmacia/comitesAdscritos/prestacionFarmaceutica/docs/20240227_CAPF_Guia_EE_definitiva.pdf
https://www.worldbank.org/en/programs/icp
https://pestadistico.inteligenciadegestion.sanidad.gob.es/publicoSNS/S
https://botplusweb.farmaceuticos.com/
https://bnf.nice.org.uk/drug/
https://www.bocm.es/boletin/CM_Orden_BOCM/2017/08/21/BOCM-20170821-1.PDF
https://boc.cantabria.es/boces/verAnuncioAction.do?idAnuBlob=320839
https://www.sanidad.gob.es/estadEstudios/portada/home.htm
https://ec.europa.eu/eurostat/data/database
https://www.england.nhs.uk/national-cost-collection/#ncc1819
https://www.leedsth.nhs.uk/assets/71432c14fa/NIHR-2020-Investigation-and-Intervention-Tariff-1-v2.2-1.pdf#
https://www.boe.es/diario_boe/txt.php?id=BOE-A-2022-560
https://boc.cantabria.es/boces/verAnuncioAction.do?idAnuBlob=320839;

ANNEX I

Supplementary material 1. Health Economics Analysis Plan (HEAP) template v1.0 Elaborated by the team of WP5.5

The full list of essential and optional items is given below, with expanded item descriptions and practical examples of how the item might appear in a HEAP. Please note that the examples are drawn from a number of different studies.[†]

Essential items

		Description	Example
Section 1: Administrative information			
1.1	Title	Title that matches protocol and which includes the phrase 'Health Economics Analysis Plan'	Health economics analysis plan for the PRUDENCE trial: a platform randomised controlled trial of care diagnostics for enhancing the quality of antibiotic prescribing for community acquired acute respiratory tract infection in ambulatory care in Europe
1.2	Trial registration number	Trial registration number and name of registry that uniquely identifies the clinical trial on a publicly accessible registry (and other relevant trial study numbers)	ISRCTN13336322 (ISRCTN registry) https://www.isrctn.com/ISRCTN13336322 IRAS Project ID: 285877
1.3	Source of funding	Name of funders for trial and economic evaluation and funder(s)' reference number(s)	This project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement No 820755
1.4	Purpose of HEAP	Brief statement of the purpose of the HEAP	The purpose of this HEAP is to describe the analysis and reporting procedure intended for the economic analyses to be undertaken. The analysis plan is designed to ensure that there is no conflict with the protocol and associated statistical analysis plan and it should be read in conjunction with them.
1.5	Trial protocol version	Trial protocol version number associated with this HEAP	This document has been written based on information contained in the trial protocol version 3.3, dated 07 November 2022
1.6	Trial Statistical Analysis Plan (SAP) version	SAP version number associated with this HEAP	SAP Version: 1.0, Date: 06 June 2022
1.7	Trial HEAP version	Sequential number and date of this version	HEAP Version: 0.1, Date: January 2023

<table>
<tr><td>1.8</td><td>HEAP revisions</td><td>Date, justification for revision and summary of changes to the HEAP. Specify the individual making any revisions/changes to the HEAP.</td><td>Each row subsequently added to the table will indicate each HEAP revision change.
<table>
<tr><td>HEAP version No</td><td>Protocol version</td><td>Section No changed</td><td>Description and reason for change</td><td>Individual making change</td><td>Date changed</td></tr>
<tr><td>V1.0</td><td></td><td></td><td></td><td></td><td></td></tr>
</table></td></tr>
<tr><td>1.9</td><td>Roles and responsibilities</td><td>Names, affiliations and roles of individuals who have significantly contributed to the HEAP</td><td>This HEAP was drafted by Claire Hawksworth, Dr Laura Flight (health economist) and Dr Dalia Dawoud (senior health economist) for discussion with work package 5 of VALUE-Dx. This HEAP was finalised by the trial health economist(s) [Fernando Atoñanzas Villar, Marino Gonzalez, Carmelo Juarez, Pim van Dorst, Thea van Asselt and Maarten Postma] who are responsible for conducting and reporting the economic evaluation in accordance with the HEAP.</td></tr>
<tr><td>1.10a</td><td>Signature(s) of person(s) writing HEAP</td><td>Signature(s) of the person(s) writing the HEAP (and date)</td><td></td></tr>
<tr><td>1.10b</td><td>Signature of senior health economist</td><td>Signature of senior health economist who is guarantor of the economic evaluation (and date)</td><td></td></tr>
<tr><td>1.10c</td><td>Signature of Chief Investigator</td><td>Signature of the Chief Investigator for the trial (and date)</td><td></td></tr>
<tr><td colspan="4">Section 2: Trial introduction & background</td></tr>
<tr><td>2.1</td><td>Trial background and rationale</td><td>Synopsis of trial background and rationale including a brief description of research question and brief justification for undertaking the trial</td><td>Tackling antimicrobial resistance is a priority. One way is to target the misuse of antibiotics. PRUDENCE is a pragmatic, platform, three arm, multicentre randomised controlled trial to determine whether point of care diagnostics improve the quality of antibiotic prescribing for community acquired acute respiratory tract infection in ambulatory care in Europe.</td></tr>
<tr><td>2.2</td><td>Aim(s) of the trial</td><td>Clearly and briefly state the main aim(s) of the trial</td><td>Briefly, PRUDENCE RCT aims to determine if there is added value provided by having a CA-ARTI diagnostic (CAARTI-Dx) test done in the surgery to give a quick result. Then the result is available when a clinician is considering, or plans to prescribe an antibiotic, which could lead to more appropriate prescribing decisions, without causing harm to patients.</td></tr>
</table>

2.3	Objectives and/or research hypotheses of the trial	Describe specific trial objectives (primary and secondary) or trial hypotheses	Primary objective: effectiveness of CA-ARTI-Dx in terms of both reductions in antibiotic prescribing, and in terms of patient recovery. Secondary objectives: to explore whether adding a CA-ARTI-DX to usual primary care has: 1) Additional effects on antibiotic prescribing 2) Effects on antibiotic use 3) Effects on patient recovery and safety, including complications and hospitalisation 4) Effects on use of medications other than antibiotics 5) Effects on clinician's decision-making process regarding diagnosis and treatment 6) Effects on patients' perceived ability to understand and cope with their illness 7) Is cost effective
2.4	Trial population	Describe the trial inclusion and exclusion criteria	Inclusion criteria: • male or female, aged 1 year or above, consulting with symptoms of lower respiratory tract infection where cough is the predominant symptom (<28 days); or, symptoms of an upper respiratory tract infection (<14 days) where sore throat is the dominant symptom; and where the clinician is considering/has decided to prescribe an antibiotic, • is able and willing to comply with all trial requirements, • participants or legal guardian(s) of a child is willing and able to give informed consent according to national regulations Exclusion criteria: • Patients with only nasal, ear or rhinosinusitis symptoms, • patients with any serious condition associated with immunocompromised (long term oral steroids or immunosuppressants, terminal cancer), • patients for whom the clinician decides on immediate hospital admission, • patients who will not be able to participate in the study because they do not understand the local language; are terminally ill; have a serious psychiatric disorder; or judgement of the recruiting clinician deems ineligible.

2.5	Intervention(s) and comparator(s)	Describe the intervention(s) and comparator(s)	Interventions: where clinicians are unsure about prescribing an antibiotic they will use a CA-ARTI-Dx to help inform prescribing choice. The intervention is therefore usual care with the additional of clinical algorithms (flow chart) including a CA-ARTI-Dx. There are different CA-ARTI-Dx POCT tests being used: 1. Abbott Afinion CRP 2. BD Veritor System. The CA-ARTI-Dx should only be used where it influences clinician certainty, not in the case of an obvious viral infection for example. Comparator: usual care of eligible patients with CA-ARTI consulting in primary care (e.g. general practices, LTCF, and primary care paediatricians) where the clinician is considering or going to prescribe an antibiotic. Clinicians will be referred to local guidelines for managing common infections where available. The trial is comparing use of a CA-ARTI-Dx POCT to usual care, rather than the specific tests.

2.6	Trial design	Briefly describe the trial design including type of trial such as cluster, crossover, etc. Can also include details of power calculation, sample size (including any separate calculations for economic endpoints), randomisation and blinding.	This is a multi-country, prospective, individually randomised, platform clinical trial in community care with a nested process evaluation. The trial uses an adaptive design. There will be a pre-specified interim analysis in between study periods. If the study questions for a given CA-ARTI-Dx are answered at the interim per pre-specified criteria, then one or both of the CA-ARTI-Dx will be replaced with a new CA-ARTI-Dx. This study design therefore has the potential to assess multiple CA-ARTI-Dx across the study recruitment period, and provides flexibility for the inclusion of additional CA-ARTI-Dx into the trial should suitable tests become available and resources permit this. Via an interim analysis, a CA-ARTI-Dx can be dropped after the end of the first recruitment period, most likely after the first winter season if recruitment is as expected, based on pre-specified criteria for either success or futility. If a CA-ARTI-Dx is dropped, it may be replaced with a new CA-ARTI-Dx to be evaluated in the second recruitment period. The nested process evaluation will capture data to understand how CA-ARTI-Dx is used in practice and how it influences patient care and experience. These data will inform implementation within the period of the trial and beyond. More details in the SAP. The adaptive nature of the trial can introduce bias into point estimates and confidence intervals. https://arxiv.org/abs/2211.15620. This will be considered if any adaptations are implemented. Randomisation takes place via an online system. The Data Monitoring and Ethics Committee (DMEC) monitor the randomisation process. Different allocation ratios are used for different combinations of cough and flu season (see SAP for further details). PRUDENCE will be an open trial. The participant and the recruiting clinician will know the participant's allocation. Therefore, no unblinding or code breaking is required in the event of a relevant emergency. The trial team and recruiting clinicians will be blind to emerging results. During the course of

			the trial, only those on the Data Monitoring Committee will have access to the unblinded data as part of the interim analysis. **Sample size justification** For the first co-primary outcome, under assumptions of 50% of participants presenting with predominantly cough, 50% of participants being enrolled during the influenza season, 5% SARS-CoV-2 positivity, and 5% of participants missing the primary endpoint, 1250 participants per winter season (2500 participants total) will provide approximately: • 266 subjects per arm per winter season for the Afinion CRP vs. usual care analysis (>99% power for a single winter season) and • 376 subjects per arm per winter season for the Veritor vs. usual care analysis (>99% power for a single winter season). For the secondary co-primary outcome, under the same assumptions as above, 1250 participants per winter season (2500 participants total) will provide approximately 247 subjects per arm per winter season for the Afinion CRP vs. usual care analysis. This corresponds to approximately 72% power for a single winter season, and 95% power for both winter seasons combined. In addition, we expect 376 subjects per arm for the Veritor vs. usual care analysis, providing approximately 85% power for a single winter season, and 99% power for both winter seasons combined. This study is not powered for any health economic outcomes.
2.7	Trial start and end dates	Trial recruitment start and end dates and the follow-up period	Recruitment started in October 2021 and is due to finish at the end of the end of the winter season 2023/24. The follow-up period is 28 days.
Section 3: Economic approach/overview			
3.1	Aim(s) of economic evaluation	Describe the aim(s) of the economic evaluation	The aim of the economic evaluation is to address the question “What are the long-term clinical, public health and economic impacts of using point of care diagnostics on AMR prevention compared to usual care in people with symptoms of respiratory tract infections?”

3.2	Objective(s) of economic evaluation	Describe the objectives (primary and secondary) of the economic evaluation	The primary aim of the health economic evaluation is to estimate the long and short-term cost-effectiveness of CA-ARTI-Dx POCT interventions compared to usual care (no CA-ARTI-Dx POCT) to inform antibiotic prescriptions in people with symptoms of respiratory tract infection in primary care.
3.3	Overview of economic analysis	Briefly outline and justify the type of economic evaluation to be undertaken, identifying the primary economic analysis and outlining the analysis plan and the methods that will be used	**Short-term (WP5.3)** The within-trial economic analysis will be performed using individual patient level data from the PRUDENCE trial. Resource use and health outcomes will be measured for each participant in the trial. The analytical approaches will take the form of cost-effectiveness analysis, based on cost per antibiotic prescription avoided, or a cost-utility analysis, based on cost per Quality Adjusted Life Years if there are sufficient data. Trial data will be used to calculate incremental cost-effectiveness ratios for care involving POCT compared to usual care. **Long-term (WO5.4)** The model-based economic analysis will be performed using an individual-based simulation model consisting of three modules (demographic, consultation, and antimicrobial resistance forecasting). The model is flexible to accommodate multiple outcome measures, including antibiotic prescriptions and QALYs. The model code is publicly available on github (https://github.com/UMCG-Global-Health/MERIAM/releases) and includes a manual describing for instance inputs, sources, and assumptions.

3.4	Jurisdiction(s)	Specify the jurisdiction(s) in which the analysis will be conducted including details of the country(s) and health system(s)	The trial is international with trial sites in Belgium, France, Georgia, Germany, Greece, Hungary, Ireland, Israel, Italy, Portugal, Spain, UK and Poland is opening soon. Some sites have recruited more than others so it is likely that the data may be slightly skewed towards the UK and Georgia. The analysis will be conducted using a pooled data set from the clinical trial. Country specific scenarios will be explored in the case of a statistically significant difference in effectiveness between countries. UK needs ability to see health system perspective Health systems of different countries will be accounted for in the economic evaluation. The UK has a national health service (NHS), providing publicly funded healthcare, primarily free of charge at the point of use.
3.5	Perspective(s)	State the perspective(s) from which the economic analysis is being conducted, such as societal perspective and/or healthcare payer perspective	**Short-term** The trial-based economic analysis will take a societal perspective. A scenario will use a NHS and personal social services (PSS) perspective as is taken in the UK. **Long-term** Different perspectives will be accounted for in the analyses. The primary economic analysis will take a societal perspective. A UK analysis using the NHS and personal social services (PSS) perspective will be done as a scenario analysis.
3.6	Time horizon(s)	State the time horizon(s) over which costs and consequences are being evaluated	**Short-term (WP5.3)** The within trial economic analysis will use a 28-day time horizon which aligns with the 28-day follow-up of the trial participant. **Long-term (WP5.4)** Preferably the long-term analysis would have a lifetime time horizon but given that there is substantial uncertainty around the AMR predictions, which increases further into the future, a time horizon of 10 years may be more feasible.
Section 4: Economic data collection & management			

4.1	Statistical software	Specify the statistical software that will be used to carry out the health economic analysis	**Short-term (WP5.3)** Excel? R? **Long-term (WP5.4)** All three modules of the model were/will be programmed in R (Core Team. R: a language and environment for statistical computing [Internet]. Vienna, Austria; 2020. https://www.R-project.org/.)
4.2	Identification of resources	Justify and describe items of resource use that will be measured as part of the trial	The following broad categories of resource are being measured: • CA-ARTI-Dx POCT • drugs (antibiotic prescriptions and days on antibiotics, antivirals, anti-cough fluids and pills, other medications) • microbiologic tests (central lab tests), biochemistry tests • GP and nurse visits • other specialists visits • hospitalisations in general wards and ICU • emergency room visits • days of work leave • days of school leave The full list of variables is in the eCRFs on the logon area of the ValueDx website: WP5-logged in area - value-dx . In the diary, also things like GP visits, A&E etc are captured.
4.3	Measurement of resource-use data	Describe the resource-use data collection method(s) (including external routine datasets) and the time points at which they will be used.	Resource-use data will be collected from day 1-14 post randomisation using the self-reported diary and from day 14- 28 via the follow up phone call using case report forms, completed by patients´reports. (?).

4.4	Valuation of resource-use data	For each resource item measured, describe how the unit cost will be derived and from which specific price year. Outline how adjustments will be made for sources from different price years and which inflation index will be used.	Unit costs will be derived in Spain using different official databases as well as from England using the NICE database on unit costs. and England using: - British National Formulary - NHS Reference Costs - PSSRU Unit Costs of Health and Social Care - Prescription Cost analysis … Costs are calculated using Euros and will be converted to GBP using purchasing power parity. In what year? Any adjustments needed for inflation? The UK has the ONS GDP deflator index and the NHS cost inflation index. The strategy for including the costs of the POCT will be agreed with the private companies that own the tests in the event that prices are not publicly available.
4.5	Identification of outcome(s)	Specify and justify the outcome(s) that will be measured	The primary economic outcome measure will be cost per antibiotic prescription avoided. Quality of life data will also be collected using an age appropriate version of the EQ-5D-5L (according to EQ5D user guide: EQ-5D-Y for ages 8-15, proxy version for ages 4-7 – for ages <4 there is no suitable alternative) to derive utility scores and a quality adjusted life year (QALY) for a cost-utility analysis. It is anticipated that there will be issues with the EQ-5D-5L data… - Children completing - High level of missing data because parents experience the questionnaire as not appropriate/relevant

4.6	Measurement of outcome(s)	Describe the outcome data collection method(s) and the time points at which they will be used	Original prescribing decision and subsequent antibiotic prescription is recorded at baseline. Differences between the trial arms can calculate the cost per antibiotic prescription avoided. Whether the POCT altered the clinician's decision on prescribing antibiotics can also be investigated. Outcome data will be collected from day 1-14 post randomisation using the self-reported diary and from day 14- 28 via the follow up phone call using case report forms, completed by the trial team. Information available in Appendix A of the CRP: Schedule of Procedures. Details of hospital admissions will be collected from patients reports beyond day 28.
4.7	Valuation of outcome(s)	For each outcome measured, describe how it will be valued and the source of these valuations	For the quality of life data: - For EQ-5D-Y there is only a few country value sets available at current
Section 5: Economic data analysis			
5.1	Analysis population	Outline the analysis population that will be used in the economic base-case analysis (such as intention to treat, per protocol)	Taken from the SAP, section 3.5: The primary analyses will include all participants in the intention to treat (ITT) population, i.e. all participants randomised and analysed in their respective randomised assignment. The analysis will study the health outcomes of the ITT population and its associated healthcare resources. It will not use a decision tree scheme.

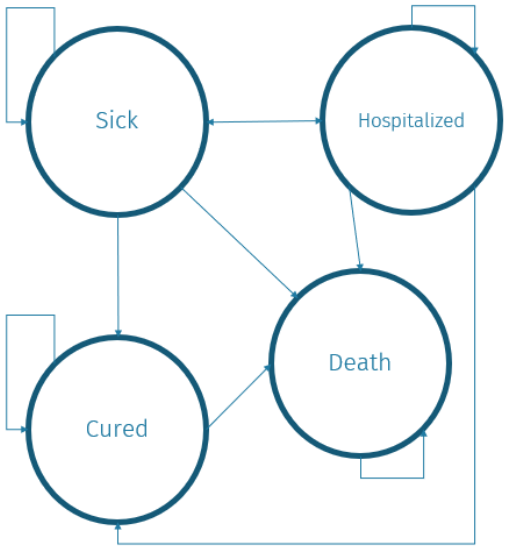
5.2	Timing of analyses	Describe the timing of all planned analyses (e.g. interim and final analyses)	The final analysis will occur after the study has been fully recruited and participants followed up for 28+ days (including retrospective hospital admissions established). The clinical trial database lock will then occur, and data can be unblinded and used for the economic analyses. The final analysis will include a within-trial analysis, taking a 28 day time horizon, not extrapolating beyond the end of the trial. An interim clinical analysis is planned for March 2023 but may not occur if the study is near fully recruited (it will proceed to final analyses). There are no planned health economic analyses using the interim data or to inform adaptations to the trial. The clinical data cut will be taken in March 2023 to provide blinded data to the economic team to develop the within trial and model based health economic analyses.
5.3	Discount rates for costs and benefits	Detail the source of, and justification for, discount rates used for costs and benefits	Discount rates for costs and benefits will not need to be considered for the trial-based economic analysis due to the short time horizon. The model-based economic evaluation will consider country specific discount rates. The UK uses a discount rate of 3.5% per annum.
5.4	Cost-effectiveness threshold(s)	Detail the cost-effectiveness threshold(s) to be used in analysis/interpretation	The primary economic evaluation is a CEA and the threshold for cost per antibiotic prescription avoided is not known. In the case of a cost-utility analysis a range of willingness to pay thresholds will be considered including the NICE threshold of £20,000-£30,000 per QALY.
5.5	Statistical decision rule(s)	Describe how inference will be drawn (e.g. significance level, confidence intervals or mean net benefit)	Mean differences in costs and outcomes between the treatment groups will be estimated with associated 95% confidence intervals.
5.6	Analysis of resource use	Describe how differences in the use of resources/services between randomised groups will be compared	Differences in the use of services between randomised groups will be described but not compared statistically.

5.7	Analysis of costs	Describe analyses of the cost data, specifying any covariates for statistical adjustment, assumptions, and alternative methods	A direct sum of the unit cost of the healthcare system resources used by each patient will be used for the cost calculation of each arm.
5.8	Analysis of outcomes	For each outcome used in the economic analysis, describe how the outcome will be analysed, specifying any covariates for statistical adjustment, assumptions, and alternative methods	The outcomes being included in the economic analysis are: - Antibiotic prescriptions - Days on antibiotic therapy (DOT) - Quality adjusted life years (QALYs) -
5.9	Data cleaning for analysis	Outline how data will be cleaned before analysis	The clinical trial team will perform all data cleaning, including for economic variables and provide this clean data to the health economic evaluation team.
5.10	Missing data	Specify the procedure for dealing with missing data	The SAP states For a given endpoint, the number of patients (% of overall sample) who's response is 'Unknown' will be recorded for each question. Should any question have >20% of missing responses, a logistic regression model will be fitted to explore factors associated with missing. The appropriate method for dealing with missing data will depend on the proportion of missing data and likely mechanism of missingness. For example, multiple imputation methods may be used if the data is missing at random (MAR). Specific to QoL 1) when someone dies during the study are their future quality of life scores and costs set to zero This reference can be consulted: https://link.springer.com/article/10.1007/s40273-014-0193-3
5.11	Analysis of cost-effectiveness	Describe the methods that will be used to summarise cost-effectiveness.	Cost and outcome data will be combined to calculate an incremental cost-effectiveness ratio (ICER), per antibiotic prescription avoided. The cost per QALY gained will be calculated as a secondary outcome if QoL data permit.

5.12	Sampling uncertainty	Describe how uncertainty around the costs and effectiveness estimates and summary cost-effectiveness measures will be explored	Things to detail here: 1) a deterministic tornado analysis 2) Cost effectiveness acceptability curve Example text from template: The nonparametric bootstrapping approach will be used to determine the level of sampling uncertainty surrounding the mean ICER by generating 10,000 estimates of incremental costs and benefits.
5.13	Subgroup analyses or analysis of heterogeneity	Describe any analyses of subgroups or heterogeneity in cost-effectiveness and the analysis methods used	The primary analyses will be pooled. A per country analysis may be relevant for decision making purposes. In this sense, we have collected unitary cost data for health resources in England and Spain for more than 20 cost items, also purchase power parities, as used by World Bank for international studies, so that national results could be provided for the two trials. Then there are two possibilities: using aggregated clinical results and country specific costs; using both health outcomes data and also specific costs per country. However, the samples of the trials were not powered to show statistical differences across countries. We suggested to test whether health outcomes are homogeneous across countries with the preliminary data and if it were the case, just use aggregated data for health outcomes. It can be considered as an empirical question.
5.14	Sensitivity analyses	Describe any sensitivity analyses and their form	Deterministic sensitivity analyses: Individually vary the following parameters: antibiotics, POCTs, hospitalisations, microbiological tests, GP visits Probabilistic sensitivity analyses will be informed using uncertainty ranges in the key input parameters from the IPD. Any scenario analyses: 1) Complete case analysis (ie including only patients with complete data) 2) Per country analyses 3) Using the quality of life data and scenarios around this
Section 6: Modelling			

6.1	Extrapolation or decision analytic modelling	Outline whether decision analytic modelling or any other extrapolation will be used to estimate cost-effectiveness results beyond the period of the trial or to introduce an additional comparator or other evidence.	MERIAM has three compartments: the demographic model, used to model the population over a long time horizon; the consultation model, used to model patients going to care with an acute respiratory tract infection and the antimicrobial resistance (AMR) model, used to forecast AMR levels and AMR-related mortality and costs. The demographic model is used to create the modelled population and simulate population changes based on Eurostat demographic data and population forecasts, including ageing, births, mortality and migration. The consultations model uses the incidence of respiratory infections (acute respiratory infections and influenza-like illness) based on consultation data from the European Centre for Disease Prevention and Control (ECDC). Considering four age categories (0–4 years, 5–15 years, 15–64 years, and ≥ 65) and the individuals from the demographic module, the incidence rates are used to simulate GP consultations. Within these consultations, the number of tests performed and the number of antibiotics are modelled based on trial data (PRUDENCE and ADEQUATE). The AMR model uses a two-step approach. First, the baseline AMR projections are generated, using an ensemble model. This is a data-driven approach where current trends are used to forecast future AMR rates. These baseline projections are then used as for the current-care scenario, where we assume current patterns in AMR will continue in the future. The second step is to incorporate the impact on antibiotic consumption from the diagnostic strategies, in the baseline AMR projections. This uses a more mechanistically-driven approach. The steps are described in more details below. A statistical forecasting method, comparable to the methods used by Hashiguchi et al. was used for this study: Several methods are available for time series forecasting but selecting a single 'best' model is challenging. Ensemble methods are an often-used technique to improve forecasts:

			instead of picking one model, several models are used simultaneously and then combined to provide an average. We developed an ensemble model, averaging three models: - An exponential smoothing (ETS) model, which forecasts future data using weighted averages of past observations. - A random forest, which aggregates many regression trees to estimate the outcome of interest (AMR rates in our case). Bagging (bootstrapping and aggregrating) is used, where each decision tree is informed by a random sample, with only a subset of the available regressors, of the original data set. The different trees are grown in parallel, i.e. new trees are not informed by previous trees. - An XGBoost model, which also combines many regression trees to estimate the outcome of interest, however, as opposed to random forests, a sequential tree growing algorithm (boosting) is used, where each new tree informs the creation of the next tree. For a detailed overview, please refer to: van der Pol, S., Jansen, D. E., van der Velden, A. W., Butler, C. C., Verheij, T. J., Friedrich, A. W., ... & van Asselt, A. D. (2022). The Opportunity of Point-of-Care Diagnostics in General Practice: Modelling the Effects on Antimicrobial Resistance. *PharmacoEconomics*, *40*(8), 823-833.
6.2	Model type	Describe the modelling approach that will be used and duration of extrapolation	The demographic and AMR modules use annual cycles, while the consultation module uses weekly incidence rates. To assess the long-term impact of large-scale testing using the hypothetical strategy, we assessed the intervention for a time horizon of 10 years For the consultation model, a micro-simulation model will be used with four health-states: sick, cured, hospitalized, death.

6.3	Model structure	Detail the model structure (where possible, include diagram of model states and transitions between them)	The model structure is shown in the figure below. To each of the health states costs and effects will be applied. 
6.4	Treatment effect beyond the end of the trial	Describe the duration and size of treatment effect in the period beyond the end of the trial	Any effect that extends beyond the trial duration will be based on literature. Long-term effects with respect to AMR will be determined by the resistance forecasting model.
6.5	Other key assumptions	List the key structural assumptions of the model	The following assumptions apply: - Current trends for antibiotic consumption can be extrapolated to future consumption. - The trial results (effects) are generalisable and can therefore be applied to other countries. - An bacterium-antibiotic specific elasticity exists between consumption and its effect on resistance. - ..
6.6	Methods for identifying and estimating parameters	For each model parameter, describe the methods and data sources that will be used to estimate the parameter (e.g. from the RCT, systematic review, meta-analysis, other published data or expert opinion)	The model parameters will be based on trial results (ADEQUATE and PRUDENCE), gathered from available literature and where necessary parameterised using formally elicited expert opinion.

6.7	Model uncertainty	Describe the methods that will be used to assess parameter uncertainty in the results. Describe sensitivity analyses for the impact of other types of uncertainty on results.	Parameter uncertainty will be assessed using univariate sensitivity analysis, probabilistic sensitivity analysis and scenario analysis.
6.8	Model validation	Describe the methods and data that will be used to check the face, internal and external validity of the model	The model is validated with an external advisory panel. Additionally, most of the short-term model parameters will be based on trial results.
6.9	Subgroup analyses/heterogeneity	Describe subgroup or heterogeneity analyses that will be executed and reported within the extrapolation or decision analytic modelling	The model will consider the population heterogeneity by the demographic model. Based on Eurostat population projections, a representative subset of country populations will be modelled. Variables that are included are Age, sex, employment status, vaccination status, migration and utilities. Additionally, care seeking behaviour will be considered.
Section 7: Reporting/publishing			
7.1	Reporting standards	Describe any guidelines that will be followed when publishing results	CHEERS 2022 guidelines will be followed when reporting the health economic evaluation, in a format appropriate to stakeholders and policy makers.
7.2	Deviations from the HEAP	Describe the procedure for reporting any deviations from the HEAP	Any deviation from HEAP will be described and justified in the final published report.
Section 8: Appendices			
8.1	Health economic collection tools	Include template examples of the resource-use data collection sheets and resource-use questionnaires	Example text from template: Data collection sheets on hospital stays, visits to A&E, and home visits are given in Appendix X and include health utility measures. The resource-use questionnaire has been deposited in DIRUM (http://www.dirum.org/).

Optional items

		Description	**Example**
Section 1: Administrative information			

O1.1	Table of contents	List of HEAP contents with page numbers	
O1.2	Abbreviations/glossary of terms/definitions	List of abbreviations and/or acronyms used within the HEAP alongside their meanings/definitions	QALY: quality-adjusted life year NHS: National Health Service
Section 4: Economic data collection & management			
O4.1	Monitoring collection of health economic data	Outline how the health economic data collected will be monitored	The trial health economist(s) will work closely with the trial team throughout the data collection period. Data collection forms will be assessed throughout the trial period to monitor quality of the data and amend any forms or procedures if necessary. Queries will be resolved through the trial management team and the health economic team.
O4.2	Database management	Outline how the economic data will be stored and managed and by whom	As documented in the PRUDENCE CRP.
O4.3	Data entry	Outline how data will be entered/handled and outline any checking systems in place	As documented in the PRUDENCE CRP. All baseline data will be entered into the case report form (CRF) by the trial research nurses at the recruitment site. Follow-up data collected from postal questionnaires will be entered by the central research team. The database will use controls to limit data entry to plausible values.

O4.4	Data archiving	State whether datasets, interim datasets and final analysis will be archived, and if so, how	As documented in the PRUDENCE CRP. A copy of health economic analysis files, derived datasets, interim datasets and final analysis will be locked and archived. Archived datasets will be held by University of Rioja and University Medical Center Groningen and will conform to the department data security policy and department data compliance and Data Protection Act policies.
Section 6: Modelling			
O6.1	Value of information analysis	Describe whether value of information analysis is planned and the type and methods that will be used to calculate value of information	From CRP: Value of perfect information analysis may also be performed to identify which sources of uncertainty should be reduced through additional research to efficiently improve decision making
Section 8: Appendices			
O8.1	Cross-referencing to other trial documents	Reference to other relevant trial documents that are adhered to and followed when writing the HEAP and any other references used when writing the HEAP	The Nuffield Department of Primary Care Health Sciences Standard Operating Procedure for Economic Evaluations (version 2.1) was followed in designing this analysis. The analysis described in this plan adheres to the University of Rioja Data Management Plan (version 1.1).

O8.2	Illustrations	Illustrations such as annotated questionnaires detailing the database fieldnames, flow charts outlining the flow of data for the economic evaluation, or template tables	The conduct and procedures for the economic evaluation will be fully integrated into the PRUDENCE trial from planning and designing the economic evaluation through to publication.

† Examples were extracted, and in some cases modified, from existing draft or final HEAPs provided to the study team. We are grateful to the trial teams from which they originated.

ANNEX II (elaborated by WP5.3, University of La Rioja (Spain).

Table 1. Database used for calculating the cost of antibiotics used in the PRUDENCE trial

Family of antibiotics	Antibiotics	Format
Tetracycline		J01AA07 (100 mg)
Narrow spectrum penicillin	Fenoximetilpeniciline	J01CE02 (250-500 mg).
Broad spectrum penicillin	Ampicillin Amoxicillin	J01CA01 (500 mg) J01CA04 (250-500-750-1000 mg).
Co-amoxiclav	Amoxiclav	J01CR02 (125-500-875 mg)
Macrolide	Erythromycin Roxithromycin Clarithromycin Azithromycin	J01FA01 (250 mg) J01FA06 (150mg) J01FA09 (250 mg) J01FA10 (500 mg)
Quinolone	Ciprofloxacin Levofloxacin Moxifloxacin	J01MA02 (250-500-750 mg) J01MA12 (500 mg) J01MA14 (400 mg).
Cephalosporin	Cephalexin Cefadroxil Cefuroxime Cefaclor Cefixime	J01DB01 (500 mg) J01DB05 (500 mg) J01DC02 (250 mg) J01DC04 (250-500mg) J01DD08 (200-400 mg)
Others	Calculated as the weighted average cost of the "AB_CLASS_OTH_SP" column in the "prescribing_20230301" sheet.	

Source: Own elaboration from Prudence databae

Table 2. Database used for calculating the cost of other prescribed drugs used in the PRUDENCE trial

Type of prescribed drugs	Active principle	Format
Inhalated	Becloforte	250 mg/inhalation (200 doses)
Antivirals	Oseltamivir	J05AH02 10 hard capsules (30-45-65-75 mg) and oral liquid (65 ml).
Antihistaminixs	Cetirizine	R06AE07 20 capsules (10 mg).
Paracetamol and other NSAID	Paracetamol	N02BE01 20 pills (500 mg).
Anticough	Codeine	R05DA04 1 jar (250 ml).
Other	Calculated as the weighted average cost of the "MED_OTH_SP" column in the "med_20230301" sheet.	

Source: Own elaboration from Prudence database

Table 3. Database used for calculating the cost of other declared drugs used in the PRUDENCE trial

Other declared drugs	Active principle	Format
Paracetamol	Paracetamol	N02BE01 20 pills (500 mg)
Ibuprofen	Ibuprofen	20 capsules (400 mg)
Other pain medication	Naproxen sodium	40 doses (550 mg)
Inhaled medication	Salbutamol sulfate	20 doses
Flu combination medication	paracetamol dextrometorfano chlorphenamine caffeine citrate ascorbic acid	200 doses 650 mg 20 mg (as hidrobromure) 4 mg (as maleato) 30 mg 250 mg
Cough medicine	60 doses willow extract 660 mg. elderberry concentrate 600 mg. propolis fluid extract 450 mg. black currant concentrate 450 mg. plantain concentrate 375 mg. propolis dry extract 300 mg. pine bud extract 300 mg. thyme concentrate 300 mg. echinacea extract 120 mg. vitamin C 80 mg. eucalyptus essential oil 45 mg. peppermint essential oil 14.4. zinc gluconate 10mg. manganese gluconate 2mg. copper gluconate 1mg.	
Lozenges. Mouth washes or gargles	Flurbiprofen	8.75 mg (16 doses
Nose spray	natural isotonic seawater solution	equivalent to 9 g/l NaCl
Ear drops	Mullein. propolis. lemon essential oil. Niaouli essential oil and Cajeput essential oil.	
Vitamins	20 doses Ginseng G115 40mg. vitamin A 640mcg 80% NRV. vitamin D 6mcg 120% NRV. vitamin E 12mg 100% NRV. vitamin C 80mg 100% NRV. thiamine 21mg 190% NRV. riboflavin 2.2mg 157% NRV. niacin 17.5mg 109% NRV. vitamin B6 2.8mg 200% NRV. folic Acid 300mcg 150% NRV. vitamin B12 3mcg 120% NRV. biotin 38mcg 76% NRV. pantothenic Acid 6.3mg 105% NRV. magnesium 77.5 mg 20% NRV. iron 8.3 mg 59% NRV. zinc 10 mg 100% NRV. copper 1 mg 100% NRV. manganese 2 mg 100% NRV. selenium 55 mcg 100% NRV.	

Source: Own elaboration from Prudence database

ANNEX III

PRUDENCE: Quality of life analysis

Value-Dx, WP5.3 University of La Rioja (Spain)

13th of June, 2024

1. Introduction

The European Quality of Life-5 Dimensions (EQ-5D), developed by the EuroQol Group, is an instrument that is highly recommended for economic evaluation studies of health technologies. It is a health measurement tool widely used in medical research and quality of life assessment.

The EQ-5D has two response formats, 3-level (EQ-5D-3L) or 5-level (EQ-5D-5L). In this study, the trial designers opted for the 5-level response format. Each patient assesses its own health status by first responding by five levels (EQ-5D-5L) of severity, ranging from "no problems" to "extreme problems", five dimensions related to fundamental aspects of health (mobility, self-care, usual activities, pain and depression) and then scoring their general condition through a Visual Analogue Scale (VAS) [1].

Based on the responses to the five dimensions, the patient's health status is determined, which could be used as an index that evaluates the impact that new medical or diagnostic interventions have on the utility perceived by the patient themselves. To convert health states into utilities, rates known as the "EQ-5D rates" are applied [1]. These rates are different between countries, as perceptions of health and quality of life vary between different cultures and societies: what is considered good health may differ between countries due to differences in cultural values, beliefs and social norms. Two patients from different countries but who have reported the same health status can obtain different utilities. Currently, EQ-5D rates have been published for several countries and are available on the EuroQol website [2]. The calculated utilities can be included in cost-effectiveness studies as they can be converted into Quality-Adjusted Life Years (QALYs).

A QALY is a measure used in economic evaluations of medical interventions. It is used to quantify the effectiveness of different medical treatments and compare their effectiveness in terms of quality-adjusted life years. The EQ-5D can be used to assign utility values to different health states. These utility values are then used to calculate QALYs. Therefore, the EQ-5D provides a measure of health-related quality of life that is used as input in QALYs calculations, allowing comparisons between different medical interventions in terms of their impacts on the quality and quantity of life.

The VAS is a tool used in medicine and research to measure the intensity of a subjective experience, such as pain, fatigue, anxiety, among others. It consists of a horizontal line of fixed length, generally 10 centimeters, on which the patient is asked to mark a point that reflects the intensity of their experience at that moment. The VAS allows the subjective experience of the patient to be captured in a quantitative way, thus allowing statistical analysis to be performed. The VAS is an integral part of the EQ-5D.

In this study, the EQ-5D form completed by PRUDENCE patients has been analyzed with the objective of knowing the effect that the new diagnostic intervention has on the patients' perceived utility and VAS score. These values are compared with those perceived

by a usual care group of patients to whom the new diagnostic technique has not been applied and who have also answered the same EQ-5D form.

2. Material and method

2.1. Data

The analysis is performed with the data included in the spreadsheets named Diary_D01, Diary_D01_LTCF, Diary_D14, Diary_D14_LTCF, Phone_LTCF, Phone_PC and Random. The last page shows the number of patients and the result of randomization to one of the two groups of the analysis: usual care group and diagnostic intervention group (Both measurements of Veritor (seen as one trial arm) + Flu measurement of Veritor (Gas not needed for cough) + CRP + Gas measurement of Veritor (Flu not needed outside flu season)). The other sheets show the answers that each of the previous patients has given to the EQ-5D-5L form at three different times: day 1, day 14 of follow-up and day 28 of follow-up. The EQ-5D-5L provides five levels of answer per dimension: 1="I have no problems in X about", 2="I have slight problems in X about", 3="I have moderate problems in X about", 4="I have severe problems in X about" and 5="I am unable to X about".

The intention to treat approach was followed in this analysis. Furthermore, during data cleaning, observations may be lost if the patient has not responded to the five dimensions of the form. To calculate the utilities perceived by each patient at the three moments analyzed, the complete health status must be available, that is, the patient has answered the five dimensions. If a patient has not completed any of the five questions, the utility cannot be calculated and that record is left out of the calculations. No data extrapolation/imputation method has been used to fill in the dimensions that remain unanswered.

Based on the health status, the utilities will be calculated applying the rates specific to each country. To calculate the gain or loss in QALYs reported by the new diagnostic innovation technique, the difference in utilities is multiplied by 14/365 (when the time between day 1 and day 14 is analyzed), by 28/365 (when the time between day 1 and day 28 is analyzed), and by 14/365 (when the time between day 14 and day 28 is analyzed), and the result is divided by two. In this way, it is assumed that the gain or loss of QALYs is progressive between the pairs of moments that are analyzed and, consequently, the QALY gain is assumed to grow uniformly within the analyzed period. Thus, we calculated the QALY gain as the area of a triangle whose base is the time period and the height the difference of utilities between day 1 as reference and the specific moment considered (day 14 or 28).

As indicated above, not all countries have published their EQ-5D rates. Countries with the value sets published can be found in EuroQol website [2]. If a patient in the database comes from a country for which there is no such information, it has been decided to use those for the United Kingdom. However, NICE is currently recommending using the 3L value set for England for NICE technology assessment submissions. EuroQol, in partnership with NICE are currently developing a new set of EQ-5D-5L values for the United Kingdom, which is currently not available [2]. Hence, in this analysis the value set for EQ-5D-5L available before its withdrawal is used, as shown by Devlin *et al* 2018 [3].

2.2. Dynamic and static analysis

Two types of analysis are performed, labelled dynamic analysis and static analysis. The

dynamic analysis allows us to know the evolution of the patients' utility from day 1 to day 14, from day 1 to day 28 and between day 14 and day 28. It is analyzed whether there are statistically significant differences within each group at the three moments analyzed. To do this, the t test for means of two paired samples is applied (since the same patient must be followed at each pair of moments). Only patients who have reported each pair of data appear here, so observations may also be lost if the patient has not responded on any of the three days analyzed.

The static analysis allows us to know if there are differences in the utilities between each group (usual care and diagnostic intervention) on day 1, on day 14 and on day 28. As it is not necessary to follow up on each patient, the responses given by all patients in each group (on each of the three days analyzed) can be grouped. To do this, the Student's t test will be applied to analyze whether there is a significant difference between the means of two different samples. First, it is checked if the variances of the two samples are equal, for this an F test is carried out. If the null hypothesis is rejected, it means that the variances of both samples are not equal. In this case, the two-sample t test is performed assuming unequal variances. If the result of the F test is opposite, it means that the variances of both samples are equal, and the t test is performed for two samples assuming equal variances. The Jarque-Bera test is used to evaluate whether a given data set follows a normal distribution. If the data set was not distributed as a normal, the appropriate non-parametric tests would be applied.

3. Results

A total of 2,664 patients (2,639 in the intention to treat patients) belonging to 13 countries appear within the spreadsheet named Random: Belgium, Germany, Spain, France, Georgia, Greece, Hungary, Ireland, Israel, Italy, Poland, Portugal and United Kingdom. Of these 13 countries, Germany, France, Spain, Hungary and the United Kingdom have their own EQ-5D rates that allow utilities to be calculated. Therefore, the UK rate is applied to data from patients in Belgium, Georgia, Greece, Ireland, Israel, Italy, Poland, Portugal and United Kingdom (Table 1).

Table 1. Number of patients by countries. Source of applied EQ-5D rates.

Country	Number of patients	EQ-5D-5L rates applied [(1)]
Belgium	118 (c: 54; d: 64)	United Kingdom [3]
Germany	180 (c: 80; d: 100)	Germany [4]
Spain	85 (c: 41; d: 44)	Spain [5]
France	26 (c: 13; d: 13)	France [6]
Georgia	498 (c: 215; d: 283)	United Kingdom [3]
Greece	350 (c: 155; d: 195)	United Kingdom [3]
Hungary	219 (c: 110; d: 109)	Hungary [7]
Ireland	227 (c: 92; d: 135)	United Kingdom [3]
Israel	30 (c: 10; d: 20)	United Kingdom [3].
Italy	102 (c: 47; d: 55)	United Kingdom [3]
Poland	271 (c: 127; d: 144)	United Kingdom [3]
Portugal	50 (c: 19; d: 31)	United Kingdom [3]
United Kingdom	483 (c: 228; d: 255)	United Kingdom [3]
Total	2,639 (c: 1,191; d: 1,448)	

c: number of patients in the usual care group; d: number of patients in the diagnostic intervention group.
[(1)] Value sets available on EuroQol [2].

Of the 2,639 patients, 1,191 patients correspond to the usual care group and 1,448 patients to the diagnostic intervention group. It would be expected that three records would appear for each patient, so that each patient would have completed the EQ-5D instrument on day 1, on day 14 and on day 28. However, of those 7,917 (2,639x3) expected records, only 2,297 rows appear in day 1, 1,965 in day 14 and 2,273 in day 28 (2,297+1,965+2,273 = 6,535).

3.1. Utilities obtained through the five dimensions of the EQ-5D. Dynamic analysis

To perform the dynamic analysis of the evolution of the patients' utility from day 1 to day 14, from day 1 to day 28 and between day 14 and day 28, it is necessary that the same patient has answered the questionnaire on the pairs of days that are analyzed. For the usual care group, there are 632 patients who have completely answered the EQ-5D on day 1 and day 14; 726 patients who completely answered the EQ-5D on day 1 and day 28; and 627 patients who completely answered the EQ-5D on day 14 and day 28 (table 2).

Table 2 shows that there are statistically significant differences in the mean utilities of the patients in the usual care group between day 1 and day 14, with an increase in utility equal to 0.12161, which represents an increase of 0.00009 QALYs. There is also a statistically significant difference in the mean utilities of the patients in the usual care group between day 1 and day 28, with an increase in utility equal to 0.12432, which represents an increase of 0.00510 QALYs. The difference between the utilities from days 14 to day 28 is not statistically significant.

Table 2. Utilities of the usual care group between day 1 and day 14; between day 1 and day 28; and between day 14 and day 28.

	Day 1	**Day 14**	**N**	**Difference** [(1)]	**QALYs** [(4)]
Utility	0.80971	0.93133	632	+0.12161 [(*)]	+0.00009
	Day 1	**Day 28**	**N**	**Difference** [(2)]	**QALYs** [(5)]
Utility	0.81227	0.93660	726	+0.12432 [(*)]	+0.00510
	Day 14	**Day 28**	**N**	**Difference** [(3)]	**QALYs** [(6)]
Utility	0.93678	0.93525	627	-0.00152	-0.00002

N= sample size

[(1)] Difference = Day 14 utility – Day 1 utility

[(2)] Difference = Day 28 utility –Day 1 utility

[(3)] Difference = Day 28 utility – Day 14 utility

[(*)] Significance level equal to or lower than 0.05

[(4)] QALYs calculate as $\frac{Difference\ x\left(\frac{14}{365}\right)}{2}$

[(5)] QALYs calculate as $\frac{Difference\ x\left(\frac{28}{365}\right)}{2}$

[(6)] QALYs calculate as $\frac{Difference\ x\left(\frac{14}{365}\right)}{2}$

For the diagnostic intervention group, there are 830 patients who have completely answered the EQ-5D on day 1 and day 14; 918 patients who completely answered the EQ-5D on day 1 and day 28; and 842 patients who completely answered the EQ-5D on day 14 and day 28 (table 3).

Table 3 shows that there are statistically significant differences in the mean utilities of the

patients in the diagnostic intervention group between day 1 and day 14, with an increase in utility equal to 0.11448, which represents an increase of 0.00219 QALYs. There is also a statistically significant difference in the mean utilities of the patients in the diagnostic intervention group between day 1 and day 28, with an increase in utility equal to 0.11326, which represents an increase of 0.00465 QALYs. The difference between the utilities from days 14 to day 28 is not statistically significant.

Table 3. Utilities of the diagnostic intervention group between day 1 and day 14; between day 1 and day 28; and between day 14 and day 28.

	Day 1	**Day 14**	**N**	**Difference** [(1)]	**QALYs** [(4)]
Utility	0.80790	0.92239	830	+0.11448 [(*)]	+0.00219
	Day 1	**Day 28**	**N**	**Difference** [(2)]	**QALYs** [(5)]
Utility	0.81157	0.92483	918	+0.11326 [(*)]	+0.00465
	Day 14	**Day 28**	**N**	**Difference** [(3)]	**QALYs** [(6)]
Utility	0.92450	0.93188	842	+0.00738	+0.00014

N= sample size

[(1)] Difference = Day 14 utility – Day 1 utility

[(2)] Difference = Day 28 utility –Day 1 utility

[(3)] Difference = Day 28 utility – Day 14 utility

[(*)] Significance level equal to or lower than 0.05

[(4)] QALYs calculate as $\frac{Difference\ x\left(\frac{14}{365}\right)}{2}$

[(5)] QALYs calculate as $\frac{Difference\ x\left(\frac{28}{365}\right)}{2}$

[(6)] QALYs calculate as $\frac{Difference\ x\left(\frac{14}{365}\right)}{2}$

3.2 **Utilities obtained through the five dimensions of the EQ-5D. Static analysis**

If the analysis is static, that is, if differences in utilities are analyzed between each group (usual care and diagnostic intervention) on day 1, on day 14 and on day 28, it is not necessary to follow-up each patient: the answers given by all the patients in each group on each of the three days analyzed can be grouped. Therefore, for this analysis we have the following number of records: on day 1, 817 patients from the usual care group and 1,033 patients from the diagnostic intervention group have completely answered the form; on day 14, 718 patients from the usual care group and 934 patients from the diagnostic intervention group have completely answered the form and on day 28 there are 905 patients from the usual care group and 1,107 from the diagnostic intervention group have completely answered the form (table 4).

Table 4 shows the results in the mean utilities of each group in the three moments analyzed. Although the usual care group shows higher utilities than the diagnostic intervention group on the days analyzed, these differences were not statistically significant.

Table 4. Differences between usual care group and diagnostic intervention group at day 1, day 14 and day 28.

		Usual care group	Diagnostic intervention group	Difference [1]	
Day 1 [4]	**Utility**	0.80700	0.80550	-0.001493	
	N	817	1,033		
		Usual care group	**Diagnostic intervention group**	**Difference** [1]	**QALYs** [2]
Day 14 [4]	**Utility**	0.92854	0.92251	-0.00602	-0.0001440
	N	718	934		
		Usual care group	**Diagnostic intervention group**	**Difference** [1]	**QALYs** [3]
Day 28 [4]	**Utility**	0.93740	0.92923	-0.00817	-0.0001852
	N	905	1,107		

N= sample size

[1] Difference = Diagnostic intervention group utility – Usual care group utility

[2] QALYs calculate as $\frac{(-0.00602+(-0.001493))\ x\left(\frac{14}{365}\right)}{2}$

[3] QALYs calculate as $\frac{(-0.00817+(-0.001493))x\left(\frac{28}{365}\right)}{2}$

[4] The variances of samples are not equal: the two-sample t test is performed assuming unequal variances

3.3 VAS results. Dynamic analysis

To perform the dynamic analysis of the evolution of the VAS answers of the patients from day 1 to day 14, from day 1 to day 28 and between day 14 and day 28, it is necessary that the same patient has answered the pairs of days that are analyzed. For the usual care group, there are 630 patients who have completely answered the VAS on day 1 and day 14; 723 patients who have completely answered the VAS on day 1and on day 28; and 663 patients who completely answered the VAS on day 14 and day 28 (table 5).

Table 5 shows that there are statistically significant differences in the VAS scores of the patients in the usual care group between day 1 and day 14, with an increase equal to 23.92. There is also a statistically significant difference in the mean scores of the patients in the usual care group between day 1 and day 28, with an increase equal to 25.52. The difference between VAS scores from days 14 to 28 is not statistically significant.

Table 5. Results of the VAS scores of the usual care group between day 1 and day 14; between day 1 and day 28; and between day 14 and day 28.

	Day 1	Day 14	N	Difference [1]
VAS score	62.39	86.32	630	+23.92 [*]
	Day 1	**Day 28**	**N**	**Difference** [2]
VAS score	63.74	89.27	723	+25.52 [*]
	Day 14	**Day 28**	**N**	**Difference** [3]
VAS score	85.56	88.78	663	+3.22

N= sample size

[1] Difference = Day 14 VAS score – Day 1 VAS score

[2] Difference = Day 28 VAS score –Day 1 VAS score

[3] Difference = Day 28 VAS score – Day 14 VAS score

[*] Significance level equal to or lower than 0.05

For the diagnostic intervention group, there are 830 patients who have completely answered the VAS on day 1 and day 14; 915 patients who completely answered the VAS on day 1 and day 28; and 856 patients who completely answered the VAS on day 14 and day 28 (table 6).

Table 6 shows that there are statistically significant differences in the VAS scores of the patients in the diagnostic intervention group between day 1 and day 14, with an increase equal to 23.19. There is also a statistically significant difference in the mean scores of the patients in the diagnostic intervention group between day 1 and day 28, with an increase equal to 25.24. The difference between VAS scores from days 14 to 28 is not statistically significant.

Table 6. Results of the VAS scores of the diagnostic intervention group between day 1 and day 14; between day 1 and day 28; and between day 14 and day 28.

	Day 1	**Day 14**	**N**	**Difference** [(1)]
VAS score	62.56	85.75	830	+23.19 [(*)]
	Day 1	**Day 28**	**N**	**Difference** [(2)]
VAS score	63.30	88.54	915	+25.24 [(*)]
	Day 14	**Day 28**	**N**	**Difference** [(3)]
VAS score	85.72	88.88	856	+3.15

N= sample size
[(1)] Difference = Day 14 VAS score – Day 1 VAS score
[(2)] Difference = Day 28 VAS score –Day 1 VAS score
[(3)] Difference = Day 28 VAS score – Day 14 VAS score
[(*)] Significance level equal to or lower than 0.05

3.3 VAS results. Static analysis

In the static analysis, differences are analyzed in the scores given in the VAS between each group (usual care and diagnostic intervention) on day 1, on day 14 and on day 28. It is not necessary to follow up each patient, therefore that the answers given by all the patients in each group on each of the three days analyzed can be grouped.

Table 7 shows the results in the average VAS scores of each group at the three moments analyzed. Although the diagnostic intervention group shows higher scores on day 14 and day 28 than the usual care group, these differences were not statistically significant. The difference between the VAS scores of both groups on day 1 was also not statistically significant.

Table 7. Differences between usual care group and diagnostic intervention group at day 1, day 14 and day 28.

		Usual care group	Diagnostic intervention group	Difference [1]
Day 1[2]	**VAS score**	63.54	63.43	-0.11
	N	1,004	1,286	
		Usual care group	**Diagnostic intervention group**	**Difference** [1]
Day 14[2]	**VAS score**	85.27	85.47	+0.19
	N	717	934	
		Usual care group	**Diagnostic intervention group**	**Difference** [1]
Day 28[2]	**VAS score**	87.98	88.01	+0.03
	N	905	1,105	

N= sample size

[1] Difference = Diagnostic intervention group VAS score – Usual care group VAS score

[2] The variances of samples are not equal: the two-sample t test is performed assuming unequal variances

4. Discussion

The dynamic analysis of the utilities obtained from the application of the EQ-5D rates and the VAS scores shows that there are statistically significant differences in both groups between day 1 and day 14, as well as between day 1 and day 28. The static analysis has not shown statistically significant differences between both groups at the three moments analyzed. In both groups (usual care and diagnostic intervention) significant benefits (in terms of higher utilities and VAS scores) were achieved after 14 days of patient follow-up.

The dynamic analysis allows us to know the evolution of the patients' utility from day 1 to day 14, from day 1 to day 28 and between day 14 and day 28. It is analyzed whether there are statistically significant differences within each group at the three moments analyzed. Several observations have also been missed since not all patients have reported each pair of data.

Concluding, note that given these results we consider that it is not appropriate to apply the cost-utility analysis to these data since there are no differences between the usual care group and the diagnostic group as shown by the EQ-5D and the VAS.

5. References

1. EuroQol Research Foundation. User guide EQ5D. 2021; Available from: https://euroqol.org/wp-content/uploads/2023/11/Userguide-EQ5D-Y-23-07.pdf

2. EuroQol. Value sets. 2024. Available from: https://euroqol.org/information-and-support/resources/value-sets/

3. Devlin NJ, Shah KK, Feng Y, Mulhern B, van Hout B. Valuing health-related quality of life: An EQ-5D-5L value set for England. Health Econ. 2018;27:7–22.

4. Ludwig K, Graf von der Schulenburg J-M, Greiner W. German Value Set for the EQ-5D-5L. PharmacoEconomics. 2018;36:663–74.

5. Ramos-Goñi JM, Craig BM, Oppe M, Ramallo-Fariña Y, Pinto-Prades JL, Luo N, et al. Handling Data Quality Issues to Estimate the Spanish EQ-5D-5L Value Set Using a Hybrid Interval Regression Approach. Value Health. 2018;21:596–604.

6. Andrade LF, Ludwig K, Goni JMR, Oppe M, de Pouvourville G. A French Value Set for the EQ-5D-5L. PharmacoEconomics. 2020;38:413–25.

7. Rencz F, Brodszky V, Gulácsi L, Golicki D, Ruzsa G, Pickard AS, et al. Parallel Valuation of the EQ-5D-3L and EQ-5D-5L by Time Trade-Off in Hungary. Value Health J Int Soc Pharmacoeconomics Outcomes Res. 2020;23:1235–45.

ANNEX IV

ADEQUATE: Quality of life analysis

Value-Dx, WP5.3 University of La Rioja (Spain)

23th of July, 2024

1. Introduction

The European Quality of Life-5 Dimensions (EQ-5D), developed by the EuroQol Group, is a highly recommended instrument for economic evaluation studies of health technologies [1]. It is a health measurement tool widely used in medical research and quality of life assessment [2].

Each patient assesses its own health status by first responding by levels of severity, ranging from “no problems” to “extreme problems”, five dimensions related to fundamental aspects of health (mobility, self-care, usual activities, pain and depression) and then scoring his/her general condition through a Visual Analogue Scale (VAS) [3].

Based on the responses to the five dimensions, the patient's health status is determined, which could be used as an index that evaluates the impact that new medical or diagnostic interventions have on the utility perceived by the patient. To convert health states into utilities, the “EQ-5D rates” are applied [3]. These rates are different between countries, as perceptions of health and quality of life vary between different cultures and societies: what is considered good health may differ between countries due to differences in cultural values, beliefs and social norms [1]. Two patients from different countries but who have reported the same health status can obtain different utilities. Currently, EQ-5D rates have been published for 48 countries [4]. The calculated utilities can be included in cost-effectiveness studies as they can be converted into Quality-Adjusted Life Years (QALYs) [4].

A QALY is a measure used in economic evaluations of medical interventions. It is used to quantify the effectiveness of different medical treatments and compare their effectiveness in terms of quality-adjusted life years [5]. The EQ-5D can be used to assign utility values to different health states. These utility values are then used to calculate QALYs. Therefore, the EQ-5D provides a measure of health-related quality of life that is used as input in QALYs calculations, allowing comparisons between different medical interventions in terms of their impacts on the quality and quantity of life [6].

The VAS is a tool used in medicine and research to measure the intensity of a subjective experience, such as pain, fatigue, anxiety, among others. It consists of a horizontal line of fixed length, generally 10 centimeters, on which the patient is asked to mark a point that reflects the intensity of their experience at that moment [7]. The VAS allows the subjective experience of the patient to be captured in a quantitative way, thus allowing statistical analysis to be performed [8].

In this study, the EQ-5D form completed by ADEQUATE patients has been analyzed with the objective of knowing the effect that the new diagnostic intervention has on the patients' perceived utility and VAS score. These values are compared with those perceived by a control group of patients to whom the new diagnostic technique has not been applied and

who have also answered the same EQ-5D form.

2. Material and method

2.1. Data

The analysis is performed with the data included in the spreadsheets named F01a_elig_rand and EQ5D_youth. The first sheet shows the number of patients and the result of randomization to one of the two groups of the analysis: control group or diagnostic intervention group. The second sheet shows the answers given by each of the previous patient to the EQ-5D form at three different times: baseline (day 1), day 14 of follow-up and day 30 of follow-up.

During data cleaning, observations may be lost if the patient has not responded to the five dimensions of the form. To calculate the utilities perceived by each patient at the three moments analyzed, the complete health status must be available, that is, the patient must have answered the five dimensions. If a patient has not completed any of the five questions, the utility cannot be calculated and that record is left out of the calculations. No data extrapolation method has been used to fill in the dimensions that remain unanswered.

Based on the health status, the utilities will be calculated applying the rates specific to each country.

To calculate the gain or loss in QALYs reported by the new diagnostic innovation technique, the difference in utilities is multiplied by the proportion of the follow-up time period with respect to the year (14/365/ or 28/365) and the result is divided by two. In this way, it is assumed that the gain or loss of QALYs is progressive between the pairs of moments that are analyzed. Thus, we calculated the QALY gain as the area of a triangle whose base is the time period and the height the difference of utilities between the baseline and the moment of reference (day 14 or 28)

As indicated above, not all countries have published their EQ-5D rates. If a patient in the database comes from a country for which there is no such information, it has been decided to use the rates validated for the United Kingdom since the National Institute for Health and Care Excellence (NICE) highlights the EQ-5D as the preferred method for utility analysis focused on the adult population.

2.2. Dynamic and static analysis

Two types of analysis are performed, called dynamic analysis and static analysis. The dynamic analysis allows us to know the evolution of the patients' utility from baseline to day 14, from baseline to day 30 and between day 14 and day 30. It is analyzed whether there are statistically significant differences within each group at the three moments analyzed. To do this, the t-test for means of two paired samples is applied (since the same patient must be followed at each pair of moments). Only patients who have reported each pair of data appear here, so observations may also be lost if the patient has not responded on any of the three days analyzed.

The static analysis allows us to know if there are differences in the utilities between each group (control and diagnostic intervention) on the baseline day, on day 14 and on day 30.

As it is not necessary to follow up on each patient, the responses given by all patients in each group (on each of the three days analyzed) can be grouped. To do this, the Student's t-test will be applied to analyze whether there is a significant difference between the means of two different samples. First, it is checked if the variances of the two samples are equal, for this an F test is carried out. If the null hypothesis is rejected, it means that the variances of both samples are not equal. In this case, the two-sample t test is performed assuming unequal variances. If the result of the F testis the opposite, it means that the variances of both samples are equal, the t-test is performed for two samples assuming equal variances.

The Jarque-Bera test is used to evaluate whether a given data set follows a normal distribution. The Mann-Whitney U test is used when we want to compare two independent groups and cannot assume that the data follow a normal distribution. Results

3. Results

A total of 524 patients belonging to five countries appear within the spreadsheet named F01a_elig_rand: Germany, Greece, Spain, Switzerland and the United Kingdom. Of these five countries, Germany, Spain and the United Kingdom have their own EQ-5D rates that allow utilities to be calculated. Therefore, the UK rate is applied to data from patients in Greece and Switzerland (Table 1). Furthermore, 7 patients in each arm were not included in the final clinical analysis, as they missed some criteria, and, consequently, we used a 510 patients database for these calculations.

Table 1. Number of patients by countries and hospitals. Source of applied EQ-5D rates.

Country	Number of patients	Hospital	EQ-5D-5L rates applied
Germany	135	GE-P-03-Tubingen University Hospital	German [9]
Greece	172	GR-P-01-Aristotole University of Thessaloniki	English [10]
Spain	79	SP-P-01-Hospital Universitario 12 de Octubre	Spanish [11]
Switzerland	71	SW-P-01-Universit„ts-Kinderspital beider Basel	English [10]
Switzerland	26	SW-P-02 Bellinzona	English [10]
United Kingdom	27	UK-P-01-University Hospital Lewisham	English [10]
Total	510		

Of the 510 patients, 255 patients correspond to the control group and 255 patients to the diagnostic intervention group. It would be expected that three records would appear for each patient, so that each patient would have completed the EQ-5D instrument on the baseline (day 1), on day 14 and on day 30. However, of those 1,530 (510x3) expected records in the spreadsheet named EQ5D_youth, the one that contains this information, only 1,164 rows appear.

3.1 Utilities obtained through the five dimensions of the EQ-5D.
3.1.1 Dynamic analysis

To performed a dynamic analysis of the evolution of the patients' utility from baseline to day 14, from baseline to day 30 and between day 14 and day 30, it is necessary that the same patient has answered the questionnaire on the pairs of days that are analyzed. For the control group, there are 28 patients who have completely answered the EQ-5D on baseline and day 14; 21 patients who completely answered the EQ-5D on baseline and day 30; and 19 patients who completely answered the EQ-5D on day 14 and day 30 (table 2). Dataset showed a Normal distribution for all cases analyzed. Hence, Student's t-tests were used for the inference study.

Table 2 shows that there are statistically significant differences in the mean utilities of the patients in the control group between the baseline and day 14, with an increase in utility equal to 0.14246, which represents an increase of 0.00585 QALYs. There is also a statistically significant difference in the mean utilities of the patients in the control group between baseline and day 30, with an increase in utility equal to 0.16295, which represents an increase of 0.00669 QALYs. The difference between the utilities from days 14 to day 30 is not statistically significant.

Table 2. Utilities of the control group between baseline and day 14; between baseline and day 30; and between day 14 and day 30.

	Baseline	**Day 14**	**N**	**Difference** [(1)]	**QALYs** [(4)]
Utility	0.83153	0.97400	28	+0.14246 [(*)]	+0.00585
	Baseline	**Day 30**	**N**	**Difference** [(2)]	**QALYs**
Utility	0.82185	0.98480	21	+0.16295 [(*)]	+0.00669
	Day 14	**Day 30**	**N**	**Difference** [(3)]	**QALYs**
Utility	0.95575	0.98270	19	+0.02695	+0.00110

N= sample size
[(1)] Difference = Day 14 utility – Baseline utility
[(2)] Difference = Day 30 utility –Baseline utility
[(3)] Difference = Day 30 utility – Day 14 utility
[(*)] Significance level equal to or lower than 0.05
[(4)] QALYs calculate as $\frac{Difference\ x\left(\frac{30}{365}\right)}{2}$

For the diagnostic intervention group, there are 26 patients who have completely answered the EQ-5D on baseline and day 14; 23 patients who completely answered the EQ-5D on baseline and day 30; and 22 patients who completely answered the EQ-5D on day 14 and day 30 (table 3).

Table 3 shows that there are statistically significant differences in the mean utilities of the patients in the diagnostic intervention group between the baseline day and day 14, with an increase in utility equal to 0.14215, which represents an increase of 0.00584 QALYs. There is also a statistically significant difference in the mean utilities of the patients in the diagnostic intervention group between baseline and day 30, with an increase in utility equal to 0.15404, which represents an increase of 0.00633 QALYs. The difference between the utilities from days 14 to day 30 is not statistically significant.

Table 3. Utilities of the diagnostic intervention group between baseline and day 14; between baseline and day 30; and between day 14 and day 30.

	Baseline	Day 14	N	Difference [1]	QALYs [4]
Utility	0.84211	0.98426	26	+0.14215 [*]	+0.00584
	Baseline	**Day 30**	**N**	**Difference [2]**	**QALYs**
Utility	0.84191	0.99595	23	+0.15404 [*]	+0.00633
	Day 14	**Day 30**	**N**	**Difference [3]**	**QALYs**
Utility	0.98345	0.99713	22	+0.01368	+0.00056

N= sample size
[1] Difference = Day 14 utility – Baseline utility
[2] Difference = Day 30 utility –Baseline utility
[3] Difference = Day 30 utility – Day 14 utility
[*] Significance level equal to or lower than 0.05
[4] QALYs calculate as $\frac{Difference\ x\ \left(\frac{30}{365}\right)}{2}$

3.1.2 Static analysis

If the analysis is static, that is, if differences in utilities are analyzed between each group (control and diagnostic intervention) on the baseline, on day 14 and on day 30, it is not necessary to follow-up each patient: the answers given by all the patients in each group on each of the three days analyzed can be grouped. Therefore, for this analysis we have the following number of records: on the baseline 179 patients from the control group and 170 patients from the diagnostic intervention group have completely answered the form; on day 14, 41 patients from the control group and 38 patients from the diagnostic intervention group have completely answered the form and on day 30 there are 33 patients from the control group and 34 from the diagnostic intervention group have completely answered the form (table 4). Dataset showed a Normal distribution for all cases analyzed. Hence, Student's t-tests were used for the inference study.Table 4 shows the results in the mean utilities of each group in the three moments analyzed. The increase in utility in the control from baseline to 30 days is notably higher at 0.17 than as compared to the corresponding increase in intervention group, although not statistically significant.

Table 4. Differences between control group and diagnostic intervention group at baseline, day 14 and day 30.

		Control group	Diagnostic intervention group	Difference [1]	QALYs [2]
Baseline[3]	**Utility**	0.80830	0.83010	+0.02179	
	N	179	170		
		Control group	**Diagnostic intervention group**	**Difference** [1]	
Day 14[3]	**Utility**	0.96965	0.98460	+0.01494	+0.00150
	N	41	38		
		Control group	**Diagnostic intervention group**	**Difference** [1]	
Day 30[3]	**Utility**	0.98145	0.99376	+0.01231	+0.00140
	N	33	34		

N= sample size
[1] Difference = Diagnostic intervention group utility – Control group utility
[2] QALYs calculate as $\frac{Difference\ of\ utilities\ at\ day\ 14\ or\ 30 + difference\ of\ utilities\ at\ baseline\ x\ \left(\frac{30}{365}\right)}{2}$
[3] The variances of samples are not equal: the two-sample t test is performed assuming unequal variances

3.2 VAS results.

3.2.1 Dynamic analysis

To performed a dynamic analysis of the evolution of the VAS answers of the patients from baseline to day 14, from baseline to day 30 and between day 14 and day 30, it is necessary that the same patient has answered the pairs of days that are analyzed. For the control group, there are 25 patients who have completely answered the VAS on the baseline and day 14; 20 patients who have completely answered the VAS on the baseline and on day 30; and 20 patients who completely answered the VAS on day 14 and day 30 (table 5).

Table 5 shows that there are statistically significant differences in the VAS scores of the patients in the control group between the baseline day and day 14, with an increase equal to 38.48. There is also a statistically significant difference in the mean scores of the patients in the control group between the baseline day and day 30, with an increase equal to 48.07. The difference between VAS scores from days 14 to 30 is not statistically significant.

Table 5. Results of the VAS scores of the control group between baseline and day 14; between baseline and day 30; and between day 14 and day 30.

	Baseline	**Day 14**	**N**	**Difference** (1)
VAS score	49.13	87.61	25	+38.48 (*)
	Baseline	**Day 30**	**N**	**Difference** (2)
VAS score	43.92	92.00	20	+48.07 (*)
	Day 14	**Day 30**	**N**	**Difference** (3)
VAS score	86.20	94.01	20	+7.80

N= sample size
(1) Difference = Day 14 VAS score – Baseline VAS score
(2) Difference = Day 30 VAS score –Baseline VAS score
(3) Difference = Day 30 VAS score – Day 14 VAS score
(*) Significance level equal to or lower than 0.05

For the diagnostic intervention group, there are 23 patients who have completely answered the VAS on baseline and day 14; 21 patients who completely answered the VAS on the baseline and day 30; and 22 patients who completely answered the VAS on day 14 and day 30 (table 6).

Table 6 shows that there are statistically significant differences in the VAS scores of the patients in the diagnostic intervention group between baseline and day 14, with an increase equal to 41.62. There is also a statistically significant difference in the mean scores of the patients in the diagnostic intervention group between the baseline day and day 30, with an increase equal to 40.98. The difference between VAS scores from days 14 to 30 is not statistically significant.

Tabla 6. Results of the VAS scores of the diagnostic intervention group between baseline and day 14; between baseline and day 30; and between day 14 and day 30.

	Baseline	Day 14	N	Difference [1]
VAS score	50.25	91.88	23	+41.62 [*]
	Baseline	**Day 30**	**N**	**Difference [2]**
VAS score	53.25	94.23	21	+40.98 [*]
	Day 14	**Day 30**	**N**	**Difference [3]**
VAS score	89.03	90.93	22	+1.89

N= sample size
[1] Difference = Day 14 VAS score – Baseline VAS score
[2] Difference = Day 30 VAS score –Baseline VAS score
[3] Difference = Day 30 VAS score – Day 14 VAS score
[*] Significance level equal to or lower than 0.05

3..2.2 Static analysis

In the static analysis, differences are analyzed in the scores given in the VAS between each group (control and diagnostic intervention) on the baseline day, on day 14 and on day 30. It is not necessary to follow up each patient, therefore that the answers given by all the patients in each group on each of the three days analyzed can be grouped.

Of the 1,164 records, 564 rows contain the VAS information. Of the 564 rows completed (unlike the EQ-5D form, which requires information from five different dimensions, the VAS only collects a single numerical value), 333 rows include information answered on the baseline (158 patient records from the control group and 175 patient records from the diagnostic intervention group), 124 rows include information answered on day 14 (59 patient records from the control group and 65 patient records from the diagnostic intervention group) and 107 rows include information answered on day 30 (50 patient records from the control group and 57 patients from diagnostic intervention group). Table 7 shows the results in the average VAS scores of each group at the three moments analyzed. Again the increase in VAS from baseline to day 30 in the control group exceeds that in the intervention group, although not statistically significant.

Table 7. Differences between control group and diagnostic intervention group at baseline, day 14 and day 30.

		Control group	Diagnostic intervention group	Difference [1]
Baseline[2]	**VAS score**	49.21	50.91	+1.69
	N	158	175	
		Control group	**Diagnostic intervention group**	**Difference [1]**
Day 14[2]	**VAS score**	86.64	90.57	+3.93
	N	59	65	
		Control group	**Diagnostic intervention group**	**Difference [1]**
Day 30[2]	**VAS score**	92.77	91.21	-1.56
	N	50	57	

N= sample size
[1] Difference = Diagnostic intervention group VAS score – Control group VAS score
[2] The variances of samples are not equal: the two-sample t test is performed assuming unequal variances

4. Discussion

The dynamic analysis of the utilities obtained from the application of the EQ-5D rates and the VAS scores shows that there are statistically significant differences in both groups between the baseline and day 14, as well as between the baseline and day 30. The static analysis has not shown statistically significant differences between both groups at the three moments analyzed. In both groups (control and diagnostic intervention) significant benefits (in terms of higher utilities and VAS scores) were achieved after 14 days of patient follow-up. The benefit was slightly higher in the control than in the intervention group.

A loss of observations has been detected, since we would expected to have for each patient (524 in total) three records, as each patient should have completed the EQ-5D instrument on the baseline (day 1), on day 14 and on day 30. However, of those 1,530 (510x3) expected records in the spreadsheet named EQ5D_youth, the one that contains this information, only 1,164 rows appear. This means that 26% of the expected information is missing, since not all patients have completed the EQ-5D form in the three indicated days. The dynamic analysis allows us to know the evolution of the patients' utility from baseline to day 14, from baseline to day 30 and between day 14 and day 30. It is analyzed whether there are statistically significant differences within each group at the three moments. Several observations have also been missed since not all patients have reported each pair of data.

As a limitation of the study, it can be highlighted that the use of the EQ-5D form in the pediatric population has been questioned [12]. This population may have difficulty understanding and answering the abstract and complex questions on the EQ-5D due to their level of cognitive development. Additionally, younger children may have limited language skills to understand and answer questions on the EQ-5D format and the form may not detect relevant changes in children's health and quality of life, especially those related to development and growth [13].

To conclude, given these results, it is not appropriate to apply the cost-utility analysis to these data since there are no differences between the control group and the diagnostic group as shown by the EQ-5D and the VAS.

3. References

1. Utilización del cuestionario European Quality of Life-5 Dimensions (EQ-5D) para valorar la variación de la calidad de vida relacionada con la salud debida a la gripe [Internet]. [cited 2024 May 6]. Available from: https://www.gacetasanitaria.org/es-pdf-S0213911108000022

2. Herdman M, Badia X, Berra S. El EuroQol-5D: una alternativa sencilla para la medición de la calidad de vida relacionada con la salud en atención primaria. Aten Primaria. 2001; 28:425–30.

3. EuroQol Research Foundation. User guide EQ5D. 2021; Available from: https://euroqol.org/wp-content/uploads/2023/11/Userguide-EQ5D-Y-23-07.pdf

4. Cabasés JM. El EQ-5D como medida de resultados en salud. Gac Sanit. 2015; 29:401–3.

5. Alvis N, Valenzuela MT. Los QALYs y DALYs como indicadores sintéticos de salud. Rev Médica Chile. 2010; 138:83–7.

6. Cañón L OI, Rodríguez CI. Introducción al uso de QALYs y EQ-5D en la evaluación de tecnologías en Colombia. Rev CIFE Lect Econ Soc. 2011; 13:33–50.

7. Guevara-López U, Covarrubias-Gómez A, Delille-Fuentes R, Hernández-Ortiz A, Carrillo-Esper R, Moyao-Garcia D. Parámetros de práctica para el manejo del dolor agudo perioperatorio.

8. Morera Salas M, Cascante Arguedas M, Elizondo Zúñiga L. Medición de la calidad de vida de un grupo de personas de la tercera edad del Hospital de Guápiles: una aplicación del instrumento Euroqol-5D. Rev Cienc Adm Financ Segur Soc. 2005; 13:33–43.

9. Ludwig K, Graf von der Schulenburg J-M, Greiner W. German Value Set for the EQ-5D-5L. PharmacoEconomics. 2018; 36:663–74.

10. Dolan P. Modeling valuations for EuroQol health states. Med Care. 1997; 35:1095–108.

11. Ramos-Goñi JM, Craig BM, Oppe M, Ramallo-Fariña Y, Pinto-Prades JL, Luo N, et al. Handling Data Quality Issues to Estimate the Spanish EQ-5D-5L Value Set Using a Hybrid Interval Regression Approach. Value Health. 2018; 21:596–604.

12. Thorrington D, Eames K. Measuring Health Utilities in Children and Adolescents: A Systematic Review of the Literature. PloS One. 2015; 10:e0135672.

13. Challenges in Measuring and Valuing Children's Health-Related Quality of Life [Internet]. ISPOR Int. Soc. Pharmacoeconomics Outcomes Res. [cited 2024 May 7]. Available from: https://www.ispor.org/publications/journals/value-outcomes-spotlight/vos-archives/issue/view/finding-the-best-and-brightest-getting-a-leg-up-on-the-race-for-talent/challenges-in-measuring-and-valuing-children-s-health-related-quality-of-life

Repository for primary data[1]

History of change	
IMI2/INT/2016-00954	Version dated 2016
IMI2/INT/2016-00954 v. 2019:	Updated version Simplification cover page Replace ‘’publishable summary” by ‘’summary.to avoid any confusion with the dissemination level of the deliverable.

[1] Suggested headings


VALUE-Dx