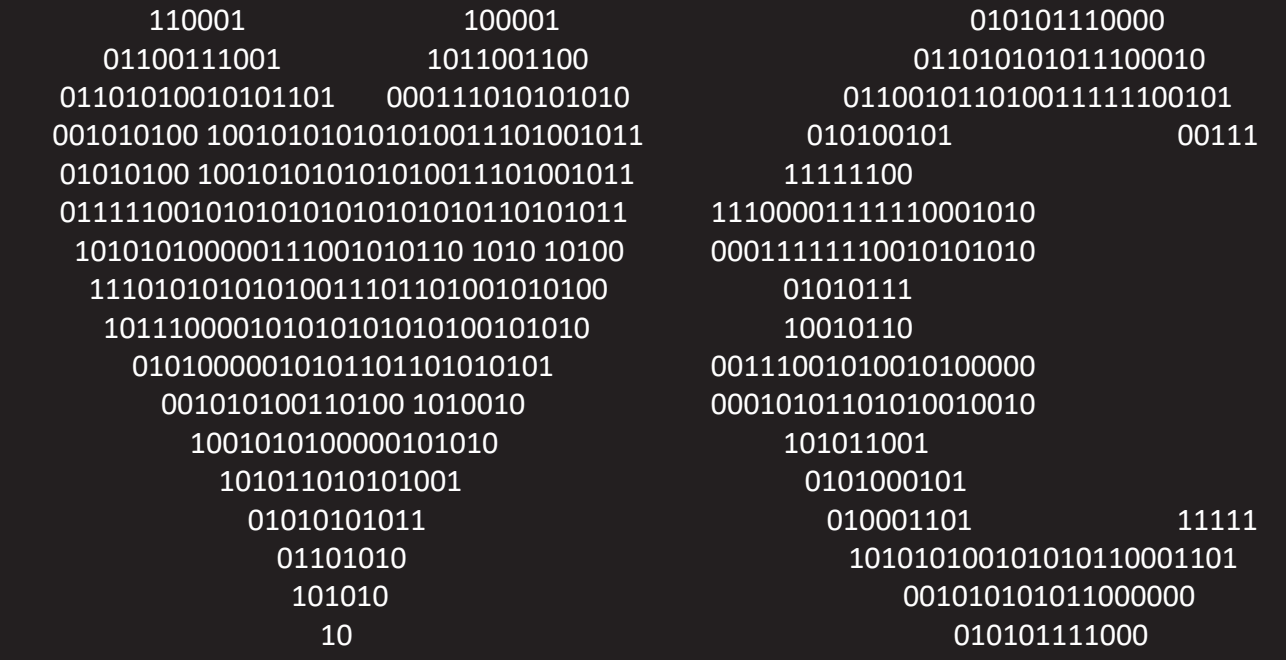
Developing Cost-Effective Analytics for Healthcare



Lytske Jantien Bakker

Developing Cost-Effective Analytics for Healthcare

Developing Cost-Effective Analytics for Healthcare

Ontwikkelen van kosteneffectieve analytics voor de zorg

Thesis

to obtain the degree of Doctor from the
Erasmus University Rotterdam
by command of the
rector magnificus

prof.dr. A.L. Bredenoord

and in accordance with the decision of the Doctorate Board.
The public defence shall be held on
9 February 2023
at 13:00 hrs

by

Lytske Jantien Bakker
born in Muscat, Oman

Erasmus University Rotterdam

Doctoral Committee:

Promotor:	prof.dr. C.A. Uyl - De Groot
Other members:	prof.dr. J.G. van der Hoeven prof.dr. P.R. Rijnbeek prof.dr. M.P.M.H. Rutten- van Mölken
Copromotors:	dr. W.K. Redekop dr. J.E.C.M. Aarts

COLOPHON

Author: Lytske Jantien Bakker
Lay-out: Anke Muijsers- Persoonlijk Proefschrift
Printing: Ipskamp Printing

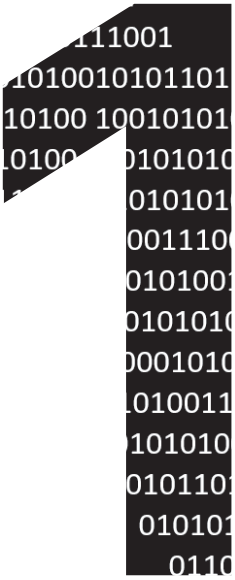
ISBN: 978-94-6421-991-3

For this PhD trajectory funding was received from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No. 644906. Printing was partially funded by the Erasmus School of Health Policy and Management

Enjoy Life. There's plenty of time to be dead - Hans Cristian Andersen

Voor Thomas

TABLE OF CONTENTS


111001
10100 10010101
0101010
010101
001110
010100
0101010
000101
010011
101010
010110
010101
0110

Introduction.

Healthcare systems are under increasing pressure to maintain the high quality of care currently provided with fewer resources. Due to the aging population, the demand for care will increase rapidly in Europe in the coming thirty years [1] while the supply of care will likely decline [2]. Many anticipate that the use of technology could decrease this gap between supply and demand [3], coinciding with a surge of projects that develop novel technologies for healthcare funded by the European Union.

In the past decade, expectations were particularly high for technologies that use analytics. Many noted that using analytics could lead to health benefits and/or cost savings for many areas of disease and clinical settings [4-17]. Alongside these high expectations, billions of dollars have been invested by both public and private entities to develop big data analytics and AI for healthcare [18-21].

However, the empirical evidence that healthcare analytics can solve an endless stream of problems remains grossly lacking. Moreover, development challenges and failures are gradually appearing in the literature probably coinciding with a spectacular loss of investment. In a recent review, researchers found that of 232 predictive analytics and 62 machine learning models for detection and prognostication of Covid-19 only 2 were eligible for large scale validation [22,23]. Further, ten years and billions of dollars after IBM Watson's winning Jeopardy, evidence of its benefits are absent, and evidence of failed development in hospitals such as the MD Anderson Cancer Center, the Memorial Sloan Kettering Cancer Center, and the University of North Carolina School of Medicine is widely reported [24-26].

In this dissertation, I aimed to explore the value of evaluating the potential of healthcare analytics alongside development to assist decision-making by developers and increase the likelihood of successful development. In this chapter, an introduction is provided for the topics addressed in more detail in the remainder of this thesis. Hereafter, the use cases are discussed that are examined in the remaining chapters of this dissertation. To conclude this first chapter, a brief overview of the remaining chapters and their contents is presented.

Analytics

Analytics has been defined as the discipline of analysis in which data is used to enable decision-making [27]. Furthermore, Bates *et al.* have defined analytics as the "discovery and communication of patterns in data" [5]. El Morr & Ali-Hassan define four types of analytics: descriptive, diagnostic, predictive and prescriptive (Table 1) [27]. Even though descriptive and diagnostic analytics offer relevant insights, we are often not only interested in knowing *which* events have happened and *why* these events have happened. Historical data can also be used to create a model which offers insights into *what will happen* in the future and how we can improve future decisions [27]. Models can be defined as '*a system of postulates, data and inferences presented as a mathematical description of an entity or state of affairs*' [28].

Table 1: Definition of different types of analytics according to Morr & Ali-Hassan [27].

Type of analytics	Definition	Example
Descriptive	Present patterns observed in data	Patients with diabetes have more hospital admissions than healthy patients
Diagnostic	Clarify the patterns observed	Patients with diabetes have more hospital admissions than patients without diabetes because they become hypoglycemic
Predictive	Knowing *what will happen* in the future	Patients with diabetes treated with sulfonylureas have a higher risk of hospitalization because of hypoglycemia
Prescriptive	Prescribe a treatment to realize an outcome	Patients with diabetes treated with sulfonylureas should carry a smart alarm to warn of hypoglycemia

Even though analytics have been around for a long time, *big* data analytics and artificial intelligence have renewed the interest in the topic (e.g., Mehta et al. 2018, Mehta et al. 2019 [4,29]). Big data analytics have previously been defined as analytics for data characterized by its complexity and the three V's (Volume, Variety and Velocity) [4]. Volume refers to the large size of the dataset, variety refers to data originating from many different sources whereas the velocity refers to the speed with which data is collected.

The term artificial intelligence (AI) was first used in the 20th century and is defined as the ability of computers to imitate human intelligence [27]. The term AI was first described by Alan Turing in 1950 but little progress was made in healthcare between its initial use and the year 2000 [30]. However, in the past decade, interest in its potential to improve healthcare has been renewed due to the progression in natural language processing, the availability of electronic data sources and improved hard- and software [30]. Moreover, expectations regarding its potential have been noted in many clinical settings and areas of healthcare [31].

Data sources

The data used to develop healthcare analytics can come from a wide variety of sources. No single classification of healthcare data is widely used although many authors have suggested ways in which to classify healthcare data. Mehta *et al.* have reported several ways in which 'big' healthcare data sources have been classified in the past [4]. Potential data sources include administrative databases (e.g., claims data, drug prescriptions), clinical data (e.g., electronic health records, laboratory information system data, imaging results, monitoring data (e.g., heart rate) and omics data) and patient generated data such as data obtained from social media, patient sensors and patient reported outcomes.

Besides the wide variety of data *sources*, data is frequently collected through two types of data *collection*: experimental and observational. The randomized controlled trial (RCT) is the golden standard intervention study if the aim is to establish the efficacy of a treatment. In an RCT, patients are randomly assigned to either the intervention or the control arm [32]. The aim of randomization is to increase internal validity by ensuring that individuals in the intervention and control arms differ only in the treatment they received, and not in any other ways; this

ensures that the observed effect cannot be attributable to confounding factors [32]. An RCT, however, is not always desirable, ethical, or feasible [33]. Important limitations of RCTs are that the generalizability of results can be limited, and follow-up is short [34]. The generalizability refers to the extent to which the treatment effect found in an RCT would also be found in daily practice. When performing an RCT, strict eligibility criteria apply and elderly patients, children, and patients with comorbidities are regularly excluded, thus limiting generalizability. Moreover, the time required to complete an RCT is not always available, thereby forcing researchers and developers to adopt a more pragmatic approach, such as an observational study.

In observational research, contrary to experimental research, the researcher does not actively assign an intervention to individuals but observes individuals in their natural setting. Most of the data sources used for development of analytics contain observational data. These often include a population representative of the general population and a prolonged duration of follow-up [34]. A limitation of observational research is that causation cannot be established [35], and results are at risk of being biased. Common forms of bias are confounding bias, information bias, and selection bias. Confounding bias occurs when the exposure and outcome variables share a common cause [36]. Selection bias occurs when selecting on a common effect [37] and information bias refers to bias caused by erroneous collection of data [38]. Thus, when using observational data, researchers should be aware of these risks.

Economic evaluations

Many authors have emphasized the potential for analytics to lead to health benefits and savings [4-17]. A means to measure the impact of a novel technology on health and financial benefits is by using economic evaluations. Economic evaluations assist decision-making of stakeholders by comparing costs and effects of alternative technologies. Several types of economic evaluations can be distinguished, depending on whether, and how, effects are measured [39]. In a cost-minimization analysis, health effects are assumed equal for the technologies compared, whereas in a cost-benefit analysis, outcomes are expressed in monetary terms. A cost-effectiveness analysis measures effects in natural units, such as mortality reduction or life years gained, while in cost-utility analyses, effects are often reported in quality adjusted life years (QALYs).

Economic evaluations are usually performed to assist market-access decisions of regulators and healthcare payers. However, they can also be used alongside development of healthcare technologies, aiding in design and investment decisions [40-44]. These ‘early’ economic evaluations may assist decision-making of developers, for instance to inform market-access and pricing strategies, or to identify relevant requirements of a technology and for go/no go decisions in the development phase [40,41,44]. Economic evaluations performed after development of technologies to assist decision-making of payers and regulators are referred to as ‘late’ economic evaluations.

Extrapolating Survival

There are multiple guidelines to assist researchers when performing an economic evaluation facilitating best-practice research [39,45]. Often, decision analytic models are used to combine input from a variety of sources to estimate cost-effectiveness. Moreover, decision analytic models enable researchers to estimate results beyond the duration of the clinical studies from which input parameters are derived. For instance, the follow-up in an RCT is 4 years, but

guidelines for economic evaluations recommend researchers estimate cost and health outcomes over the lifetime of patients instead of for that short 4-year period [45,46]. This can be done by fitting parametric models to the data available and using these models to estimate long-term outcomes (Figure 1). In Figure 1, the black Kaplan-Meier survival curve (KM data cut) reflects the true survival for a subset of patients for a short follow-up period. Here the red Kaplan-Meier curve represents the true long-term survival for the same subset of patients. In economic evaluations, the short-term survival data is used to model long-term survival represented in Figure 1 by the smoothed curves fitted for a time horizon exceeding 12 years.

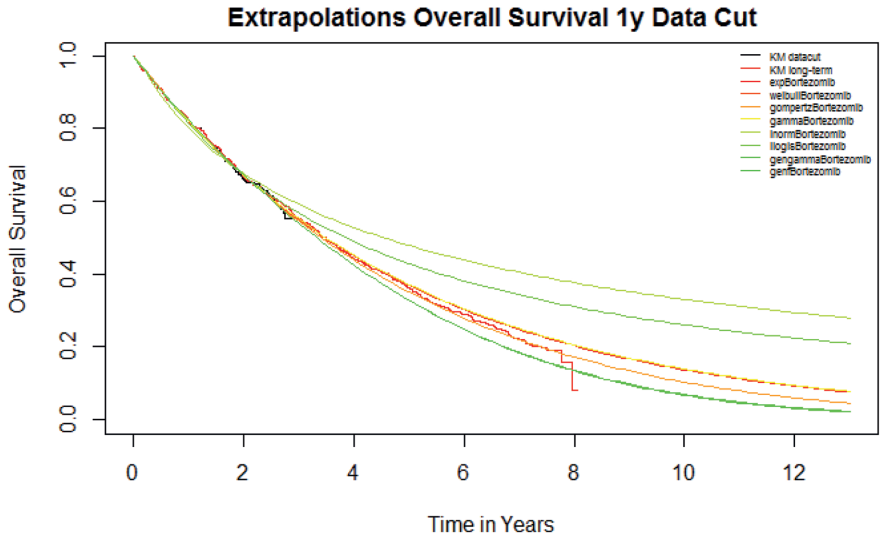


Figure 1: Here the Kaplan-Meier estimate refers to the empirical evidence available for patients with multiple myeloma treated with bortezomib. The smoothed curves represent survival estimated for 12 years using several standard parametric models

However, there is often considerable uncertainty surrounding long-term survival in any economic evaluation and different models can result in very different outcomes. In earlier studies, authors found that different models were most accurate for different durations of follow-up (a.k.a. 'data cuts') [47] which likely coincided with higher percentages censored and a lower number of absolute events. Moreover, the number of patients for which the time to event (i.e., death due to disease or other causes) was *not* reported (i.e., censored patients) was associated with increased error in survival estimates [48]. Insight into the impact using of shorter follow-up for extrapolations is relevant for economic evaluations that assess technologies where the follow-up of the patients included is relatively short.

Clinical Use Cases

There are many types of healthcare problems for which analytics can be developed and for which economic evaluations can be performed. In this dissertation use cases were derived from the AEGLE project. In this project, use cases were selected based on the variety of data sources available and the characteristics of the use case. The use cases differed in the type of data they

included (e.g., next generation sequencing (NGS), electronic health records (EHRs), monitoring data), the characteristics of the data (e.g., volume, velocity with which the data was collected and the variety of the data sources) as well as the type of disease addressed (e.g., acute care, non-malignant chronic disease, and hematological malignancies).

Intensive Care Unit

First, the intensive care unit is a fast-moving environment where patients can deteriorate rapidly. Decisions must be made quickly, and early detection of deterioration is considered essential to reduce the impact an intensive care unit (ICU) admission has on a patient's remaining life. Although the need for an ICU admission is not very common, the consequences are severe, not only in terms of lost health, but also in terms of costs. The average costs of an ICU Day exceed €2,000 [49,50] and thus improving outcomes and reducing length of stay for these patients may result in considerable savings.

The poor outcomes and high costs of patients in the ICU makes it an interesting setting for which to develop analytics that aim to improve provision of ICU care. Many application domains for big data analytics and AI have been suggested for the ICU, including predictions to optimize resource use (i.e., length of stay, readmissions), predictions of progression, sepsis, complications and mortality and analytics to optimize interaction between patients and mechanical ventilators [7,8]. A constant stream of patient level data is collected using electronic health records, bio-signal monitors, and mechanical ventilators (Figure 2). For instance, for mechanical ventilation alone, there are 236 variables intensivists should monitor [8] and the sheer volume of the data renders it impossible for health care professionals to process all variables without analytics.

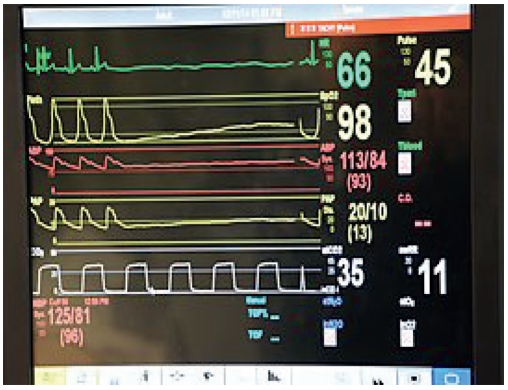


Figure 2: A display containing a few of the vital signs monitored in an intensive care environment. Source: Vital Signs Monitor Display, Petty Officer 1st Class James Stenberg [Internet] https://commons.wikimedia.org/w/index.php?search=hospital+monitor&title=Special:MediaSearch&go=Go&type=image. Taken February 11, 2014. Public domain.

Diabetes Mellitus

Care for patients with diabetes type 2 could be considered the opposite of ICU care. Diabetes is a chronic disease in which the inability to process or produce insulin results in elevated blood sugar levels [51]. It often takes many years for the disease to develop but the consequences can be

severe. Prolonged elevated blood sugar can lead to serious complications such as cardiovascular disease, neuropathy, foot ulceration and amputation, retinopathy, and kidney damage. In 2019, almost 9% of the European population was living with diabetes; 90% of them has type 2 diabetes [52]. Its financial impact is substantial, with annual European expenditures in 2019 exceeding 160 billion USD [52] and both the prevalence and expenditures are expected to steadily increase in the coming twenty years. Many treatments are available for type 2 diabetes that can be prescribed in various combinations and sequences. The many treatment options have resulted in large practice variation and uncertainty about the relative effectiveness of the different options.

This use case focused on data collected slowly and routinely in electronic health records in a hospital setting in Northern Ireland. Because of this variation in treatment, observational data from EHRs is considered a valuable source of information. The potential ways in which to support conclusions about effectiveness of treatments for diabetes patients using routinely collected data (e.g., EHRs, registries) has been emphasized by researchers [34,53]. However, there is also a need for caution since many challenges can arise when using EHR data [34,53] and the large practice variation requires access to big data sets to enable any meaningful analyses.

Two hematological malignancies

The last clinical use cases were two hematological malignancies. The first is chronic lymphocytic leukemia (CLL) which is characterized by its heterogeneous nature. CLL is the most common hematologic malignancy in the western world [54,55] with more than 12,000 new patients in the Europe each year [56]. Some of these new patients are treated upon diagnosis and have a short life expectancy while 40% of patients never require treatment and die long after diagnosis due to causes unrelated to the disease [55]. The treatments available for patients progressing are often costly [57] and given the variation in outcomes, optimally allocating these treatments is essential. When treatment is required, the drug administered will depend on patient characteristics (i.e., their overall 'fitness') and the chromosomal alterations the patient has.

For this use case the aim was to use data from next-generation sequencing (NGS) to develop prognostic indexes that enable clinicians to stratify patients according to their risk of progression and treatment response. With NGS, many genes are examined simultaneously, resulting in large, complex data sets [13]. Generating results valuable for clinical practice requires analytics and computing power due to the size and complexity of the data.

The second hematological malignancy, multiple myeloma (MM), is a rare, incurable oncological malignancy of plasma cells [58]. In Europe, roughly 40,000 people are diagnosed with MM annually and this number will increase to almost 46,000 by 2025 [59]. The treatment of patients with MM has evolved considerably in the past two decades. Where in 2004 the majority of patients were treated with chemotherapy-based regimens (e.g., melphalan), novel drugs have become available since then, such as immunomodulatory drugs (e.g., thalidomide, lenalidomide), followed by a proteasome inhibitor (bortezomib) [60], the use of autologous peripheral blood stem cell transplantation has offered substantial improvement for younger patients [61], and recently chimeric antigen receptor T cells therapies may offer better outcomes following a diagnosis with MM [58].

Where the previous use cases focused on the development of analytics and how economic evaluations can be used to estimate their impact, the final use case focused on the data required when performing these economic evaluations. The Netherlands Cancer Registry (NCR) has collected data on treatment and survival of patients with hematological malignancies for many years. The NCR has data on the care provided (e.g., treatments) for MM patients but also includes the long-term survival of these individuals. Databases such as the NCR offer unique opportunities to assess the accuracy of different models used to extrapolate survival, an essential component of many economic evaluations.

Thesis Aim

The aim of this dissertation was to assess the potential of using economic evaluations to assist decision-making of developers of healthcare analytics. When initiating this dissertation, it was apparent that, despite the many promises, performing good quality economic evaluations for adopting health information technologies in clinical practice was not common practice [13,62-64]. Therefore, my aim was to address this gap and increase the likelihood that future development and implementation of technologies that use healthcare analytics may succeed. I assessed how economic evaluations may assist decision-making of developers of healthcare analytics. First, the current use of economic evaluations to evaluate healthcare analytics was explored. Hereafter the ways in which economic evaluations can assist decisions-making of analytics development were analyzed and recommendations were formulated how they should be performed alongside development.

Thesis Outline

In the first part of this thesis, the limited evidence on cost-effectiveness of novel analytics is discussed (Chapter 2). For Chapter 2, the current use of economic evaluations to assess healthcare analytics was explored. In addition to examining the studies that have been performed, areas for improvement were identified. Hereafter, in Chapter 3, I presented the results from an early cost-effectiveness analysis in which the potential of analytics for the intensive care by identifying suboptimal interaction between a patient and their mechanical ventilator is explored. The availability of routinely collected data from sources such as electronic health records has increased the possibilities for observational research. However, conducting observational research using EHRs can be challenging, and results are at risk of being biased. In Chapter 4, I discussed how target trial emulation can assist researchers using observational data to identify and assess the ability to adjust for confounding and other forms of bias while considering the limitations of the dataset such as missing data. Hereafter, for Chapter 5 a framework was developed to assist decision makers when using economic evaluations to guide development. As discussed in Chapter 4, there are important limitations to historical data sources, and these should be considered at an early stage during development. In Chapter 6, I discussed the challenges that might occur when extrapolating long-term survival from short-term data. For analytics, RCTs are rare, and long-term follow-up data for extrapolating survival is lacking. Therefore, insights into the consequences of extrapolating using shorter duration of follow-up can assist decision making on future collection of data on efficacy and effectiveness of analytics. For Chapter 7, the main findings of this dissertation were summarized. In this discussion I reflected on how economic evaluations can assist decision-making of developers and how they should be used during the process of development.

REFERENCES

1 Corselli-Nordblad & Strandell. Ageing Europe — looking at the lives of older people in the EU — 2020 edition. Published 2020, Eurostat, Luxembourg
2 SER. 2020. Zorg voor de toekomst: over de toekomstbestendigheid van de zorg. Pg 65. [Internet] https://www.ser.nl/-/media/ser/downloads/adviezen/2020/zorg-voor-de-toekomst.pdf Accessed September 30, 2022.
3 WRR. 2021. Kiezen voor houdbare zorg Mensen, middelen en maatschappelijk draagvlak
4 Mehta N, Pandit A. Concurrence of big data analytics and healthcare: a systematic review. Int J Med Inform. 2018;114:57–65.
5 Bates DW, Saria S, Ohno-Machado L, et al. Big data in health care: Using analytics to identify and manage high-risk and high-cost patients. Health Aff (Millwood). 2014;33:1123–31.
6 Celi LA, Mark RG, Stone DJ, et al. "Big data" in the intensive care unit. Closing the data loop. Am J Respir Crit Care Med. 2013;187(11):1157.
7 Sanchez-Pinto LN, Luo Y, Churpek MM. Big data and data science in critical care. Chest. 2018;154(5):1239–48.
8 Gutierrez G. Artificial intelligence in the intensive care unit. Critical Care. 2020 Dec;24:1-9
9 Janke AT, Overbeek DL, Kocher KE, et al. Exploring the Potential of Predictive Analytics and Big Data in Emergency Care. Ann Emerg Med. 2016;67:227–36.
10 Hemingway H, Asselbergs FW, Danesh J, et al. Big data from electronic health records for early and late translational cardiovascular research: challenges and potential. Eur Heart J. 2017;39(16):1481–95.
11 Rumsfeld JS, Joynt KE, Maddox TM. Big data analytics to improve cardiovascular care: promise and challenges. Nat Rev Cardiol. 2016;13(6):350.
12 Hofmann-Apitius M. Is dementia research ready for big data approaches? BMC Med. 2015;13:145.
13 Phillips KA, Trosman JR, Kelley RK, et al. Genomic sequencing: assessing the health care system, policy, and big-data implications. Health Aff (Millwood). 2014;33(7):1246-1253.
14 Roski J, Bo-Linn GW, & Andrews TA. Creating value in health care through big data: opportunities and policy implications. Health Aff (Millwood). 2014;33(7):1115-1122.
15 Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential Review. Health Inf Sci Syst. 2014;2:3.
16 Yang C, Kong G, Wang L, et al. Big data in nephrology: Are we ready for the change? Nephrology. 2019;24:1097–102.
17 Kayyali B, Knott D, Van Kuiken S. The big-data revolution in US health care: Accelerating value and innovation. Mc Kinsey & Company. 2013 Apr;2(8):1-3.
18 Morello L, Guglielmi G. US science agencies set to win big in budget deal. Nature. 2018;555(7698).
19 Banks, MA. Sizing up big data. Nat Med 2020;26:5–6
20 Kisner J. Creating Shareholder Value with AI? Not so Elementary, My Dear Watson. 2017.
21 Furlow, B. ASCO Announces Funding for CancerLinQ Clinical Data Analysis Initiative. 2012; [Internet] https://www.cancertherapyadvisor.com/home/cancer-topics/general-oncology/asco-announces-funding-for-cancerlinq-clinical-data-analysis-initiative/. Accessed September 29, 2022
22 Roberts M, Driggs D, Thorpe M, et al. Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans. Nat Mach Intell. 2021 Mar;3(3):199-217.
23 Wynants L, Van Calster B, Collins GS, et al. Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. BMJ. 2020 Apr 7;369.
24 Advisory Board 2021. 10 years ago, IBM's Watson threatened to disrupt health care. What happened? [Internet] https://www.advisory.com/daily-briefing/2021/07/21/ibm-watson Accessed September 29, 2022
25 LPB. 2018. Overpromising and Underdelivering at Sloan Kettering: is AI Still 'Human' After All? [Internet] https://d3.harvard.edu/platform-rctom/submission/overpromising-and-underdelivering-at-sloan-kettering-is-ai-still-human-after-all/ Accessed September 29, 2022.
26 Pringle S. Scoop: IBM tries to sell Watson Health again, [Internet] https://www.axios.com/ibm-tries-to-sell-watson-health-again-82f691a4-ab81-4b2b-a5bb-13a7556c8ef1.html Accessed March 23, 2022

27 El Morr C, Ali-Hassan H. Analytics in Healthcare: A Practical Introduction. Springer; 2019 Jan 21.
28 Merriam Webster. Model [Internet] https://www.merriam-webster.com/dictionary/model Accessed May 18, 2022.
29 Mehta N, Pandit A, Shukla S. 2019. Transforming healthcare with big data analytics and artificial intelligence: A systematic mapping study. J Biomed Inform, 100: 103311.
30 Kaul V, Enslin S, Gross SA. The history of artificial intelligence in medicine. Gastrointest Endosc. 2020 Jun 18.
31 Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med; 2019 Jan;25(1):44-56
32 Kirkwood BR, Sterne JA. Essential medical statistics. John Wiley & Sons; 2010 Sep 16.
33 García-Albéniz X, Hsu J, Hernan M. 2017. The value of explicitly emulating a target trial when using real world evidence: an application to colorectal cancer screening. Eur J Epidemiol, 32.6: 495-500.
34 Farmer R, Mathur R, Bhaskaran K, Eastwood SV, Chaturvedi N, Smeeth L. Promises and pitfalls of electronic health record analysis. Diabetologia. 2018 Jun 1;61(6):1241-8.
35 Dickerman BA, García-Albéniz X, Logan RW, Denaxas S, Hernán MA. Avoidable flaws in observational analyses: an application to statins and cancer. Nat Med. 2019 Oct;25(10):1601-6.
36 Hernán MA, Robins JM. Causal Inference: What If. Chapman & Hall/CRC, Boca Raton; 2020. [cited 2020 Nov 1] Available from: https://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/
37 Hernán MA, Hernández-Díaz S, Robins JM. A structural approach to selection bias. Epidemiology 2004 Sep 1:15(5):615-25
38 Tripepi G, Jager KJ, Dekker FW, Zoccali C. Selection bias and information bias in clinical research. Nephron Clin Pract. 2010;115(2):c94-9.
39 Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. Methods for the economic evaluation of health care programmes. Oxford university press; 2015 Sep 25.
40 Buisman LR, Rutten-van Mölken MPMH, Postmus D, et al. The early bird catches the worm: early cost-effectiveness analysis of new medical tests. Int J Technol Assess Health Care. 2016;32(1–2):46–53.
41 Pietzsch JB, Paté-Cornell ME. Early technology assessment of new medical devices. Int J Technol Assess Health Care. 2008;24(1):36–44.
42 Sculpher M, Drummond M, Buxton M. The iterative use of economic evaluation as part of the process of health technology assessment. J Health Serv Res Policy; 1997 Jan;2(1):26-30.
43 Grutters JP, Govers T, Nijboer J, Tummers M, Van Der Wilt GJ, Rovers MM. Problems and promises of health technologies: the role of early health economic modeling. Int J Health Policy Manag. 2019 Oct;8(10):575.
44 Hartz S, John J. Contribution of economic evaluation to decision making in early phases of product development: a methodological and empirical review. Int J Technol Assess Health Care. 2008 Oct;24(4):465-72.
45 National Health Care Institute. Guideline for Economic Evaluations in Health care. Diemen: National Health Care Institute the Netherlands. 2015.
46 Sharma D, Aggarwal AK, Downey LE, Prinja S. National healthcare economic evaluation guidelines: a cross-country comparison. Pharmacoecon Open. 2021 Sep;5(3):349-64.
47 Bullement A, Willis A, Amin A, Schlichting M, Hatswell AJ, Bharmal M. Evaluation of survival extrapolation in immuno-oncology using multiple pre-planned data cuts: learnings to aid in model selection. BMC Med Res Methodol. 2020 Dec;20(1):1-4.
48 Everest L, Blommaert S, Chu RW, Chan KK, Parmar A. Parametric Survival Extrapolation of Early Survival Data in Economic Analyses: A Comparison of Projected Versus Observed Updated Survival. Value Health. 2021 Nov 24.
49 Hakkaart-van Roijen L, van der Linden N, Bouwmans C, Kanters T, Tan SS. Bijlage 1. Kostenhandleiding: methodologie van kostenonderzoek en referentieprijzen voor economische evaluaties in de gezondheidszorg. Institute for Medical Technology Assessment, Erasmus Universiteit Rotterdam, Rotterdam; 2015: 1-120 (in Dutch).
50 McLaughlin AM, Hardt J, Canavan JB, Donnelly MB. Determining the economic cost of ICU treatment: a prospective “micro-costing” study. Intensive Care Med. 2009 Dec;35(12):2135-40.

51 World Health Organization. Diabetes [Internet] https://www.who.int/news-room/fact-sheets/detail/diabetes Accessed May 6, 2021
52 International Diabetes Federation Report 2019 [Internet] https://www.diabetesatlas.org/upload/resources/material/20200302_133351_IDFATLAS9e-final-web.pdf Accessed September 30, 2022.
53 Riddle MC, Blonde L, Gerstein HC, et al. Diabetes Care Editors' Expert Forum 2018: managing big data for diabetes research and care.
54 Eichhorst B, Robak T, Montserrat E, et al. Chronic lymphocytic leukemia: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol. 2020. In Press.
55 Baliakas P, Mattsson M, Stamatopoulos K, Rosenquist R. Prognostic indices in chronic lymphocytic leukemia: where do we stand how do we proceed? J Intern Med. 2016 Apr;279(4):347-57.
56 Scarfò L, Ferreri AJ, Ghia P. Chronic lymphocytic leukaemia. Crit Rev Oncol Hematol. 2016 Aug 1;104:169-82.
57 Chen Q, Jain N, Ayer T, et al. Economic burden of chronic lymphocytic leukemia in the era of oral targeted therapies in the United States. J Clin Oncol. 2017 Jan 10;35(2):166.
58 Mikkilineni L, Kochenderfer JN. CAR T cell therapies for patients with multiple myeloma. Nat Rev Clin Oncol. 2021 Feb;18(2):71-84.
59 Myeloma Patients Europe. Incidence of Myeloma. [Internet] https://www.mpeurope.org/about-myeloma/incidence-of-myeloma/ Accessed September 30, 2022.
60 Cavo M, Terpos E, Bargay J, et al. The multiple myeloma treatment landscape: international guideline recommendations and clinical practice in Europe. Expert Rev Hematol. 2018 Mar 4;11(3):219-37.
61 Minnema MC, Van der Spek E, Van de Donk NW, Lokhorst HM. New developments in the treatment of patients with multiple myeloma. Neth J Med. 2010 Jan 1;68(1):24-32.
62 Bassi J, Lau F. Measuring value for money: a scoping review on economic evaluation of health information systems. J Am Med Inform Assoc. 2013 Jul 1;20(4):792-801.
63 O'Reilly D, Tarride JE, Goeree R, Lokker C, McKibbon KA. The economics of health information technology in medication management: a systematic review of economic evaluations. J Am Med Inform Assoc. 2012 May 1;19(3):423-38.
64 Gospodarevskaya E, Westbrook JI. Correspondence: Call for discussion about the framework for categorizing economic evaluations of health information systems and assessing their quality. J Am Med Inform Assoc. 2014 Jan;21(1):190.


0001110101010101
0101010011101001011
01010101011
1010101011
001010100
100
101010
01010101
100 1010010
0000101010
01010100
0101011
101010
01010

Economic evaluations of big data analytics for clinical decision-making: a scoping review.

Bakker L, Aarts J, Uyl-de Groot C, Redekop W.

Journal of the American Medical Informatics Association. 2020 Sep;27(9):1466-75.

ABSTRACT

Objective: Much has been invested in big data analytics to improve health and reduce costs. However, it is unknown whether these investments have achieved the desired goals. We performed a scoping review to determine the health and economic impact of big data analytics for clinical decision-making.

Materials and Methods: We searched Medline, Embase, Web of Science and the National Health Services Economic Evaluations Database for relevant articles. We included peer-reviewed papers that report the health economic impact of analytics that assist clinical decision-making. We extracted the economic methods and estimated impact, and also assessed the quality of the methods used. In addition, we estimated how many studies assessed 'big data analytics' based on a broad definition of this term.

Results: The search yielded 12,133 papers but only 71 studies fulfilled all eligibility criteria. Only a few papers were full economic evaluations; many were performed during development. Papers frequently reported savings for healthcare payers but only 20% also included costs of analytics. Twenty studies examined 'big data analytics' and only 7 reported both cost-savings and better outcomes.

Discussion: The promised potential of big data is not yet reflected in the literature, partly since only a few full and properly performed economic evaluations have been published. This and the lack of a clear definition of 'big data' limit policymakers and healthcare professionals from determining which big data initiatives are worth implementing.

INTRODUCTION

Extracting valuable knowledge from big healthcare data has been an important aim of many research endeavors and commercial entities. While no clear definition for big data is available, it is often described according to its complexity and the characteristics of the data such as the size of a dataset (Volume), the speed with which data is retrieved (Velocity) and the fact that the data comes from many different sources (Variety)[1]. Bates et al. [2] emphasize that big data comprises *both* the data with its large volume, variety and velocity, as well as the use of analytics. In this respect, analytics are the 'discovery and communication of patterns in data'.

Big data's potential to assist clinical decision-making has been expressed for a variety of clinical fields such as the intensive care [3,4], emergency department [2,5], cardiovascular diseases [6,7], dementia [8], diabetes [9], oncology [10-12], and asthma [13]. Big data analytics could also lead to economic benefits [1,2,14-17]. Annual savings for the United States (US) healthcare system of providing timely, personalized care have been estimated to exceed US$140 billion [18].

Over the years, much has been invested to achieve the promised benefits of big data. For instance, the US has invested millions in their Big Data to Knowledge centers [19]. While in Europe, many calls and projects in Europe's Horizon 2020 program have focused on the use of Big Data for better healthcare (e.g., AEGLE, OACTIVE, BigMedylitics). In 2018, US$290 million was allocated to The All of Us initiative which aims to personalize care using a wide variety of data sources (e.g., genomic data, monitoring data, electronic health record data) from one million US citizens [20]. The investments by governments are far exceeded by the investments in 'big data technologies' in the commercial sector [21]. For example, IBM has already invested billions of dollars in 'Dr. Watson' and big data analytics [22], and Roche purchased FlatIron Health for US$1.9 billion in 2018 [11].

For optimal spending of scarce resources, economic evaluations can be used to assess the (potential) return on investment of novel technologies. Economic evaluations are comparative analyses of the costs and consequences of alternative courses of action [23]. Economic evaluations that provide evidence on the health and economic impact of a technology can assist decision-making and justify further investments required to achieve a technology's potential. Despite the promise that big data analytics can lead to savings, it is unclear whether this promise is corroborated by good evidence. *Therefore, we aimed to determine the health and economic impact of big data analytics to support clinical decision-making.* Given the absence of a clear definition for big data, we first determined how analytics impacted clinical practice. We then considered which of these analytics could be classified as big data analytics.

METHODS

The study follows the Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) Checklist [24].

Search strategy and study inclusion

Since there is no consensus on the definition of big data [1], we widened the scope of our search to identify economic evaluations of a variety of analytics. An information specialist from the

Academic Library at the Erasmus University Medical Centre was consulted when developing the search strategy (Supplementary Appendix A). In the search strategy, we included MesH and title/abstract terms related to (big data) analytics, economic evaluations, and health care. These terms for (big data) analytics included artificial intelligence, tools used to extract patterns from big data such as machine learning, and generic tools that use analytics to enable decision-making such as clinical decision support. We combined these with terms such as economic evaluations and cost-effectiveness and terms to exclude studies that had no relation to healthcare (e.g., veterinary care).

All major databases were searched (Embase, Medline, Web of Science, and the NHS Economic Evaluations Database). We included all English, peer-reviewed, primary research papers and limited our search to studies of humans. The primary search was performed in March 2018 and updated in December 2019. Initial screening was performed by one author (LB). Hereafter, all studies about which there was uncertainty regarding their inclusion were discussed with two other authors (JA, WR). Studies were included if they met the following criteria: a) the study reported pattern discovery, interpretation, and communication to assist decision-making of clinical experts at the individual patient level; b) the study implemented analytics in clinical practice using computerized technology; and c) the study reported a monetary estimation of the potential impact of the analytics. Application of these three criteria led to the exclusion of studies that only reported time or computation savings and studies that did not assist clinical experts at the individual patient level. Thus, we did not include studies that informed guidelines or policymakers. We also excluded analytics that produced results that could be easily printed on paper for use in clinical practice (e.g., Ottawa Ankle Rules) and studies that simply used data mining technologies to extract records from an electronic health record (EHR) but not to perform any analyses of the extracted data.

Data extraction

Data extraction was performed by one author (LB). For a random 10% of papers data was extracted by a second author (KR) to check for concordance. In the end, there were no significant differences in the results. We extracted the following data for each study; patient population, description of the technology in which the analytics are embedded (i.e., clinical decision support systems), the analytics used for discovery and communication of patterns in data, description of the data, the intervention and the comparator in the economic evaluation, the perspective, outcomes, and costs included, results, recommendations, and conflicts of interest. Conflicts of interest included those reported in the paper (related and unrelated), commercial employment of authors and funding by industry.

We also reported the type of economic evaluation (e.g., full, or partial) that was used. A full economic evaluation compares two or more alternatives and includes both costs and consequences. Partial economic evaluations do not contain a comparison or exclude either costs or consequences [23]. Thus, when a study reported cost estimates but no health outcomes they were classified as partial. For full economic evaluations we reported the ratio of costs over effects, also known as the incremental cost-effectiveness ratio (ICER). Furthermore, economic evaluations can offer valuable insights for decision-makers at many different stages in the development process (e.g., during and after development) [25]. After development, they can assist healthcare payers when choosing novel technologies in which to invest their constrained

budget. During development, an 'early' economic evaluation can assist developers by identifying minimal requirements of the technology, areas for further research, and viable exploitation and market access strategies [25-27].

In our results, we also distinguished in which stage of development the economic evaluation was performed. If a study provided recommendations for developing a technology that did not exist, it was categorized as 'before' development. Studies were categorized as 'during' development when the economic evaluation was performed and presented alongside development unless the aim of the study was to inform purchasing decisions of funding bodies (i.e., perspective of the National Healthcare Services) or when the analytics were already implemented in clinical practice. All remaining studies were categorized as being performed 'after' development.

We also performed an analysis to identify economic evaluations that might be classified as 'big' data analytics. We used broad criteria to select the highest possible number of papers to sketch a best-case scenario. We defined these criteria based on the volume, variety, and velocity of the data. We classified papers as having big volume when they utilized next generation sequencing (NGS) data, EHR records or claims data with a sample size of more than 100,000 units (e.g., patients, admissions), and all imaging papers published after 2013. Papers were included because of their variety when they combined multiple datatypes (e.g., structured, and unstructured data, combining multiple data sources). All papers that used monitoring data published after 2013 were included because they might fulfil the velocity criteria.

RESULTS

The initial search yielded 12,133 records of which seventy-one papers were included in the final analysis after title/abstract and full-text screening (Figure 1). Important exclusion criteria for full-text papers were that no monetary estimates were included and that no analytics were used.

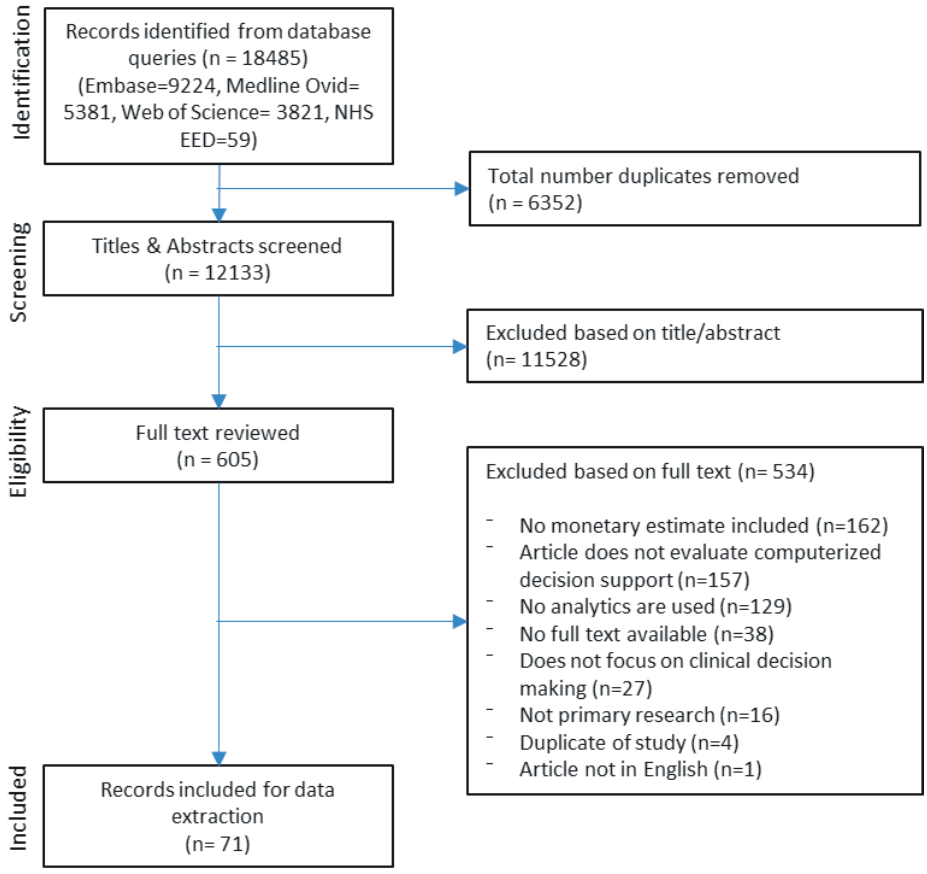


Figure 1: PRISMA flowchart

Summary of papers

We found that all papers could be classified into four categories according to the type of data that was used; medical history databases (e.g., data from EHRs, clinical trial databases, claims databases), imaging data, monitoring data (e.g., continuous data collection using sensors), and omics data (e.g., proteomics, genomics, transcriptomics, metabolomics) (Table 1). Almost all papers originated from North America and Europe (87%). The US was well represented with 39 papers mainly focusing on the use of medical history and omics data. The number of papers originating from Europe was considerably lower (n=20) while few or no papers originated from South America, Australia, and Africa. There has been a clear increase in the number of publications from 2016 onwards (Figure 2).

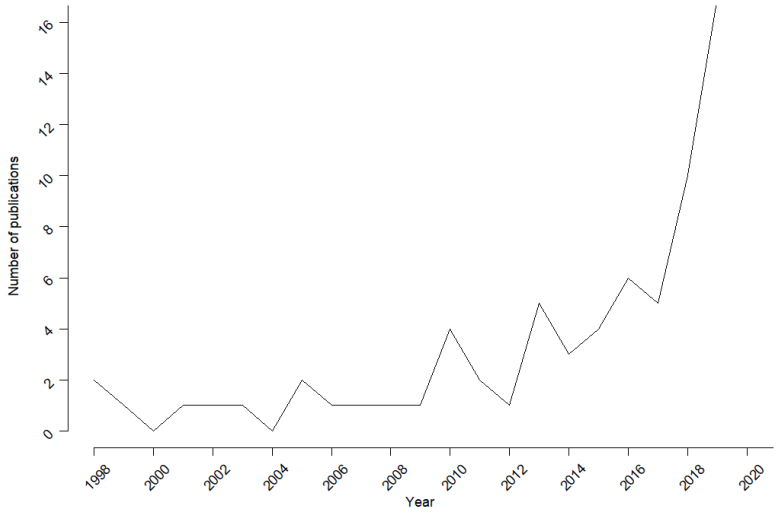


Figure 2: Number of publications according to the year of publication

Most studies were partial economic evaluations and found that analytics may improve outcomes and generate savings. A perspective was not often reported, and no study reported a societal perspective. Almost all partial economic evaluations reported savings compared to half of the studies reporting results from full economic evaluations. When grouped according to conflict of interest, no significant differences were found in the percentage of studies that reported savings and improved health. For economic evaluations without a conflict of interest, 61% were performed during development compared to 22% with no conflict of interest. All but one reported savings.

In the following paragraphs we will discuss economic results for all four data types. An overview of the economic results for all papers can be found in Supplementary Appendix B. A detailed description of all analytics and data used can be found in Supplementary Appendix C.

Table 1: Summary of all records according to data type used

	Total	Medical history	Imaging	Monitoring	Omics
Total	71	44	8	8	11
Continent					
North America	42	27	3	3	9
Europe	20	11	2	5	2
Asia	7	5	2	-	-
Africa	1	-	1	-	-
South America	1	1	-	-	-
Australia	-	-	-	-	-
Type of economic evaluation					
Full	22	8	5	2	7
Partial	49	36	3	6	4
Perspective					
Payer perspective	7	3	-	-	4
National healthcare system	8	3	1	2	2
Provider perspective	3	1	-	1	-
Other	2	-	2	-	-
No perspective reported	52	37	5	5	5
Stage of development					
Before development	1	1	-	-	-
During development	33	31	2	-	-
After development	37	12	6	8	11
Measure of effectiveness					
QALYs and Life Years	15	5	4	2	4
Model Performance	29	27	2	-	-
Other	20	10	2	3	5
Not included	7	2	-	3	2
Incremental health effects					
Decrease in effects	5	2	1	-	2
No difference	5	3	-	2	-
Increase in effects	41	23	7	4	7
Not included	20	16		2	2
Incremental costs					
Savings	54	39	5	5	5
No difference	5	2	-	3	-
Increase in costs	12	3	3	-	6

Table 1: *Continued.*

	Total	Medical history	Imaging	Monitoring	Omics
Include costs of implementing analytics	22	2	5	5	10
Recommendations for research & development					
Focus development on improving the analytics	30	23	2	2	3
Validation and feasibility of implementation	19	13	3	2	1
Development for other clinical areas or subgroups	11	8	2	1	-
Pricing and economics of the analytics	9	2	3	-	4
Cost-effectiveness research	5	4	-	1	-
Development of the intervention that follows	3	3	-	-	-
Multidisciplinary collaboration	2	2	-	-	-
Refer to big data in the text	6	6	-	-	-
Potential to be classified as big data analytics	20	8	5	4	3
QALYs= Quality Adjusted Life Years					

Analytics for medical history data

The first category consisted of studies that used historic databases containing information on patient demographics and medical history (e.g., test results and drug prescriptions) (n=44) [28-71]. All papers presented predictive or prescriptive analytics that assist clinical decision-making using a variety of techniques (regression, support vector machines, Markov decision processes). The risk of readmission (n=9) and problems pertaining to the emergency department (n=5) were most often examined and one study addressed pediatric care [42]. Both structured data such as demographics and laboratory results, as well as unstructured data such as free text messages (n=4) [37,40,44,50] were used and the sample size varied from N=80 patients [65] to more than 800,000 urine samples [68]. This was the only category in which authors referred to the term 'big data' (n=6) [32,35,37,40,50,60].

Most of the studies in this category were partial economic evaluations (n=36) and most were conducted during development (n=31). Results were often limited to model performance (e.g., classification accuracy, area under the curve) and were rarely translated into health benefits such as quality-adjusted life-years. Almost all studies found that the analytics could lead to monetary savings, yet only two papers included implementation costs of the analytics [33,62]. These costs could for instance consist of licensing costs and costs of implementing analytics within a hospital system. Authors often recommended to continue development and focus on improving the analytics. Furthermore, the need for further validation prior to implementation was frequently emphasized.

Analytics for imaging data

Eight studies presented predictive analytics for seven different types of imaging data (CT, MRI, Chest radiographs, digital cervical smears, mammographies, digital photographs and ventilation-

perfusion lung scans) [72-79]. The number of full economic evaluations [73,75-77,79], and studies performed after development [73-76,78] were both higher than the first group of papers that used medical history data. Four studies measured effects in (quality-adjusted) life-years [73,75-77], and more than half of the studies included implementation costs of analytics [73-77]. The number of studies that found the analytics could lead to cost-savings was once again quite high (63%) [72-74,78,79]. Just like the studies that used medical history data, authors of studies in this category emphasized the need for further validation prior to implementation. However, several studies also emphasized the balance between the requirements of the technologies (e.g., test sensitivity) and potential health benefits and cost-savings [75,76,79].

Analytics for monitoring data

Monitoring data collected with a variety of devices and sensors (e.g., airflow monitoring, continuous glucose monitoring, continuous performance tests, infrared cameras, vital signs monitors) was used in eight studies [80-87]. Five of these studies reported descriptive analytics that monitored patient outcomes and compared this to a range or reference value [81,83-85,87]. This group of papers differed from those using imaging and medical history data since most analytics were implemented in a medical device. All technologies were evaluated after development of which many were partial economic evaluations. Roughly half of the studies resulted in more effects [82-84,86], savings [82,84-87], and included costs of the device and/or analytics [81,86,87].

Analytics for omics data

Eleven papers reported the potential impact of predictive and prescriptive analytics of omics data, often with the aim of applying them as a test in clinical practice [88-98]. Only two of these papers focused on the use of Next Generation Sequencing data [94,96], and one paper combined multiple types of data (pharmacogenomics, literature, medical history) [89]. The remaining papers utilized microarray data and all the analytics that were adopted as a test were used in oncology (n=9) [88,90-93,95-98].

Compared to the other categories, the percentage of full economic evaluations was high [90,92,93,95-98]. In half of the studies the perspective used were that of the payer or the healthcare system. Furthermore, just like the studies that used monitoring data, all economic evaluations were performed after development. Seven studies reported increased effects [88,90-93,96,97], and six studies reported that use of analytics would increase costs [90,93-95,97,98]. All but one study included the costs of the analytics or the test in which the analytics were implemented [89]. Moreover, unlike the other categories, several papers discussed price thresholds at which the analytics or the test would be cost-neutral, dominant (i.e., more effects and lower costs) or thresholds at which the analytics or test would be cost-effective (i.e., where the ICER would be below a specific cost-effectiveness threshold).

Big Data Analytics

We found that less than a third of all papers (n=20) might fulfil criteria for classification as 'big data analytics' (Table 2). Most papers were included because their volume might be large enough to be considered big data (e.g., N>100,000, imaging data) and studies that used monitoring data were included because of the potential speed with which the data is collected (velocity). Eight of these papers used medical history data [32,37,40,44,45,50,60,68], five used imaging data

[72-74,76,78], four used monitoring data [80,82,83,87], and three used omics data [89,94,96]. Most were partial economic evaluations (n=15) and twelve were performed after development. All but five [44,76,80,83,94] corroborated expectations that big data analytics could result in cost-savings, varying from US$126 per patient [89] to more than US$500 million for the entire US healthcare system [72]. However, only a handful of papers included the costs of the analytics [73,74,76,87,94,96].

Table 2: Classification of papers that could be defined as 'big data' studies based on the criteria of volume, velocity, and variety. These papers represent a subset of the initial 71 papers.

	Volume			Velocity	Variety
Article	**Next generation sequencing data**	**Medical history data with n > 100,000**	**Imaging data published after 2013**	**Monitoring data published after 2013**	**Combines multiple datatypes**
Burton 2019		X			
Duggal 2016					X
Golas 2018					X
Hunter-Zinck 2019		X			X
Jamei 2017		X			
Lee 2015		X			X
Rider 2019		X			
Wang 2019		X			
Carballido-Gamio 2019			X		
Crespo 2019			X		
Philipsen 2015			X		
Sato 2014			X		
Sreekumari 2019			X		
Brocklehurst 2018				X	
Calvert 2017				X	
Hollis 2018				X	
Sánchez-Quiroga 2018				X	
Brixner 2016					X
Mathias 2016	X				
Nicholson 2019	X				
Total	**2**	**6**	**5**	**4**	**5**

DISCUSSION

In this review, we aimed to determine the health and economic impact of big data analytics for clinical decision-making. We found that expectations of big data analytics with respect to savings and health benefits are not yet reflected in the academic literature. Most studies are partial economic evaluations and the costs of implementing analytics are scarcely included in the calculations. To ensure optimal decision-making, guidelines recommend a full economic evaluation that includes all relevant costs for payers (e.g., costs of analytics). Our results align with earlier research noting deployment costs are rarely considered while these costs can be a major barrier to successfully implementing analytics [99].

We found that a small subset might be classified as big data analytics. We adopted a broad definition of big data to maximize the number of studies that would be considered as studies of big data. Therefore, the actual number of studies would be even lower if papers were to be assessed by a panel of experts. This corroborates a previous study from 2018 which found that quantified benefits of big data analytics are scarce [1].

The studies were grouped into four categories according to the data sources used, which were similar to those reported by Mehta et al [1]. Two main differences were that we grouped all databases that reported information relating to a patient's medical history (instead of separating claims and EHR data) and we included a category that evaluated analytics for monitoring data generated in the hospital. This category was not available in the classification used by Mehta et al. However, they reported some categories (e.g., social media and wearable sensors) that are not yet represented in the literature on economic evaluations. None of the studies evaluated technologies that used patient generated data collected using different methods such as healthcare trackers.

Recommendations for future economic evaluations

Good policymaking decisions about the use of analytics requires knowledge of the impact that the analytics will have on costs and health outcomes. With this in mind, policymakers could provide incentives to developers of analytics to perform good-quality economic evaluations. Economic evaluations of analytics are still scarce and the studies that were available often did not adhere to best-practice guidelines, thereby limiting their value to inform decision-making. Often a partial instead of a full economic evaluation was performed, costs of purchasing and implementing the analytics were excluded, or only intermediate outcomes were reported. For payers and policymakers, excluding for instance the costs of the analytics could result in an underestimation of the investment needed to implement the technology or an overestimation of its financial benefits. By means of incentives, policymakers could stimulate developers to adhere to guidelines and best practice recommendations (e.g., Drummond [23], Buisman [26], Morse [99]). This could improve the quality of results and thus their ability to inform decision-making.

We found a relatively high number of studies that performed an economic evaluation of analytics during development, compared to other fields (e.g., drug or medical device development) [100,101]. A possible explanation for this is the high costs of validating and deploying analytics which are known to be an important barrier of implementation [11,99]. Few artificial intelligence and big data analytics solutions have been implemented successfully [3,11]. To overcome this

challenge, Frohlich et al. recommend the use of pilot trials to illustrate the potential effectiveness and efficiency of analytics. These results can then be used to find new investors for clinical research [11]. In our results, we also saw that those without a conflict of interest (e.g., academia) were more inclined to publish during development which might be explained by the need to attract funders for further development.

Defining big data to assist evaluation

Without consensus on a definition, no objective assessment can be made as to whether investments following the introduction of big data in healthcare have realized expectations, whether they can be considered good value for money and whether future investments should be stimulated. In our analysis, we found that it is likely that a small number of studies have performed an economic evaluation of big data analytics. However, this absolute number is uncertain since a clear definition of 'big data' is still lacking almost ten years after its introduction in healthcare. For policymakers and those that wish to practice evidence-based medicine, it is essential to know where and how big data analytics would result in health and financial benefits before investing in products described in mainstream media as 'big data' technologies (e.g., Afirma GSC, YouScript) [102,103]. This remains a challenging task if there is no consensus on its definition. Therefore, we recommend experts in the field to reconsider the possibility of generating a quantitative definition of big data in healthcare.

Defining big data is no easy task and we think that a definition will only be accepted by the healthcare field if it is developed by a multidisciplinary collaboration of experts from academia, healthcare organizations, insurers, federal entities, policymakers, and commercial parties. Many authors have described the term in slightly different words [1,104], some have tried to quantify [105], and others have purposefully refrained from doing so [14]. Auffray et al [14] stated in 2016 that a single definition of big data would probably be '*too abstract to be useful*' and proposed the use of a workable definition in which big data covers the high volume and diversity of data sources managed with best-practice technologies such as advanced analytics solutions. However, descriptions such as 'best-practice', 'advanced' [14], or 'traditional' [106] are time-dependent. What is 'traditional' in 2014 is not necessarily 'traditional' in 2020. Thus, perhaps a definition of big data should quantify the 'data' element, include a concrete list of analytics that are considered advanced or best practice, be time-dependent, and be updated regularly. We recognize that it might be extremely difficult to achieve wide consensus and we do not think this can be realized without support from academic, clinical, policy, federal and commercial stakeholders.

Limitations

One limitation of our research is that economic evaluations do not always describe the analytics element of the intervention that was being evaluated. For instance, in studies of omics data the papers generally referred to the tool (e.g., Afirma GSC) but did not describe the analytics used in this tool. One way to ensure that economic evaluations that assess a big data technology are included in future reviews would be to specify explicit tools that might contain big data analytics (e.g., Afirma GSC) for each data type in a search strategy. However, such a list is likely to be very long, and this will also be challenging without a definition of big data. Research into the economic value of big data analytics might also be facilitated by better reporting in economic evaluations on the data and analytics used for development. Another limitation is that studies

that did not refer to cost estimations in their title/abstract were excluded. This could have led to exclusion of studies that perform a cost estimation but do not report this as a primary outcome in the abstract. A possible solution for future research would be to include studies for full-text screening when one of the authors is a health economist or employed in a health policy or economics department.

Also, since our review included only published economic evaluations, it is possible that our results are influenced by the absence of an incentive to submit an academic paper and by publication bias. Commercial developers do not always have an incentive to publish but do have an incentive to market their products using the results of economic analyses. If these studies do not include costs of analytics in their estimation of benefits, this would only underline the importance of our recommendations. It is also possible that studies that do not find a technology cost-effective include costs of analytics more often and are rejected for publication because of negative results.

Methodological limitations were that study selection and data extraction were performed by a single reviewer due to the size of the hits from the search strategy, and the fact that Business Review Complete (BSC) was not included in the literature search. While this may have resulted in the exclusion of some relevant studies, we expect this number to be small. Moreover, this does not affect the conclusions of our study. Our search was limited to analytics for decision-making of clinical experts at the individual patient level. There are many other ways in which analytics could improve health such as managing epidemics and policy making to improve population health that were beyond the scope of this paper. To conclude, it is possible that developers sometimes have a valid reason for not including costs of analytics which we did not consider in this study.

CONCLUSION

This is the first study to assess the health and economic impact of big data analytics for clinical decision-making. At present the potential benefits of big data analytics for clinical practice cannot yet be corroborated with academic literature despite high expectations. We found that economic evaluations were sometimes used to estimate the potential of analytics. However, many studies were partial economic evaluations and did not include costs of implementing analytics. Therefore, economic evaluations that adhere to best practice guidelines should be encouraged. This and the lack of an appropriate definition of big data complicates justification of future expenses and makes it exceedingly difficult to determine whether expectations of big data analytics have thus far been realized. Therefore, we recommend key experts in the field of data science in healthcare to reconsider the possibility to define big data analytics for healthcare.

REFERENCES

1 Mehta N, Pandit A. Concurrence of big data analytics and healthcare: A systematic review. Int J Med Inform. 2018;114:57–65.
2 Bates DW, Saria S, Ohno-Machado L, et al. Big data in health care: Using analytics to identify and manage high-risk and high-cost patients. Health Aff (Millwood). 2014;33:1123–31.
3 Sanchez-Pinto LN, Luo Y, Churpek MM. Big data and data science in critical care. Chest. 2018;154(5):1239–48.
4 Celi LA, Mark RG, Stone DJ, et al. "Big data" in the intensive care unit. Closing the data loop. Am J Respir Crit Care Med. 2013;187(11):1157.
5 Janke AT, Overbeek DL, Kocher KE, et al. Exploring the Potential of Predictive Analytics and Big Data in Emergency Care. Ann Emerg Med. 2016;67:227–36.
6 Hemingway H, Asselbergs FW, Danesh J, et al. Big data from electronic health records for early and late translational cardiovascular research: challenges and potential. Eur Heart J. 2017;39(16):1481–95.
7 Rumsfeld JS, Joynt KE, Maddox TM. Big data analytics to improve cardiovascular care: promise and challenges. Nat Rev Cardiol. 2016;13(6):350.
8 Hofmann-Apitius M. Is dementia research ready for big data approaches? BMC Med. 2015;13:145.
9 Alyass A, Turcotte M, Meyre D. From big data analysis to personalized medicine for all: challenges and opportunities Review. BMC Med Genomics. 2015;8:33.
10 Beckmann JS, Lew D. Reconciling evidence-based medicine and precision medicine in the era of big data: challenges and opportunities Review. Genome Med. 2016;8:134.
11 Fröhlich H, Balling R, Beerenwinkel N, et al. From hype to reality: Data science enabling personalized medicine. BMC Med. 2018;16.
12 Phillips KA, Trosman JR, Kelley RK, et al. Genomic sequencing: assessing the health care system, policy, and big-data implications. Health Aff (Millwood). 2014;33(7):1246-1253.
13 Prosperi M, Min JS, Bian J, et al. Big data hurdles in precision medicine and precision public health. BMC Med Inform Decis Mak. 2018;18(1):139.
14 Auffray C, Balling R, Barroso I, et al. Making sense of big data in health research: towards an EU action plan. Genome Med. 2016;8(1):71.
15 Roski J, Bo-Linn GW, & Andrews TA. Creating value in health care through big data: opportunities and policy implications. Health Aff (Millwood). 2014;33(7):1115-1122.
16 Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential Review. Health Information Science and Systems. 2014;2:3.
17 Yang C, Kong G, Wang L, et al. Big data in nephrology: Are we ready for the change? Nephrology. 2019;24:1097–102.
18 Groves P, Kayyali B, Knott D, et al. The 'big data' revolution in healthcare: Accelerating value and innovation. 2016.
19 Big Data to Knowledge. 2019; [Internet] https://commonfund.nih.gov/bd2k. Accessed January 8, 2019.
20 Morello L, Guglielmi G. US science agencies set to win big in budget deal. Nature. 2018;555(7698).
21 Banks, MA. Sizing up big data. Nat Med 2020;26:5–6
22 Kisner J. Creating Shareholder Value with AI? Not so Elementary, My Dear Watson. 2017.
23 Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. Methods for the economic evaluation of health care programmes. Oxford university press; 2015.
24 Tricco AC, Lillie E, Zarin W et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann. Intern. Med. 2018;169:467-73
25 Ijzerman MJ, Steuten LMG. Early assessment of medical technologies to inform product development and market access. Appl Health Econ Health Policy. 2011;9(5):331–47.
26 Buisman LR, Rutten-van Mölken MPMH, Postmus D, et al. The early bird catches the worm: early cost-effectiveness analysis of new medical tests. Int J Technol Assess Health Care. 2016;32(1–2):46–53.
27 Pietzsch JB, Paté-Cornell ME. Early technology assessment of new medical devices. Int J Technol Assess Health Care. 2008;24(1):36–44.

28 Ashfaq A, Sant'Anna A, Lingman M, et al. Readmission prediction using deep learning on electronic health records. J Biomed Inform. 2019;97.
29 Bennett CC, Hauser K. Artificial intelligence framework for simulating clinical decision-making: A Markov decision process approach. Artif Intell Med. 2013;57:9–19.
30 Bremer V, Becker D, Kolovos S, et al. Predicting Therapy Success and Costs for Personalized Treatment Recommendations Using Baseline Characteristics: Data-Driven Analysis. J Med Internet Res. 2018;20:e10275.
31 Brisimi TS, Xu T, Wang T, et al. Predicting diabetes-related hospitalizations based on electronic health records. Stat Methods Med Res. 2019;28:3667–82.
32 Burton RJ, Albur M, Eberl M, Cuff SM. Using artificial intelligence to reduce diagnostic workload without compromising detection of urinary tract infections. BMC Med Inform Decis Mak. 2019;19:171.
33 Chae YM, Lim HS, Lee JH, et al. Development of an intelligent laboratory information system for community health promotion center. Medinfo. 2001;10:425–8.
34 Chi CL, Street WN, Katz DA. A decision support system for cost-effective diagnosis. Artif Intell Med. 2010;50:149–61.
35 Choi HS, Choe JY, Kim H, et al. Deep learning based low-cost high-accuracy diagnostic framework for dementia using comprehensive neuropsychological assessment profiles. BMC Geriatr. 2018;18:234.
36 Cooper GF, Abraham V, Aliferis CF, et al. Predicting dire outcomes of patients with community acquired pneumonia. J Biomed Inform. 2005;38:347–66.
37 Duggal R, Shukla S, Chandra S, et al. Predictive risk modelling for early hospital readmission of patients with diabetes in India. International Journal of Diabetes in Developing Countries. 2016;36:519–28.
38 Elkin PL, Liebow M, Bauer BA, et al. The introduction of a diagnostic decision support system (DXplainTM) into the workflow of a teaching hospital service can decrease the cost of service for diagnostically challenging Diagnostic Related Groups (DRGs). Int J Med Inform. 2010;79:772
39 Escudero J, Ifeachor E, Zajicek JP, et al. Machine learning-based method for personalized and cost-effective detection of alzheimer's disease. IEEE Trans Biomed Eng. 2013;60:164–8.
40 Golas SB, Shibahara T, Agboola S, et al. A machine learning model to predict the risk of 30-day readmissions in patients with heart failure: a retrospective analysis of electronic medical records data. BMC Med Inform Decis Mak. 2018;18:44.
41 Govers TM, Rovers MM, Brands MT, et al. Integrated prediction and decision models are valuable in informing personalized decision making. J Clin Epidemiol. 2018;104:73–83.
42 Grinspan ZM, Patel AD, Hafeez B, et al. Predicting frequent emergency department use among children with epilepsy: A retrospective cohort study using electronic health data from 2 centers. Epilepsia. 2018;59:155–69.
43 Harrison D, Muskett H, Harvey S, et al. Development and validation of a risk model for identification of non-neutropenic, critically ill adult patients at high risk of invasive Candida infection: the Fungal Infection Risk Evaluation (FIRE) Study Review. Health technology assessment (Winchester, England). 2013;17:1–156.
44 Hunter-Zinck HS, Peck JS, Strout TD, et al. Predicting emergency department orders with multilabel machine learning techniques and simulating effects on length of stay. J Am Med Inform Assoc. 2019;26:1427–36.
45 Jamei M, Nisnevich A, Wetchler E, et al. Predicting all-cause risk of 30-day hospital readmission using artificial neural networks. PLoS One. 2017;12.
46 Javitt JC, Steinberg G, Locke T, et al. Using a claims data-based sentinel system to improve compliance with clinical guidelines: results of a randomized prospective study. Am J Manag Care. 2005;11:93–102.
47 Kocbek P, Fijacko N, Soguero-Ruiz C, et al. Maximizing Interpretability and Cost-Effectiveness of Surgical Site Infection (SSI) Predictive Models Using Feature-Specific Regularized Logistic Regression on Preoperative Temporal Data. Comput Math Methods Med. 2019;2019.
48 Kofoed K, Zalounina A, Andersen O, et al. Performance of the TREAT decision support system in an environment with a low prevalence of resistant pathogens. J Antimicrob Chemother. 2009;63:400–4.
49 Ladabaum U, Wang G, Terdiman J, Blanco A, Kuppermann M, Boland CR, Ford J, Elkin E, Phillips KA. Strategies to identify the Lynch syndrome among patients with colorectal cancer: a cost-effectiveness analysis. Ann Intern Med. 2011:155:69-79.

50 Lee EK, Atallah HY, Wright MD, et al. Transforming Hospital Emergency Department Workflow and Patient Care. Interfaces. 2015;45:58–82.
51 Lee HK, Jin R, Feng Y, et al. An Analytical Framework for TJR Readmission Prediction and Cost-Effective Intervention. IEEE J Biomed Health Inform. 2018;23:1760–72.
52 Lin Y, Huang S, Simon GE, et al. Cost-effectiveness analysis of prognostic-based depression monitoring. IISE Transactions on Healthcare Systems Engineering. 2019;9:41–54.
53 Ling CX, Sheng VS, Yang Q. Test strategies for cost-sensitive decision trees. IEEE Transactions on Knowledge and Data Engineering. 2006;18:1055–67.
54 Marble RP, Healy JC. A neural network approach to the diagnosis of morbidity outcomes in trauma care. Artif Intell Med. 1999;15:299–307.
55 Monahan M, Jowett S, Lovibond K, et al. Predicting out-of-office blood pressure in the clinic for the diagnosis of hypertension in primary care: An economic evaluation. Hypertension. 2018;71:250–61.
56 Murtojärvi M, Halkola AS, Airola A, et al. Cost-effective survival prediction for patients with advanced prostate cancer using clinical trial and real-world hospital registry datasets. Int J Med Inform. 2020;133.
57 Parekattil SJ, Fisher HAG, Kogan BA. Neural network using combined urine nuclear matrix protein-22, monocyte chemoattractant protein-1 and urinary intercellular adhesion molecule-1 to detect bladder cancer. J Urol. 2003;169:917–20.
58 Petersen ML, LeDell E, Schwab J, et al. Super learner analysis of electronic adherence data improves viral prediction and may provide strategies for selective HIV RNA monitoring. J Acquir Immune Defic Syndr. 2015;69:109–18.
59 Pölsterl S, Singh M, Katouzian A, et al. Stratification of coronary artery disease patients for revascularization procedure based on estimating adverse effects. BMC Med Inform Decis Mak. 2015;15:9.
60 Rider NL, Miao D, Dodds M, et al. Calculation of a primary immunodeficiency "risk vital sign" via population-wide analysis of claims data to aid in clinical decision support. Frontiers in Pediatrics. 2019;7.
61 Ruff C, Koukalova L, Haefeli WE, et al. The role of adherence thresholds for development and performance aspects of a prediction model for direct oral anticoagulation adherence. Frontiers in Pharmacology. 2019;10.
62 Shazzadur Rahman AAM, Langley I, Galliez R, et al. Modelling the impact of chest X-ray and alternative triage approaches prior to seeking a tuberculosis diagnosis. BMC Infect Dis. 2019;19.
63 Shi J, Su Q, Zhang C, et al. An intelligent decision support algorithm for diagnosis of colorectal cancer through serum tumor markers. Comput Methods Programs Biomed. 2010;100:97–107.
64 Singh S, Nosyk B, Sun H, et al. Value of Information of a clinical prediction rule: Informing the efficient use of healthcare and health research resources. Int J Technol Assess Health Care. 2008;24:112–9.
65 Tan BK, Lu G, Kwasny MJ, et al. Effect of symptom-based risk stratification on the costs of managing patients with chronic rhinosinusitis symptoms. Int Forum Allergy Rhinol. 2013;3:933–40.
66 Teferra RA, Grant BJB, Mindel JW, et al. Cost minimization using an artificial neural network sleep apnea prediction tool for sleep studies. Ann Am Thorac Soc. 2014;11:1064–74.
67 Walsh CG, Sharman K, Hripcsak G. Beyond discrimination: A comparison of calibration methods and clinical usefulness of predictive models of readmission risk. J Biomed Inform. 2017;76:9–18.
68 Wang HY, Hung CC, Chen CH, et al. Increase Trichomonas vaginalis detection based on urine routine analysis through a machine learning approach. Sci rep. 2019;9:11074.
69 Warner JL, Zhang P, Liu J, et al. Classification of hospital acquired complications using temporal clinical information from a large electronic health record. J Biomed Inform. 2016;59:209–17.
70 Webb BJ, Sorensen J, Mecham I, et al. Antibiotic Use and Outcomes After Implementation of the Drug Resistance in Pneumonia Score in ED Patients With Community-Onset Pneumonia. Chest. 2019;156:843–51.
71 Zhang C, Paolozza A, Tseng PH, et al. Detection of children/youth with fetal alcohol spectrum disorder through eye movement, psychometric, and neuroimaging data. Frontiers in Neurology. 2019;10.
72 Carballido-Gamio J, Yu A, Wang L, et al. Hip Fracture Discrimination Based on Statistical Multi-parametric Modeling (SMPM). Ann Biomed Eng. 2019;47:2199–212.
73 Crespo C, Linhart M, Acosta J, et al. Optimisation of cardiac resynchronisation therapy device selection guided by cardiac magnetic resonance imaging: Cost-effectiveness analysis. Eur J Prev Cardiol. 2019: 2047487319873149.

74 Philipsen RH, Sánchez CI, Maduskar P, et al. Automated chest-radiography as a triage for Xpert testing in resource-constrained settings: a prospective study of diagnostic accuracy and costs. Sci Rep. 2015;5:12215.
75 Radensky PW, Mango LJ. Interactive neural network-assisted screening: An economic assessment. Acta Cytol. 1998;42:246–52.
76 Sato M, Kawai M, Nishino Y, et al. Cost-effectiveness analysis for breast cancer screening: Double reading versus single + CAD reading. Breast Cancer. 2014;21:532–41.
77 Scotland GS, McNamee P, Fleming AD et al. Costs and consequences of automated algorithms versus manual grading for the detection of referable diabetic retinopathy. Br J Ophthalmol. 2010;94:712-9.
78 Sreekumari A, Shanbhag D, Yeo D, et al. A deep learning-based approach to reduce rescan and recall rates in clinical MRI examinations. AJNR Am J Neuroradiol. 2019;40:217–23.
79 Tourassi GD, Floyd CE, Coleman RE. Acute pulmonary embolism: Cost-effectiveness analysis of the effect of artificial neural networks on patient care. Radiology. 1998;206:81–8.
80 Brocklehurst P, Field D, Greene K, et al. Computerised interpretation of the fetal heart rate during labour: A randomised controlled trial (INFANT). Health Technology Assessment. 2018;22:1–218.
81 Brüggenjürgen B, Israel CW, Klesius AA, et al. Health services research in heart failure patients treated with a remote monitoring device in Germanya retrospective database analysis in evaluating resource use. J Med Econ. 2012;15:737–45.
82 Calvert J, Hoffman J, Barton C, et al. Cost and mortality impact of an algorithm-driven sepsis prediction system. J Med Econ. 2017;20:646–51.
83 Hollis C, Hall CL, Guo B, et al. The impact of a computerised test of attention and activity (QbTest) on diagnostic decision-making in children and young people with suspected attention deficit hyperactivity disorder: single-blind randomised controlled trial. J Child Psychol Psychiatry. 2018;59:1298–308.
84 Levin RI, Koenig KL, Corder MP, et al. Risk stratification and prevention in chronic coronary artery disease: Use of a novel prognostic and computer-based clinical decision support system in a large primary managed-care group practice. Disease Management. 2002;5:197–213.
85 Marchetti A, Jacobs J, Young M, et al. Costs and benefits of an early-alert surveillance system for hospital inpatients. Curr Med Res Opin. 2007;23:9–16.
86 Salzsieder E, Augstein P. The Karlsburg Diabetes Management System: Translation from research to eHealth application. J Diabetes Sci Technol. 2011;5:13–22.
87 Sánchez-Quiroga MA, Corral J, Gómez-de-Terreros FJ, et al. Primary care physicians can comprehensively manage patients with sleep apnea a noninferiority randomized controlled trial. Am J Respir Crit Care Med. 2018;198:648–56.
88 Abeykoon JP, Mueller L, Dong F, et al. The Effect of Implementing Gene Expression Classifier on Outcomes of Thyroid Nodules with Indeterminate Cytology. Horm Cancer. 2016;7:272–8.
89 Brixner D, Biltaji E, Bress A, et al. The effect of pharmacogenetic profiling with a clinical decision support tool on healthcare resource utilization and estimated costs in the elderly exposed to polypharmacy. J Med Econ. 2016;19:213–28.
90 D'Andrea E, Choudhry NK, Raby B, et al. A bronchial-airway gene-expression classifier to improve the diagnosis of lung cancer: Clinical outcomes and cost-effectiveness analysis. Int J Cancer. 2020;146:781-790.
91 Green N, Al-Allak A, Fowler C. Benefits of introduction of Oncotype DX® testing. Ann R Coll Surg Engl. 2019;101:55–9.
92 Labourier E. Utility and cost-effectiveness of molecular testing in thyroid nodules with indeterminate cytology. Clin Endocrinol (Oxf). 2016;85(4):624-3193 Lobo JM, Trifiletti DM, Sturz VN, et al. Cost-effectiveness of the Decipher Genomic Classifier to Guide Individualized Decisions for Early Radiation Therapy After Prostatectomy for Prostate Cancer. Clin Genitourin Cancer. 2017;15:e299–309.
94 Mathias PC, Tarczy-Hornoch P, Shirts BH. Modeling the costs of clinical decision support for genomic precision medicine. AMIA Summits on Translational Science Proceedings. 2016;2016:60–4.
95 Nelson RE, Stenehjem D, Akerley W. A comparison of individualized treatment guided by VeriStrat with standard of care treatment strategies in patients receiving second-line treatment for advanced non-small cell lung cancer: A cost-utility analysis. Lung Cancer. 2013;82:461–8.
96 Nicholson KJ, Roberts MS, McCoy KL, et al. Molecular Testing Versus Diagnostic Lobectomy in Bethesda III/IV Thyroid Nodules: A Cost-Effectiveness Analysis. Thyroid. 2019;29:1237–43.

97 Seguí MÁ, Crespo C, Cortés J, et al. Genomic profile of breast cancer: Cost-effectiveness analysis from the Spanish National Healthcare System perspective. Expert Rev Pharmacoecon Outcomes Res. 2014;14:889–99.
98 Shapiro S, Pharaon M, Kellermeyer B. Cost-effectiveness of Gene Expression Classifier Testing of Indeterminate Thyroid Nodules Utilizing a Real Cohort Comparator. Otolaryngol Head Neck Surg. 2017;157:596–601.
99 Morse, KE, Bagely, SC, Shah, NH. Estimate the hidden deployment cost of predictive models to improve patient care. Nat Med 2020:26;18–19.
100 Markiewicz K, Van Til JA, IJzerman MJ. Medical devices early assessment methods: systematic literature review. Int J Technol Assess Health Care. 2014;30(2):137–46.
101 Tu Hat, Han D, Pechlivanoglou P, et al. Early economic evaluation of emerging health technologies: protocol of a systematic review. Syst Rev. 2014;3(1):81.
102 Analytical Performance of Afirma GSC: A Genomic Sequencing Classifier for Cytology-Indeterminate Thyroid Nodule FNA Specimens. [Internet] https://www.afirma.com/physicians/ata-2017/qa-3/ . Accessed January 2020.
103 Miller K. Big Data and Clinical Genomics. [Internet] .https://www.clinicalomics.com/magazine-editions/volume-6-issue-number-2-march-april-2019/big-data-and-clinical-genomics/. Accessed January 2020
104 Shilo S, Rossman H, Segal E. Axes of a revolution: challenges and promises of big data in healthcare. Nat Med. 2020;26(1):29-38.
105 Baro E, Degoul S, Beuscart R, et al. Toward a literature-driven definition of big data in healthcare. Biomed Res Int. 2015;2015.
106 Frost & Sullivan. Drowning in big data? Reducing information technology complexities and costs for healthcare organizations 2012

APPENDIX

APPENDIX A - Search strategies

Search strategy embase.com

('economic evaluation'/de OR 'economic aspect'/de OR cost/de OR 'cost control'/de OR 'health care cost'/de OR 'economic model'/de OR 'economics'/de OR investment/de OR funding/de OR 'device economics'/de OR 'resource allocation'/de OR 'health care financing'/de OR 'hospital purchasing'/de OR 'cost benefit analysis'/exp OR 'cost effectiveness analysis'/exp OR 'cost minimization analysis'/exp OR 'cost utility analysis'/exp OR 'biomedical technology assessment'/de OR 'research and development'/de OR 'Markov chain'/de OR 'device approval'/de OR 'product development'/de OR 'diagnostic test approval'/de OR 'strategic planning'/exp OR 'return on investment'/de OR ((econom* NEAR/3 evaluat*) OR ((cost OR costs OR expenditure* OR economic*) NEAR/6 (benefit* OR effectiv* OR utili* OR minimi* OR implement* OR instal* OR operat* OR development* OR analy* OR implication* OR associat* OR perform* OR optim* OR reduc* OR avoid* OR save OR saving* OR increase* OR decrease* OR health* OR medical* OR consider* OR impact* OR control*)) OR funding OR (business* NEAR/3 perform*) OR (value NEAR/3 money) OR (technolog* NEAR/3 assessment*) OR (research NEAR/3 development) OR headroom* OR head-room* OR Markov OR ((device* OR product* OR diagnostic-test*) NEAR/3 (approv* OR develop* OR economic*)) OR (strateg* NEAR/3 plan*) OR (return NEAR/3 invest*)):ab,ti,kw) AND ('big data'/exp OR 'clinical data repository'/de OR 'clinical decision support system'/de OR 'computerized provider order entry'/exp OR 'alarm monitor'/de OR 'alarm monitoring'/de OR 'artificial intelligence'/de OR 'clinical prediction rule'/de OR (('decision making'/de OR 'medical decision making'/de OR 'clinical decision making'/de) AND ('computer assisted diagnosis'/de OR 'computer assisted therapy'/de)) OR ('machine learning'/exp NOT 'hidden Markov model'/de) OR (('information technology'/de OR automation/de OR 'medical informatics'/de OR 'electronic medical record'/de OR 'information processing'/de OR 'hospital information system'/de OR 'medical information system'/de) AND 'decision support system'/de) OR ('big data' OR (clinical NEAR/6 data NEAR/6 repositor*) OR (clinical* NEAR/6 decision* NEAR/6 (system OR systems OR support* OR automat* OR computer* OR technolog* OR algorith* OR tool*)) OR (computer* NEAR/6 (provider* OR order*) NEAR/6 entr*) OR ((alarm* OR alert* OR warning) NEAR/3 (monitor* OR system*)) OR (electronic* NEAR/3 (ordering* OR prescri*)) OR E-prescri* OR (clinical* NEAR/3 predict* NEAR/3 (rule* OR model*)) OR ((computer* OR automat* OR technolog* OR algorith*) NEAR/6 (decision* OR protocol*) NEAR/6 (diagnos* OR therap* OR surg*)) OR 'machine learning' OR 'artificial intelligence' OR (data NEAR/3 mining) OR datamining OR (mining NEAR/3 (health* OR patient* OR medical*) NEAR/3 record*)):ab,ti,kw) AND ('health care facilities and services'/exp OR 'health'/exp OR 'medicine'/exp OR 'diseases'/exp OR 'health care personnel'/exp OR 'health care organization'/exp OR patient/exp OR (hospital* OR clinic* OR medic* OR health* OR disease* OR practitioner* OR physician* OR doctor* OR patient* OR diagnos* OR therap*):ab,ti,kw) NOT ([Conference Abstract]/lim OR [Letter]/lim OR [Note]/lim OR [Editorial]/lim) AND [english]/lim

Search strategy Medline

("Costs and Cost Analysis"/ OR Health Care Costs/ OR Models, Economic/ OR Economics/ OR Economics, Medical/ OR Hospital Costs/ OR Health Expenditures/ OR Diagnostic Test Approval/ OR Investments/ OR Resource Allocation/ OR Purchasing, Hospital/ OR exp Cost-Benefit Analysis/

OR Technology Assessment, Biomedical/ OR Markov Chains/ OR Device Approval/ OR Strategic Planning/ OR Cost Savings/ OR ((econom* ADJ3 evaluat*) OR ((cost OR costs OR expenditure* OR economic*) ADJ6 (benefit* OR effectiv* OR utili* OR minimi* OR implement* OR instal* OR operat* OR development* OR analy* OR implication* OR associat* OR perform* OR optim* OR reduc* OR avoid* OR save OR saving* OR increase* OR decrease* OR health* OR medical* OR consider* OR impact* OR control*)) OR funding OR (business* ADJ3 perform*) OR (value ADJ3 money) OR (technolog* ADJ3 assessment*) OR (research ADJ3 development) OR headroom* OR head-room* OR Markov OR ((device* OR product* OR diagnostic-test*) ADJ3 (approv* OR develop* OR economic*)) OR (strateg* ADJ3 plan*) OR (return ADJ3 invest*)).ab,ti,kf.) AND (Decision Support Systems, Clinical/ OR Medical Order Entry Systems/ OR Clinical Alarms/ OR Decision Making, Computer-Assisted/ OR ((Decision Making/ OR Clinical Decision-Making/) AND (Diagnosis, Computer-Assisted/ OR Therapy, Computer-Assisted/)) OR Artificial Intelligence/ OR exp Machine Learning/ OR ((Information Technology/ OR Automation/ OR Medical Informatics/ OR Medical Informatics Applications/ OR Electronic Health Records/ OR Automatic Data Processing/ OR Hospital Information Systems/) AND Decision Support Techniques/) OR (big data OR (clinical ADJ6 data ADJ6 repositor*) OR (clinical* ADJ6 decision* ADJ6 (system OR systems OR support* OR automat* OR computer* OR technolog* OR algorith* OR tool*)) OR (computer* ADJ6 (provider* OR order*) ADJ6 entr*) OR ((alarm* OR alert* OR warning) ADJ3 (monitor* OR system*)) OR (electronic* ADJ3 (ordering* OR prescri*)) OR E-prescri* OR (clinical* ADJ3 predict* ADJ3 (rule* OR model*)) OR ((computer* OR automat* OR technolog* OR algorith*) ADJ6 (decision* OR protocol*) ADJ6 (diagnos* OR therap* OR surg*)) OR machine learning OR artificial intelligence OR (data ADJ3 mining) OR datamining OR (mining ADJ3 (health* OR patient* OR medical*) ADJ3 record*)).ab,ti,kf.) AND (exp Health Care Facilities, Manpower, and Services/ OR exp health/ OR exp Medicine/ OR exp "Diseases (Non MeSH)"/ OR exp Health Personnel/ OR exp Patients/ OR (hospital* OR clinic* OR medic* OR health* OR disease* OR practitioner* OR physician* OR doctor* OR patient* OR diagnos* OR therap*).ab,ti,kf.) NOT (letter* OR news OR comment* OR editorial* OR congres* OR abstract* OR book* OR chapter* OR dissertation abstract*).pt. AND english.la.

Search strategy Web of science

TS=((((econom* NEAR/2 evaluat*) OR ((cost OR costs OR expenditure* OR economic*) NEAR/5 (benefit* OR effectiv* OR utili* OR minimi* OR implement* OR instal* OR operat* OR development* OR analy* OR implication* OR associat* OR perform* OR optim* OR reduc* OR avoid* OR save OR saving* OR increase* OR decrease* OR health* OR medical* OR consider* OR impact* OR control*)) OR funding OR (business* NEAR/2 perform*) OR (value NEAR/2 money) OR (technolog* NEAR/2 assessment*) OR (research NEAR/2 development) OR headroom* OR head-room* OR Markov OR ((device* OR product* OR diagnostic-test*) NEAR/2 (approv* OR develop* OR economic*)) OR (strateg* NEAR/2 plan*) OR (return NEAR/2 invest*))) AND ((“big data” OR (clinical NEAR/5 data NEAR/5 repositor*) OR (clinical* NEAR/5 decision* NEAR/5 (system OR systems OR support* OR automat* OR computer* OR technolog* OR algorith* OR tool*)) OR (computer* NEAR/5 (provider* OR order*) NEAR/5 entr*) OR ((alarm* OR alert* OR warning) NEAR/2 (monitor* OR system*)) OR (electronic* NEAR/2 (ordering* OR prescri*)) OR E-prescri* OR (clinical* NEAR/2 predict* NEAR/2 (rule* OR model*)) OR ((computer* OR automat* OR technolog* OR algorith*) NEAR/5 (decision* OR protocol*) NEAR/5 (diagnos* OR therap* OR surg*)) OR “machine learning” OR “artificial intelligence” OR (data NEAR/2 mining) OR datamining OR (mining NEAR/2 (health* OR patient* OR medical*) NEAR/2 record*))) AND

((hospital* OR clinic* OR medic* OR health* OR disease* OR practitioner* OR physician* OR doctor* OR patient* OR diagnos* OR therap*))) AND DT=(article) AND LA=(english)

NHS EED via https://www.crd.york.ac.uk/CRDWeb/ 59

((((econom* NEAR2 evaluat*) OR ((cost OR costs OR expenditure* OR economic*) NEAR5 (benefit* OR effectiv* OR utili* OR minimi* OR implement* OR instal* OR operat* OR development* OR analy* OR implication* OR associat* OR perform* OR optim* OR reduc* OR avoid* OR save OR saving* OR increase* OR decrease* OR health* OR medical* OR consider* OR impact* OR control*)) OR funding OR (business* NEAR2 perform*) OR (value NEAR2 money) OR (technolog* NEAR2 assessment*) OR (research NEAR2 development) OR headroom* OR headroom* OR Markov OR ((device* OR product* OR diagnostic-test*) NEAR2 (approv* OR develop* OR economic*)) OR (strateg* NEAR2 plan*) OR (return NEAR2 invest*))) AND ((“big data” OR (clinical NEAR5 data NEAR5 repositor*) OR (clinical* NEAR5 decision* NEAR5 (system OR systems OR support* OR automat* OR computer* OR technolog* OR algorith* OR tool*)) OR (computer* NEAR5 (provider* OR order*) NEAR5 entr*) OR ((alarm* OR alert* OR warning) NEAR2 (monitor* OR system*)) OR (electronic* NEAR2 (ordering* OR prescri*)) OR E-prescri* OR (clinical* NEAR2 predict* NEAR2 (rule* OR model*)) OR ((computer* OR automat* OR technolog* OR algorith*) NEAR5 (decision* OR protocol*) NEAR5 (diagnos* OR therap* OR surg*)) OR “machine learning” OR “artificial intelligence” OR (data NEAR2 mining) OR datamining OR (mining NEAR2 (health* OR patient* OR medical*) NEAR2 record*))) AND ((hospital* OR clinic* OR medic* OR health* OR disease* OR practitioner* OR physician* OR doctor* OR patient* OR diagnos* OR therap*)))

APPENDIX B: RESULTS FROM ECONOMIC EVALUATIONS IN THE RECORDS INCLUDED (N=71)

Article	Intervention	Comparator	Outcome	Costs	Full or partial	Stage of development	Key economic results	ICER
Medical History Data								
Ashfaq 2019[1]	Multiple predictive models for 30-day hospital readmission combined with a hypothetical preventative telemonitoring intervention program	Current care without the model and intervention	Model performance	Intervention Readmission	Partial	During	Savings depend on the success of the intervention and the intervention costs (presented irrespective of model performance). Deep learning model with all important features had the best model performance and highest potential savings.	NR
Bennet 2013[2]	Multiple models (e.g., Markov decision process) that decide on the optimal treatment plan.	Treatment as usual	% Patients receiving maximum number of treatment sessions Outcome rating scale used to estimate the costs per unit changed	Treatment costs to estimate costs per unit of outcome change	Partial	During	The Markov decision making process was superior to the simpler models (e.g., Raw Effect) with better outcomes and lower costs per unit change	NR
Bremer 2018[3]	A predictive model that personalizes treatment recommendations	Care without a predictive model	Model performance QALYs	Medical Productivity losses treatment	Full	During	Ridge regression was the best performing model. For an ICER of €25,000 per QALY gained the model resulted in savings of 5.42% but also decreased QALYs (1.98%).	€25,000 per QALY gained
Brisimi 2019[4]	Multiple models that predict the risk of hospitalization combined with a preventative intervention for those at high risk	Care without a predictive model and no intervention	Model performance	Intervention Hospitalization	Partial	During	The best model had a 89% accuracy, resulted in savings of $34 per patient and ensured interpretability for clinicians.	NR

Article	Intervention	Comparator	Outcome	Costs	Full or partial	Stage of development	Key economic results	ICER
Burton 2019[5]	Care with a model that predicts which urine samples are most likely to contain urinary tract infection	No predictive analytics	Model performance Workload reduction	Culture agar (in discussion)	Partial	During	Three Extreme Gradient Boosting (XGBoost) algorithms were preferred with +- 95% sensitivity and annual savings (£800,000-£5,000,000) to identify urinary tract infections in pregnant patients, children and other patients.	NR
Chae 2001[6]	Care with an intelligent laboratory information system	Care without an intelligent laboratory information system	Model performance	Analytics Testing Laboratory processing Reporting errors and misdiagnosis	Partial	After	The intelligent laboratory system resulted in a higher overall predictive rate and generated savings (US $21,660).	NR
Chi 2010[7]	Several models that optimize on speed or costs that assist decision-making by providing the optimal sequence of tests when diagnosing a patient.	A classifier trained with all available tests in the dataset	Model performance Number of tests	Tests	Partial	During	All functions reduced costs, the number of tests used and had similar model performance compared to baseline.	NR
Choi 2018[8]	A two-step diagnostic framework for early detection of dementia that combines a deep learning neural network with results from the MMSE and KLOSCAD-N	1: Deep learning of results from the MMSE 2: Deep learning of results from the KLOSCAD-N 3: Results from the MMSE 4: Results from the KLOSCAD-N	Model performance	Mini mental status examination Korean Longitudinal Study on Cognitive Aging and Dementia Neuropsychological Assessment Battery	Partial	During	The two-stage diagnostic framework that combines results from the MMSE, KLOSCAD-N and a deep learning neural networks had the best model performance and reduced costs compared to the other KLOSCAD-N strategies	NR

Cooper 2005[9]	For costs: A decision model to predict poor outcomes + inpatient care for high risk For accuracy multiple models were compared.	Current care without the model	Model performance	Inpatient Outpatient	Partial	During	The best model in terms of model performance (an artificial neural network) has the potential to generate savings compared to current care by reducing the number of patients that need to be admitted.	NR
Duggal 2016[10]	Intervention 1: Naïve Bayes algorithm Intervention 2: Logical regression Intervention 3: Random Forest Intervention 4: Adaboost Intervention 5: Neural Network	For costs: Care without a predictive algorithm	Model performance	Readmission Intervention to avoid readmission	Partial	During	The random forest model had the best model performance. Naïve Bayes algorithm resulted in the most true positives which lead to large savings that outweighted the high number of false positives and thus maximized savings (INR 15.92 million).	NR
Elkin 2010[11]	Dxplain: a computerised education, reference and clinical decision support system that offers differential diagnoses	Treatment as usual	X	All costs and charges reported in institutions	Partial	After	Total charges and costs per admission reduced with US$1,281 and US$990 after introducing DXplain compared to before introduction of the system.	NR
Escudero 2013[12]	Personalized classifiers that minimize the costs of biomarkers or the number of biomarkers for: - Patients with early biomarkers for Alzheimer's Disease from healthy patients. - Patients with mild cognitive impairment that progress to Alzheimer's Disease after 12 months from patients to do not progress.	Models (Logistic regression or weighted logistic regression) that use all biomarkers to detect Alzheimer's Disease	Model performance Number of biomarkers needed for classification	Biomarker measurement	Partial	During	By minimizing the number of biomarkers and costs, the personalized classifiers resulted in much lower costs and similar model performance compared to including all biomarkers.	NR

Article	Intervention	Comparator	Outcome	Costs	Full or partial	Stage of development	Key economic results	ICER
Golas 2018[13]	Predictive models for 30-day hospital readmission combined with heart failure telemonitoring	For costs: No intervention to prevent readmission	Model performance	Intervention Readmission	Partial	During	The deep unified networks had the best model performance and resulted in the maximum net savings of 3.403 ± 0.536 million US dollars	NR
Govers 2018[14]	1: Existing (imperfect) prediction models for personalized decision making in oral cancer of patients 2: Perfect prediction models for personalized decision making in oral cancer of patients	Population based approach All include the following interventions: 1: elective neck dissection 2: watchful waiting 3: gene expression profiling followed by neck dissection or watchful waiting 4: sentinel lymph node procedure followed by neck dissection or watchful waiting 5: gene expression profiling followed by sentinel lymph node procedure followed by neck dissection or watchful waiting	QALYs	Costs unclear, no access to supplementary data in which the model input is reported	Full	During	The imperfect models used in this study must be improved before they can be implemented in clinical The perfect prediction model approach resulted in more QALYs and cost savings compared to the population-based approach.	NR

Grinspan 2018[15]	Models that predict emergency department use + a preventative intervention	For costs: Care without the algorithm and intervention	Model performance	Emergency department visit Unplanned inpatient admission	Partial	During	A relatively simple 3 variable model performed equally good or better than the machine learning algorithms. If the emergency department visits could be reduced by 10%, the per-patient-per-year breakeven points for the two hospitals are $958 ($568-$1390) and $1086 ($886-$1320).	NR
Harrison 2013[16]	Predictive algorithms to determine the risk of invasive candida infection (based on logistic regression) + prophylaxis treatment	Current care without algorithms + prophylaxis	QALYs	Treatment Hospitalization	Full	After	The algorithm resulted in a slight reduction in costs and a slight increase in QALYs. For a cost-effectiveness threshold of £20,000 the probability that any of the risk models was cost-effective did not exceed 30%	INB varied from positive to negative depending on time of risk assessment
Hunter-Zinck 2019[17]	Model to determine required tests shortly after emergency department arrival	Care without the model	Model performance LOS	Tests	Partial	During	The median costs of orders per visit increased (US$21-US$45) while the median LOS per visit decreased.	NR
Jamei 2017[18]	An artificial Neural Network model that predicts all-cause 30-day readmission risk + an intervention for the high risk	Readmission risk according to LACE, the industry-standard scoring model + an intervention for the high risk	Model performance	Readmission Intervention	Partial	During	The neural network model outperformed the industry standard (LACE). The classifier's threshold (which determines the number of patients needing an intervention) could be determined according to the expected savings. When the intervention rate is +- 9-10% savings were highest.	NR
Javitt 2005[19]	A system that issued clinical recommendations for clinicians in primary care	Care without a system that provided recommendations	Hospitalizations Mortality at 6 months Hospital LOS	Hospital costs based on charges	Partial	After	Hospital admissions and charges in the intervention group were significantly lower compared to the control group. The system yielded an 8-fold return on investment	NR

Article	Intervention	Comparator	Outcome	Costs	Full or partial	Stage of development	Key economic results	ICER
Kockbek 2019[20]	Predictive models to assess the risk of surgical site infections after gastrointestinal surgery Costs only: Price penalized models restricted to 20 features and unrestricted	Full lasso model	Model performance	Tests	Partial	During	The full lasso model restricted to 50 features achieved the highest AUC When a restricted price penalized model was used total testing costs reduced from 25,783,747 NOK to 12,401,982 compared to full Lasso	NR
Kofoed 2009[21]	Decision support system to assist clinicians on which antimicrobial therapy to prescribe.	Clinician decides the antimicrobial therapy	Number with appropriate antimicrobial treatment Type of antimicrobial treatment	Antimicrobials Drug administration Bed day Future resistance	Partial	After	The number of patients that received antibiotics was significantly higher with TREAT compared to without TREAT. There was no difference in the average costs per patient that received antibiotics. Costs of side effects were lower with TREAT	NR
Ladabaum 2011 [22]	Screening using: 1. Prediction algorithms 2. Tumor testing 3. Up-front germline mutation testing 4. Clinical criteria All followed by tailored screening and risk-reducing surgery.	No active screening of Lynch syndrome	Life Years Cancer cases Mortality	Genetic and clinical testing Colorectal cancer treatment Risk-reducing surgery	Full	After	All screening strategies improved outcomes (reduced mortality) and increased costs. Tumor testing strategies were preferred.	ICER varied from $30,600-$293,000 p. LY gained depending on the strategy adopted
Lee 2015[23]	Emergency department decision support system that couples machine learning, simulation, and optimization to address improvement goals.	Before (baseline) implementation of the emergency department decision support system	Length of stay Waiting time Annual throughput (number of patients treated) Avoidable 72hr and 30day readmissions	30day readmission penalties Total emergency department costs Revenues	Partial	After	- Decreased 72-hr and 30-day readmissions - Reduced readmission penalties (- US$7.5 million) - Increased revenues (US$190 million)	NR

Lee 2018[24]	A predictive model to identify patients at risk of readmission after joint replacement combined with a preventative intervention for high risk patients	For costs: Current care	Model performance	Readmission / penalty Intervention	Partial	During	Illustrate the potential for an intervention yet to be developed based on performance of the risk prediction model.	NR
Lin 2019[25]	1. Models in the prognostic monitoring framework that identifies high risk patients and personalizes monitoring strategies for patients 2. Routine monthly monitoring (status quo) 3. Routine 2-monthly monitoring (status quo) 4. PHQ-9 based strategy that predicts depression severity and adaptively monitors patients with high scores	Routine 3-monthly monitoring (status quo)	Number of true positives	Monitoring visit	Full	During	The natural history matching is best at distinguishing high from low risk (highest AUC and correlation) The Markov-based collaborative model and the rule based model are better for individual prediction (low rMSE) Markov-based model has the highest monitoring accuracy Implementing most of the predictive models could lead to savings compared to status quo.	Dominated-$ 2,133 per true positive
Ling 2006[26]	Novel algorithms that optimize test selection for coronary heart disease.	Existing algorithms that optimize test selection for coronary heart disease. 1. Lazy decision tree learning algorithm 2. Naïve Bayesian-based cost sensitive learning algorithm.	X	Tests Misclassification	Partial	During	Emphasize the need for including cost consequences when developing algorithms. Illustrate that the proposed analytics could lead to savings in terms of better diagnoses compared to alternative analytics.	NR

Article	Intervention	Comparator	Outcome	Costs	Full or partial	Stage of development	Key economic results	ICER
Marble 1999[27]	A neural network model to predict sepsis with improved sensitivity	Care as described by earlier studies	Model performance	Antibiotics	Partial	During	Adopting an improved algorithm could generate savings due to reduced use of antibiotics (US$103,000 per 1000 trauma cases). Report 100% sensitivity and 96.5% specificity for the novel predictive model	NR
Monahan 2018[28]	Ambulatory blood pressure monitoring using the PROOF-BP algorithm	1: Clinic blood pressure monitoring 2: Home blood pressure monitoring 3: Ambulatory blood pressure monitoring	QALYs	Diagnosis Blood pressure monitoring devices Stroke, TIA, angina, myocardial infarction Treatment	Full	After	Use of the PROOF-BP algorithm was cost-effective (> costs, > QALYs) in all age and sex groups given a willingness to pay threshold below £20,000.	Age and sex dependent ICERs varied: Dominant to £11,363 per QALY gained
Murtojärvi 2020[29]	Model I to predict survival to assist clinical decision making in patients with mCRPC with varying features selected	Model II to predict survival to assist clinical decision making in patients with mCRPC with varying features selected	Model performance	Tests	Partial	During	Illustrate that costs can be reduced when selecting less features coinciding with a small or no loss in model performance	NR
Parekattil 2003[30]	Novel algorithm to detect bladder cancer based on tumor markers	For model performance: 1: Individual tumor markers 2: Hematuria 3: Cytology For costs: Current bladder cancer evaluation	Model performance	Cytology Office cystoscopy Combined assay	Partial	During	Model performance was better for the novel algorithm. The costs for all patients were $61,054 in current care and $34,450 in care with the assay.	NR

Petersen 2015[31]	Several prediction models for virological failure to improve adherence	For costs: Testing HIV RNA plasma every 3 months	Model performance	Outpatient primary care visit HIV RNA test	Partial	During	Classification of possible virological failure improved when the Super Learner algorithm included MEMS data, CD4 T-cell counts and ART regimen data. Health care payers can spend $16-$29 per person-month on an entire system for real-time treatment adherence monitoring system.	NR
Pölsterl 2015[32]	Subgroup-specific classifiers to predict the risk of adverse events for different treatment options	1: Current clinical practice 2: Population wide regression	Adverse Events defined as the probabilities of (restenosis)+ (hazardous events defined as death, myocardial infarction and thrombosis)	Treatment 1-year dual antiplatelet therapy Treatment myocardial infarction and stroke	Partial	During	Using the model to personalize treatment could reduce the number of adverse events and costs(US$693-$794 per patient) compared to current care.	NR
Rider 2019[33]	SPIRIT analyzer screening tool to identify children with potential primary immunodeficiency disease	Care without the screening tool	Individuals diagnosed	Hospitalization	Partial	After	The algorithm identified 98 individuals with primary immunodeficiency disorder. Early identification of the 98 PI patients in the cohort could save $7.7 million dollars.	NR
Ruff 2019[34]	Predictive algorithm to determine the risk of non-adherence to treatment with oral anticoagulants	Care without the algorithm	Risk of an event	Stroke Intervention	Partial	Before	A higher specificity is important for clinical utility and cost-effectiveness. A higher adherence threshold is recommended to ensure cost-effectiveness	NR
Shazzadur Rahman 2019[35]	Triage based on neural network algorithm Triage based on theoretical products not yet developed (optimal or target)	No triage based on clinical symptoms (cough)	Case detection rate	Testing Treatment Analytics	Full	After	Detecting using the artificial neural network resulted in a higher detection rate than microscopy and lower costs than Xpert for all.	ICER: Depended on strategy and whether an X-ray was available varied from dominant to dominated

Article	Intervention	Comparator	Outcome	Costs	Full or partial	Stage of development	Key economic results	ICER
Shi 2010[36]	Algorithm to determine optimal cut-off points and combinations of serum tumor markers for diagnosing colorectal cancer	For costs: Care without the algorithm	Model performance	Serum tumor markers	Partial	During	Compared to traditional serial testing: - the number of serum tumor markers tested could be reduced - the costs of diagnosis could be reduced (56.08%). - The predictive performance was better	NR
Singh 2008[37]	Clinical prediction model for early disposition within 3 hours after presentation at the emergency department	Current care	Number of inappropriate emergency department discharges	Hospitalization Surgical procedures Physician visits Diagnostic tests	Full	After	Initial estimates indicate that using the technology is cost-effective (increased effects, reduced costs (Canadian 36,564 per 1000 patients$)).	ICER: $2,999 per inappropriate acute coronary syndrome discharge
Tan 2013[38]	Algorithm to assist diagnostic decision-making for chronic rhinosinusitis	For costs: Individual or combinations of symptoms to diagnose chronic rhinosinusitis. For all these diagnostic combinations the benefits of empiric medical therapy were compared to an upfront CT scan.	Model performance	Diagnostic tests Treatment Follow-up	Partial	During	Without the algorithm, use of an upfront CT is preferred when available. Use of the algorithm identified groups for whom an upfront CT was preferred and patients for whom empiric medical therapy was preferred	NR

Teferra 2014[39]	1: OSUnet; an artificial neural network to determine who is likely to have obstructive sleep apnea. 2: MNC algorithm 3: STOP-BANG algorithm	For costs: Current care	Model performance	In lab PSG Home sleep test	Partial	During	- OSUNet had a higher positive likelihood ratio than existing predictive rules. - OSUnet could result in savings when the prevalence of OSA is below 40%	NR
Walsh 2017[40]	Four classes of model training and calibration methods that predict 30-day readmission - Model training to optimize discrimination - Model recalibration - Model extension	Standard method (logistic regr. Trained on all data) that predicts 30-day readmission	Model performance	Readmission Intervention	Partial	During	- The costs depended on the threshold adopted at which an intervention was triggered (i.e., more false positives or more false negatives). - Algorithms which were not properly calibrated resulted in financial losses.	NR
Wang 2019[41]	Predictive model to increase detection of Trichomonas vaginalis in urine samples	Current care	Model performance	Urine sediments (Monetary estimate in discussion)	Partial	During	The costs per positive case were lower for both men and women (US$170.1, US$29.7) than the annual GDP per capita and therefore the strategy was considered cost-effective.	NR
Warner 2016[42]	Perfect predictive model for early identification of hospital acquired complications	No prevention	Model performance	Hospital costs (Monetary estimate in conclusion)	Partial	During	Model performance was good especially for medical complications (AUC=0.845). If all hospital acquired complications could be prevented, $10 billion could be saved annually in the US.	NR
Webb 2019[43]	Prediction tool (Drug-Resistance in Pneumonia) to identify pneumonia with drug-resistant pathogens implemented into a CDSS	Care before implementation of the prediction tool	Broad spectrum AB use, 30-day mortality, LOS, Model performance	Direct costs	Partial	After	No difference was found in costs, mortality and LOS but the prediction score was associated with inadequate antibiotics prescriptions and use of broad-spectrum antibiotics	NR

Article	Intervention	Comparator	Outcome	Costs	Full or partial	Stage of development	Key economic results	ICER
Zhang 2019[44]	Annual screening for children at risk of having fetal alcohol spectrum disorders	For costs: No annual screening resulting in progression of the disease from mild to severe	Model performance	Testing Difference severe or mild disease	Partial	During	Adopting the screening protocol can lead to savings (US$3-4 million per 1000 patients)	NR
Imaging data								
Carballido-Gamio 2019[45]	A predictive model for assessing the risk of hip fractures from quantitative computed tomography using statistical multi-parametric modelling techniques	A predictive model for assessing the risk of hip fractures based on mean femoral neck areal bone mineral density, age, height and weight	Model performance	Unclear (in discussion)	Partial	During	Most SMPM LASSO models that included volumetric bone mineral density features and cortical bone thickness features had better model performance. Potential savings of 382-503 million US dollars by using a SMPM model.	NR
Crespo 2019[46]	Algorithms that use cardiac magnetic resonance imaging data to determine whether patients should receive an implantable cardiac resynchronisation pacemaker or an implantable cardiac resynchronisation cardioverter-defibrillator	Current clinical practice	QALYs	Analytics (implementation, battery etc.) surgery Device related complications Specialist visits Drug treatment heart transplant Hospital stay	Full	After	Algorithm one dominated algorithm II and current care since it reduced long term costs and increased QALYs.	Dominant

Philipsen 2015[47]	Automated digital chest radiography as a pre-screening tool prior to molecular (Xpert) testing for diagnosing tuberculosis	1: Xpert is used for all patients 2: Radiographs are read by humans.	Model performance Throughput	Xpert Analytics	Partial	After	Automated digital chest radiography reduced costs compared to Xpert and increased the number of patients screened but also resulted in several TB cases that were missed.	NR
Radensky 1998[48]	Interactive neural network for rescreening negative smears for cervical cancer	Unassisted manual examination of cervical smears	Life Years	Analytics Inpatient management Outpatient management	Full	After	INNA results in increased costs(US$197-US$289) and increased health outcomes (25.492-26.44)	ICER: $35,000-$80,000 per life year saved
Sato 2014[49]	Computer-aided detection of breast cancer from mammographies with a single reader	Double reading of mammographies with two clinicians to diagnose breast cancer	Life expectancy	Mammography + first reading Second reading Analytics Diagnosis costs at outpatient facility Treatment Follow-up	Full	After	The use of computer aided detection increases costs (2,704 yen) and life expectancy (0.0087). The ICER was below the threshold used in Japan to determine cost-effectiveness.	ICER: 310,805 Yen per life year gained
Scotland 2010[50]	3. Automated detection of diabetic retinopathy from images combining quality assessment, detection of microaneurysms, blot haemorrhages and exudates	1. Automated detection of diabetic retinopathy from images using quality assessment and microaneurysms/ dot haemorrhages 2. Manual assessment of images to detect diabetic retinopathy	Primary: Cases of diabetic retinopathy Secondary: QALYs	Costs of analytics Costs of diagnosis Costs of grading	Full	After	The novel automated detection algorithm resulted in increased costs(£7,759) and an increase in cases detected (+113). Compared to manual grading the novel algorihtm reduced costs(£212,695) but decreased referable cases (123).	ICER 3vs1: £68 per correct case ICER 3 vs 2: £1,727 per correct case

Article	Intervention	Comparator	Outcome	Costs	Full or partial	Stage of development	Key economic results	ICER
Sreekumari 2019[51]	Algorithm to determine the need for a rescan of an MRI image	The clinician decides whether a rescan is needed without an algorithm	Ratio of rescans and recalls	Brain examination	Partial	During	The reduction in rescan and recall with the algorithm depended on the clinical application (i.e., recommended for MS) and expertise of the radiologist in current care. The algorithm could potentially reduce costs ($24,000 per scanner per year).	NR
Tourassi 1998[52]	Computer-assisted diagnosis of pulmonary embolism from ventilation-perfusion lung scans 1. To decide who should receive anticoagulant medication without pulmonary angiography. 2. To decide who does not receive treatment. 3. With risk score which is used to make a treatment decision. 4. Angiography for all	5. Only ventilation perfusion lung scan and no treatment	Mortality at 6 months Morbidity at 6 months	Angiography including a 2day hospital stay Heparin and long-term treatment with anticoagulants	Full	After	- Reduced costs, mortality and morbidity rates when using the ANN compared to angiography for all. - Results depend on the thresholds used in the ANN	ICER 1 vs 5: $154,872-$181,992 per extra life saved ICER 2 vs 5: $34,613-$75,437 per extra life saved ICER 3 vs 5: $46,220-$48,394 per extra life saved ICER 4 vs 5: $107,180 per extra life saved

Monitoring data								
Brocklehurst 2018[53]	Care with the continuous fetal heart rate monitoring and decision support system (INFANT)	Care with no decision support	X	Direct costs to mothers and babies	Partial	After	No difference in health and cost outcomes between the two study arms	No difference
Brüggenjürgen 2012[54]	Care with a cardiac resynchronization therapy device linked to a fluid status monitor that provides alerts (OptiVol System) to avoid pulmonary congestion, hospitalization and progression of heart failure	1. Conventional cardiac resynchronization therapy 2. Implantable cardioverter defibrillator 3. No intervention	X	Analytics Medication Adjuvants Remedies Sick pay Rehabilitation Transportation Social services Home nursing Additional expenses	Partial	After	Implementation of the cardiac resynchronization therapy + alert was more costly than the comparators but healthcare costs in the year that followed were lower.	NR
Calvert 2017[55]	Early sepsis identification and treatment using a predictive model (Insight)	1: SOFA score for sepsis screening 2: SIRS score for sepsis screening	Mortality	Intensive Care	Partial	After	Depending on the number of ICU beds, InSight generated cost savings ($340,000 - $5,500,000) and saved lives (47-752 per year) compared to SOFA and SIRS.	NR
Hollis 2018[56]	A computerised test that assesses attention and activity using a continuous performance test and a motion-tracking infrared camera to diagnose	Assessment as usual	Number and duration of appointments until diagnosis Clinicians confidence Days until diagnosis Stability in diagnosis Technology performance QALYs	Diagnosis	Full	After	The Qbtest reduced costs (small difference thus considered cost neutral by authors), increased QALYs and reduced days until diagnosis compared to current care.	Dominant
Levin 2002[57]	A prognostic technology that provides real-time recommendations for personalizing treatment/prevention in patients with chronic coronary artery disease	Care before introduction of the technology	Number of cardiovascular events and procedures	cardiovascular care (hospitalization and services), treatment, Analytics	Partial	After	The number of adverse cardiovascular events was lower, and costs were reduced with 17% after introduction of the technology	NR

Article	Intervention	Comparator	Outcome	Costs	Full or partial	Stage of development	Key economic results	ICER
Marchetti 2007[58]	Care with an early alert system that detects deleterious events such as falls or cardiopulmonary arrest	Care without an early alert system	X	Fall with injury, ICU, Cardiopulmonary resuscitation, Analytics	Partial	After	Estimated daily savings of $14.59 per patient resulting in a break-even for costs of the system.	NR
Salzsieder 2011[59]	A telemedicine assisted personalized decision support system. The system offers personalized glycaemic control and specifies treatment recommendations for primary care physicians.	Care before implementation of the decision support	Change in HbA1c Physician acceptance	Medication and treatment devices Clinical care Analytics	Partial	After	The technology resulted in lower costs and lower HbA1c measurements	NR
Sánchez-Quiroga 2018[60]	Apnealink software for automatic detection of apnea in the primary health care setting	In-laboratory specialized management.	Blood pressure Health related quality of life QALYs	Analytics visits Treatment Nurse tutorial Patient costs	Full	After	The automated detection in primary care resulted in lower costs and similar (-0.03) QALYs.	€16,449 per QALY gained.
Omics data								
Abeykoon 2016[61]	GEC + Afirma BRAF test to determine the need for surgical resection	Care before Afirma GEC implementation	Rate of surgical recommendations Incidence of thyroid malignancy diagnosed by surgery	Afirma GEC analysis/Analytics Outpatient surgery	Partial	After	Care with Afirma GEC decreased surgical recommendation rate, increased malignant surgical pathology and resulted in average savings of $1,048 per patient.	NR

Brixner 2016[62]	Pharmacogenomic testing + YouScript, a CDST that enables personalized prescribing	Care without YouScript+ pharmacogenetic profiling	X	Hospitalization Emergency department Outpatient visits	Partial	After	Potential cost savings varied from an average $126 to $218 per patient. Costs of the test are nearly or completely offset by the savings generated.	NR
D'Andrea 2020[63]	1. Bronchoscopy+ bronchial GEC to diagnose lung cancer for a subset of patients depending on the location of the lesion. Needle aspiration or biopsy for patients with only peripheral nodules. 2. Bronchoscopy+ bronchial GEC to diagnose lung cancer for all patients with moderate/intense F-FDG nodule uptake.	Current standard of care	QALYs	Direct medical Analytics	Full	After	Strategy 1 reduced the rate of invasive procedures, resulted in similar QALYs while cost neutral. Strategy 2 reduced the rate of invasive procedures, resulted in increased QALYs(0.07) and increased costs($708).	1. Dominated current care if the price of the GEC did not exceed $3,000 2. ICER of $10,109 per QALY gained
Green 2019[64]	21-gene assay (Oncotype DX) to assist treatment decision-making in breast cancer patients	Care prior to introduction of OncotypeDX	Number of patients receiving chemotherapy	Administration Drug Adverse events Analytics	Partial	After	With oncotype DX the number of patients receiving (unnecessary) chemotherapy was reduced resulting in financial savings (£25,356).	NR
Labourier 2016[65]	1: GEC combined with Afirma BRAF for all patients to determine the need for surgical resection 2: Mutation testing combined with miRNA testing for all patients to determine the need for surgical resection	Standard of care according to the American Thyroid Association guidelines of 2015.	Number of unnecessary surgeries avoided	Diagnostic/ Analytics Follow-up Surgical	Full	After	Mutation testing combined with miRNA testing dominates (i.e., less unnecessary surgeries and financial savings) GEC combined with Afirma BRAF and the current standard of care GEC combined with Afirma BRAF resulted in an increase in effects as well as costs of the current standard of care	1. $5,070 per unnecessary surgery 2. Dominant

Article	Intervention	Comparator	Outcome	Costs	Full or partial	Stage of development	Key economic results	ICER
Lobo 2017[66]	Decipher genomic classifier to assist adjuvant radiation therapy	1: 100% usage of adjuvant radiation therapy 2: Usual care (approximately 7% rate of adjuvant radiation therapy usage and 4% rate of radiation therapy combined with hormonal therapy)	QALYs Biochemical recurrence Metastasis	Radiotherapy Hormonal therapy Treatment of associated side effects Annual care GC assay/ Analytics	Full	After	No differences were found in life years gained GC increased QALYs compared to both strategies Costs of GC care were lower compared to strategy 1 and costs increased compared to strategy 2.	ICER 1. Dominant when the GC assay costs was below $11,402 ICER 2. $90,833 per QALY gained when the GC assay costs were $4,000.
Mathias 2016[67]	A genomic clinical decision support system	X	X	Design Implementation Maintenance Analytics	Partial	After	Including alerts for 8 genetic markers in a genetic clinical decision support system in an EHR can results in an estimated $4,600 per decision support event.	NR
Nelson 2013[68]	1: Individualized treatment based on VeriStrat 2: All patients receive chemo	3: All patients receive EGFR inhibitor 4: Patients receive individualized treatments based on performance status.	QALYs	Treatment Administration Follow-up Adverse events Analytics	Full	After	- Survival and costs were highest when all patients were treated with chemotherapy. - For a WTP threshold below $125,000/QALY treat all with Erlotinib was most likely to be cost-effective. - When the threshold exceeded $125,000/QALY this changed to chemotherapy for all.	ICER 2 vs.1: $105,616 p QALY gained ICER 1 vs 3: $91,111 (vs.E-all) p. QALY gained ICER 1 vs 4: $8462 p. QALY gained
Nicholson 2019[69]	1: ThyroSeq v3 to determine the need for surgical resection 2: Afirma GSC to determine the need for surgical resection	3: Diagnostic lobectomy	Number of patients correctly diagnosed	Costs incurred by third party payer (treatment, complication, analytics, Follow-up costs)	Full	After	ThyraSeq v3 was preferred over the Afirma GSC in both costs and effects. The results were robust for large variations in the pricing of the classifier tests.	ICER 1 vs 2&3: Dominant

Segui 2014[70]	1: Online algorithm (Adjuvant! Online) with 70-gene signature (Mammaprint) to assist treatment decision making 2: Online algorithm (Adjuvant! Online) with 21-gene assay (Oncotype DX) to assist treatment decision making	3: Online algorithm (Adjuvant! Online) to assist treatment decision making	QALYs	Direct medical costs (Analytics, treatment, Follow-up, recurrence)	Full	After	Mammaprint resulted in increased QALYs compared to OncotypeDX and the online algorithm alone. Costs of the Mammaprint arm were lower compared to Oncotype DX but were higher compared to online algorithm alone.	ICER Mammaprint vs Adjuvant!Online: € 1,457 p QALY ICER MammaPrint vs OncotypeDX: Dominant
Shapiro 2017[71]	Routine GEC testing of FNA-indeterminate nodules to determine the need for surgical resection	Patients managed without routine molecular testing in a tertiary care center	Reduction in surgery	Costs of surgery Anaesthesia Hospital costs Radiographic Pathologic procedures Analytics	Full	After	The number of surgeries performed reduced with 13% whilst costs increased (with $2399) when using molecular testing compared to current care.	NR

ABBREVIATIONS

ICER= Incremental Cost Effectiveness Analysis, NR= Not reported, QALYs= Quality Adjusted Life Years, MMSE= Mini-Mental State Examination, KLOSCAD-N= Korean Longitudinal Study on Cognitive Aging and Dementia Neuropsychological Assessment Battery, INR= Indian Rupee, LOS=Length of Stay, AUC=Area Under the Curve, mCRPC= metastatic castration resistant prostate cancer, RNA= Ribonucleic acid, HIV= Human Immunodeficiency virus, MEMS= Microelectromechanical systems, ART= Antiretroviral therapy, CT=computed tomography, PSG=Polysomnography, GDP= Gross Domestic Product, US= United States, AB= antibiotics, MRI, MRI= Magnetic Resonance Images, ANN=Artificial Neural Network, ICU= Intensive Care Unit, GEC= Gene Expression Classifier, miRNA= Micro ribonucleic acid, GC=Genomic Classifier, EHR = Electronic health record data, GSC= Gene Sequencing Classifier

APPENDIX C: DESCRIPTION OF THE ANALYTICS AND DATA OF THE RECORDS INCLUDED (N=71)

Article	Patients	Technology	Analytics	Data	Mention of BDA
Medical History Data					
Ashfaq 2019[1]	Congestive heart failure patients eligible for hospital discharge	Prediction algorithm for 30-day hospital readmission	Predictive analytics (deep learning network)	N development= 7655 patients Structured data from EHR	N
Bennet 2013[2]	Assist treatment decision-making. The framework is non-disease specific.	A framework that optimizes treatment decisions of clinicians	Prescriptive analytics (partially observable Markov decision process simulation framework)	N development= 5807 patients EHR data from patients with major clinical depression used for evaluation.	N
Bremer 2018[3]	Patients with depression	A model for personalizing treatment recommendations	Predictive analytics (support vector regression, regression tree, ridge regression and lasso regression)	N development=350 patients Database from an earlier study	N
Brisimi 2019[4]	Type II diabetes patients	A model that predicts the risk of hospitalization in diabetes patients	Predictive analytics (support vector machines, random forest, gradient tree boosting, sparse logistic regression).	N development= 40,921 patients Structured data from EHR	N
Burton 2019[5]	Patients at risk of urinary tract infection	A model to predict which urine samples are most likely to contain urinary tract infection	Predictive analytics (random forest, neural network, extreme gradient boosting)	N development= 212,554 urine reports Laboratory measurements and notes	Y
Chae 2001[6]	Patients receiving laboratory tests in a community health centre	An intelligent laboratory information system that consists of a result interpretation module and a therapy advisor module	Predictive analytics (rule-based reasoning model)	N development=170 patients Data on demographics, clinical history	N
Chi 2010[7]	Assist decision-making when diagnosing patients. Non-disease specific	A system that assists decision-making by providing the optimal sequence of tests when diagnosing a patient.	Prescriptive analytics (lazy support vector machines, non-linear knowledge and optimization techniques)	Development: N heart = 303 patients N thyroid = 3772 patients N diabetes = unclear N hepatitis = unclear Patient demographics and results from clinical and laboratory testing	N

Choi 2018[8]	Patients at risk of having dementia	A two-step diagnostic framework for early detection of dementia	Predictive analytics (deep neural networks, support vector machines, random forest)	N development= 3,101 patients Results from neuropsychological diagnostic tests (MMSE and KLOSCAD-N)	Y
Cooper 2005[9]	Patients reporting at the emergency department with community acquired pneumonia	A predictive model for dire outcomes	Predictive analytics (Artificial neural network)	N development= 2,287 patients EHR data	N
Duggal 2016[10]	Patients with diabetes who experience a hospital encounter	Predictive algorithm for risk of readmission	Predictive analytics (naïve Bayesian, Logistic Regression, Random Forest, Adaboost and Neural Network)	N= 9,381 Structured and unstructured (text mining) EHR data	Y
Elkin 2010[11]	Patients attending the General Medicine wards	Dxplain: a computerised education, reference and clinical decision support system that offers differential diagnoses	Prescriptive analytics (a modified form of Bayesian logic with 65,000 relationships in underlying knowledge base)[72]	N observational study = 323 consultations in the system Uses clinical information as input	N
Escudero 2013[12]	Patients with early Alzheimer's Disease	Personalization of diagnosing Alzheimer's disease based on the optimal sequence of biomarker testing	Prescriptive analytics (instance-based classifiers)	N = 86 and N = 71 Biomarkers (N=11) and several patient characteristics registered in the ADNI database.	N
Golas 2018[13]	Heart failure patients	A predictive model for 30-day hospital readmission	Predictive analytics (deep unified networks, Gradient boosting, maxout networks, logistic regression)	N= 11,510 patients Structured and unstructured EHR data	Y
Govers 2018[14]	Patients with early stage oral cavity squamous cell carcinoma with lymph node metastases in the neck	Prediction models for the risk of occult metastases, utility of patients, survival	Predictive analytics (Cox proportional hazards model, logistic regression, linear regression)	metastases N development=235 utility N development=68 & N development=83 survival N development=1,369 Data from EHRs and existing databases containing demographics and relevant outcomes	N
Grinspan 2018[15]	Children with epilepsy	Algorithm that predicts risk of emergency department use	Predictive analytics (Random Forest, adaBoost, support vector machines, Lasso logistic regression, logistic regression).	N development= 2730 and N development= 784 (2 centres) Electronic health records	N

Article	Patients	Technology	Analytics	Data	Mention of BDA
Harrison 2013[16]	Critical care patients at risk of invasive candida infection	Predictive algorithms to determine the risk of invasive candida infection	Predictive analytics (Logistic regression model)	N development=40,000 admissions N random validation=5000 admission N temporal validation= 10,000 admissions N geographic validation=5,000 admission EHR data including demographics and care history	N
Hunter-Zinck 2019[17]	Patients reporting at the emergency department	Analytics to predict and recommend test ordering in the emergency department	Prescriptive analytics (e.g., random forest, support vector machine, partial least squares and multilayer perceptron)	N development= 140,855 emergency department visits Structured and unstructured EHR data including patient demographics, vital signs, test results	N
Jamei 2017[18]	Patients at risk of 30-day readmission.	Algorithm to predict 30-day all-cause readmission	Predictive analytics (artificial neural network model based on Google's TensorFlow library)	N development= 335,815 hospital visits EHR data combined with block-level census data for social determinants	N
Javitt 2005[19]	Patients reporting in a primary care practice	A sentinel system that detects and mitigates errors in care	Prescriptive analytics (decision matrices that provide recommendations)	N in RCT = 39,462 patients Data from the EHR/claims database	N
Kockbek 2019[20]	Patients at risk of surgical site infection after gastrointestinal surgery	Predictive models to assess the risk of surgical site infections after gastrointestinal surgery	Predictive analytics (Generalized linear model, XGBoost model)	N development = 1,137 patients Results from blood test	N
Kofoed 2009[21]	Patients with systemic inflammation and suspected community-acquired infection	A decision support system that provides prescribing guidance for empirical antimicrobial therapy	Prescriptive analytics (causal probabilistic network).	N observational study = 161 patients data from EHR and laboratory, radiological, microbiological results	N

Ladabaum 2011 [22]	Patients with newly diagnosed colorectal cancer at risk of Lynch syndrome, an autosomal dominant cancer predisposition syndrome and their relatives	Screening to identify families with Lynch syndrome using predictive algorithms	Predictive analytics (Regression, Mendelian, Bayesian analysis).	MMRPro N development= 279 PREMM N development= 4,539 MMRPredict N development= 875 Data on genotype, phenotype, clinical and demographic background [73]	N
Lee 2015[23]	Population of uninsured patients and diverse socioeconomic groups that come to the emergency department	Predictive analytic framework to predict patient complaints, admission and readmission patterns.	Predictive analytics (machine-learning analytic framework that utilizes pattern recognition and text mining)	N development= 45,983 data fields from 2,509 patients through chart review, time-motion studies, interviews, vital statistics from hospital N development= 42,456 readmission status N development= >40,000 individual service times for laboratory N development= 16,217 EHR data sets (admissions, discharge, assessment, laboratory results etc.) Includes both structured and unstructured (text mining) data	Y
Lee 2018[24]	Patients at risk of a total joint replacement readmission	Analytics to predict patient 90-day admission and readmission patterns.	Predictive analytics (random under sampling boost, decision trees)	N development = 525 Demographic, preoperative, operative, postoperative and discharge variables	N
Lin 2019[25]	Patients at high risk of major depression	A prognostic monitoring framework that identifies high risk patients and personalizes monitoring strategies for patients	Prescriptive analytics (logistic regression, rule-based methods, a Markov-based collaborative model and natural history matching)	N development = 965 patients Longitudinal questionnaire and demographics	N
Ling 2006[26]	Patients with heart disease as a case study	Algorithm to optimize test selection for coronary heart disease.	Prescriptive analytics (Lazy decision tree)	N development= 296 patients[74] Patient demographics and results from clinical test and lab results	N
Marble 1999[27]	Patients at risk of sepsis	A model for diagnosing sepsis	Predictive analytics (artificial neural network)	N development= 515 patients EHR chart review	N

Article	Patients	Technology	Analytics	Data	Mention of BDA
Monahan 2018[28]	Patients with potential hypertension	Calculator to estimate PROOF-BP to predict out-of-office blood pressure	Predictive analytics (linear regression)	N= 991 for derivation N=1172 for validation Existing clinical database 14 variables included[75]	N
Murtojärvi 2020[29]	Patients with metastatic castration resistant prostate cancer	Algorithms to predict survival in patients with metastatic castration resistant prostate cancer to assist clinical decision making	Predictive analytics (several feature selection methods cost-specified greedy algorithm & LASSO regularization combined with penalized Cox regression)	N development=2181 RCT and EHR data including laboratory results of patients	N
Parekattil 2003[30]	Patients with a history of bladder cancer	Algorithm to detect bladder cancer	Predictive analytics (neural network model)	N development=253 patients Urine tumor markers from sandwich-enzyme linked immunosorbent assay kits.	N
Petersen 2015[31]	HIV- positive patients that receive antiretroviral therapy	Prediction models for virological failure to improve adherence	Predictive analytics (Superlearner)	N development= 1478 patients Patients in the MACH14 cohort. Demographic and laboratory variables and data from Medication Event Monitoring Systems (MEMS)	N
Pölsterl 2015[32]	Patients eligible for treatment of coronary atherosclerosis	Algorithm to predict the risk of adverse events for different treatment options	Predictive analytics (leaf node logistic regression)	N development = 2,377 Three sets of features based on clinical variables and biomarkers. Data collected (manually) in an observational study between 1999-2006.	N
Rider 2019[33]	Children at risk of having primary immunodeficiency disease	Screening tool (SPIRIT) to identify children with potential primary immunodeficiency disease	Predictive analytics (Analytics assign risk points and estimate whether patients are low-, medium- or high- risk. Unclear how these categories were established)	N = 185,892 patients Health claims data	Y

Ruff 2019[34]	Patients receiving direct oral anticoagulants	Predictive algorithm to determine the risk of non-adherence to treatment	To be developed	To be developed Potential predictors mentioned are those extracted from EHRs	N
Shazzadur Rahman 2019[35]	Patients at risk of tuberculosis	Triage tools prior to using the Xpert system to detect tuberculosis. The tools consisted of an artificial neural network or a theoretical product.	Predictive analytics (Artificial Neural Networks, classification and regression trees, support vector machines)[76]	N development=136 patients Demographics, results from physical examination and history of treatment. [76]	N
Shi 2010[36]	Patients who may have colorectal cancer	Algorithm to determine optimal cut-off points and combinations of serum tumor markers for diagnosing colorectal cancer	Prescriptive analytics (genetic algorithm, case-based reasoning)	N development= 578 patients Data: Biomarker measurements	N
Singh 2008[37]	Patients presenting at the emergency department	Clinical prediction model for early disposition within 3 hours after presentation at the emergency department	Predictive analytics (recursive partitioning) [77]	N development= 769 patients Demographics, results from physical examination, ECG findings, laboratory results, cardiovascular history, medication history[77]	N
Tan 2013[38]	Patients at risk of chronic rhinosinusitis	Algorithm to assist diagnostic decision-making for chronic rhinosinusitis	Predictive analytics (Classification and regression tree models)	N development= 80 patients Survey data on symptoms and results from clinical examination	N
Teferra 2014[39]	Patients at risk of obstructive sleep apnea	Prediction tool to determine who should undergo home sleep test and who should undergo an in-laboratory polysomnogram based on who is likely to have obstructive sleep apnea.	Predictive analytics (artificial neural network model)	N derivation= 383 patients N validation = 149 patients Polysomnography outcome, demographics and results from questionnaire	N
Walsh 2017[40]	Patients at risk of 30-day readmission	Predictive models for the risk of 30-day readmission	Predictive analytics (L1-regularized logistic regression (LASSO estimator))	N development= 92,530 patients Variables extracted from the electronic health record.	N

Article	Patients	Technology	Analytics	Data	Mention of BDA
Wang 2019[41]	Patients with at risk of having Trichomonas vaginalis infection	Predictive model to increase detection of Trichomonas vaginalis in urine samples	Predictive analytics (Logistic regression, support vector machine, random forest)	N development=410,952 urine samples from women, 428,203 urine samples from men. Urinalysis test data	N
Warner 2016[42]	Patients at risk of having hospital acquired complications	Predictive model for early identification of hospital acquired complications	Predictive analytics (Tree-augmented naïve Bayes algorithm, AdaBoost	N development= 8183 admissions Data from MIMIC II (lab results, demographics, prescriptions	N
Webb 2019[43]	Patients with community-onset pneumonia presenting at the emergency department	Prediction tool (Drug-Resistance in Pneumonia) to identify pneumonia with drug-resistant pathogens	Predictive analytics (Logistic regression)[78]	N development= 200 patients N validation= 200 patients Patient demographics, treatment history[78]	N
Zhang 2019[44]	Patients at risk of having fetal alcohol spectrum disorders	A screening protocol for children at risk of having fetal alcohol spectrum disorders	Predictive analytics (Classification analysis (e.g., naïve Bayes), Multilinear regression)	N development =207 patients Distinct results from imaging, eye movement tests, psychometric tests. The technology focusses on the decision after all results from the test are available. Not on the analysis of imaging	N
Imaging Data					
Carballido-Gamio 2019[45]	Patients at risk of a hip fracture	A model that predicts the risk of hip fractures	Predictive analytics (principal component analysis, logistic regression LASSO)	N development= 143 patients Quantitative computed tomography images	N
Crespo 2019[46]	Patients with symptomatic chronic heart failure	Algorithms that use cardiac magnetic resonance imaging-based algorithms when deciding whether to implement a pacemaker or cardioverter-defibrillator device.	Predictive analytics (Semiautomated segmentation algorithm for identifying border zone channels in scar tissue in ceMRI.)[79,80]	N development= 22 patients[79] cardiac magnetic resonance imaging data	N

Philipsen 2015[47]	Patients at risk of tuberculosis	Automated chest radiography, which consists of digital chest radiography with automated interpretation by computer software.	Predictive analytics (supervised machine learning)[81]	N prospective study = 388 Chest radiographs	N
Radensky 1998[48]	Women eligible for cervical screening for cancer.	Screening of magnified cervical smear images for cervical cancer	Predictive analytics (interactive, neural network–assisted screening employs advanced pattern recognition)	N development/validation = 203 smears[82] Magnified, digital cervical smear images.	N
Sato 2014[49]	50-year-old women eligible for breast cancer screening	Computer-aided detection of mammography	Predictive analytics (Unknown, depend on algorithm used)	N development= unknown Modelling input from a systematic review which combines multiple studies that use imaging data combined with analytics (e.g. neural networks).	N
Scotland 2010[50]	Patients at risk of having diabetic retinopathy	Automated image quality assessment and detection of microaneurysms and dot haemorrhages	Predictive analytics (k nearest neighbors classifiers)[83]	N development quality assessment = 489 patients [84] N development automated detection = 422 patients[83] Photographic Images from digital cameras.	N
Sreekumari 2019[51]	Patients at risk of having a neurological disorder receiving an MRI	Algorithm to determine the need for a rescan of an MRI image (i.e., due to poor quality of the image)	Predictive analytics (convolutional neural network)	N development= 803 series, 17,166 slices Magnetic Resonance Brain Images	N
Tourassi 1998[52]	Patients suspected of having pulmonary embolism	Computer-assisted diagnosis of pulmonary embolism from ventilation-perfusion lung scans	Predictive analytics (artificial neural networks)	N development= 1,064 patients[85] Ventilation-perfusion lung scans.[85]	N

Article	Patients	Technology	Analytics	Data	Mention of BDA
Monitoring Data					
Brocklehurst 2018[53]	Pregnant women	INFANT system: combines a rule base and inference engine to generate relevant alerts based on intrapartum cardiotocograph monitoring and information on risk factors	Prescriptive analytics (neural networks and numerical algorithms used for intrapartum cardiotocograph feature extraction)[86]	N RCT= 234 patients N development= 31 patients Intrapartum cardiotocographs[86]	N
Brüggenjürgen 2012[54]	Heart failure patients	Cardiac resynchronization therapy device linked to a fluid status monitor that provides alerts to avoid pulmonary congestion, hospitalization and progression of heart failure	Descriptive analytics (intra-thoracic impedance measurement is compared to a personalized reference value every 6hr)[87]	N development= 33 patients[87] N observational study = 3076 Sensor data on intra-thoracic impedance[87]	N
Calvert 2017[55]	Intensive care unit patients	Insight: Algorithm for early sepsis prediction	Predictive analytics (machine learning algorithm for early sepsis prediction)[88]	N development=1,394 patients Time series data from the MIMIC II Clinical Database including the following variables: - Systolic blood pressure - Pulse pressure - Heart rate - Temperature - Respiration rate - White blood cell count - PH - Blood oxygen saturation - Age[88]	N
Hollis 2018[56]	Children with suspected attention deficit hyperactivity disorder	A computerised test that assesses attention and activity using a continuous performance test and a motion-tracking infrared camera	Descriptive analytics (Deviation from a normative data set)	N RCT= 135 patients Using continuous performance test and infrared camera that detects motor activity	N

Levin 2002[57]	Patients with chronic coronary artery disease	A prognostic technology that provides real-time recommendations for personalizing treatment/ prevention	Descriptive analytics (Assessment of deviation from targets using ECG data and clinical factors)	N observational study=1628 Combines ECG monitoring, patient risk factors and compares to targets	N
Marchetti 2007[58]	Patients with a medical-surgical hospitalization at risk of falling or cardiopulmonary arrest	Early alert system that detects deleterious events such as falls or cardiopulmonary arrest	Descriptive analytics (Deviation from a normative range)	N development =unknown Modelling input based on the literature and the tool sensor data to detect motion and bed departures. Physiological data on heart and respiratory measurements	N
Salzsieder 2011[59]	Patients with diabetes and cardiovascular disease	A telemedicine assisted personalized decision support system. The system offers personalized glycaemic control and specifies treatment recommendations for primary care physicians.	Prescriptive analytics (Patented algorithm consisting of coupled differential equation & weighted least squares regression)[89]	N development= unknown N validation=34, 18, 49 N retrospective study=538 Continuous glucose monitoring data and self-control data (e.g., treatment received, exercise performed) and demographics[89]	N
Sánchez-Quiroga 2018[60]	Patients with an intermediate to high risk of obstructive sleep apnea	Apnealink software for automatic detection of apnea	Descriptive analytics (analyzes flow limitation by comparing the flow/time curve to specific range)[90]	N RCT=307 patients N validation = 59[90] Airflow monitoring data[90]	N
Omics Data					
Abeykoon 2016[61]	Patients with indeterminate thyroid nodule cytology	GEC + Afirma BRAF test (GEC) to determine the need for surgical resection	Predictive analytics (Proprietary algorithm for classification based on support vector machines)[91]	N development= 247,186 transcripts in 315 thyroid nodules[91] N observational study= 299 RNA expression analysis using microarrays[91]	N
Brixner 2016[62]	Elderly patients exposed to polypharmacy	Pharmacogenomic testing + YouScript, a clinical decision support system that enables personalized prescribing	Prescriptive analytics (YouScript proprietary algorithm)	N observational study= 216 patients Combines results from pharmacogenomic data (PCR data) + literature+ drug prescriptions	N
D'Andrea 2020[63]	Patients at risk of lung cancer with indeterminate pulmonary nodules	GEC for diagnosing lung cancer	Predictive analytics (penalized logistic regression and empirical Bayes linear model) [92]	N development= 299 patients RNA expression analysis using microarrays[92]	N

Article	Patients	Technology	Analytics	Data	Mention of BDA
Green 2019[64]	Women with oestrogen receptor positive, human epidermal growth factor receptor 2 negative, node negative, early invasive breast cancer	Oncotype DX: 21-gene RT-PCR assay to assist treatment decision-making	Predictive analytics (Cox regression)	N development=447 patients Results from RT-PCR [93] N observational study = 101 patients Results from RT-PCR	N
Labourier 2016[65]	Patients with indeterminate thyroid nodule cytology	GEC + Afirma BRAF test or Mutation testing + ThyraMIR Thyroid miRNA Classifier to determine the need for surgical resection	Predictive analytics (Proprietary algorithm for classification based on support vector machines)[91] (Diagonal linear discriminant analysis, k-nearest neighbours, Logistic regression, Neural network, Random forest, Quadratic discriminant analysis, Support vector machine:linear kernel, Support vector machine:radial kernel)[93]	GEC N development= 247,186 transcripts in 315 thyroid nodules RNA expression analysis using microarrays[91] ThyraMIR N development= 534 thyroid lesions miRNA specimens from reverse rt-PCR[94]	N
Lobo 2017[66]	Prostate cancer patients eligible for adjuvant radiation therapy after radical prostatectomy	Decipher genomic classifier	Predictive analytics (random forest classifier)[95]	N development= 545 patients Data: RNA expression analysis using microarrays[95]	N
Mathias 2016[67]	No specific group defined	Genomic clinical decision support implemented in the EHR system	Prescriptive analytics (results from whole exome sequencing compared to known variants)[95].	N observational study= 94 patients Results from whole exome sequencing compared to known variants for certain clinical conditions[96].	N
Nelson 2013[68]	60-year-old patients with advanced NSCLC despite treatment with first-line chemotherapy	Veristrat: assesses the serum protein expression profile to assist treatment decisions in lung cancer	Predictive analytics (k-nearest neighbors (kNN) classification)[97]	N development= 139 patients[97] Analysis of serum proteomic data using mass spectrometry to develop Veristrat. [97]	N

Nicholson 2019[69]	Patients with indeterminate thyroid nodule cytology	ThyroSeq v3 to determine the need for surgical resection Afirma GSC to determine the need for surgical resection	Predictive analytics ThyroSeq v3: Classification (proprietary sequencer for NGS data)[98] Afirma GSC: Classification (Classifier based on elastic net logistic regression and support vector machines)[99]	Thyroseq v3: N development = 238 tissue samples and 175 FNA samples[98] GSC: N development = 634 patients NGS data[99]	N
Segui 2014[70]	Patients with potential breast cancer	Oncotype DX: 21-gene RT-PCR assay to assist treatment decisions MammaPrint: 70-gene signature to assist treatment decisions	Predictive analytics MammaPrint: Clustering and classification (agglomerative hierarchical clustering algorithm)[100] Oncotype DX: Prediction (Cox regression) [93]	MammaPrint: N development=117 patients DNA microarray data[100] Oncotype DX: N development=447 patients Results from RT-PCR[93] Adjuvant! Online: -	N
Shapiro 2017[71]	Patients with indeterminate thyroid nodule cytology	GEC + Afirma BRAF test to determine the need for surgical resection	Predictive analytics (Proprietary algorithm for classification based on support vector machines)[91]	N observational study= 96 patients N development= 247,186 transcripts in 315 thyroid nodules RNA expression analysis using microarrays[91]	N

ABBREVIATIONS

BDA =Big data analytics, EHR = Electronic health record data, MMSE= Mini-Mental State Examination, KLOSCAD-N= Korean Longitudinal Study on Cognitive Aging and Dementia Neuropsychological Assessment Battery, ADNI= Alzheimer's Disease Neuroimaging Initiative, RCT= Randomized Controlled Trial, ceMRI= Contrast-Enhanced Magnetic Resonance Imaging, HIV= Human Immunodeficiency virus, MACH14= Multi-site Adherence Collaboration 14, MRI= Magnetic Resonance Images, ECG= electrocardiogram, RNA= Ribonucleic acid, RT-PCR= Reverse transcription polymerase chain reaction, GEC= Gene Expression Classifier, GSC= Gene Sequencing Classifier, NGS= Next Generation Sequencing

REFERENCES

1 Ashfaq A, Sant'Anna A, Lingman M, et al. Readmission prediction using deep learning on electronic health records. J Biomed Inform. 2019;97.
2 Bennett CC, Hauser K. Artificial intelligence framework for simulating clinical decision-making: A Markov decision process approach. Artif Intell Med. 2013;57:9–19.
3 Bremer V, Becker D, Kolovos S, et al. Predicting Therapy Success and Costs for Personalized Treatment Recommendations Using Baseline Characteristics: Data-Driven Analysis. J Med Internet Res. 2018;20:e10275.
4 Brisimi TS, Xu T, Wang T, et al. Predicting diabetes-related hospitalizations based on electronic health records. Stat Methods Med Res. 2019;28:3667–82.
5 Burton RJ, Albur M, Eberl M, Cuff SM. Using artificial intelligence to reduce diagnostic workload without compromising detection of urinary tract infections. BMC Med Inform Decis Mak. 2019;19:171.
6 Chae YM, Lim HS, Lee JH, et al. Development of an intelligent laboratory information system for community health promotion center. Medinfo. 2001;10:425–8.
7 Chi CL, Street WN, Katz DA. A decision support system for cost-effective diagnosis. Artif Intell Med. 2010;50:149–61.
8 Choi HS, Choe JY, Kim H, et al. Deep learning based low-cost high-accuracy diagnostic framework for dementia using comprehensive neuropsychological assessment profiles. BMC Geriatr. 2018;18:234.
9 Cooper GF, Abraham V, Aliferis CF, et al. Predicting dire outcomes of patients with community acquired pneumonia. J Biomed Inform. 2005;38:347–66.
10 Duggal R, Shukla S, Chandra S, et al. Predictive risk modelling for early hospital readmission of patients with diabetes in India. International Journal of Diabetes in Developing Countries. 2016;36:519–28.
11 Elkin PL, Liebow M, Bauer BA, et al. The introduction of a diagnostic decision support system (DXplainTM) into the workflow of a teaching hospital service can decrease the cost of service for diagnostically challenging Diagnostic Related Groups (DRGs). Int J Med Inform. 2010;79:772
12 Escudero J, Ifeachor E, Zajicek JP, et al. Machine learning-based method for personalized and cost-effective detection of alzheimer's disease. IEEE Trans Biomed Eng. 2013;60:164–8.
13 Golas SB, Shibahara T, Agboola S, et al. A machine learning model to predict the risk of 30-day readmissions in patients with heart failure: a retrospective analysis of electronic medical records data. BMC Med Inform Decis Mak. 2018;18:44.
14 Govers TM, Rovers MM, Brands MT, et al. Integrated prediction and decision models are valuable in informing personalized decision making. J Clin Epidemiol. 2018;104:73–83.
15 Grinspan ZM, Patel AD, Hafeez B, et al. Predicting frequent emergency department use among children with epilepsy: A retrospective cohort study using electronic health data from 2 centers. Epilepsia. 2018;59:155–69.
16 Harrison D, Muskett H, Harvey S, et al. Development and validation of a risk model for identification of non-neutropenic, critically ill adult patients at high risk of invasive Candida infection: the Fungal Infection Risk Evaluation (FIRE) Study Review. Health technology assessment (Winchester, England). 2013;17:1–156.
17 Hunter-Zinck HS, Peck JS, Strout TD, et al. Predicting emergency department orders with multilabel machine learning techniques and simulating effects on length of stay. J Am Med Inform Assoc. 2019;26:1427–36.
18 Jamei M, Nisnevich A, Wetchler E, et al. Predicting all-cause risk of 30-day hospital readmission using artificial neural networks. PLoS One. 2017;12.
19 Javitt JC, Steinberg G, Locke T, et al. Using a claims data-based sentinel system to improve compliance with clinical guidelines: results of a randomized prospective study. Am J Manag Care. 2005;11:93–102.
20 Kocbek P, Fijacko N, Soguero-Ruiz C, et al. Maximizing Interpretability and Cost-Effectiveness of Surgical Site Infection (SSI) Predictive Models Using Feature-Specific Regularized Logistic Regression on Preoperative Temporal Data. Comput Math Methods Med. 2019;2019.
21 Kofoed K, Zalounina A, Andersen O, et al. Performance of the TREAT decision support system in an environment with a low prevalence of resistant pathogens. J Antimicrob Chemother. 2009;63:400–4.
22 Ladabaum U, Wang G, Terdiman J, Blanco A, Kuppermann M, Boland CR, Ford J, Elkin E, Phillips KA. Strategies to identify the Lynch syndrome among patients with colorectal cancer: a cost-effectiveness analysis. Ann Intern Med. 2011:155:69-79.

23 Lee EK, Atallah HY, Wright MD, et al. Transforming Hospital Emergency Department Workflow and Patient Care. Interfaces. 2015;45:58–82.
24 Lee HK, Jin R, Feng Y, et al. An Analytical Framework for TJR Readmission Prediction and Cost-Effective Intervention. IEEE J Biomed Health Inform. 2018;23:1760–72.
25 Lin Y, Huang S, Simon GE, et al. Cost-effectiveness analysis of prognostic-based depression monitoring. IISE Transactions on Healthcare Systems Engineering. 2019;9:41–54.
26 Ling CX, Sheng VS, Yang Q. Test strategies for cost-sensitive decision trees. IEEE Transactions on Knowledge and Data Engineering. 2006;18:1055–67.
27 Marble RP, Healy JC. A neural network approach to the diagnosis of morbidity outcomes in trauma care. Artif Intell Med. 1999;15:299–307.
28 Monahan M, Jowett S, Lovibond K, et al. Predicting out-of-office blood pressure in the clinic for the diagnosis of hypertension in primary care: An economic evaluation. Hypertension. 2018;71:250–61.
29 Murtojärvi M, Halkola AS, Airola A, et al. Cost-effective survival prediction for patients with advanced prostate cancer using clinical trial and real-world hospital registry datasets. Int J Med Inform. 2020;133.
30 Parekattil SJ, Fisher HAG, Kogan BA. Neural network using combined urine nuclear matrix protein-22, monocyte chemoattractant protein-1 and urinary intercellular adhesion molecule-1 to detect bladder cancer. J Urol. 2003;169:917–20.
31 Petersen ML, LeDell E, Schwab J, et al. Super learner analysis of electronic adherence data improves viral prediction and may provide strategies for selective HIV RNA monitoring. J Acquir Immune Defic Syndr. 2015;69:109–18.
32 Pölsterl S, Singh M, Katouzian A, et al. Stratification of coronary artery disease patients for revascularization procedure based on estimating adverse effects. BMC Med Inform Decis Mak. 2015;15:9.
33 Rider NL, Miao D, Dodds M, et al. Calculation of a primary immunodeficiency "risk vital sign" via population-wide analysis of claims data to aid in clinical decision support. Frontiers in Pediatrics. 2019;7.
34 Ruff C, Koukalova L, Haefeli WE, et al. The role of adherence thresholds for development and performance aspects of a prediction model for direct oral anticoagulation adherence. Frontiers in Pharmacology. 2019;10.
35 Shazzadur Rahman AAM, Langley I, Galliez R, et al. Modelling the impact of chest X-ray and alternative triage approaches prior to seeking a tuberculosis diagnosis. BMC Infect Dis. 2019;19.
36 Shi J, Su Q, Zhang C, et al. An intelligent decision support algorithm for diagnosis of colorectal cancer through serum tumor markers. Comput Methods Programs Biomed. 2010;100:97–107.
37 Singh S, Nosyk B, Sun H, et al. Value of Information of a clinical prediction rule: Informing the efficient use of healthcare and health research resources. Int J Technol Assess Health Care. 2008;24:112–9.
38 Tan BK, Lu G, Kwasny MJ, et al. Effect of symptom-based risk stratification on the costs of managing patients with chronic rhinosinusitis symptoms. Int Forum Allergy Rhinol. 2013;3:933–40.
39 Teferra RA, Grant BJB, Mindel JW, et al. Cost minimization using an artificial neural network sleep apnea prediction tool for sleep studies. Ann Am Thorac Soc. 2014;11:1064–74.
40 Walsh CG, Sharman K, Hripcsak G. Beyond discrimination: A comparison of calibration methods and clinical usefulness of predictive models of readmission risk. J Biomed Inform. 2017;76:9–18.
41 Wang HY, Hung CC, Chen CH, et al. Increase Trichomonas vaginalis detection based on urine routine analysis through a machine learning approach. Sci rep. 2019;9:11074.
42 Warner JL, Zhang P, Liu J, et al. Classification of hospital acquired complications using temporal clinical information from a large electronic health record. J Biomed Inform. 2016;59:209–17.
43 Webb BJ, Sorensen J, Mecham I, et al. Antibiotic Use and Outcomes After Implementation of the Drug Resistance in Pneumonia Score in ED Patients With Community-Onset Pneumonia. Chest. 2019;156:843–51.
44 Zhang C, Paolozza A, Tseng PH, et al. Detection of children/youth with fetal alcohol spectrum disorder through eye movement, psychometric, and neuroimaging data. Frontiers in Neurology. 2019;10.
45 Carballido-Gamio J, Yu A, Wang L, et al. Hip Fracture Discrimination Based on Statistical Multi-parametric Modeling (SMPM). Ann Biomed Eng. 2019;47:2199–212.
46 Crespo C, Linhart M, Acosta J, et al. Optimisation of cardiac resynchronisation therapy device selection guided by cardiac magnetic resonance imaging: Cost-effectiveness analysis. Eur J Prev Cardiol. 2019: 2047487319873149.

47 Philipsen RH, Sánchez CI, Maduskar P, et al. Automated chest-radiography as a triage for Xpert testing in resource-constrained settings: a prospective study of diagnostic accuracy and costs. Sci Rep. 2015;5:12215.
48 Radensky PW, Mango LJ. Interactive neural network-assisted screening: An economic assessment. Acta Cytol. 1998;42:246–52.
49 Sato M, Kawai M, Nishino Y, et al. Cost-effectiveness analysis for breast cancer screening: Double reading versus single + CAD reading. Breast Cancer. 2014;21:532–41.
50 Scotland GS, McNamee P, Fleming AD et al. Costs and consequences of automated algorithms versus manual grading for the detection of referable diabetic retinopathy. Br J Ophthalmol. 2010;94:712-9.51 Sreekumari A, Shanbhag D, Yeo D, et al. A deep learning-based approach to reduce rescan and recall rates in clinical MRI examinations. AJNR Am J Neuroradiol. 2019;40:217–23.
52 Tourassi GD, Floyd CE, Coleman RE. Acute pulmonary embolism: Cost-effectiveness analysis of the effect of artificial neural networks on patient care. Radiology. 1998;206:81–8.
53 Brocklehurst P, Field D, Greene K, et al. Computerised interpretation of the fetal heart rate during labour: A randomised controlled trial (INFANT). Health Technology Assessment. 2018;22:1–218.
54 Brüggenjürgen B, Israel CW, Klesius AA, et al. Health services research in heart failure patients treated with a remote monitoring device in Germanya retrospective database analysis in evaluating resource use. J Med Econ. 2012;15:737–45.
55 Calvert J, Hoffman J, Barton C, et al. Cost and mortality impact of an algorithm-driven sepsis prediction system. J Med Econ. 2017;20:646–51.
56 Hollis C, Hall CL, Guo B, et al. The impact of a computerised test of attention and activity (QbTest) on diagnostic decision-making in children and young people with suspected attention deficit hyperactivity disorder: single-blind randomised controlled trial. J Child Psychol Psychiatry. 2018;59:1298–308.
57 Levin RI, Koenig KL, Corder MP, et al. Risk stratification and prevention in chronic coronary artery disease: Use of a novel prognostic and computer-based clinical decision support system in a large primary managed-care group practice. Disease Management. 2002;5:197–213.
58 Marchetti A, Jacobs J, Young M, et al. Costs and benefits of an early-alert surveillance system for hospital inpatients. Curr Med Res Opin. 2007;23:9–16.
59 Salzsieder E, Augstein P. The Karlsburg Diabetes Management System: Translation from research to eHealth application. J Diabetes Sci Technol. 2011;5:13–22.
60 Sánchez-Quiroga MA, Corral J, Gómez-de-Terreros FJ, et al. Primary care physicians can comprehensively manage patients with sleep apnea a noninferiority randomized controlled trial. Am J Respir Crit Care Med. 2018;198:648–56.
61 Abeykoon JP, Mueller L, Dong F, et al. The Effect of Implementing Gene Expression Classifier on Outcomes of Thyroid Nodules with Indeterminate Cytology. Horm Cancer. 2016;7:272–8.
62 Brixner D, Biltaji E, Bress A, et al. The effect of pharmacogenetic profiling with a clinical decision support tool on healthcare resource utilization and estimated costs in the elderly exposed to polypharmacy. J Med Econ. 2016;19:213–28.
63 D'Andrea E, Choudhry NK, Raby B, et al. A bronchial-airway gene-expression classifier to improve the diagnosis of lung cancer: Clinical outcomes and cost-effectiveness analysis. Int J Cancer. 2020;146:781-790.
64 Green N, Al-Allak A, Fowler C. Benefits of introduction of Oncotype DX® testing. Ann R Coll Surg Engl. 2019;101:55–9.
65 Labourier E. Utility and cost-effectiveness of molecular testing in thyroid nodules with indeterminate cytology. Clin Endocrinol (Oxf). 2016;85(4):624-3166 Lobo JM, Trifiletti DM, Sturz VN, et al. Cost-effectiveness of the Decipher Genomic Classifier to Guide Individualized Decisions for Early Radiation Therapy After Prostatectomy for Prostate Cancer. Clin Genitourin Cancer. 2017;15:e299–309.
67 Mathias PC, Tarczy-Hornoch P, Shirts BH. Modeling the costs of clinical decision support for genomic precision medicine. AMIA Summits on Translational Science Proceedings. 2016;2016:60–4.
68 Nelson RE, Stenehjem D, Akerley W. A comparison of individualized treatment guided by VeriStrat with standard of care treatment strategies in patients receiving second-line treatment for advanced non-small cell lung cancer: A cost-utility analysis. Lung Cancer. 2013;82:461–8.
69 Nicholson KJ, Roberts MS, McCoy KL, et al. Molecular Testing Versus Diagnostic Lobectomy in Bethesda III/IV Thyroid Nodules: A Cost-Effectiveness Analysis. Thyroid. 2019;29:1237–43.

70 Seguí MÁ, Crespo C, Cortés J, et al. Genomic profile of breast cancer: Cost-effectiveness analysis from the Spanish National Healthcare System perspective. Expert Rev Pharmacoecon Outcomes Res. 2014;14:889–99.
71 Shapiro S, Pharaon M, Kellermeyer B. Cost-effectiveness of Gene Expression Classifier Testing of Indeterminate Thyroid Nodules Utilizing a Real Cohort Comparator. Otolaryngol Head Neck Surg. 2017;157:596–601.
72 Barnett GO, Cimino JJ, Hupp JA, et al. DXplain: an evolving diagnostic decision-support system. JAMA. 1987;258(1):67–74.
73 Kastrinos F, Balmaña J, Syngal S. Prediction models in Lynch syndrome. Fam Cancer. 2013;12:217-28.
74 Turney PD. Cost-sensitive classification: Empirical evaluation of a hybrid genetic decision tree induction algorithm. Journal of artificial intelligence research. 1994;2:369-409.
75 Sheppard JP, Stevens R, Gill P, et al. Predicting Out-of-Office Blood Pressure in the Clinic (PROOF-BP) Derivation and Validation of a Tool to Improve the Accuracy of Blood Pressure Measurement in Clinical Practice. Hypertension. 2016 May;67(5):941-50.
76 João Filho B de O, de Seixas JM, Galliez R, et al. A screening system for smear-negative pulmonary tuberculosis using artificial neural networks. Int J Infect Dis. 2016;49:33–9.
77 Christenson J, Innes G, McKnight D, et al. A clinical prediction rule for early discharge of patients with chest pain. Ann Emerg Med. 2006;47(1):1–10.
78 Webb BJ, Dascomb K, Stenehjem E, et al. Derivation and multicenter validation of the drug resistance in pneumonia clinical prediction score. Antimicrob Agents Chemother. 2016;60(5):2652–63.
79 Serra L, i Ventura RMF, Planes X, et al. Identification of border-zone corridors in the left ventricle using the core expansion method. In: 12th International Symposium on Medical Information Processing and Analysis. International Society for Optics and Photonics; 2017. p. 1016011.
80 Acosta J, Fernández-Armenta J, Borràs R, et al. Scar characterization to predict life-threatening arrhythmic events and sudden cardiac death in patients with cardiac resynchronization therapy: the GAUDI-CRT Study. JACC Cardiovasc Imaging. 2018;11(4):561–72.
81 Maduskar P, Muyoyeta M, Ayles H, et al. Detection of tuberculosis using digital chest radiography: automated reading vs. interpretation by clinical officers. Int J Tuberc Lung Dis. 2013;17(12):1613–20.
82 Koss LG, Lin E, Schreiber K, et al. Evaluation of the PAPNETTM cytologic screening system for quality control of cervical smears. Am J Clin Pathol. 1994;101(2):220–9.
83 Fleming AD, Philip S, Goatman KA et al. Automated microaneurysm detection using local contrast normalization and local vessel detection. IEEE Trans Med Imaging. 2006;25:1223-32.
84 Fleming AD, Philip S, Goatman KA et al. Automated assessment of diabetic retinal image quality based on clarity and field definition. Invest Ophthalmol Vis Sci. 2006;47:1120-5.
85 Tourassi GD, Floyd CE, Sostman HD, et al. Acute pulmonary embolism: artificial neural network approach for diagnosis. Radiology. 1993;189(2):555–8.
86 Keith RDF, Greene KR. Development, evaluation and validation of an intelligent system for the management of labour. Bailliere's clinical obstetrics and gynaecology. 1994;8(3):583–605.
87 Yu C-M, Wang LI, Chau E, et al. Intrathoracic impedance monitoring in patients with heart failure: correlation with fluid status and feasibility of early warning preceding hospitalization. Circulation. 2005;112(6):841–8.
88 Calvert JS, Price DA, Chettipally UK, et al. A computational approach to early sepsis detection. Comput Biol Med. 2016;74:69–73.
89 Salzsieder E, Vogt L, Kohnert K-D, et al. Model-based decision support in diabetes care. Comput Methods Programs Biomed. 2011;102(2):206–18.
90 Erman MK, Stewart D, Einhorn D, et al. Validation of the ApneaLinkTM for the screening of sleep apnea: a novel and simple single-channel recording device. J Clin Sleep Med. 2007;3(04):387–92.
91 Chudova D, Wilde JI, Wang ET, et al. Molecular classification of thyroid nodules using high-dimensionality genomic data. J Clin Endocrinol Metab. 2010;95(12):5296–304.
92 Whitney DH, Elashoff MR, Porta-Smith K, et al. Derivation of a bronchial genomic classifier for lung cancer in a prospective study of patients undergoing diagnostic bronchoscopy. BMC Med Genomics. 2015;8(1):18.

93 Paik S, Shak S, Tang G, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. N Engl J Med. 2004;351(27):2817–26. 94 Wylie D, Beaudenon-Huibregtse S, Haynes BC, et al. Molecular classification of thyroid lesions by combined testing for miRNA gene expression and somatic gene alterations. J Pathol Clin Res. 2016;2(2):93–103.95 Erho N, Crisan A, Vergara IA, et al. Discovery and validation of a prostate cancer genomic classifier that predicts early metastasis following radical prostatectomy. PLoS One. 2013;8(6).

96 Nishimura AA, Shirts BH, Dorschner MO, et al. Development of clinical decision support alerts for pharmacogenomic incidental findings from exome sequencing. Genet Med. 2015;17(11):939.

97 Gautschi O, Dingemans A-M, Crowe S, et al. VeriStrat® has a prognostic value for patients with advanced non-small cell lung cancer treated with erlotinib and bevacizumab in the first line: Pooled analysis of SAKK19/05 and NTR528. Lung cancer. 2013;79(1):59–64.

98 Nikiforov YE, Carty SE, Chiosea SI, Coyne C, Duvvuri U, Ferris RL, et al. Highly accurate diagnosis of cancer in thyroid nodules with follicular neoplasm/suspicious for a follicular neoplasm cytology by ThyroSeq v2 next-generation sequencing assay. Cancer. 2014;120(23):3627–34.

99 Patel KN, Angell TE, Babiarz J et al. Performance of a genomic sequencing classifier for the preoperative diagnosis of cytologically indeterminate thyroid nodules. JAMA Surg. 2018;153(9):817–24.

100 Veer LJ, Dai H, Van De Vijver MJ, et al. Gene expression profiling predicts clinical outcome of breast cancer. Nature. 2002;415(6871):530–6.

The potential of real-time analytics to improve care for mechanically ventilated patients in the intensive care unit: an early economic evaluation.

Bakker L, Vaporidi K, Aarts J, Redekop W.

Cost Effectiveness and Resource Allocation. 2020 Dec;18(1):1-1.

ABSTRACT

Background: Mechanical ventilation services are an important driver of the high costs of intensive care. An optimal interaction between a patient and a ventilator is therefore paramount. Suboptimal interaction is present when patients repeatedly demand, but do not receive, breathing support from a mechanical ventilator (>30 times in 3 min), also known as an ineffective effort event (IEEV). IEEVs are associated with increased hospital mortality prolonged intensive care stay, and prolonged time on ventilation and thus development of real-time analytics that identify IEEVs is essential. To assist decision-making about further development we estimate the potential cost-effectiveness of real-time analytics that identify ineffective effort events.

Methods: We developed a cost-effectiveness model combining a decision tree and Markov model for long-term outcomes with data on current care from a Greek hospital and literature. A lifetime horizon and a healthcare payer perspective were used. Uncertainty about the results was assessed using sensitivity and scenario analyses to examine the impact of varying parameters like the intensive care costs per day and the effectiveness of treatment of IEEVs.

Results: Use of the analytics could lead to reduced mortality (3% absolute reduction), increased quality adjusted life years (0.21 per patient) and cost-savings (€264 per patient) compared to current care. Moreover, cost-savings for hospitals and health improvements can be incurred even if the treatment's effectiveness is reduced from 30% to 10%. The estimated savings increase to €1,155 per patient in countries where costs of an intensive care day are high (e.g., the Netherlands). There is considerable headroom for development and the analytics generate savings when the price of the analytics per bed per year is below €7,307. Furthermore, even when the treatment's effectiveness is 10%, the probability that the analytics are cost-effective exceeds 90%.

Conclusions: Implementing real-time analytics to identify ineffective effort events can lead to health and financial benefits. Therefore, it will be worthwhile to continue assessment of the effectiveness of the analytics in clinical practice and validate our findings. Eventually, their adoption in settings where costs of an intensive care day are high and ineffective efforts are frequent could yield a high return on investment.

BACKGROUND

Annual intensive care costs in the United States represent more than 13% of all hospital costs [1]. The costs of an intensive care unit (ICU) day per patient are high (e.g., €5,695) and an important factor that contributes to these high daily costs is whether or not patients receive mechanical ventilation [2]. Therefore, better management of mechanically ventilated patients could be a worthwhile investment when it reduces length of stay and their time on ventilation support.

One way to achieve better outcomes in the intensive care is by using analytics to process the huge amounts of monitoring data that are continuously collected in order to improve clinical decision-making [3]. Ventilation monitors in the ICU generate a wealth of data on a patient's status and patient-monitor interaction. Ideally, this data can be used to help clinicians intervene promptly when the interaction between the patient and the monitor is poor. One example of poor interaction is when a patient tries but does not receive a breath. These so-called 'ineffective efforts' are reflected in the airway pressure and airflow data from the monitor [4]. When many ineffective efforts occur in a short period of time (>30 ineffective efforts in 3 min.) it is referred to as an ineffective effort event (IEEV) which have been associated with higher hospital mortality, an increase in ICU length of stay of almost 10 days and prolonged time on mechanical ventilation [5]. Timely identification of ineffective effort events is crucial and early-warning systems using big data analytics have been portrayed as an important means to improve care for mechanically ventilated patients [4, 6] since the complexity and velocity required to process this data in real-time are beyond the capacities of humans such as healthcare professionals.

Real-time analytics of ventilation data would enable clinicians to identify IEEVs and intervene accordingly thereby shortening their duration and potentially reducing mortality risks and healthcare costs. Several types of interventions are recommended to improve the interaction between a patient and a mechanical ventilator such as, adjustment of ventilator settings, reducing sedation when managing pain and anxiety [7,8] and adjustments in the management of bronchodilation [8]. Developing real-time analytics that identify IEEVs would enable clinicians to adopt these interventions currently already recommended when other forms of suboptimal interaction are present, identified manually for instance through waveform graphics [9]. However, large investments will need to be made in further research and development before these analytics could be implemented in clinical practice; the need for these investments can pose a major barrier for their development and future success. We aim to assist future development and clinical trial plans by identifying the performance requirements of the technology such as maximum costs or minimum efficacy. We performed a cost-effectiveness analysis in which we estimated how analytics that identify IEEVs in real-time could generate health improvements and/or financial savings.

METHODS

We used a decision tree model to assess the potential cost-effectiveness of analytics to detect IEEVs. Short term effects were estimated, such as hospital mortality and length of stay, but also long-term outcomes such as life years gained, and quality adjusted life years gained (QALYs). Where policy makers involved with national reimbursement decisions would be familiar with outcomes such as life years gained and QALYs developers of analytics and hospitals deciding on

their acquisition may be less familiar with these outcomes but interested in mortality and length of stay. The target population consisted of patients who receive assisted modes of ventilation in a Greek ICU. In current care, IEEVs are not detected in these patients, which means that clinicians do not intervene to stop them. We compared current care with the intervention in which IEEVs are detected with analytics that process data from mechanical ventilators in real-time. Their detection would enable clinicians to provide treatment to reduce duration of the IEEV.

Decision tree model

We developed a decision tree model that compared the health and cost outcomes of current care to the use of analytics for early detection of IEEVs (Figure 1). In the intervention arm, data from ventilation monitors is analyzed in real-time and an alarm is generated when a patient has an IEEV (branch 1-5). An alarm sounds when patients are labelled as having IEEVs (branch 1, 2 & 3) while no alarm sounds when patients are labelled as not having IEEVs (branch 4 & 5). When the alarm sounds, a clinician will carry out a treatment that may or may not be successful (branch 1 vs branch 2). The other arm in the decision tree represents current care (branch 6). Since IEEVs are currently not identified, no treatment is performed.

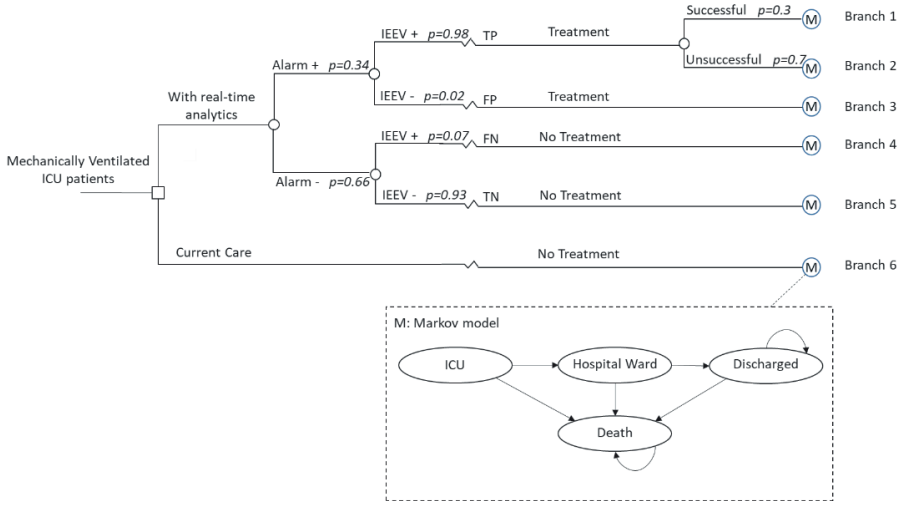


Figure 1: Cost-effectiveness model structure comparing use of real-time analytics to current care. All probabilities were estimated using the sensitivity, specificity, and prior probability of having an IEEV reported in Table 1. Legend: ICU=Intensive Care Unit, IEEV= Ineffective Effort Event, FN= False Negative, TP= True Positive, FP=False Positive, TN= True Negative, M=Markov Model.

Fig. 1 also shows a Markov model with four states ('ICU', 'hospital ward', 'discharged' and 'death'), which was used to estimate the long-term outcomes of IEEV detection and treatment. At the start of this model, all patients start in the ICU. At the end of the first cycle, patients transition to the general 'hospital ward' or 'death'; the cycle length equals the median length of ICU stay. Within the data used to model results, no patients were readmitted to the ICU after ICU discharge. Therefore, we excluded the possibility to transition back to the ICU from the hospital ward.

At the end of the second cycle, all patients in the 'hospital ward' transition to either 'discharged' or 'death'; this cycle's length equals the median length of hospital stay (following ICU discharge). For the remainder of the cycles, patients can remain in the 'discharged' state or die; the length of these cycles was one year. Because it was uncertain as to where in the cycle patients transitioned, a half-cycle correction was applied assuming patients transitioned on average in the middle of the cycle. Without the correction, patients would either be assumed to transition at the start or end of a cycle incurring more or less of the costs they should be assigned. The time horizon was lifetime, and we adopted a healthcare payer perspective including only direct medical costs. Since Greece does not have a national guideline for performing economic evaluations, health outcomes and costs were discounted at a rate of 3.5%. Key model assumptions can be found in Table A.1 (Appendix) and the model was built in R v.3.3.1.

Analytics and treatment parameters

Table 1 shows the values and distributions of the input parameters used in the model. Identifying IEEVs and the subsequent treatment can be complex and to estimate its potential several parameters need to be combined. First, ineffective efforts need to be identified from airway pressure and airflow data. In the Greek ICU a prototype monitor was used to identify ineffective efforts. Data from this 'ineffective effort monitor' can be used to calculate ineffective effort events. The sensitivity and specificity of the algorithm that identified ineffective efforts were derived from the literature [10]. Real-time analytics would use the data from the prototype monitor to identify clusters of ineffective effort events [5]. The prior probability of IEEVs was 38% [5]. When an IEEV is detected, the clinician can perform one of the following treatments: adjust the ventilator settings, reduce sedation when managing pain and anxiety [7,8], or change the management of secretions and bronchodilation [8].

There is evidence that patients experiencing ineffective effort events have worse outcomes such as increased hospital mortality and prolonged ICU stay [5]. However, assessing the probability that treatments are effective when IEEVs occur can only be done once these real-time analytics are available. Therefore, we assessed the impact on health and cost benefits when varying the probability of effective treatment from 0 to 50%. Because the treatment was performed shortly after an IEEV occurred (3 minutes) while the median duration of the events was 21 minutes [5] we assumed that an effective treatment would lead to an outcome similar to those without IEEVs.

Health parameters

Long term health benefits were quantified in life years gained and QALYs gained. QALYs are estimated by multiplying the life years gained by the quality of life in those years. Therefore, if a patient lives two extra years but in suboptimal health, the QALYs gained will be less than two.

We used patient data on current care from a medical-surgical ICU in Greece (the University hospital of Heraklion (PAGNI)) [5] to estimate life years gained and QALYs. The study was approved by the hospital's ethics committee and detailed results from the observational study can be found elsewhere [5]. All 110 patients in that study received assisted modes of mechanical ventilation for >12 hours (total of 4,456,537 breaths).

Life years gained were estimated by combining patient level data with results from the literature. The probability of surviving the ICU was considerably higher - although not statistically significant-

amongst patients without IEEVs compared to patients with IEEVs (75% vs 63% (p=0.249)). The probability of surviving the hospital was statistically significantly higher for patients without IEEVs compared to patients with IEEVs (67% vs 41% (p=0.025). Life years gained after discharge were estimated using the post-discharge hazard ratio of mortality for ICU patients [11] combined with a baseline hazard of the Greek general population [12,13].

Unsurprisingly, no research is available on quality of life of patients during ICU stay. Therefore, using a value set from the United Kingdom, quality of life for those in the ICU whilst on mechanical ventilation was assumed to be 0.297. This corresponds with an EQ-5D state of individuals who have extreme problems with mobility and self-care, cannot perform their usual activities but no pain, discomfort or anxiety. QALYs during a hospital stay were estimated using utility estimates derived from the literature [14]. Quality of life after discharge was estimated using the mean age of the patients and the time since ICU discharge [15].

Resource Use and Unit Costs

To estimate costs, we obtained time on mechanical ventilation and length of stay from the patient level data from PAGNI. For patients with IEEVs, median ICU length of stay was longer than for patients without IEEVs (26 vs 17 days (p=0.017)), as was the median time on mechanical ventilation (16 vs 11 days (p=0.02)). We assumed annual licensing costs for the analytics (€1,918) to estimate the costs of the analytics per ICU day [16]. This estimate was varied extensively in uncertainty analyses. The costs included for treatment when IEEVs occur were assumed to be low since the interventions currently performed to improve interaction between a patient and the mechanical ventilator are easy and cheap to perform (e.g., adjusting sedation, adjustment of ventilator settings). Base case estimates for the costs per ICU day [17] and costs per hospital day [18] were derived from micro-costing studies conducted in Greece. There was a considerable amount of uncertainty in especially the ICU costs per day and these were therefore varied extensively in the univariate uncertainty analyses. These daily ICU costs were decreased by 10% for patients who remained in the ICU but were successfully weaned. All costs were adjusted to 2019 euros.

Cost-effectiveness analysis

We determined the incremental costs, life years gained, quality adjusted life years and the incremental cost-effectiveness ratio of using analytics to identify IEEVs compared to current care. First, base-case estimates for all outcomes were calculated using the most likely input values based on patient-level data and the literature. We then performed univariate sensitivity analyses in which one input parameter at a time was varied to determine how they affected the cost-effectiveness results. Costs of an ICU day are much higher in countries such as the Netherlands compared to the parameter values used in the base case [19]. Therefore, we assessed the impact of increasing this value to the Dutch estimate (€2,153) on the cost-effectiveness results. Finally, we also examined a 'worst case' scenario and 'best case' scenario using the highest and lowest estimates presented in Table 1. In the 'worst case' scenario the analytics and the treatment were expensive, whilst the number of people with IEEVs, the probability of effective treatment, and the sensitivity and specificity of the ineffective effort algorithm were all low. For the 'best case' scenario, the analytics and intervention costs were reduced whilst the probability of having IEEVs, the probability of an effective treatment, sensitivity and specificity were all high.

Table 1: Input parameters for the cost-effectiveness model.

Parameter	Base Case estimate	Lowest estimate	Highest estimate	Distribution	Source
Discount rate costs (%)	3.5	3	5	-	
Discount rate health benefits (%)	3.5	1	5	-	
Sensitivity prototype monitor (%)	88	79	94	Beta	[10]
Specificity prototype monitor (%)	99	80	100	Beta pert	[10]
Prior probability of IEEVs (%)	38	10	50	Beta pert	[5]
Treatment's effectiveness (%)	30	0	50	-	[Expert opinion]
ICU survival (%)					
with IEEVs	63	48	77	Beta	Hospital data
without IEEVs	75	63	84	Beta	Hospital data
Hospital survival (%)					
with IEEVs	41	27	57	Beta	Hospital data
without IEEVs	67	55	77	Beta	Hospital data
Hazard ratio of death after ICU admission vs no. admission	2.01	1.64	2.46	Normal	[11]
Quality of Life (utilities)					
ICU	0.297	0.24	0.36	Beta	Assumed
Hospital	0.60	0.53	0.67	Beta	[14]
Year 1 post discharge	0.67	0.62	0.71	Beta	[15]
Year 2-10 post discharge	0.70	0.65	0.75	Beta	[15]
Year>10 post discharge	0.68	0.62	0.74	Beta	[15]
Resource Use					
ICU LOS (days)					
with IEEVs	28	23	34	Gamma	Hospital data
without IEEVs	22	18	27	Gamma	Hospital data
Time on MV (days)					
with IEEVs	21	17	27	Gamma	Hospital data
without IEEVs	15	12	17	Gamma	Hospital data
Hospital LOS post-ICU discharge (days)	17.3	14	21	Gamma	[20]

Table 1: *Continued.*

Parameter	Base Case estimate	Lowest estimate	Highest estimate	Distribution	Source
Unit costs (in 2019 Euros)					
Analytics licensing (per bed, per year)	1918	100	20,000	-	[16]
Treatment	100	57	155	Gamma	[Expert opinion]
ICU day	686	392	1060	Gamma	[17]
Hospital day	298	170	460	Gamma	[18]
Reduction in ICU costs when patients no longer receive MV (%)	10	0	35	Beta pert	[Expert opinion]

IEEVs = Ineffective Effort Events, ICU = Intensive Care Unit, LOS = Length of stay, MV = Mechanical Ventilation, Hospital data = Patient level data from the intensive care unit of PAGNI in Greece

Probabilistic sensitivity analysis and headroom analysis

In a probabilistic sensitivity analysis (PSA) we varied all parameters simultaneously with the exception of the price and the probability of the treatment's effectiveness. In the PSA we performed 10,000 simulations during which random parameter values for all input parameters were simultaneously drawn from their underlying distributions. We ran the PSA three times using different levels for the probability that the treatment is effective (10%, 30% and 50%). The results were shown using cost-acceptability curves, which display the probability that using the analytics is cost-effective given various willingness-to-pay thresholds. We also estimated the headroom per patient which is the maximum price that could be charged for the analytics per patient or per bed given a fixed willingness-to-pay and can be estimated as follows;

$$Headroom = N + \lambda * Q$$

Where N are the savings given a price of zero for the analytics per bed, λ is the threshold used and Q refers to the incremental QALYs gained [21]. We assumed the device would be sold to a hospital on a per bed basis and that patients needed the device for an average of 17 days. Since no official willingness-to-pay threshold is used in Greece, we adopted three alternative thresholds. The first two were based on opportunity costs proposed by Woods et al resulting in thresholds of €4,946 and €7,758 [22]. Alternatively, we also used a threshold of €30,000 which has been adopted in the past in Greek economic evaluations [23, 24].

RESULTS

Cost-effectiveness analytics

We found that the analytics could reduce hospital mortality (3% absolute reduction), increase QALYs (0.21 per person) and lead to cost-savings (€246 per person) when the probability of the treatment's effectiveness is 30% (Table 2). Even if the probability that the treatment is effective is small (10%) health improvements and cost-savings were gained. Long-term health outcomes (QALYs and life years) were influenced by hospital survival and the discount rate of

health benefits. Incremental costs were greatly influenced by the costs of the analytics, the prevalence of IEEVs, the probability the treatment is effective, and the costs of an ICU day (Figure 2). Increasing sensitivity and specificity of the monitor that identifies ineffective efforts had a limited effect on costs; but when sensitivity increased so did health gains. In the base-case scenario, when the price of the analytics was €1,918, cost-savings were generated (Figure 3). When the costs of the analytics exceeded €7,307 per year, using the analytics was more expensive than current care. Moreover, when costs of an ICU day were high (e.g., €2,153), savings increased from €183 to €1,155 per patient. In the 'best case' scenario, the analytics resulted in greater health benefits (0.50 QALYs), reduced mortality (6% absolute reduction) and higher cost-savings than the base case scenario (€831). However, in the 'worst case' scenario, using the analytics offered no health benefits and increased average costs per patient (€895).

Table 2: Discounted results from the base case analysis and the worst and best case scenarios.

Scenario	Costs €	Length of ICU[f] stay	Hospital Mortality	Life Years	QALYs [e]
Base Case					
Current Care	19,501	24.28	0.43	6.87	4.72
With Analytics	19,255,	23.68	0.40	7.18	4.93
Incremental	-264	-0.6	-0.03	0.31	0.21
ICER [a]		Dominant	Dominant	Dominant	Dominant
Worst Case [b]					
Current Care	18,474	22.6	0.36	7.73	5.31
With Analytics	19,369	22.6	0.36	7.73	5.31
Incremental	895	0	0	0	0
ICER		-	-	-	-
Best Case [c]					
Current Care	19,942	25.00	0.46	6.50	4.46
With Analytics	19,111	23.59	0.40	7.22	4.96
Incremental	-831	-1.41	-0.06	0.72	0.50
ICER		Dominant	Dominant	Dominant	Dominant
High ICU day costs [d]					
Current Care	53,520	24.28	0.43	6.87	4.72
With Analytics	52,366	23.68	0.40	7.18	4.93
Incremental	-1,155	-0.6	-0.03	0.31	0.21
ICER		Dominant	Dominant	Dominant	Dominant

[a] ICER = Incremental Cost Effectiveness Ratio, [b] High costs of the analytics (€20,000) and treatment intervention (€155), Low probability of IEEVs (0.1), sensitivity (0.79), specificity (0.8) and an unsuccessful treatment intervention (0), [c] Low costs of the analytics (€100) and the intervention (€57), High probability of IEEVs (0.5), sensitivity (94%), specificity (1) and probability of successful intervention (0.5), d High costs of an ICU day, [e] QALYs = Quality Adjusted Life Years, [f] ICU = Intensive Care Unit

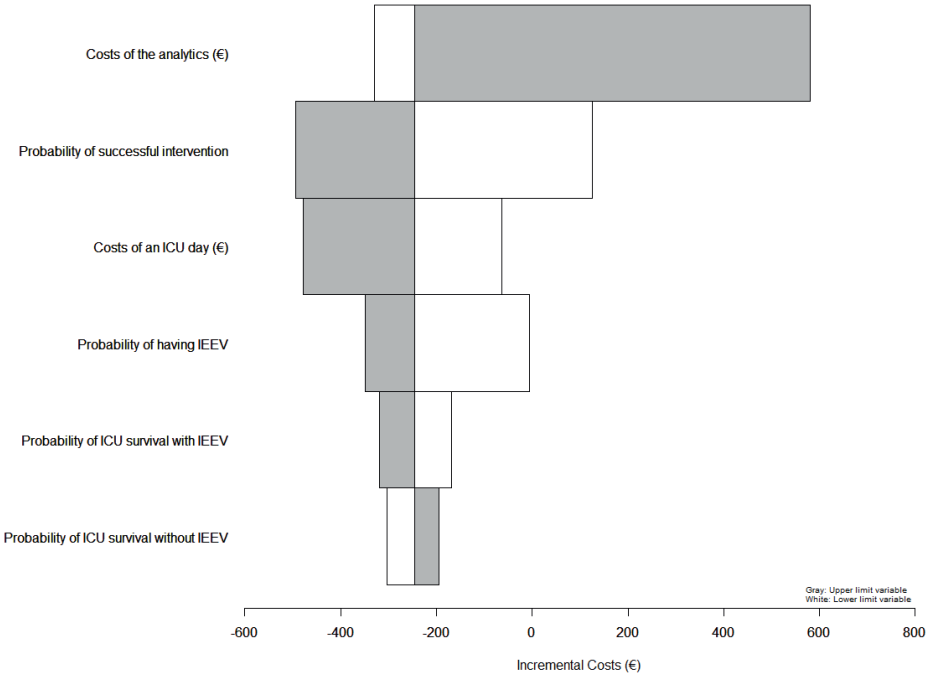


Figure 2: Tornado diagram illustrating the influence of individual parameters on the incremental costs.

Legend: ICU=Intensive Care Unit, IEEV= Ineffective Effort Event

Probabilistic sensitivity analysis and headroom analysis

Fig. 4 shows a cost-effectiveness plane that illustrates the degree of uncertainty surrounding the differences in costs and effectiveness between using real-time analytics and current care. Three scatterplots are shown, one for each of the scenarios. This figure shows us that a greater probability that the treatment is effective increases the degree of cost-savings and health gain from using real-time analytics. The cost-effectiveness acceptability curves shown in Fig. 5 present the probability that the analytics are considered cost-effective for a range of willingness-to-pay thresholds. We presented three different acceptability curves each with their own probability of the treatment's effectiveness. Fig. 5 illustrates that for a low willingness-to-pay threshold (€4,946), the probability that the analytics for IEEVs are cost-effective exceeds 90% even when the probability that the treatment is effective is 10%. The headroom was €1,963 per patient (equivalent to €41,468 per bed), for a willingness-to-pay threshold of €7,758. Moreover, for a threshold of €30,000 the headroom per patient was much higher (€6,634 per patient equivalent to €140,128 per bed).

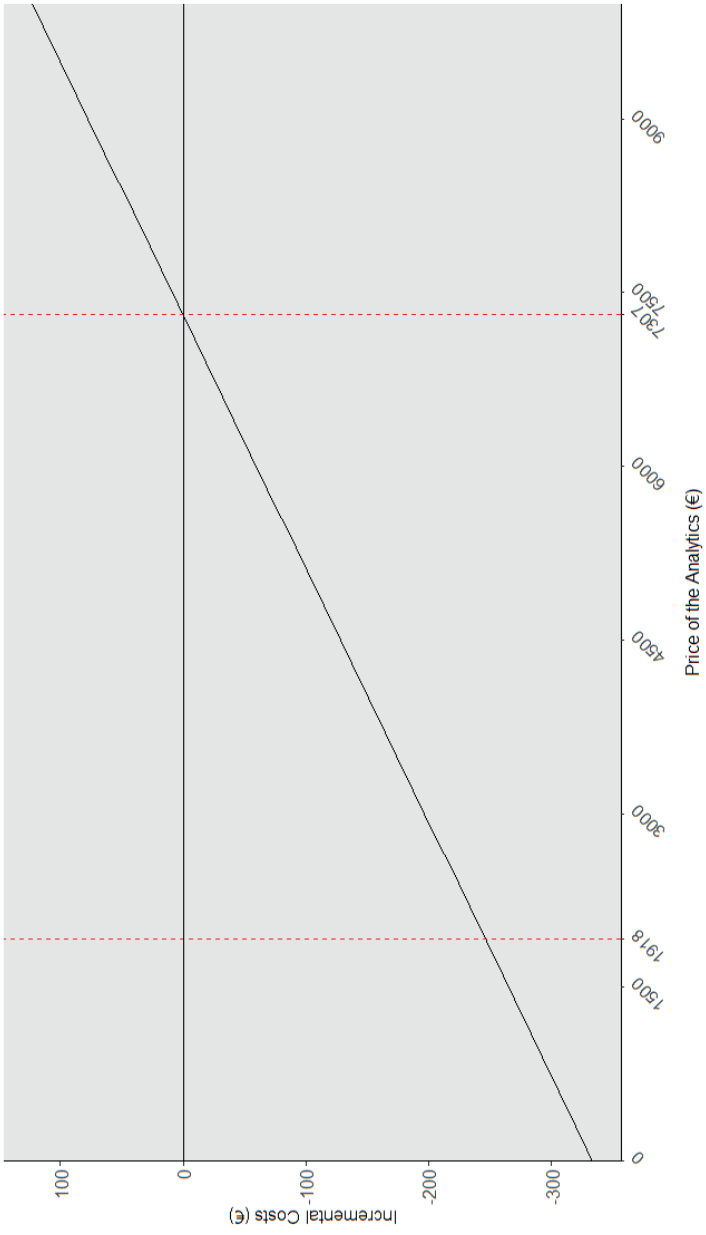


Figure 3: Impact of the price of the analytics on incremental costs.

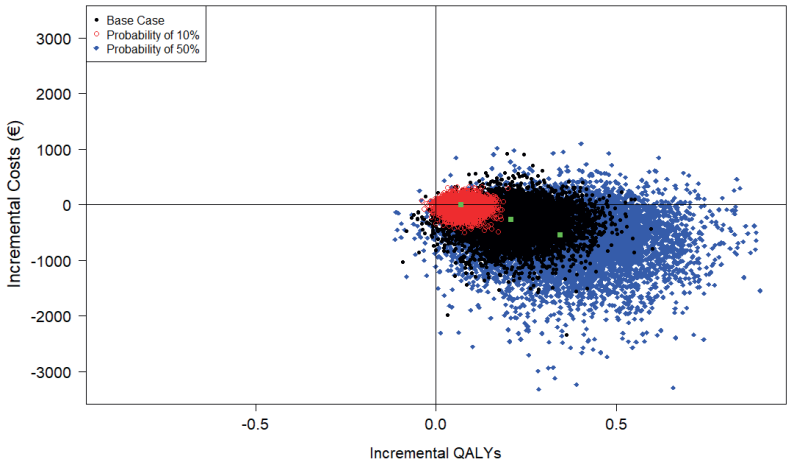


Figure 4: Cost-effectiveness plane for real-time analytics of an ineffective effort event. Results are presented for three probabilities of a successful treatment; 10%, 30% (base case) and 50%.

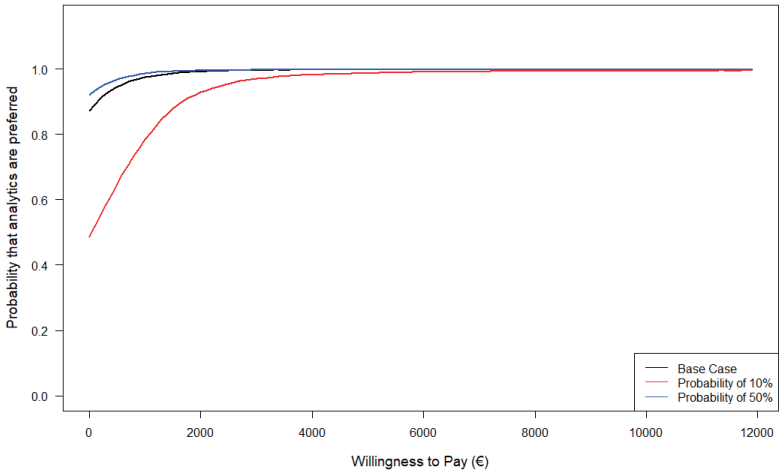


Figure 5: Cost-effectiveness acceptability curves for real-time analytics of an ineffective effort event. Results are presented for three probabilities of a successful treatment; 10%, 30% (base case) and 50%.

DISCUSSION

We estimated the potential cost-effectiveness of real-time analytics that identify ineffective effort events in mechanically ventilated ICU patients. Even when the probability that the treatment is effective is low, use of real-time analytics could still lead to health benefits for patients (0.21 QALYs per person) and savings (€264 per person) for healthcare payers. Moreover, there is considerable headroom for development since the maximum price that can be charged per bed varies from €28,994 to €140,128 depending on the willingness-to-pay threshold used.

This is the first study to examine the cost-effectiveness of analytics that detect IEEVs. These estimates are important to stimulate further development of analytics that detect IEEVs in real-time since patients with IEEVs have much poorer outcomes compared to those without IEEVs. Previous studies have emphasized that patients with IEEVs have a longer time on mechanical ventilation compared to those without IEEVs and authors have reported that health and economic benefits can be gained by reducing time on mechanical ventilation [25, 26]. Moreover, Marchuk et al. found that those patients with many ineffective efforts in a brief timeframe had reduced oxygen saturation [4]. This further confirms that using analytics that enable timely identification of IEEVs are essential since this allows clinicians to intervene rapidly to improve their oxygen saturation. The underlying assumption that an intervention is successful in at least a small subset of these patients is an important one in the analysis and we cannot be sure that this assumption is valid without further research. However, the results available thus far suggest that it is more likely that an intervention improves outcomes compared to the possibility that the intervention has no or a negative effect. First, we see that patients with IEEVs are severely worse off compared to patients without IEEVs suggesting that there is a lot of room for improvement [4,5]. Second, IEEVs can be identified after 3 minutes while their median duration at present is 21 minutes leaving a large time window in which a clinician can intervene to stop their continuation [5]. This is very important because the potential interventions are relatively easy to perform, are straightforward and are unlikely to lead to any adverse effects. In the unlikely case that there would be absolutely no effect of an intervention whatsoever, we expect purchasers would lose money, but patients would not necessarily be worse off. Since the probability of successful treatment influences the health benefits and savings from using real-time analytics, we recommend further development of these analytics for clinical practice and performing a prospective clinical trial to assess their true impact. This study should provide more information about the percentage of patients with IEEVs, and the effectiveness of treating them.

Transferability of our findings to other countries and hospitals could be influenced by the cost estimates used in our analyses. Especially ICU costs had a large influence on the results, and we therefore varied these costs by 25% in the univariate sensitivity analysis. Moreover, we also performed a scenario analysis using the ICU costs of the Netherlands as an example for other western countries. The benefits for hospitals also depend on the reimbursement system in place. Diagnostic related groups in which hospitals receive a fixed payment for patients with a specific diagnosis can stimulate hospitals to reduce length of stay which could in turn lead to financial savings for hospitals. However, if services are reimbursed on a fee-for-service basis in which the hospital is reimbursed for each additional day in the hospital, there could be perverse incentives to increase length of stay. Either way, the aim of healthcare providers should be to maximize the health outcomes of their patients which makes use of analytics to detect IEEVs desirable. We excluded the possibility that alarms generated by the analytics might sometimes be ignored because of alert fatigue which could lead to lower benefits than estimated here. We also excluded the possibility that patients are readmitted to the ICU and excluded any side effects of treatments to stop an ineffective effort event. Even though no patients were readmitted in the observational study and experts thought that side effects did not necessarily occur, both should be verified in a clinical trial.

Our results are not generalizable to all ICU patients receiving mechanical ventilation, since we only considered patients who were expected to remain on proportional assisted mechanical

ventilation for a longer period of time (>24h). Furthermore, a small subset of patients can have IEEVs a couple of days after initiation of ventilation support. Our assumption that all treatments are performed on the first day could therefore have led to an overestimation of the benefits of using the analytics. Even though few patients had IEEVs after the first day, additional research on the estimated number and timing of IEEVs could improve the estimate of the benefits. A final limitation is that we did not include any benefits of reducing any delays in ICU admission of other patients. Since there is a shortage of ICU beds in Greece, reducing length of stay for patients with IEEVs could reduce health losses incurred by other patients because of delays in admitting them to the ICU. Therefore, the true benefits could be higher than presented here.

Although clinical experts have emphasized the relevance of developing analytics to detect IEEVs [5, 27] their adoption is uncertain and compromised by constrained budgets and competing investments. Our results provide developers with estimates of the potential benefits of these analytics, which they can show to healthcare payers. There is a considerable market that could benefit from analytics that identify IEEVs since the number of critical care beds in Europe has been previously estimated at 75,585 [28]. Sixty percent of all ICU patients receive mechanical ventilation, of which 30% will receive prolonged ventilation [29,30]. Therefore, the analytics would be relevant for 18% of ICU patients. In Greece, there is a shortage of ICU beds and because of this all ICU beds are constantly occupied. If this is also the case in other European countries, the analytics would be relevant for 18% of these 75,585 beds in Europe alone. Based on our results, the analytics should first be assessed in countries where ICU costs are high, such as the United States or The Netherlands, where the potential financial benefits of the analytics would be considerably higher.

CONCLUSION

Real-time analytics to identify ineffective effort events have the potential to improve patient outcomes and generate financial savings for healthcare payers even when the probability of an effective treatment is low. There is considerable headroom for development, and this should therefore be encouraged. Exploitation in countries where the costs of an ICU day are high could yield a higher return on investment. One important next step is to obtain additional clinical evidence of using these analytics in settings where there is a high frequency of IEEVs.

REFERENCES

1 Halpern NA, Pastores SM. Critical care medicine in the United States 2000-2005: an analysis of bed numbers, occupancy rates, payer mix, and costs. Crit Care Med 2010 Jan;38(1):65-71.

2 Dasta JF, McLaughlin TP, Mody SH, Piech CT. Daily cost of an intensive care unit day: The contribution of mechanical ventilation. Crit Care Med 2005; 33:1266-71.

3 Sanchez-Pinto LN, Luo Y, Churpek MM. Big Data and Data Science in Critical Care. Chest 2018; 154: 1239-1248.

4 Marchuk Y, Magrans R, Sales B, Montanya J, et al Predicting Patient-ventilator Asynchronies with Hidden Markov Models. Sci Rep. 2018;8.

5 Vaporidi K, Babalis D, Chytas A, et al. Clusters of ineffective efforts during mechanical ventilation: Impact on outcome. Intensive Care Med 2017; 43:184-91.

6 Chouvarda I, Babalis D, Papaioannou V, Maglaveras N, Georgopoulos D. Multiparametric modeling of the ineffective efforts in assisted ventilation within an ICU. Med Biol Eng Comput 2016;54(2-3):441-451.

7 Pierson DJ. Patient-ventilator interaction. Respir Care 2011; 56:214-28.

8 Alex CG, Fahey PJ, Tobin MJ. What a Clinician Should Do When a Patient "Fights the Ventilator". In Mechanical Ventilation and Weaning 2003 (pp. 100-117). Springer, Berlin, Heidelberg.

9 Mellott KG, Grap MJ, Munro CL, Sessler CN, Wetzel PA. Patient-ventilator dyssynchrony: clinical significance and implications for practice. Critical care nurse. 2009 Dec;29(6):41-55.

10 Younes M, Brochard L, Grasso S, et al. A method for monitoring and improving patient: Ventilator interaction. Intensive Care Med 2007; 33:1337-46.

11 Williams TA, Dobb GJ, Finn JC, et al. Determinants of long-term survival after intensive care. Crit Care Med 2008; 36:1523-30.

12 World Health Organization: WHO Mortality Database: Query the Online Database. [Internet] http://apps.who.int/healthinfo/statistics/mortality/causeofdeath_query/start.php. Accessed June 12, 2017.

13 Index Mundi: Greece Age structure. [Internet] http://www.indexmundi.com/greece/age_structure.html. Accessed June 12, 2017

14 Cox CE, Carson SS, Govert JA, Chelluri L, Sanders GD. An economic evaluation of prolonged mechanical ventilation. Crit Care Med 2007; 35:1918-27.

15 Cuthbertson BH, Elders A, Hall S, et al. Mortality and quality of life in the five years after severe sepsis. Crit Care 2013; 17:1.

16 Datamed: EKG Format Translators. Product Price List: Effective 4/1/2017. [Internet] http://www.datamed.com/docs/datamed_product_price_list.pdf. Accessed June 12, 2017

17 Leftakis A, Geitona M. Cost analysis and estimation of thoracic surgical patients with lung cancer in Greece: The case of Sotiria ICU. Intensive Crit Care Nurs 2001; 17:322-30.

18 Geitona M, Hatzikou M, Steiropoulos P, Alexopoulos EC, Bouros D. The cost of COPD exacerbations: A university hospital–based study in Greece. Respir Med 2011; 105:402-9.

19 Hakkaart-van Roijen L, van der Linden N, Bouwmans C, Kanters T, Tan SS. Bijlage 1. Kostenhandleiding: methodologie van kostenonderzoek en referentieprijzen voor economische evaluaties in de gezondheidszorg. Institute for Medical Technology Assessment, Erasmus Universiteit Rotterdam, Rotterdam; 2015: 1-120 (in Dutch).

20 GiViT, Gruppo Italiano per la Valutazione degli Interventi In Terapia Intensiva. Report PROSAFE project; 2014. Report nr Centre GR001.

21 Girling A, Lilford R, Cole A, Young T. Headroom approach to device development: Current and future directions. Int J Technol Assess Health Care. 2015;31(5):331-8.

22 Woods B, Revill P, Sculpher M, Claxton K. Country-level cost-effectiveness thresholds: initial estimates and the need for further research. Value Health 2016;19(8):929-935.

23 Makras P, Athanasakis K, Boubouchairopoulou N, et al. Cost-effective osteoporosis treatment thresholds in Greece. Osteoporosis Int 2015;26(7):1949-1957.

24 Geitona M, Kousoulakou H, Kalogeropoulou M, Mitsiki E, Panitti E, Steiropoulos P. Cost-effectiveness analysis of the fixed combination indacaterol/glycopyrronium vs. tiotropium and salmeterol/fluticasone in the management of COPD in Greece. Value Health 2015;18(7):A500.

25 Cox CE, Carson SS. Medical and Economic Implications of Prolonged Mechanical Ventilation and Expedited Post–Acute Care. Semin Respir Crit Care Med. 2012; 33:357-361.
26 Cox CE, Carson SS, Lindquist JH, Olsen MK, Govert JA, Chelluri L. Differences in one-year health outcomes and resource utilization by definition of prolonged mechanical ventilation: A prospective cohort study. Crit Care 2007; Feb;11(1):1-1.
27 Jaber S, Bellani G, Blanch L, et al. The intensive care medicine research agenda for airways, invasive and noninvasive mechanical ventilation. Intensive Care Med 2017; 1-14.
28 Rhodes A, Ferdinande P, Flaatten H, Guidet B, Metnitz P, Moreno R. The variability of critical care bed numbers in Europe. Intensive Care Med 2012;38(10):1647-1653.
29 Tan SS, Bakker J, Hoogendoorn ME, et al. Direct cost analysis of intensive care unit stay in four European countries: applying a standardized costing methodology. Value Health 2012;15(1):81-86.
30 Heunks LM, Van Der Hoeven, Johannes G. Clinical review: The ABC of weaning failure-a structured approach. Crit Care 2010;14(6):245.

APPENDIX

Table A.1: Overview of key assumptions underlying the model

Assumption	Potential Impact
We assumed patients do not transition back to the ICU during their initial hospital visit.	If the number of patients transitioning back to the ICU during the initial visit is high, we may have underestimated the costs in care with the analytics. With the analytics more patients survive their initial ICU visit and thus more of these patients are eligible for a readmission and thus higher costs. However, this would only be the case if the intervention performed when IEEVs occur has no impact the risk of readmission
We assumed a base case effectiveness estimate of 0.3	This estimate was varied very extensively in the uncertainty analysis. Therefore, the potential impact of this assumption is clearly demonstrated throughout the paper. Of course, it is crucial to perform a clinical trial once real-time analytics have been developed.
We assumed patients in the ICU have a utility estimate of 0.297 which was based on no pain, discomfort, or anxiety	It is possible that when sedation is adjusted when an IEEV occurs, a clinical expert has difficulty finding the balance between a patient's ability to trigger the ventilator (so no excessive sedation) while avoiding pain/anxiety. However, if pain or anxiety occurs (thus the balance has not been found) it is unlikely that patients will experience this for very long because the clinician will continue adjusting medication until this balance is found.
We assumed costs of the analytics to be €1918	This estimate was varied very extensively in the uncertainty analysis. Therefore, the potential impact of this assumption is clearly demonstrated throughout the paper.
We assumed costs of the intervention to be €100	At present all interventions that are applied to improve the interaction between a patient and a mechanical ventilator are quite simple and relatively low in costs. We therefore consider this assumption reasonable. Would these costs be higher in reality we may have slightly overestimated the savings.
We assumed patients used the analytics for 17 days on average when estimating the headroom per bed and that all beds were constantly occupied (thus a shortage of ICU beds).	This estimate was based on the results from the observational study performed in PAGNI and the shortage of ICU beds in Greece. It is possible that in other European countries the headroom would be slightly lower because there is no shortage of ICU beds. Furthermore, we expect that ICU stay could be shorter in other countries for instance because of differences in quality of care and patient characteristics.


1101 00011
10101010101001
0101010 010100
0101010 010101
0000011 010110
1010101 101101
1110000 010101
010100 101101
100 101
1001010100000101010
101011010101001
1011
010
10

Analyzing electronic health records: the benefits of target trial emulation.

Bakker L, Goossens L, O'Kane M, Uyl-de Groot C, Redekop W.

Health Policy and Technology. 2021 Sep 1;10(3):100545.

ABSTRACT

Aim: Electronic health records (EHRs) are increasingly used in effectiveness and safety research. However, these studies are often at risk of bias. This study demonstrates the relevance, and discusses challenges, of using target trial emulation to avoid bias, such as selection bias, immortal time bias and confounding when performing observational research with EHRs.

Methods: Target trial emulation can be used to identify and address some of the drawbacks of observational research in a systematic way. Potential sources of bias are identified by describing key components of an ideal randomized controlled trial and comparing this to the observational study actually performed. The methods were applied to assess treatment response to antidiabetic treatment using EHRs from patients with diabetes treated in secondary care.

Results: Using target trial emulation ensured prevalent users were excluded and patients were not included based on information generally not available when initiating a clinical trial. Furthermore, applying these methods demonstrated how the number of records eligible for use can rapidly decrease. Hereafter, adjustments were performed to address potential sources of bias and it was shown that missing variables essential for adjustment can be an important issue.

Conclusions: Using target trial emulation, sources of selection bias and confounding were identified and adjusted for accordingly when analyzing treatment response in patients with type 2 diabetes. However, when using EHR data to emulate a target trial, samples containing sufficient information on outcome measures and variables to adjust for confounding and selection bias are essential given the risk of missing data.

INTRODUCTION

The randomized controlled trial (RCT) is the preferred method for assessing the efficacy of novel treatments. Randomly assigning patients to either the treatment group or the comparator enables researchers to isolate the effect the treatment has on the outcome. However, RCTs are often costly to perform, have a limited generalizability and may pose ethical challenges [1,2]. Furthermore, for diseases such as diabetes, where many different treatment combinations are possible, performing an RCT for each possible antidiabetic treatment combination is often not feasible. The use of observational research and electronic health records (EHR) to assess the real-world effectiveness of these treatments has sometimes been suggested as an alternative [1]. However, observational studies can be challenging to perform due to missing data and the risk of bias [1,3-7]. Even though best practice methods have been emphasized for observational research in a chronic disease such as diabetes [8], recent systematic reviews have found that many observational studies are still at risk of bias such as selection bias, immortal time bias and confounding [1,5,7].

Target trial emulation can be used to identify and address some of the drawbacks of observational research in a systematic way [2]. In target trial emulation, potential sources of bias are identified by describing an ideal trial and comparing this to the observational study that is designed to emulate this target trial [9]. If shortcomings of the observational study that have a large impact on the quality of the study cannot be overcome, researchers can adjust their design or find additional data [9]. Elements of target trial emulation (e.g., eligibility criteria) are sometimes already presented and used in observational research [2,8,9]. However, these elements are often not used in a systematic way and the necessity of using them, and the challenges that can be expected, may not always be apparent to researchers [2]. In this paper, we aim to offer practical guidance for those performing real-world effectiveness and safety research using EHRs. We demonstrate the value of target trial emulation but also discuss several challenges that can be expected by emulating a target trial using EHRs of patients with diabetes.

METHODS

Identifying bias

Throughout the remainder of the paper, we will discuss the components of target trial emulation based on Hernán et al. [2] and demonstrate how these can be used to systematically reduce the risk of bias. There are many ways in which results from observational research are at risk of bias. While the definitions of these types of bias may differ between research disciplines, they can be clearly illustrated using directed acyclic graphs (DAGs). A DAG presents all assumed potential causal relationships between variables using directional arrows [10]. Thus, when an arrow is drawn from the exposure variable (e.g., treatment) to the outcome variable (e.g., Hemoglobin A1C (HbA1c) response) the arrow cannot also go the other way (Figure 1a). Moreover, DAGs are acyclic, thus no variable can cause itself. DAGs may contain both measured and unmeasured causal variables as well as common causes and effects of exposure and outcome [11]. If the rules for drawing DAGs are followed, the graphs can be used to determine which statistical adjustments are required. Statistically, an exposure and outcome are associated when the DAG contains an open connection ('path') between them [10]. These paths can be closed or opened by applying statistical adjustments, conditioning, or altering the research design.

Target trial emulation

In target trial emulation, the seven components of the target trial protocol are compared to the observational study to be performed. The first component of the protocol is to define eligibility criteria using only data that would be available prior to treatment initiation [2]. In the target trial, information on follow-up would not be available prior to starting treatment and using this information for patient selection could induce selection bias. Selection bias can occur when conditioning on a common effect of outcome and exposure (Figure 1c & 1d) [11]. This opens a path between exposure and outcome, which had been closed due to the presence of the common effect ('collider'). Studies are at risk of selection bias when future information is used to define inclusion and exclusion criteria. In research that uses EHRs, a frequently seen example of this is the selection of patients based on the availability of outcome data (Figure 1c) [1,6,12]. It is possible that there is no difference in HbA1C reduction between treatment with a sodium glucose transporter-2 inhibitors (SGLTs) vs a dipeptidyl-peptidase 4 inhibitors (DPP4s) but patients receiving SGLTs are at risk of more severe side effects of treatment, which may reduce attendance at follow-up visits [13]. Meanwhile, depression may independently result in loss to follow-up and higher HbA1C values [14,15]. Missing patient data in a patient prescribed an SGLT may therefore relate both to side effects of treatment and to depression. Thus, when including only those with data at follow-up we would open the backdoor path from treatment to HbA1C at follow-up through depression. We would find that patients included in follow-up receiving DPP4s have lower HbA1C values while actually there is no difference in HbA1C.

The second component of target trial emulation describes the treatment strategies, preferably including only new users [2]. By including all events that occur early after drug initiation, a new-user design reduces the risk of selection bias [16]. This form of selection bias, also referred to as prevalent user or survival bias, occurs when patients are included in the study that were already prescribed the drug prior to the start of follow-up (Figure 1d). Thus, if we were to include patients already prescribed treatment by the general practitioner prior to hospital referral then some patients that already failed treatment early after initiating therapy would be excluded. If DPP4s are more often prescribed by the general practitioner than SGLTs, this could lead to an underestimation of the benefits of SGLTs over DPP4s.

The third component contains the assignment procedures to reduce the risk of confounding. Where random assignment is often the preferred strategy in the target trial, adjustment for potential (post-) baseline confounders is required in observational research [2]. Confounding is present when the exposure and outcome share a common cause (Figure 1b) [10]. In Figure 1b, it is assumed that treatment and HbA1c response are both (partially) influenced by a patient's weight, measured before treatment. This opens a path between treatment and outcome that does not represent the causal effect of treatment. Failing to adjust for weight in such an analysis could result in incorrect conclusions that treatment improves HbA1c while in truth weight causes both the exposure and the outcome.

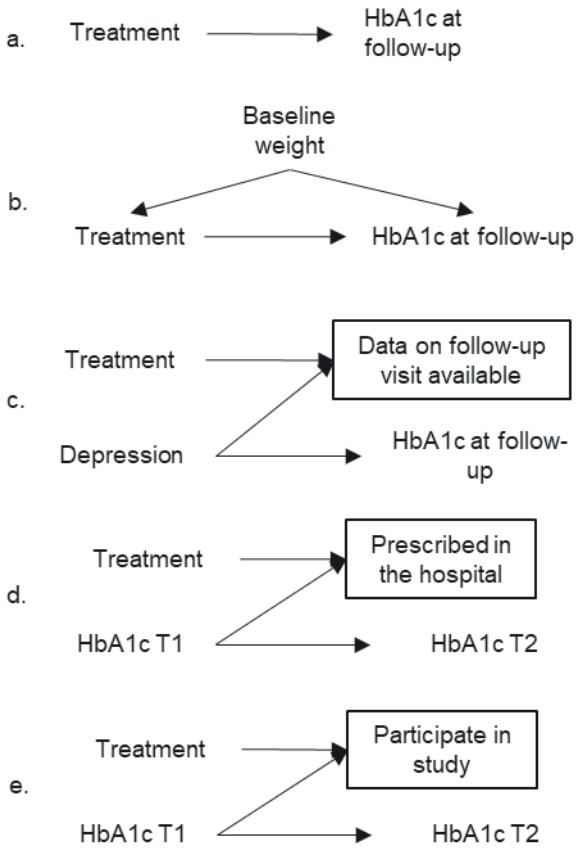


Figure 1: Directed acyclic graphs that represent the causal relationships between treatment and treatment response (HbA1c at follow-up). A) Assumes that the effect of treatment on HbA1c response is only decided by treatment. B) Confounding: Assumes that the selected treatment and HbA1c response share a common cause, namely baseline weight. C) Selection bias: Patients are included when follow-up data is present. If in truth no relation between treatment and HbA1c exists, a spurious relation would be induced by selecting on a collider thus opening the backdoor path through depression. D) Prevalent user bias: Here we assume no difference exists in HbA1c at follow-up between dipeptidyl-peptidase 4 inhibitors (DPP4s) and sodium glucose transporter-2 inhibitors (SGLTs). However, if DPP4s are more often prescribed prior to hospital referral and only those users with a good response are still on DPP4s when referred to the hospital, this would open the backdoor path through the initial HbA1c measurement 'HbA1c T1' by selecting on the collider; 'Prescribed in hospital'. E) Immortal time bias: In truth no relation exists between treatment and time to progression defined as elevated HbA1c measurements. If inclusion criteria require patients on an SGLT receive treatment for at least 1 month but follow-up starts at treatment assignment, results would be biased upwards for SGLTs. HbA1c T1= HbA1c at time 1, HbA1c T2= HbA1c at time 2.

The next component contains the follow-up period defined by the start and the end of follow-up. When initiation of follow-up is not aligned with eligibility criteria and treatment assignment, immortal time bias can influence results [17]. Immortal time is the time in which the outcome cannot occur [6,18]. Suppose we wish to compare time to progression according to elevated HbA1c in patients prescribed an SGLT compared to DPP4s. If inclusion criteria require patients

on SGLTs receive treatment for at least 1 month because of the intensification schedule, but follow-up starts when the drug is first prescribed, then results are at risk of immortal time bias. The time to progression will be higher for patients in the SGLT arm since they receive additional time in which they cannot progress at the beginning of follow-up (Figure 1e). Excluding prevalent users and avoiding use of future information can ensure alignment of inclusion criteria, treatment assignment and follow-up, aiming to avoid immortal time bias [6].

This is followed by a description of the relevant outcome for which blinded measurements are often preferred [2], but often not possible in the observational study. The causal contrast of interest is then selected in the following component, which often includes the intention-to-treat and/or the per-protocol effect [2]. An intention-to-treat analysis includes all patients in the analysis within the arm to which they were originally randomized, irrespective of whether they completed treatment. A per-protocol effect includes only those patients that completed treatment. Both contrasts may require specific statistical adjustments.

In the final component, the analysis plan reports the analyses that need to be performed to properly estimate the causal contrasts of interest. Statistical techniques such as inverse probability weighting, imputation, stratification, g-methods, and regression can be used to adjust for (postbaseline) selection bias and confounding [2,10,12].

We applied target trial emulation to an example from diabetes care by assessing the effectiveness of SGLTs compared to DPP4s added to insulin therapy in patients with type 2 diabetes.

Patient level data

To demonstrate several benefits and challenges of target trial emulation, hospital EHRs were used from patients with diabetes treated between June 2012 and December 2017 at the Western Health and Social Care Trust in Northern Ireland. The analyses were performed as part of the AEGLE project and ethical approval was granted by the Chelsea Research Ethics Committee (REC reference 16/LO/2018). The EHR was a specific diabetes management system: Diamond (Hicom, Surrey, UK). EHR patient demographic information was populated from a patient administration system. Laboratory test results were transferred automatically from the local Laboratory Information Management System. Clinical data were entered live at the time of the patient consultation by the clinical medical or nursing staff; this included the recording of updated anthropometric data, medication changes or the development of new clinical problems. In the records, any active ingredient and/or trade name mentioned were given an individual identifier. These were automatically classified into subgroups (e.g., antidiabetic drugs, antibiotics etc.). This list was checked manually to identify any drugs wrongly classified and active antidiabetic drugs were grouped according to drug class. The initiation date of a drug combination was the first date on which the combination was recorded. The end date was the first date of the record on which the combination was no longer recorded after initiation.

Target Trial Emulation Applied to EHRs in Diabetes Care

In Table 1, we present the target trial protocol alongside the design of the observational study.

Table 1: A description of the protocol components of target trial emulation, the protocol of the target trial and the observational study. All steps are based on the description as provided by Hernán et al. [2].

Protocol	Target trial	Observational study using EHR data
Eligibility criteria	- Patients with type II diabetes treated in secondary care - Patients receiving insulin ± an oral antidiabetics (metformin/sulfonylurea)] for 84 days or more. - Patients do not reach their (personalized) HbA1C target on current therapy	- Patients with type II diabetes treated in secondary care - Patients received insulin ± an oral antidiabetics (metformin/sulfonylurea) for at least 84 days.
Treatment strategies	An SGLT compared to a DPP4 inhibitor with insulin ± an oral antidiabetics (metformin/sulfonylurea).	Similar to target trial
Assignment procedures	Patients are randomly assigned to an SGLT or a DPP4 in combination with insulin ± an oral antidiabetics (metformin/sulfonylurea).	Decided by physicians and their patients on a case-by-case basis. Adjusted for several confounding variables or their proxies measured at baseline e.g. weight, HbA1c, duration of diabetes, age, insulin regimen, number of non-diabetes drugs.
Follow-up period	Starts at randomization. Ends at 6 months of follow-up, when lost to follow-up or death	Started on the prescription date for the SGLT or DPP4. Ended at first HbA1c measurement at least 3 months after drug initiation, when lost to follow-up or death.
Outcome	Blinded measurement of HbA1c 6 months after randomization.	Blinded first measurement of HbA1c > 3 months after treatment initiation. Measurement was blinded as this was not performed by the prescribing clinician.
Causal contrast of interest	Both intention-to-treat effect and per protocol effect.	Analog of the intention-to-treat effect
Statistical Analysis plan	Intention-to-treat analysis. Per-protocol-analysis with correction for baseline variables and correction for variables associated with loss-to follow-up.	Intention-to-treat analysis (as-started). Generalized linear model was fitted after inverse probability weighting and after multiple imputation to correct for post-'randomization' selection bias due to loss of follow-up.

EHR= Electronic Health Record, SGLT= sodium glucose transporter-2 inhibitors, DPP4= dipeptidyl-peptidase 4 inhibitors

Eligibility Criteria

The target trial is an RCT including patients with type 2 diabetes managed in secondary care who do not achieve their personalized HbA1C target after at least 12 weeks of treatment with insulin with or without oral antidiabetics (metformin/sulfonylurea). Thus, these are generally not newly diagnosed patients with diabetes but more complex, progressed patients seen by specialists in secondary care. The eligibility criteria in the observational study were identical except that we explicitly assumed that all those receiving intensification do not achieve their personalized targets and that the personalized target itself does not influence the treatment prescribed. We refrained from using the availability of Hba1c follow-up measurements as a criterion for inclusion since this would have also been unknown upon inclusion in the target trial.

These criteria were applied to the EHRs available, and Figure 2 illustrates how the number of patients eligible for inclusion can greatly reduce with each criteria. The initial dataset consisted of 183,570 prescription records of 7,927 patients. This number reduced with almost 90% after applying selection criteria. Only 57% of patients remained when only patients with type 2 diabetes were included, 21.3% received treatment with an insulin ± an oral antidiabetic and 12.8% received this treatment for at least 84 days. The treatment that followed varied greatly since a wide variety of diabetes treatment combinations are available and treatment decisions depend on the patient and hospital. Of course, the number of patients for whom relevant baseline and follow-up data was missing is not yet considered here but this is known issue when using EHR data.

Treatment Strategies

In both the target trial and the observational study, eligible patients are randomly assigned to an SGLT or a DPP4 added to insulin with or without an oral antidiabetic (metformin/sulfonylurea). In this example, we did not limit the strategy to a specific dose. However, if the aim is to evaluate specific dosing schedules it should be considered that missing information on dose prescribed can be an important issue [19].

In Table 2, we present the sample of patients that fulfilled eligibility criteria and received the relevant treatment strategy. By including only patients that were on a treatment with insulin ± an oral antidiabetic for at least 84 days we did not include any prevalent users. Patients that received an SGLT were generally younger, had a shorter duration of diabetes and were heavier. HbA1c within 6 months prior to treatment initiation was similar for both groups however, the number of missing values was high. The time between measuring weight and HbA1c prior to treatment and initiating treatment was similar between the two groups. Furthermore, patients that received an SGLT were more often treated with a combination of basal and bolus insulin compared to patients that received a DPP4.

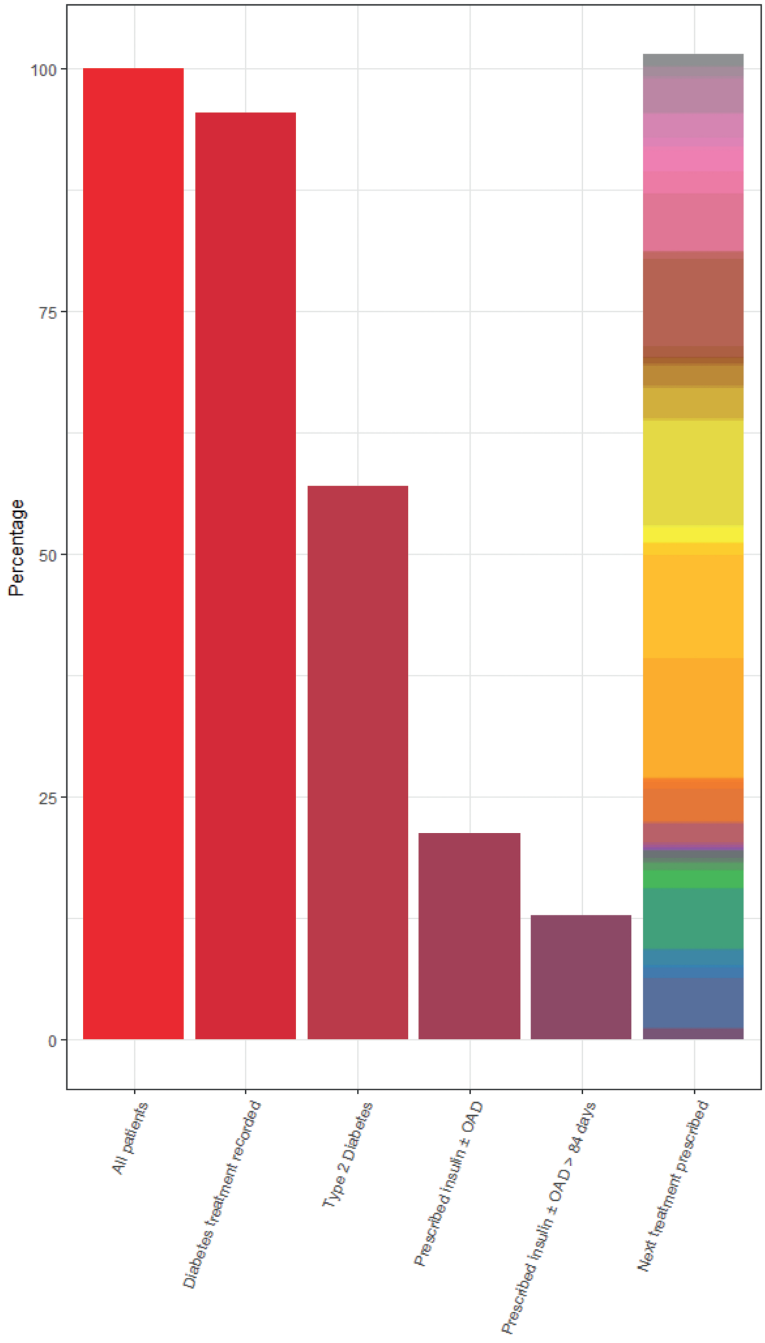


Figure 2: The number of prescriptions that remain when applying relevant selection criteria to the initial 7,927 patients. The column on the right hand side illustrates the variation in the next perscribed treatment for the remaining 10% that received insulin ± an oral antidiabetic after applying selection criteria. OAD = Oral antidiabetic drug

Table 2: Patients included in the analysis.

	Received DPP4 (N = 160)	Received SGLT (N = 123)
Sex, n(%)		
Female	77 (48)	51 (41)
Male	83 (52)	72 (59)
Age, n, mean(sd)	134; 70.29 ± 10.10	110; 57.30 ± 10.62
Duration of diabetes in years, n, mean(sd)	134; 17.54 (7.89)	110; 13.22 (6.19)
Weight in kg, n, mean(sd)	116; 87.64 (20.01)	112; 105.00 (21.69)
Days from baseline weight to treatment, n, median(IQR)	116; 36.50 (20-112)	112; 33.50 (20-56.25)
Baseline HbA1c in mmol/mol, n, mean(sd)	82; 78 (17)	73; 80 (20)
Baseline HbA1c in %, n, mean(sd)	82; 9.3 (1.6)	73; 9.5 (1.8)
Days from baseline HbA1c to treatment, n, median(IQR)	82; 36.50(6.25-70.25)	73; 33.50 (20-56.25)
HbA1c at follow-up in mmol/mol, n, mean(sd)	77; 72 (16)	74; 74 (17)
HbA1c at follow-up in %, n, mean(sd)	77; 8.7 (1.5)	74; 8.9 (1.6)
Days to HbA1c measurement, mean(sd)	77; 196 ± 83	74;221 ± 75
Receive antidepressants, (%)		
Yes	6 (4)	9 (7)
No	154 (96)	114 (93)
Number of non-diabetes drugs, mean(sd)	1.54 (1.98)	1.71 (1.97)
Insulin Type, n(%)		
Basal and bolus	111 (69)	104 (85)
Basal or bolus	49 (31)	19 (15)
Sulfonylurea, n(%)		
Yes	26 (16)	21 (17)
No	134 (84)	102 (83)

DPP4= dipeptidyl-peptidase 4 inhibitor, SGLT= sodium glucose transporter-2 inhibitor, IQR= interquartile range, sd= standard deviation

Assignment Procedures

While in the target trial, patients are randomly assigned to either treatment, in the observational study statistical adjustments, selected based on subject knowledge, would be used to emulate randomization. When drawing a DAG to determine for which variables should be adjusted, all potential confounders should be included. Figure 3 shows an example of a DAG of the assumed causal relationship between treatment and HbA1c response. Here potential baseline confounders would be duration of diabetes, HbA1c, age, glomerular filtration rate (eGFR), frailty, insulin resistance and weight. It is evident that in the observational study missing data on confounders can be an important issue (Table 2) in addition to variables being absent altogether. Sometimes

confounders are not available in the data and alternative measures to acquire the data such as linking records from different settings have not yielded sufficient input. In that case, proxies can sometimes be used to adjust for bias [20]. For instance, in the EHRs from which the sample was selected, eGFR, insulin resistance and frailty were unmeasured confounders. Here type of insulin (basal vs bolus vs basal-bolus insulin) [21], weight and duration of diabetes [22] could be included as proxy variables for insulin resistance in subjects with type 2 diabetes. The number of non-diabetes drugs a patient received prior to treatment initiation could serve as a proxy for frailty [23]. For recent eGFR results, no suitable proxy would be available and thus this would remain an unmeasured confounder. eGFR is an unmeasured confounder since patients with an eGFR between 30-60 ml/min would be eligible for receiving a reduced dose of some DPP4s but not SGLTs. At this point it can then be considered whether it is worth continuing with the study, given the absence of such an important confounder.

Another important point when including baseline confounders is the time at which these variables are measured. Ideally, a baseline HbA1c measurement is available from a date as close as possible before the date of novel treatment initiation. Therefore, an important difference between the observational study and target trial would therefore be that baseline HbA1c is defined as the most recent measurement within 6 months prior to treatment initiation, and weight is defined as the most recent measurement within 365 days before treatment initiation. It would be expected that a recent HbA1c measurement is available given that an elevated HbA1c will often be a reason to switch treatments. However, in the records included in the example, routine HbA1C measurements could also be performed in primary care and therefore not directly available in secondary care EHRs. For HbA1c, only 14% of available measurements were obtained more than 3 months before initiating treatment.

Follow-up Period

In the target trial, follow-up starts at randomization and ends at six months of follow-up, loss to follow-up or death. The follow-up period of the observational study is similar and starts on the date that the first prescription of the treatment combination is recorded and ends at the first follow-up visit more than 3 months after treatment initiation, at 12-months after treatment initiation or death.

Outcome

In the target trial, researchers are blinded when measuring the outcome, HbA1c in mmol/mol at 6 months after randomization. However, in the observational study the follow-up visit is unlikely to take place at 6 months after treatment initiation and a wider time interval would be adopted, including the first follow-up visit recorded between 3-12 months after treatment initiation. Such differences in time to measurement may influence results if there are differences between the two treatment arms in time to first follow-up. For a disease such as diabetes HbA1C values generally tend to increase over time and thus if one of the arms has a longer time to first measurement it could be that these measurements are biased upwards simply because of the time to measurement. In the patient sample in Table 2, the time between treatment initiation and follow-up measurement is similar between the two arms.

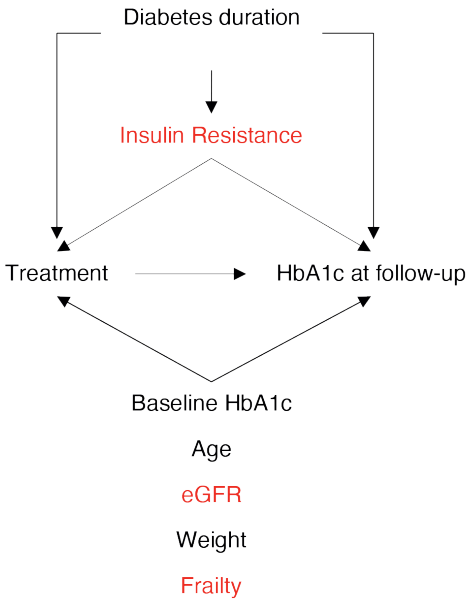


Figure 3: Directed Acyclic Graph representing the underlying causal structure of the difference in response between patients receiving a sodium glucose transporter-2 inhibitor or a dipeptidyl-peptidase 4 inhibitoradded to insulin ± oral antidiabetics. Variables in red were not measured and when available, proxy variables were used. eGFR = glomerular filtration rate

Causal Contrast of Interest

In a target trial, often the intention-to-treat effect as well as the per-protocol effect are estimated. Here in the observational study, only the estimation of the intention-to-treat effect would be possible. In the EHR, only prescription data is available and thus there is no insight concerning whether patients received and ingested their medication.

Statistical Analysis Plan

When estimating the intention-to-treat effect in the observational study, adjustments would be needed for baseline confounding (Figure 3) and selection bias due to loss to follow-up (Figure 1c). HbA1C measurements between 3-12 months would be included as the outcome variable and the type of drug prescribed (SGLT or DPP4) would be the exposure variable. Confounders would include the following baseline variables: HbA1C, weight, type of insulin, number of non-diabetes drugs, age, and diabetes duration and whether or not a patient receives antidepressants would be used to adjust for selection bias.

Adjusting for confounding could be achieved by for instance regression and adjusting for selection bias due to loss to follow-up could be achieved by inverse probability weighting [24] or multiple imputation [12]. Inverse probability weighting has been recommended because it is valid and less complex to perform, whereas multiple imputation is considered complex but also efficient and has been previously recommended when analyzing EHR data [1,12]. When using inverse probability weighting to adjust for loss to follow-up related to depression, weights can be based on a logistic regression model that estimates the inverse of the probability of being

lost to follow-up while including whether or not a patient receives antidepressants at baseline. The 95% confidence intervals can be estimated using a bootstrap variance estimator.

The second option would be to use multiple imputation which is generally accepted as one of the better practices for handling missing data [1,12,25,26]. In multiple imputation, the distribution of the observed data is used to impute realistic values for the missing data [25]. Multiple imputation using chained equations is often used in which a model is estimated for each variable containing missing data [25]. These missing variables are then repeatedly imputed, and Rubin's rules can be used for pooling the imputed datasets [26]. Sensitivity analyses can then be performed in which imputed variables are varied for instance up to 50%.

All variables included in the analysis model should also be included in the imputation model as well as variables used to adjust for selection bias. Auxiliary variables (e.g., sex and baseline height) can be included in the model to improve model fit. Auxiliary variables are those that are not included in the final analysis but are used for imputation because they provide information on the missing variables, increasing the likelihood that the 'missing at random' assumption holds [25,27]. The fraction of missing information can be used to decide which auxiliary variables to include. The fraction of missing information is high when the variables in the dataset provide limited information to impute the missing values [27]. When performing multiple imputation using chained equations, predictive mean matching is often used for larger sample sizes and imputation of continuous variables [25]. However, other methods such as logistic regression and weighted predictive mean matching (MIDAStouch) can be selected for specific types of variables (e.g., nominal) or with smaller samples [26]. The number of imputations (*m*) can be estimated depending on the fraction of missing information and the proportion of missing data. With for instance the observational study, data are assumed to be missing at random and missingness therefore depends on covariates measured in the sample. Furthermore, in the example the sample size is small, the fraction of missing information is high, and the proportion of missing data is high, therefore, the number of imputations should be high (m=70) and a method for small samples such as MIDAStouch can be used [26,28].

After adjusting for selection bias using either multiple imputation or inverse probability weighting regression can be performed to adjust for confounding. To estimate effects for the observational study a generalized linear model with a gamma distribution and log link can be used, results are then presented as the relative change in HbA1c when receiving an SGLT compared to a DPP4 (Table 3). Thus, if the HbA1c after 6 months of follow-up for a patient is 9.2% (77 mmol/mol) when receiving a DPP4, a coefficient of -10% would imply that this same patient would have had an outcome of 8.5% (69 mmol/mol) had they received an SGLT. In the example, no evidence of a clinically meaningful difference in effect can be found, estimated as a 1.7% relative increase in HbA1c when receiving an SGLT compared to a DPP4 with considerable uncertainty (95%CI: -5.4%;9.4%) (Table 3). When adjusting for confounding, small differences can be seen after correcting for loss to follow-up using inverse probability weighting (1.0% change, 95%CI: -8.5%;11.5%) and after multiple imputation (-0.3% change, 95%CI: -7.5%;7.5%). Both report small differences between the two drugs with wide confidence intervals. The effect estimated using multiple imputation varies from -5% to 3% throughout sensitivity analyses and including the use of a sulfonylurea and smoking as auxiliary variables does not lower the fraction of missing information.

Table 3: Relative change in HbA1C from the generalized linear model after adjustment for baseline confounders and selection bias due to loss to follow-up (i.e., when patients receive a sodium glucose transporter-2 inhibitor (SGLT) compared to a dipeptidyl-peptidase 4 inhibitor (DPP4), the relative increase in HbA1C is 1.0%). The results are presented as the exponentiated coefficients which translates to the relative change in HbA1c when receiving an SGLT compared to when a patient receives a DPP4. Thus, if the HbA1c after 6 months of follow-up for a patient is 9.2% (77 mmol/mol) when receiving a DPP4, a coefficient of -10% would imply that this same patient would have had an outcome of 8.5% (69 mmol/mol) had they received an SGLT.

	Relative change in HbA1C	95% Confidence Interval
No adjustment	1.7%	-5.4%; 9.4%
Adjustment for confounding, No adjustment for loss to follow-up	1.4%	-7.2%; 10.8%
Adjustment for confounding, Adjustment for loss to follow-up using IPW	1.0%	-8.5%; 11.5%
Adjustment for confounding, Adjustment for loss to follow-up using MI	-0.3%	-7,5%; 7,5%

IPW= Inverse probability weighting, MI= Multiple Imputation

DISCUSSION

The use of EHRs has resulted in many opportunities for real-world effectiveness and safety research for chronic diseases such as diabetes. However, using these EHRs for research purposes remains challenging. Missing data can limit feasibility of research and bias (e.g., selection bias and confounding) is often not properly addressed [1,5-7] which limits the value of results for clinical practice. Here, target trial emulation can be used to systematically identify potential sources of bias and adjust results accordingly.

We discussed the use of target trial emulation and apply this to an example for diabetes research. By identifying differences between the target trial and the observational study performed, important sources of bias can be identified. Frequently recurring pitfalls, such as including patients based on future information and including prevalent users, can be avoided by comparing eligibility criteria of the target trial with those of the observational study. Furthermore, by identifying differences in the assignment procedures and outcomes between the target trial and observational study, potential sources of bias can be identified and adjusted where needed. However, despite its advantages, target trial emulation should be used thoughtfully. Some sources of bias, though identified, cannot be addressed (i.e., absence of eGFR data) and sometimes decisions to reduce bias and improve internal validity can limit generalizability and reduce sample size. Where strict inclusion criteria reduce the risk of bias, they also reduce generalizability. Furthermore, correcting for certain confounders in the treatment assignment step requires sufficient data to be available on confounders. Use of imputation could address issues with missing confounders but is complex to apply and could also lead to bias when used incorrectly. As with all observational studies, it is unlikely that researchers are able to exclude all potential sources of bias. Therefore, they will carefully need to consider whether results are still of clinical value despite the potential of residual bias to be present.

The amount of missing data is an important problem that influences many steps of target trial emulation using EHRs. Missing data are an unavoidable fact of life in the realm of EHRs and registries. Target trial emulation stimulates researchers to critically consider which variables are essential to generate valid inferences when drafting the protocol for the target trial. In the example, data on a relevant confounder (eGFR) was absent and it is uncertain whether continuing is worthwhile in the presence of residual confounding and the study could also be halted until additional data is collected. To improve validity of results, we recommend researchers formulate expectations concerning the proportion of missing data as well as the availability of variables needed for imputation or weighting (e.g., outcome variables, exposure variables, confounders, causes of selection bias and potential auxiliary variables) when using EHR data. When large amounts of missing data are to be expected, collecting additional data would be the preferred option. Ideally, EHRs are available from for instance initiatives such as the Clinical Practice Research Datalink in which only those primary care practices are included that adhere to certain quality standards in terms of reporting and linking to for instance data from secondary care is possible.

Selecting variables to include in DAGs can be challenging and depends on the setting and time [29]. For instance, physicians in the US are likely to consider a patient's social economic status (SES) in treatment selection whereas in Northern Ireland no out-of-pocket payments are required to receive SGLTs or DPP4s. Therefore, SES was excluded as a confounder in the analyses. Medication adherence was also not included but could be an important confounder in many other settings. When including variables such as medication adherence as a confounder across various treatments, the value of this confounder will depend on the time of measurement [5]. When adjusting for time-dependent confounders, appropriate statistical analyses should be used such as g-estimation [6]. It should be recognized that obtaining information on medication adherence from EHRs can be challenging [30], and it might therefore often be included as an unmeasured confounder. In our analyses, eGFR was considered an important unmeasured confounder but sometimes proxies can be used. In the example in this study, we included proxies for insulin resistance and frailty. However, when interpreting results, it should be recognized that proxies are often imperfect and therefore residual bias can remain. Furthermore, proxies should be used with care since their use can also open backdoor paths, thus introducing bias [20].

However, generally some data will remain missing despite efforts of additional data collection and researchers will need to perform analyses while taking into account this missing data. Multiple imputation when analyzing EHR data is often recommended since it is one of the few methods that is both efficient and effective in reducing bias, if used correctly [1,12,25]. However, results from multiple imputation can also be biased when the likelihood of model misspecification is large for instance when the proportion of missing data is high [31]. In the example provided in this study, we illustrated that the large amounts of missing data that can be present in EHRs might limit the value of imputation when other (auxiliary) variables have a limited predictive value and are themselves also missing. The proportion of missing data and the fraction of missing information were high while correlations with many of the variables included were low. When auxiliary variables contain many missing values, the risk of bias in regression estimates increases because of the higher proportion of missingness and ratio of variables to complete cases [32]. In such scenarios, inverse probability weighting might be preferred [31]. However, both strategies assume data are missing at random and when this assumption does

not hold, and data are missing not at random both methods could lead to biased estimates [31]. When collecting additional information or performing statistical adjustment such as inverse probability weighting or multiple imputation are not possible, researchers can consider whether continuation is worthwhile. For instance, in the example used in this study, the absence of information on eGFR could be a motivation to cease the study altogether. In other cases when adjustments cannot be made for all sources of bias, but their impact is considered small, acknowledging their presence could assist interpretation of results.

In this study, we used EHRs of the Western Health and Social Care Trust to discuss several benefits and challenges when using target trial emulation. The dataset presented several challenges one of which was the small size of the sample after applying eligibility criteria. The required sample size to assess significance of the effect found in this study is 868 patients per arm assuming an alpha of 0.05 and a beta of 0.8. Thus, if the same amount of missing data is expected, EHRs of roughly ten times as many patients (n=71,673) would be required before applying eligibility criteria. Furthermore, it should be noted that there are many more aspects that should be considered when performing an observational study using EHRs relating to for instance data extraction, data pre-processing and data validation not addressed in this study [33].

Observational research that uses EHRs to assess real-world effectiveness and safety is crucial for informing clinical practice. Target trial emulation is a useful tool for conducting these studies since it enables researchers to avoid frequently recurring problems such as selection bias, immortal time bias and confounding. However, researchers should consider that emulating a target trial using EHRs can be challenging when large variations in treatments prescribed are expected and the amount of missing data is large. Therefore, target trial emulation can only be used to improve care when sufficient information on outcome measures and variables to adjust for bias is collected.

REFERENCES

1 Farmer R, Mathur R, Bhaskaran K, Eastwood SV, Chaturvedi N, Smeeth L. Promises and pitfalls of electronic health record analysis. Diabetologia. 2018 Jun 1;61(6):1241-8.

2 Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. Am J Epidemiol. 2016 Apr 15;183(8):758-64.

3 Flory JH, Roy J, Gagne JJ et al. Missing laboratory results data in electronic health databases: implications for monitoring diabetes risk. J Comp Eff Res. 2017 Jan;6(1):25-32.

4 Petersen I, Welch CA, Nazareth I et al. Health indicator recording in UK primary care electronic health records: key implications for handling missing data. Clin Epidemiol. 2019 Feb 11;11:157-167. doi: 10.2147/CLEP.S191437

5 Farmer RE, Ford D, Forbes HJ et al. Metformin and cancer in type 2 diabetes: a systematic review and comprehensive bias evaluation. Int J Epidemiol 2017 Apr 1;46(2):728-44.

6 Patorno E, Patrick AR, Garry EM et al. Observational studies of the association between glucose-lowering medications and cardiovascular outcomes: addressing methodological limitations. Diabetologia. 2014 Nov 1;57(11):2237-50.

7 Ripamonti E, Azoulay L, Abrahamowicz M, Platt RW, Suissa S. A systematic review of observational studies of the association between pioglitazone use and bladder cancer. Diabet Med. 2019 Jan;36(1):22-35.

8 Yang W, Zilov A, Soewondo P, Bech OM, Sekkal F, Home PD. Observational studies: going beyond the boundaries of randomized controlled trials. Diabetes Res Clin Pract. 2010 May 1;88:S3-9.

9 Labrecque JA & Swanson SA. Target trial emulation: teaching epidemiology and beyond. Eur J Epidemiol 2017 Jun 1;32(6):473-5.

10 Hernán MA, Robins JM. Causal Inference: What If. Chapman & Hall/CRC, Boca Raton; 2020. [cited 2020 Nov 1] Available from: https://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/

11 Hernán MA, Hernández-Díaz S, Robins JM. A structural approach to selection bias. Epidemiology 2004 Sep 1:15(5):615-25

12 Seaman SR, White IR. Review of inverse probability weighting for dealing with missing data. Stat Methods Med Res. 2013 Jun;22(3):278-95.

13 Jin J, Sklar GE, Oh VM, Li SC. Factors affecting therapeutic compliance: A review from the patient's perspective. Ther Clin Risk Manag. 2008 Feb;4(1):269.

14 Fitzpatrick SL, Jeffery R, Johnson KC. Baseline predictors of missed visits in the Look AHEAD study. Obesity. 2014 Jan;22(1):131-40.

15 Holt RI, De Groot M, Golden SH. Diabetes and depression. Curr Diab Rep. 2014 Jun 1;14(6):491.

16 Ray WA Evaluating medication effects outside of clinical trials: new-user designs. Am J Epidemiol. 2003 Nov 1;158(9):915-20.

17 Hernán MA, Sauer BC, Hernández-Díaz S, Platt R, Shrier I. Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses. J Clin Epidemiol. 2016 Nov 1;79:70-5.

18 Suissa S. Immortal time bias in observational studies of drug effects. Pharmacoepidemiol Drug Saf. 2007 Mar;16(3):241-9.

19 Sullivan SD, Bailey TS, Roussel R et al. Clinical outcomes in real-world patients with type 2 diabetes switching from first-to second-generation basal insulin analogues: comparative effectiveness of insulin glargine 300 units/mL and insulin degludec in the DELIVER D+ cohort study. Diabetes Obes Metab. 2018 Sep;20(9):2148-58.

20 VanderWeele TJ. Principles of confounder selection. Eur J Epidemiol. 2019 Mar 15;34(3):211-9.

21 Seufert J, Borck A, Bramlage P. Addition of a single short-acting insulin bolus to basal insulin-supported oral therapy: a systematic review of data on the basal-plus regimen. BMJ Open Diabetes Res Care. 2019 Oct 1;7(1):e000679

22 Dennis JM, Shields BM, Hill AV. Precision medicine in type 2 diabetes: clinical markers of insulin resistance are associated with altered short-and long-term glycemic response to DPP-4 inhibitor therapy. Diabetes Care. 2018 Apr 1;41(4):705-12.

23 Gutiérrez-Valencia M, Izquierdo M, Cesari M, Casas-Herrero Á, Inzitari M, Martínez-Velilla N. The relationship between frailty and polypharmacy in older people: A systematic review. Br J Clin Pharmacol. 2018 Jul;84(7):1432-44.
24 Mansournia MA, Altman DG. Inverse probability weighting. BMJ. 2016 Jan 15;352:i189.
25 White IR, Royston P, Wood AM. Multiple imputation using chained equations: issues and guidance for practice. Stat Med. 2011 Feb 20;30(4):377-99
26 Van Buuren S. Flexible imputation of missing data. Chapman & Hall/CRC, Boca Raton. 2018. [Cited 2020 June 18]. Available from: https://stefvanbuuren.name/fimd/
27 Madley-Dowd P, Hughes R, Tilling K, Heron J. The proportion of missing data should not be used to guide decisions on multiple imputation. J Clin Epidemiol. 2019 Jun 1;110:63-73.
28 Gaffert P, Meinfelder F, Bosch V. Towards an mi-proper predictive mean matching. 2016 Jan 25. Available from: https://www. Uni-bamberg. De/fileadmin/uni/fakultaeten/sowi_lehrstuehle/andomiza/Personen/Dateien_Florian/properPMM.
29 Schneeweiss S, Patorno E. Conducting real-world evidence studies on the clinical outcomes of diabetes treatments. Endocrine Reviews. 2021 Mar 12.
30 Galozy A, Nowaczyk S, Sant'Anna A, Ohlsson M, Lingman M. Pitfalls of medication adherence approximation through EHR and pharmacy records: Definitions, data and computation. International journal of medical informatics. 2020 Apr 1;136:104092.
31 Hughes RA, Heron J, Sterne JA, Tilling K. Accounting for missing data in statistical analyses: multiple imputation is not always the answer. Int J Epidemiol. 2019 Aug 1;48(4):1294-304.
32 Hardt J, Herke M, Leonhart R. Auxiliary variables in multiple imputation in regression with missing X: a warning against including too many in small sample research. BMC Med Res Methodol. 2012 Dec 1;12(1):184.
33 Critical Data MI. Secondary analysis of electronic health records. Springer Nature; 2016.


101
00111
0010100000
11111
0001101
010101111000

How can we discover the most valuable types of big data and artificial intelligence-based solutions? A methodology for the efficient development of the underlying analytics that improve care.

Bakker L, Aarts J, Uyl-de Groot C, Redekop W.

BMC Medical Informatics and Decision Making. 2021 Dec;21(1):1-2.

ABSTRACT

Background: Much has been invested in big data and artificial intelligence-based solutions for healthcare. However, few applications have been implemented in clinical practice. Early economic evaluations can help to improve decision-making by developers of analytics underlying these solutions aiming to increase the likelihood of successful implementation, but recommendations about their use are lacking. The aim of this study was to develop and apply a framework that positions best practice methods for economic evaluations alongside development of analytics, thereby enabling developers to identify barriers to success and to select analytics worth further investments.

Methods: The framework was developed using literature, recommendations for economic evaluations and by applying the framework to use cases (chronic lymphocytic leukemia (CLL), intensive care, diabetes). First, the feasibility of developing clinically relevant analytics was assessed and critical barriers to successful development and implementation identified. Economic evaluations were then used to determine critical thresholds and guide investment decisions.

Results: When using the framework to assist decision-making of developers of analytics, continuing development was not always feasible or worthwhile. Developing analytics for progressive CLL and diabetes was clinically relevant but not feasible with the data available. Alternatively, developing analytics for newly diagnosed CLL patients was feasible but continuing development was not considered worthwhile because the high drug costs made it economically unattractive for potential users. Alternatively, in the intensive care unit, analytics reduced mortality and per-patient costs when used to identify infections (-0.5%, -€886) and to improve patient-ventilator interaction (-3%, -€264). Both analytics have the potential to save money but the potential benefits of analytics that identify infections strongly depend on infection rate; a higher rate implies greater cost-savings.

Conclusions: We present a framework that stimulates efficiency of development of analytics for big data and artificial intelligence-based solutions by selecting those applications of analytics for which development is feasible and worthwhile. For these applications, results from early economic evaluations can be used to guide investment decisions and identify critical requirements.

BACKGROUND

With the increasing ability to collect healthcare data, billions of dollars have been invested in (big) data analytics and artificial intelligence (AI) by private (e.g., IBM, Google, hospitals) and public institutions worldwide (e.g., Agency for Healthcare Research and Quality, the Patient-Centered Outcomes Research Institute, European Commission) [1-9]. Analytics can be applied in many ways, and it has often been suggested that they can improve care for a wide variety of clinical fields [10-15]. Bates et al. define big data analytics as the discovery and communication of patterns in datasets that are extremely complex due to their size (volume), rapid collection (velocity) and/or the need to combine multiple data sources (variety) [14]. The term Artificial Intelligence was first mentioned many years ago and is defined as the ability of computers to mimic or simulate the human mind [16]. However, despite many publications on the potential of big data analytics and AI, few analytics have been implemented [6,17-20] and resulted in health benefits and/or cost savings [21,22,23].

Data availability can be an important barrier to the development of analytics that improve healthcare [4,12,17,24-26]. The datasets required to develop machine learning models should be large and, depending on the method used, should contain sufficient data on relevant features [11, 27]. Data-related problems mentioned in the literature include limited sample size [4,24-26,28], a short duration of follow-up [24], validity of results with heterogeneous patient populations and selection bias [4,13,17,24,28,29] and bias due to missing data [12,24,29,30]. Moreover, successful development does not mean easy implementation; important barriers to implementation include the need for prospective validation [4,24,28] and the high costs of validation and implementation [4,19,24,31-33].

For other healthcare technologies, such as drugs, medical devices and diagnostic tests, economic evaluations are used to assess the potential impact of anticipated barriers early on during development [34-38]. In economic evaluations, the health benefits and costs of novel technologies are compared to the benefits and costs of an alternative such as current care. Use of these economic evaluations alongside development is recommended to assist decision-making by developers, to analyze the impact of uncertainty in performance of the technology on outcomes, and to identify critical requirements (e.g., price) for successful market access and dissemination [36,37]. A key aim of this approach is to increase the likelihood of successful market uptake and avoid wasting investments due to failed implementation.

Very few economic evaluations of analytics exist [13,17,20-23,39,40] and the ones that do have omitted relevant costs [19,22]. Moreover, recommendations on how and when to perform economic evaluations of analytics do not exist, even though their use could improve development efficiency by identifying analytics with the greatest potential health impact. In this paper, we present a framework that can assist developer decision-making by selecting applications of analytics that are not only worth developing but also feasible.

METHODS

We present a framework that efficiently selects analytics that are relevant, feasible and capable of generating important health and economic benefits (Figure 1). The framework was developed

based on challenges of analytics development defined in the literature and best practice recommendations for economic evaluations. It was then further refined by applying it in three clinical use cases. The use cases were selected from a European Horizon 2020 funded project (AEGLE) that aimed to develop a cloud-based big data analytics platform. The three use cases focused on chronic lymphocytic leukemia (CLL), the intensive care unit (ICU) and diabetes.

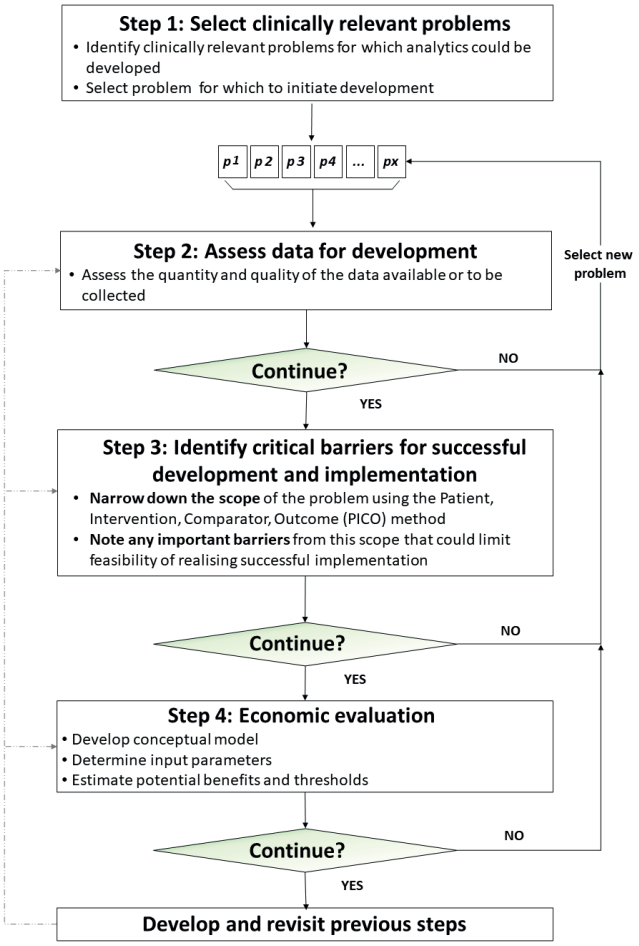


Figure 1: Flowchart for assessing health economic benefits of novel analytics alongside development. *p=problem*

Step 1: Select clinically relevant problems

This first step involves selecting relevant clinical problems. Whether problems are considered clinically relevant depends on the setting for which analytics are developed and the experts involved. When analytics are developed for a local hospital (e.g., for a learning health system), local experts should be consulted to identify relevant problems. When the aim is to develop

analytics for a wider audience such as clinical experts in different countries or continents, then interviews with multiple potential users are recommended alongside a review of guidelines and the literature. Needless to say, a multidisciplinary approach throughout this step is crucial [10,41].

Step 2: Assess data for development

After relevant problems are selected, it is necessary to assess whether the data available, or to be collected, is of sufficient quantity and quality to address the problem. Such an assessment may include careful scrutiny of the sample size, duration of follow-up, expected frequency of missing data, potential sources of bias and heterogeneity in care practices between sites. Moreover, the timing of data collection and the types of outcomes collected during follow-up may differ between clinical sites.

Step 3: Identify critical barriers to realizing successful development and implementation

The scope of the problem should be narrowed down and used to identify critical barriers prior to estimating costs and benefits. Narrowing down the scope is a critical step in any economic evaluation [37] and one way to achieve this is through the Population (or Patient), Intervention, Comparator, and Outcomes (PICO) method [37]. First, the target population (P) is defined, which can include a description of the setting and the population size. The intervention (I) should include a description of the care pathways involved, including the analytics to be developed, the additional software and hardware needed to use the analytics, and the actions that follow from use of the analytics. The description of the comparator (C) entails a discussion on treatments available and relevant software and hardware elements used in current care. The final component of outcomes(O), refers not just to clinical outcomes but all outcomes considered relevant by users and purchasers, including mortality, life years gained, quality-adjusted life years gained (QALYs) and economic benefits. Ideally, they should go beyond diagnostic performance metrics like Area Under the Curve (AUC) [4,17,42,43] and include outcomes related to health benefits, satisfaction, and costs.

The detailed description of the scope, formulated using the PICO method, can then be used to identify potential barriers to successful development and implementation of the analytics. An example of a critical barrier is whether the health information system currently used in a health center is sufficient to support the analytics or whether major upgrades are needed. If the examination of possible barriers does not reveal any insurmountable barriers, the health and economic benefits can be estimated. When continuing development seems risky, for instance because of the limited availability of required software and hardware elements in current practice, a developer can decide to select a new problem or cease development altogether.

Step 4: Economic evaluation

The next step is to perform an economic evaluation of the analytics that are considered feasible to develop. An evaluation starts by developing a conceptual model and collecting input data. A conceptual model can be developed in different ways, including the estimation of the number needed to treat [44], decision curve analysis [42,43], decision trees, and Markov models. Depending on the stage of development, the models may vary from very simple to very complex. The validity of the model should be assessed according to best practice guidelines [37,45]. Information on relevant input parameters required to populate the model can be collected

alongside model development from sources such as patient-level data and the literature, but are sometimes limited to expert opinion or assumptions, particularly in the early stages of development. Uncertainty surrounding parameter estimates generally decreases as development progresses and more information becomes available [36,38].

Base case estimates of potential benefits can then be determined using the most likely parameter values. Results can be presented using the incremental cost-effectiveness ratio (ICER) but more importantly; results should be presented such that they are understandable to the target audience (investors, future users, and purchasers). The uncertainty in these point estimates should always be analyzed using uncertainty analyses. Uncertainty analyses can include scenario analyses and sensitivity analyses, but also analyses to determine critical thresholds of relevant parameters, such as accuracy and pricing thresholds needed to realize health and economic benefits. The headroom can also be estimated according to the following formula:

$$\text{Headroom} = N + \lambda * Q$$

Here N refers to the potential savings where the costs of the technology are set to zero, λ is the willingness to pay threshold and Q are the health effects gained [46]. Moreover, probabilistic sensitivity analyses can be used to estimate the impact of uncertainty in all parameters simultaneously. For each parameter, random estimates are drawn many times (e.g., n=1000) from their underlying distribution. For these estimates, the costs and effects are calculated and presented using a cost-effectiveness plane and a cost-effectiveness acceptability curve. In a cost-effectiveness acceptability curve, the probability that an intervention is cost-effective is plotted against a range of willingness to pay thresholds.

Iterative Approach

When a developer decides to continue development, the different steps (assess data for development, critical barriers to realizing success, and the economic evaluation) should be revisited as needed throughout development, represented by the dotted line in Figure 1.

Clinical use cases

Chronic lymphocytic leukemia

The first clinical use case, focused on developing cloud-based analytics using next generation sequencing (NGS) data of CLL patients from three clinical sites across Europe (Sweden, Italy & Greece). CLL is characterized by considerable heterogeneity in disease progression [47,48] and after diagnosis, the majority of CLL patients are followed according to a 'watch and wait' (W&W) strategy. Roughly 60% of these patients progress to having active disease requiring treatment [47]. The treatment they receive depends on their molecular profile and general fitness as well as on treatment approval and availability [47].

Intensive care

In the second use case, the aim was to develop analytics for ICU care using routinely collected data. Data from electronic health records (EHRs) and mechanical ventilators of patients from a Greek ICU was available for development. There are many ways in which analytics can improve ICU care and a variety of applications have been suggested in the literature [10,11]; these include

analytics to determine readmission risk, predict length of stay, diagnose sepsis, and improve the interaction between patients and mechanical ventilators [11].

Diabetes Mellitus (diabetes type 2)

Many diabetes treatments are available, and these can often be combined to improve effectiveness. However, evaluating efficacy for all combinations, types of patients and treatment lines in randomized controlled trials would not be feasible, and using EHRs to evaluate effectiveness of treatment combinations has previously been suggested [30]. In this third use case, the aim was to develop analytics using EHRs in the United Kingdom to personalize diabetes treatment for patients.

RESULTS

The framework was applied to three clinical use cases (e.g., CLL, intensive care and diabetes) (Table 1). The results for each case are described one by one.

Table 1: The methodology applied to address problems in care for chronic lymphocytic leukemia, the intensive care and diabetes.

	CLL Problem 1	CLL Problem 2	ICU Problem 1	ICU Problem 2	Diabetes
Clinically relevant problem	Variations in treatment response to 1st and 2nd line	Imperfect algorithms for identifying newly diagnosed, high-risk CLL patients	Identifying patients with ineffective efforts at risk of poor outcomes	Diagnosing catheter related bloodstream infections (CRBSI)	Unknown variation in response to treatment with SGLTs+ GLPs
Assess data for development	- NGS data available - Follow-up probably sufficient - Large variation in treatments	- NGS data available - Follow-up sufficient	- Monitoring & EHR data available - Sufficient sample size, sufficient follow-up, limited missing data.	- EHR & biosignal data available & continued prospectively - Limited missing data anticipated.	- EHR data available from secondary care. - Large amounts of missing follow-up data.

Table 1: *Continued.*

	CLL Problem 1	CLL Problem 2	ICU Problem 1	ICU Problem 2	Diabetes
Identify critical barriers for successful development and implementation	-	**P:** Newly diagnosed CLL patients without treatment indication. **I:** Analytics that identify high risk patients followed by treatment with ibrutinib. **C:** Stratification using clinical symptoms without receiving treatment. **O:** Costs, LYG, QALYs **Barriers:** - Site-specific validation required. - Reimbursement of novel treatment.	**P:** Patients on assisted mechanical ventilation. **I:** Identify patients at risk of poor outcomes with analytics and intervene to avoid ineffective efforts **C:** Care in which ineffective efforts are not identified **O:** Mortality, LOS, costs, LYG, QALYs **Barriers**: - Availability of monitor that identifies ineffective efforts. - Site-specific validation.	**P:** Patients with central venous catheter **I:** Early identification of CRBSI, catheter removal & antibiotics. **C:** Late identification of CRBSI, catheter removal & antibiotics. **O:** Mortality, LOS, costs, LYG, QALYs **Barriers**: - Varying prevalence of CRBSI. - Integration of analytics in an EHR. - Site-specific validation.	-
Economic Evaluation	-	Benefits: 0.13 QALYs, +€89,985	Benefits: -3% mortality, 0.21 QALYs, -€264 [58]	Benefits: -0.5% mortality, +0.06 QALYs, -€886	-
Continue development	Not feasible. Sample size too small and large variations in prescribing practices.	Not feasible. High costs of treatment offset benefits gained.	Feasible. Invest in research into the effectiveness of intervention and the price of the analytics [58].	Feasible. If the target market extends beyond Greece the impact of the prevalence of CRBSI on benefits should be considered.	Not feasible. Small sample size and large amount of missing follow-up data.

CLL= Chronic Lymphocytic Leukemia, ICU= Intensive Care Unit, NGS= Next generation sequencing, SGLTs= sodium glucose transporter-2 inhibitors, GLPs= glucagon-like peptide-1 agonists, CRBSI= Catheter related bloodstream infection, EHR= Electronic Health Record, LOS= Length of Stay, LYG= life years gained, QALY=quality-adjusted life years gained

Case 1: CLL

Because of the heterogeneous nature of CLL progression and treatment response, stratifying patients according to their expected prognosis could improve care [47]. In discussions with clinical experts, problems were selected based on the three decision points suggested by Baliakas et al. The first is upon diagnosis, when clinicians want to determine which patients are likely to progress to active disease. The second decision point is the moment when patients have active disease, and a first-line treatment needs to be selected. The third is the decision point when first-line treatment has failed, and a decision needs to be made about which second-line treatment is best for a patient [47]. CLL experts stated that decision points two and three were the most clinically relevant.

Regarding decision point one, developing analytics to improve stratification for these patients was considered feasible with the data available (Table 1). In contrast, the feasibility regarding decision point two was limited because of large variations between countries in the treatments prescribed. For decision point three, development of analytics to improve decision-making would not be feasible because it was expected that few patients in the data set received second-line treatment, which therefore meant a small sample size. Consequently, the first decision point was considered the best choice for analytics development.

When defining critical barriers, the scope included newly diagnosed Swedish CLL patients. In current care, these patients are not treated, but are regularly seen by the hematologist and undergo a blood test. When developing the analytics in 2015-2016, no treatment was available for patients with a high risk of progression. The only possible changes in care available at the time was the ability to personalize the intensity of follow-up and the ability to inform patients about their risk. These very limited options of 'treatment' can be considered a critical barrier for success since it is likely that costs of NGS and analytics are high while health benefits could only be expected through the reduction in a patient's uncertainty (and anxiety) regarding prognosis. Therefore, at the time, analytics development did not continue beyond research purposes. However, a recent publication has suggested that early treatment of intermediate- and high-risk patients with ibrutinib could delay time to next treatment. Given these new findings, we updated results for this application, including the possibility of treatment with ibrutinib as part of the intervention.

After the PICO question was formulated, input parameters (probabilities, utilities, unit costs and resource use) were derived from the literature, Swedish guidelines, and expert opinion (Table S1). A four state Markov model (Figure S1) was used to estimate costs, life years and quality-adjusted life years adopting a lifetime time horizon and a healthcare payer perspective. Long-term survival was estimated by combining results on time to next treatment from Condoluci et al [49] with the hazard ratio reported in preliminary results from a randomized controlled trial comparing early ibrutinib treatment with current care [50]. More details on the model structure and input parameters used to estimate the health and economic benefits can be found in the Supplementary File. Even if an effective treatment is available, it is unlikely that analytics to improve stratification of newly diagnosed watch and wait CLL patients would be considered cost-effective: use of analytics would lead to a substantial cost increase (€89,985) but only a modest gain in health (0.13 QALYs) (Table 2). We demonstrated the relevance of univariate uncertainty analyses to assess the impact of parameter uncertainty (Figure S2). In

univariate uncertainty analyses, the impact of an individual parameter is assessed by varying its estimate while keeping all other parameters constant. Here, the high costs of the treatment in the intervention arm are decisive in the incremental costs. The relevance of scenario analyses is demonstrated in Table 2 where even in the best-case scenario, analytics are unlikely to be cost-effective, since the incremental cost-effectiveness ratio exceeds thresholds used in Sweden. When varying all parameters simultaneously in the probabilistic sensitivity analyses, most of the estimates are in the upper right and left quadrant (Figure 2). This means that most estimates reflect higher costs and either higher or lower QALYs. When these results are shown on a cost-effectiveness acceptability curve, we can see that better stratification of watch and wait patients and subsequent treatment with ibrutinib has an extremely low chance of being cost-effective (Figure S3).

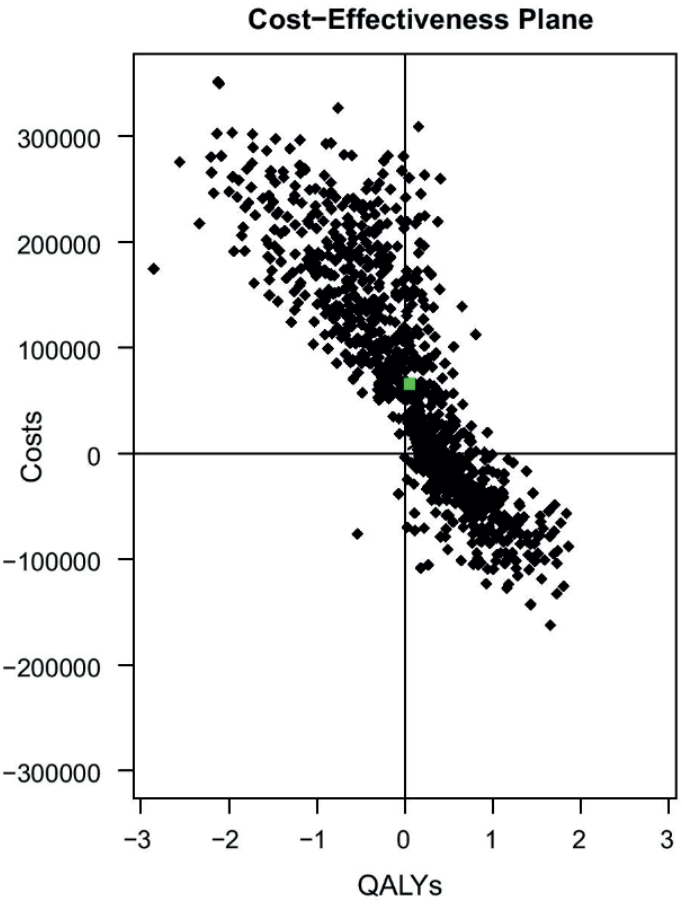


Figure 2: Cost-effectiveness plane reporting the quality-adjusted life years and costs (€) from the probabilistic sensitivity analysis.

Table 2: Results from the base case and best case scenario for analytics to improve stratification of watch and wait patients in chronic lymphocytic leukemia compared to current care.

	Costs	Life Years	QALYs
Base Case			
Current Care	€103,947	11.18	8.57
Care with analytics	€193,932	11.51	8.69
Incremental	€89,985	0.34	0.13
ICER	-	€268,373	€708,192
Best Case Scenarioa			
Current Care	€98,458	11.18	8.57
Care with analytics	€155,667	11.58	8.91
Incremental	€57,209	0.41	0.34
ICER	-	€141,972	€166,879

ICER= Incremental Cost-Effectiveness Ratio, [a]Best Case Scenario= Low HR of time to next treatment with early ibrutinib treatment (0.11), 50% reduction in costs of ibrutinib per cycle (€2,542), 50% reduction of costs of venetoclax with 50% (€2,731), low costs of analytics and genomic and genetic testing (€100), High quality of life for those receiving early treatment with ibrutinib (0.78).

Case 2: The intensive care unit

For the intensive care, relevant problems were identified through discussions with an intensivist at the Greek hospital that was involved in development.

Catheter Related Bloodstream Infection

The first ICU-related problem selected, was that infections caused by central venous catheters were often diagnosed only after they are severe. Catheter related bloodstream infections (CRBSIs) are considered an important issue in the ICU since infected patients have an increased mortality and prolonged length of stay compared to other ICU patients [51]. The aim was to use analytics to diagnose CRBSI in an early stage to reduce disease severity, risk of death and costs.

EHR and biosignal data were available to develop the analytics (N=2000) and additional records were to be collected prospectively. The required follow-up was short, and the relevant parameters needed to develop the analytics and evaluate outcomes (e.g., mortality, length of stay) were routinely collected. Missing data was expected to be present but manageable.

No insurmountable barriers were identified when narrowing down the scope in the early stages of development. An example of a potential barrier for the CRBSI analytics is the uncertainty in the probability of CRBSI. The frequency of CRBSI varies tremendously across countries and sites. In Western European countries, the reported incidence of CRBSI is low [52]. However, for the Greek hospital for which analytics were developed 7.5% of patients developed CRBSI during their ICU stay [53] and in other Greek hospitals reported even higher percentages (22.4%) [54]. If the target market for the analytics would have been limited to the US and western European countries, obtaining better estimates of the frequency of CRBSI would have been recommended prior to continuing with an economic evaluation. Another barrier might have been the need for

EHRs to enable the analytics. However, since most Greek and European hospitals have adopted EHRs this was not expected to be an issue. Additional validation when adopting results in other hospitals would probably be required and feasible but would need to be taken into account in the economic evaluation. Based on these barriers, continuing with the economic evaluation was recommended.

A detailed description of the model and input parameters used to estimate health and economic benefits can be found in Figure S4 and Table S2. A decision tree was combined with a four state Markov model (Figure S4), adopting a lifetime time horizon, and including only direct medical costs. Input parameters were derived from the literature, hospital reports, and expert opinion. The effect of earlier intervention on ICU mortality and ICU length of stay were derived from a study reporting the effect of earlier prescription of antibiotics [55]. Initial estimates demonstrated that continuing development was worthwhile since analytics could reduce mortality (0.5%), improve QALYs (0.06) and lead to cost-savings (€886) per patient. All input parameters were varied extensively in uncertainty analyses but the probability of CRBSI had substantial influence on the results. When the price of the technology was below €19,216 per bed, the analytics could reduce costs compared to current care. This meant that the headroom to achieve cost-neutrality with the intervention was €19,216 per bed, which meant there was sufficient room for costs of analytics, validation, and implementation. Given the large potential for the analytics to generate savings it was considered relevant to continue with development. However, the key factor that influenced benefits was the prevalence of CRBSI (Figure 3). In this case, it was worthwhile to closely monitor site-specific prevalence throughout development and carefully consider the appropriate target market given the large variation in prevalence across sites.

Ineffective Effort Events

The second ICU-related problem to be addressed with analytics, was suboptimal interaction between patients and their mechanical ventilator. One form of suboptimal interaction relates to ineffective efforts where a patient tries, but fails, to trigger the mechanical ventilator into providing a breath. Several studies have found that ineffective efforts could be associated with worse outcomes [56,57]. Here the aim was to enable clinicians to intervene in those patients with ineffective efforts, who are therefore at risk of having worse outcomes.

EHR records were available for all patients and once again relevant parameters were routinely collected and missing data was expected to be manageable. Furthermore, recordings of > 24hrs for more than 100 patients were available from a prototype monitor detecting patient-ventilator interaction.

When assessing feasibility, no barriers were considered insurmountable (Table 1). An important barrier was the need to have a monitor capable of measuring ineffective efforts in addition to analytics that could identify patients with ineffective efforts at risk of having worse outcomes. The prototype monitor available in the Greek ICU would need to be purchased in order to use the analytics. Furthermore, costs of site-specific validation would need to be included in the economic evaluation.

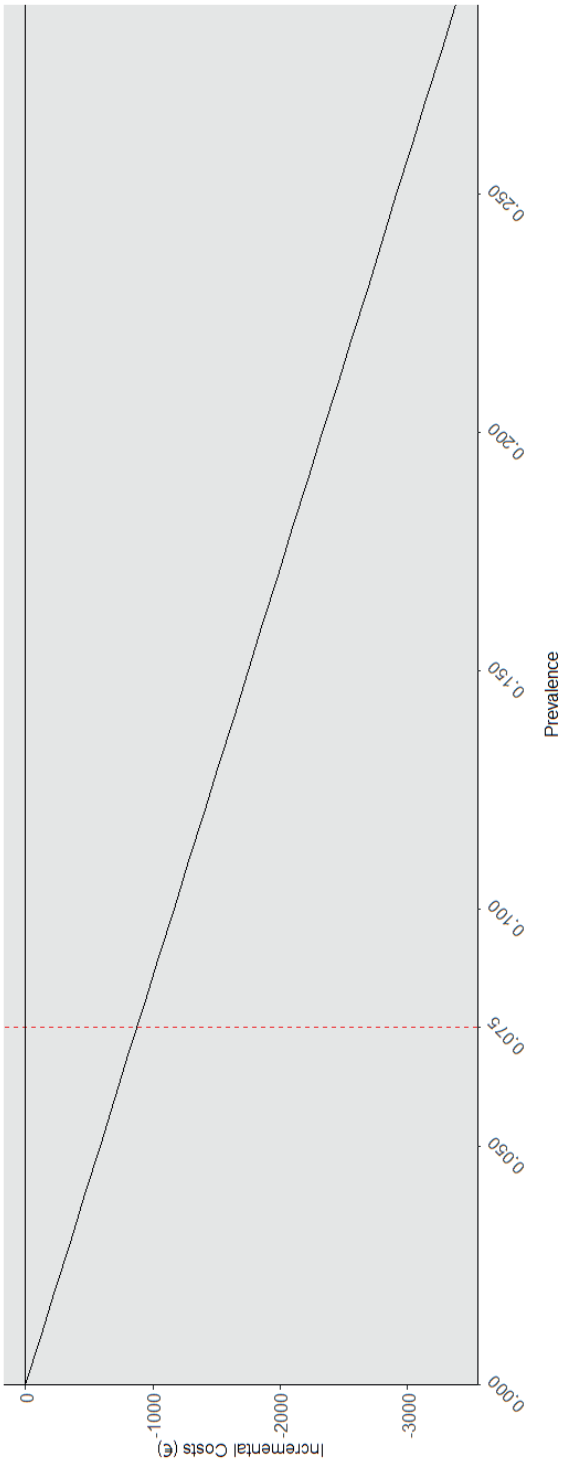


Figure 3: Impact of the prevalence of catheter related bloodstream infection in the intensive care unit on incremental savings.

The model and input parameters used to estimate the health and economic benefits have been previously reported [58]. The potential impact of analytics that identify patients with ineffective efforts at risk of having worse outcomes also suggests that continuing further development is worthwhile [58] since it can reduce mortality by 3%, increase QALYs by 0.21 and reduce costs (€264) [58]. Furthermore, it was demonstrated that even if the effectiveness of intervening was varied extensively, benefits could still be achieved [58]. The headroom for the analytics to generate savings (€7,307) was considered sufficient to cover relevant hardware costs and additional costs of site-specific validation. Thus, further development was considered both relevant and feasible and the potential impact of the analytics was considered substantial.

Case 3: Diabetes

For diabetes, clinicians indicated that a highly relevant problem was to determine predictors of response to treatment with sodium glucose transporter-2 inhibitors combined with glucagon-like peptide-1 agonists. EHR data was available from diabetes patients treated in secondary care in the United Kingdom. However, a small sample size and substantial missing follow-up data raised questions about the feasibility of development, which resulted in the decision not to assess critical barriers and conduct an economic evaluation.

DISCUSSION

In this paper, we present a framework that aims to promote the efficient development of high potential analytics by rapidly assessing whether it is feasible and worthwhile to continue development. The use cases demonstrate the value of first assessing the feasibility of development and identifying relevant barriers before estimating the potential health and economic benefits of analytics. Examples were presented for CLL and diabetes where development was not feasible given the data available. Furthermore, the essence of critically narrowing down the scope is demonstrated for CLL and the ICU where the absence of actionable output is an important barrier to realizing success and disease prevalence strongly influences benefits.

Early economic evaluations of analytics can assist decision-making of developers and stimulates them to develop those analytics with the greatest potential benefits. These evaluations allow developers to assess the influence of certain requirements of analytics (e.g., the costs of the technology, validation, and implementation) on their potential health and economic impact. In our use cases, we see risks that could strongly influence widespread adoption, such as the prevalence of CRBSI and the high drug costs for CLL. Early economic evaluations can also be used to strengthen the business case of developers seeking funding for prospective validation and evaluation. This is especially relevant since the high costs of validation and implementation are important barriers to successful use of analytics in clinical practice [4,19,24,31-33]. During implementation, data and tools used to perform early economic evaluations alongside development can be reused to perform a ‘late’ economic evaluation to convince payers that the analytics are worth purchasing. Elements covered in this framework align with key economic information sought by payers such as the UK’s National Institute for Health and Clinical Excellence [59].

However, for efficient development, economic evaluations should only be initiated for those applications deemed feasible and after ensuring that there are no critical barriers to success. Often multiple analytics can be developed for a single setting, disease or using a single dataset [27,60]. For instance, for the ICU [11] and diabetes care [61] many more types of EHR-based analytics have been suggested than the ones presented here. This is an important difference compared to when early economic evaluations are used to assist decision-making during development of a technology with one or few applications (e.g., diagnostics). Since it is often unrealistic to evaluate - all potential applications of a particular type of analytics, our framework stimulates developers to select which applications are worthy of additional resources. Where feasibility is clearly a problem for the diabetes use case, the lack of an actionable output is the shortcoming for CLL; an issue often reported in the literature [10,15,24,25]. The initial analyses performed in the early economic evaluation can be very simple at first but can become more complex as development progresses; this corresponds with recommendations that analytics development and validation should also be iterative [4,62]. However, as with analytics for CLL and CRBSI, it is sometimes worthwhile to invest more time in adding additional details at an early stage, since it is better to fail fast when limited investments have been made. Using early economic evaluations in an iterative manner and providing a detailed definition of the scope aligns with best practices for early economic evaluations of other healthcare technologies such as diagnostic tests [34-38]. The recommendations provided by others such as Drummond et al [63], or Buisman et al. [37] regarding the selection of a model structure (e.g., decision tree, Markov model), estimation of input parameters, and calculating outcomes (such as the ICER) are likely to be applicable when estimating benefits. We demonstrate in the CLL and diabetes use cases how the framework may assist developers in selecting those applications that are likely to succeed, before investing additional resources in performing an economic evaluation. Similar to other papers [e.g.,4,12,17,24-26], we found the data available for development to be a barrier to success in the CLL and diabetes case studies. Analytics for artificial intelligence are 'data hungry' and therefore require large datasets [11,27]. Furthermore, the quality of the data is an important issue when developing and using AI. Roberts et al. have emphasized in their review of AI for the diagnosis and prognostication of secondary pneumonia, that many AI analytics were hampered by poor quality data [64]. Our framework aligns with recommendations by Vollmer et al. who include critical questions regarding the data used as part of their framework to inform design and evaluate AI analytics [65]. Reviewing the data quality ensures developers select those applications of analytics for which development is most likely to succeed. For instance, rapid checks of potential sample sizes have been previously suggested [66]. For analytics with adequate data quality, additional resources can then be invested to perform an economic evaluation.

In this study, the framework was applied to three clinical use cases. Therefore, validation in other use cases is recommended. Other use cases can include different clinical areas (e.g., psychiatric disorders) but also other data sources such as data from patient devices (e.g., Fitbits), imaging and social media. Additional research could also assess criteria to value the quality of unstructured data. Furthermore, the framework presented could be easily adopted alongside initiatives such as RE-AIM used to translate research into practice [67]. This framework pays particular attention to the timing of economic evaluations intended to assist development considering relevant elements in the 'Reach', 'Effectiveness', 'Adoption', 'Implementation' and 'Maintenance' steps.

Since many factors can influence the successful implementation and adoption of analytics, we may have adopted a somewhat narrow approach by solely focusing on the value of economic evaluations to support developer decision-making. A wider form of decision support can be achieved through a broader evaluation of analytics, for instance using health technology assessment, which includes social, and ethical elements besides the health and economic impact [68]. Moreover, elicitation of stakeholder preferences such as patients and clinicians could ensure that potential barriers to development, acceptability and implementation are addressed [69].

In recent years, there has been an increased interest in the ethical challenges that we face relating to the adoption of artificial intelligence [70]. In this paper, we discuss that factors such as the risk of bias and small sample sizes, should be assessed at an early stage of development prior to performing an economic evaluation. Trocin et al emphasize the severity of the consequences of failing to do so. Some of the challenges relating to the data quality mentioned in this paper have also been emphasized by Trocin et al. Moreover, these authors also provide research questions that need to be answered to ensure the responsible adoption of AI related technologies [70]. Many answers to these questions could be very relevant for future improvements of the flowchart. Depending on the setting and type of analytics, for instance, the quality of the data can be assessed according to the risk of selection bias in the data [4,13], or the absence of ethnic variation in the data which could limit generalizability of machine learning models [4,17,28].

CONCLUSION

This is the first study providing recommendations on the use of economic evaluations to support development decisions of analytics for big data and artificial intelligence-based solutions. Many types of analytics can be developed within a specific clinical setting or disease or using a particular dataset. The framework presented in this study stimulates efficiency of development by selecting those applications worth further investment after assessing the feasibility of development and identifying critical barriers. For these applications, early economic evaluations can assist decision-making of analytics developers by estimating for instance requirements of effectiveness and the headroom for pricing, validation, and implementation.

REFERENCES

1 Morello L, Guglielmi G. US science agencies set to win big in budget deal. Nature; 2018;555(7698).
2 Banks, MA. Sizing up big data. Nat Med; 2020;26:5–6
3 Kisner J. Creating shareholder value with AI? Not so elementary, My Dear Watson. 2017.
4 Fröhlich H, Balling R, Beerenwinkel N, et al. From hype to reality: Data science enabling personalized medicine. BMC Med; 2018;16.
5 Furlow, B. ASCO announces funding for CancerLinQ Clinical Data Analysis Initiative. 2012; [Internet] https://www.cancertherapyadvisor.com/home/cancer-topics/general-oncology/asco-announces-funding-for-cancerlinq-clinical-data-analysis-initiative/. Accessed December 12, 2020
6 McLachlan S, Dube K, Johnson O, et al. A framework for analysing learning health systems: Are we removing the most impactful barriers?. Learn Health Syst; 2019 Oct;3(4):e10189.
7 Mandl KD, Kohane IS, McFadden D, et al. Scalable collaborative infrastructure for a learning healthcare system (SCILHS): architecture. J Am Med Inform Assoc; 2014 Jul 1;21(4):615-20.
8 Big Data to Knowledge. 2019; [Internet] https://commonfund.nih.gov/bd2k. Accessed December 17, 2020.
9 Agency for Healthcare Research and Quality. 2020; [Internet] https://www.ahrq.gov/funding/training-grants/lhs-k12.html. Accessed December 17, 2020
10 Sanchez-Pinto LN, Luo Y, Churpek MM. Big data and data science in critical care. Chest; 2018;154(5):1239–48.
11 Gutierrez G. Artificial intelligence in the intensive care unit. Crit Care. 2020 Dec;24:1-9
12 Hemingway H, Asselbergs FW, Danesh J, et al. Big data from electronic health records for early and late translational cardiovascular research: challenges and potential. Eur Heart J; 2017;39(16):1481–95.
13 Rumsfeld JS, Joynt KE, Maddox TM. Big data analytics to improve cardiovascular care: promise and challenges. Nat Rev Cardiol; 2016;13(6):350.
14 Bates DW, Saria S, Ohno-Machado L, et al. Big data in health care: Using analytics to identify and manage high-risk and high-cost patients. Health Aff (Millwood); 2014;33:1123–31.
15 Phillips KA, Trosman JR, Kelley RK, et al. Genomic sequencing: assessing the health care system, policy, and big-data implications. Health Aff (Millwood); 2014;33(7):1246-1253.
16 El Morr C, Ali-Hassan H. Analytics in Healthcare: A Practical Introduction. Springer; 2019 Jan 21.
17 Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. BMC Med; 2019 Dec 1;17(1):195.
18 Magrabi F, Ammenwerth E, McNair JB, et al. Artificial intelligence in clinical decision support: Challenges for evaluating AI and practical implications: A position paper from the IMIA Technology Assessment & Quality Development in Health Informatics Working Group and the EFMI Working Group for Assessment of Health Information Systems. Yearb Med Inform; 2019 Aug;28(1):128.
19 Morse, KE, Bagely, SC, Shah, NH. Estimate the hidden deployment cost of predictive models to improve patient care. Nat Med; 2020:26;18–19.
20 Budrionis A, Bellika JG. The learning healthcare system: where are we now? A systematic review. J Biomed Inform; 2016 Dec 1;64:87-92.
21 Mehta N, Pandit A. Concurrence of big data analytics and healthcare: A systematic review. Int J Med Inform; 2018;114:57–65.
22 Bakker L, Aarts J, Uyl-de Groot C, Redekop W. Economic evaluations of big data analytics for clinical decision-making: a scoping review. J Am Med Inform Assoc; 2020 Sep;27(9):1466-75.
23 Wolff J, Pauling J, Keck A, Baumbach J. The economic impact of artificial intelligence in health care: systematic review. J Med Internet Res; 2020;22(2):e16866.
24 Shilo S, Rossman H, Segal E. Axes of a revolution: challenges and promises of big data in healthcare. Nat Med; 2020;26(1):29-38.
25 Prosperi M, Min JS, Bian J, et al. Big data hurdles in precision medicine and precision public health. BMC Med Inform Decis Mak; 2018;18(1):139.
26 Beckmann JS, Lew D. Reconciling evidence-based medicine and precision medicine in the era of big data: challenges and opportunities Review. Genome Med; 2016;8:134.

27 Obermeyer Z, Emanuel EJ. Predicting the future—big data, machine learning, and clinical medicine. The New England journal of medicine; 2016 Sep 29;375(13):1216.
28 Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med; 2019 Jan;25(1):44-56
29 Goldstein BA, Navar AM, Pencina MJ, Ioannidis J. Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review. J Am Med Inform Assoc; 2017 Jan 1;24(1):198-208.
30 Farmer R, Mathur R, Bhaskaran K, Eastwood SV, Chaturvedi N, Smeeth L. Promises and pitfalls of electronic health record analysis. Diabetologia; 2018 Jun;61(6):1241-8.
31 Marsolo K, Margolis PA, Forrest CB, Colletti RB, Hutton JJ. A digital architecture for a network-based learning health system: integrating chronic care management, quality improvement, and research. eGEMs. 2015;3(1).
32 Bhandari RP, Feinstein AB, Huestis SE, et al. Pediatric-Collaborative Health Outcomes Information Registry (Peds-CHOIR): a learning health system to guide pediatric pain research and treatment. Pain; 2016 Sep;157(9):2033.
33 Feeley TW, Sledge GW, Levit L, Ganz PA. Improving the quality of cancer care in America through health information technology. J Am Med Inform Assoc; 2014 Sep 1;21(5):772-5.
34 Sculpher M, Drummond M, Buxton M. The iterative use of economic evaluation as part of the process of health technology assessment. J Health Serv Res Policy; 1997 Jan;2(1):26-30.
35 Annemans L, Genesté B, Jolain B. Early modelling for assessing health and economic outcomes of drug therapy. Value Health; 2000 Nov 1;3(6):427-34.
36 Pietzsch JB, Paté-Cornell ME. Early technology assessment of new medical devices. Int J Technol Assess Health Care; 2008;24(1):36–44.
37 Buisman LR, Rutten-van Mölken MPMH, Postmus D, et al. The early bird catches the worm: early cost-effectiveness analysis of new medical tests. Int J Technol Assess Health Care; 2016;32(1–2):46–53.
38 Ijzerman MJ, Steuten LMG. Early assessment of medical technologies to inform product development and market access. Appl Health Econ Health Policy; 2011;9(5):331–47.
39 Shah NH, Milstein A, Bagley SC. Making machine learning models clinically useful. JAMA; 2019 Oct 8;322(14):1351-2.
40 Simpson, L., Dr Lisa Simpson Interview, T. Foley and F. Fairmichael, Editors. 2015, The Learning Healthcare Project: Web.www.thelearninghealthcareproject.org
41 Wiens J, Saria S, Sendak M, et al. Do no harm: a roadmap for responsible machine learning for health care. Nat Med; 2019 Sep;25(9):1337-40.
42 Vickers AJ, Van Calster B, Steyerberg EW. Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. BMJ; 2016 Jan 25;352:i6.
43 Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Making; 2006 Nov;26(6):565-74.
44 Liu VX, Bates DW, Wiens J, Shah NH. The number needed to benefit: estimating the value of predictive analytics in healthcare. J Am Med Inform Assoc; 2019 Dec;26(12):1655-9.
45 Eddy DM, Hollingworth W, Caro JJ, Tsevat J, McDonald KM, Wong JB. Model transparency and validation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force–7. Med Decis Making; 2012 Sep;32(5):733-43.
46 Girling A, Lilford R, Cole A, Young T. Headroom approach to device development: Current and future directions. Int J Technol Assess Health Care. 2015;31(5):331-8.
47 Baliakas P, Mattsson M, Stamatopoulos K, Rosenquist R. Prognostic indices in chronic lymphocytic leukemia: where do we stand how do we proceed? J Intern Med; 2016 Apr;279(4):347-57.
48 Eichhorst B, Robak T, Montserrat E, et al. Chronic lymphocytic leukemia: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol; 2020. In Press.
49 Condoluci A, Terzi di Bergamo L, Langerbeins P, et al. International prognostic score for asymptomatic early-stage chronic lymphocytic leukemia. Blood. 2020 May 21;135(21):1859-69.
50 Langerbeins P, Bahlo J, Rhein C, et al. Ibrutinib versus placebo in patients with asymptomatic, treatment-naïve early stage CLL: Primary endpoint results of the phase 3 double-blind randomized CLL12 trial. Hematol Oncol. 2019 Jun;37:38-40

51 Blot SI, Depuydt P, Annemans L, et al. Clinical and economic outcomes in critically ill patients with nosocomial catheter-related bloodstream infections. Clin Infect Dis; 2005;41(11):1591-1598.

52 Blot S, Poulakou G, Timsit JF. Catheter-associated bloodstream infection rates: how low can you go? Intensive Care Med; 2019;45(6):896-897

53 GiViT, Gruppo Italiano per la Valutazione degli Interventi In Terapia Intensiva. Report PROSAFE project. 2014. 2014; Centre GR001.

54 Apostolopoulou E, Raftopoulos V, Filntisis G, et al. Surveillance of device-associated infection rates and mortality in 3 greek intensive care units. Am J Crit Care; 2013;22(3):e12-20.

55 Ferrer R, Martin-Loeches I, Phillips G, et al. Empiric antibiotic treatment reduces mortality in severe sepsis and septic shock from the first hour: Results from a guideline-based performance improvement program. Crit Care Med. 2014;42(8):1749-1755.

56 De Wit M, Miller KB, Green DA, Ostman HE, Gennings C, Epstein SK. Ineffective triggering predicts increased duration of mechanical ventilation. Crit Care Med; 2009 Oct 1;37(10):2740-5.

57 Georgopoulos D. Ineffective efforts during mechanical ventilation: the brain wants, the machine declines. Intensive Care Med; 2012; 38: 738–740

58 Bakker L, Vaporidi K, Aarts J, Redekop W. The potential of real-time analytics to improve care for mechanically ventilated patients in the intensive care unit: an early economic evaluation. Cost Eff Resour Alloc; 2020 Dec;18(1):1-1.

59 National Institute for Health and Care Excellence. Evidence Standards Framework for Digital Health Technologies; NHS England 2019

60 Zeitoun JD, Ravaud P. Artificial intelligence in health care: value for whom?. Lancet Digit Health; 2020 Jul 1;2(7):e338-9.

61 Riddle MC, Blonde L, Gerstein HC, et al. Diabetes Care Editors' Expert Forum 2018: managing big data for diabetes research and care.

62 Newton KM, Peissig PL, Kho AN, et al. Validation of electronic medical record-based phenotyping algorithms: results and lessons learned from the eMERGE network. J Am Med Inform Assoc; 20(e1), e147-e154.

63 Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. Methods for the economic evaluation of health care programmes. Oxford university press; 2015 Sep 25.

64 Roberts M, Driggs D, Thorpe M, et al. Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans. Nat Mach Intell; 2021 Mar;3(3):199-217.

65 Vollmer S, Mateen BA, Bohner G, et al. Machine learning and artificial intelligence research for patient benefit: 20 critical questions on transparency, replicability, ethics, and effectiveness. BMJ; 2020 Mar 20;368.

66 Wallace PJ, Shah ND, Dennen T, Bleicher PA, Crown WH. Optum Labs: building a novel node in the learning health care system. Health Aff (Millwood); 2014 Jul 1;33(7):1187-94.

67 RE-AIM. 2021; [Internet] https://www.re-aim.org/ Accessed June 30,2021

68 Kristensen FB, Lampe K, Chase DL, et al. Practical tools and methods for health technology assessment in Europe: Structures, methodologies, and tools developed by the European network for Health Technology Assessment, EUnetHTA. Int J Technol Assess Health Care. 2009 Dec;25(S2):1-8.

69 Salloum RG, Shenkman EA, Louviere JJ, Chambers DA. Application of discrete choice experiments to enhance stakeholder engagement as a strategy for advancing implementation: a systematic review. Implement Sci. 2017 Dec;12(1):1-2.

70 Trocin C, Mikalef P, Papamitsiou Z, Conboy K. Responsible AI for digital health: a synthesis and a research agenda. Inf Syst Front. 2021 Jun 26:1-9.

APPENDIX

Model & Input parameters case study 1: Chronic lymphocytic leukemia

Several risk scores for newly diagnosed chronic lymphocytic leukemia (CLL) patients are available that combine clinical, laboratory and/or molecular data to stratify patients according to risk of progression and time to treatment [1,2,3]. After diagnosis, patients without clinically active disease are monitored during frequent follow-up visits also referred to as the 'watch & wait' phase. During the watch and wait phase, prognostic scores are currently used to identify patients with a higher risk of progressing eligible for enrollment in clinical trials and to personalize their frequency of follow-up. Even though at present, no early treatment is prescribed to patients at higher risk of progressing, preliminary results from the CLL-12 study suggest that some patients without active disease might benefit from early treatment with ibrutinib [4].

In light of these recent results, we estimated the potential cost-effectiveness of using next generation sequencing data to improve prognostic algorithms for assessing the risk of progressing to needing treatment of patients diagnosed with CLL. Even though development to improve available risk scores is recommended [1], the potential health and economic benefits of using them have not yet been assessed. Prior to continuing development of analytics, it can be estimated whether further research into this area could be considered a worthwhile investment given this novel treatment available. The patient population consisted of patients newly diagnosed with CLL in a Swedish healthcare setting. Currently, these patients can be classified as high, intermediate or low-risk by assessing their unmutated immunoglobulin heavy variable gene status (IGHV), the absolute lymphocyte count, and the presence of palpable lymph nodes [1]. These patients are followed through a watch & wait strategy in which they receive frequent follow-up visits but do not receive treatment until the disease becomes clinically active. At present, no additional genomic or genetic testing is performed. In care with the novel analytics, those with a high and intermediate risk score would receive early treatment with ibrutinib.

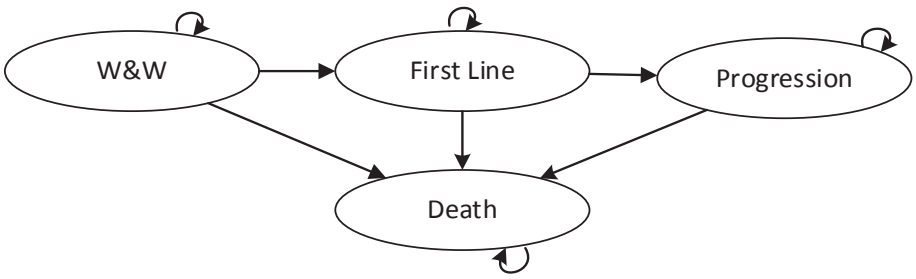


Figure S1: Markov model used to estimate costs and effects of using analytics to estimate the risk of progression of watch and wait patients with chronic lymphocytic leukemia compared to current care. W&W= watch and wait patients

We used a Markov model with 4 states (Figure S1) to estimate costs, life years gained, and quality adjusted life years gained (QALYs). Tunnel states in the first-line health state were used to vary costs according to the respective time on treatment for the different treatments prescribed. A lifetime time horizon was adopted, and the cycle length was 28 days. Even though

a societal perspective has been recommended in Swedish guidelines for performing economic evaluations, this is not reflected in recent reimbursement decisions for CLL treatments. Here, cost-effectiveness was assessed by the TLV, The Dental and Pharmaceutical Benefits Agency in Sweden that decides on reimbursement decisions, without considering non-medical costs [5]. Therefore, we adopted a healthcare payer perspective including only direct medical costs. A discount rate of 3% was used for both costs and effects.

Transition Probabilities

Time to first treatment in current care was estimated using the survival curves presented in the supplemental figures of Condoluci et al [1]. Individual patient data was reconstructed according to Guyot et al. using Digizeit [6]. Care with the analytics assumed perfect stratification of patients where those progressing within 3 years were considered high risk, those progressing within 3-7 years are intermediate risk and those progressing after 7 years would be considered low risk. Background mortality in Sweden was used for the transition probabilities from watch and wait to death and first-line to death [7,8]. Sylvan et al. found that for 80% of patients in Sweden, treatments were prescribed in accordance with national guidelines [9]. Therefore, the first-line treatment in current care depended on the prevalence of IGHV mutations, Tp53 mutations and age and fitness of patients, in accordance with Swedish guidelines [10]. For the treatments prescribed in first- and second-line, the probability of requiring novel treatment over the first 24 months was derived from the time to next treatment curves and converted to rates to estimate probabilities in accordance with a 28-day cycle length. For second-line treatment in current care, overall survival curves were used to estimate the 28-day transition probability of death.

Utilities

Utility of watch and wait patients was derived from a study by Holtzer-Goor et al. [11] while utility values from Kosmas et al. [12] were used for utility of watch and wait patients receiving oral treatment and for patients with progressive disease.

Costs

When available, costs were based on estimates from studies and reports for the Swedish health care setting. The majority of unit costs for treatment in the first-line were obtained from a recent Swedish drug approval report for venetoclax from the Swedish HTA organization [5]. Based on recommendations from the Svenska KLL Gruppen, it was assumed that in the progression state in current care 50% of patients received Ibrutinib, 20% received treatment with FCR and 30% received monotherapy with venetoclax [10]. Costs of progression in the intervention arm (€1,572) were based on rituximab treatment since patients have already received ibrutinib and venetoclax. Costs of analytics were obtained from a micro-costing study performed by Swarzche et al reporting the costs of genomic testing [13]. All costs were reported in 2019 euros.

Cost-effectiveness analysis

The base case input parameters were used to estimate the incremental cost effectiveness ratio. The incremental cost-effectiveness ratio is the incremental costs of the novel intervention compared to current care divided by the incremental effects. Hereafter, input parameters were varied extensively in univariate sensitivity analyses and scenario analyses. All parameters were varied simultaneously in the probabilistic sensitivity analysis using a beta distribution for probabilities and utilities and a gamma distribution for costs. R v3.6.3 was used for the model according to best practice modelling recommendations [14].

Table S1: Input parameters used to estimate costs and effects of using analytics to estimate the risk of progression of watch and wait patients with chronic lymphocytic leukemia compared to current care. The values for all input parameters were obtained from the literature.

Parameter	Base case	Distribution	Parameters	Lower	Upper	Source
Probabilities						
Patients classified in current care		Dirichlet	Shape=0.24, 0.38, 0.38			[1]
Low risk	0.24					
Intermediate risk	0.38					
High risk	0.38					
Patients classified with analytics		Dirichlet	Shape= 0.57, 0.17, 0.26			Perfect stratification assumed
Low risk	0.57					
Intermediate risk	0.17					
High risk	0.26					
HR Time to Next Treatment Ibrutinib in W&W vs placebo	0.21	Betapert	upper=0.39, lower=0.11, mode=0.21	0.11	0.39	[4]
TP 1st-line FCR	0.0050 Based on time to next treatment. Included as the first-line in current care for 24% of patients with mutated IGHV, fit, young patients and 8% of patients with unmutated IGHV, fit, young patients	Beta	se=0.0002	0.0044	0.0054	[1,15,16]
TP 1st-line ChlO	0.0130 Based on time to next treatment. Included as the first-line in current care for 44% of patients mutated IGHV, elderly, unfit patients	Beta	se=0.0007	0.012	0.014	[1,16,17]
TP 1st-line VO	0.0067 Based on time to next treatment. Included as the first-line in current care for 15% of patients in current care with unmutated IGHV, elderly, unfit patients	Beta	se=0.0003	0.0061	0.0074	[1,16, 18]

TP 1st-line Ibrutinib	0.0061 Based on time to next treatment. Included as the first-line in current care for 9% of patients current care with 17pdel / TP53 mutation	Beta	se=0.0003	0.0056	0.0068	[10,19,20]
TP 1st-line VR	0.0056 Based on time to next treatment. Assumed applicable to all patients. First-line for all patients in intervention arm	Beta	se=0.0003	0.0050	0.0061	[21,22]
TP progression current care	0.0107 Based on 50% of patients treated with ibrutinib, 30% of patients treated with venetoclax and 20% of patients treated with FCR	Beta	se=0.0005	0.0097	0.0118	[10,23,24,25]
TP progression intervention	0.075	Beta	se=0.0038	0.068	0.083	[26]
Utilities						
Utility W&W	0.81	Beta	se=0.04	0.72	0.88	[11]
Utility W&W on treatment	0.71	Beta	se=0.04	0.64	0.78	[12]
Utility 1st line	0.71	Beta	se=0.04	0.64	0.78	[12]
Utility Progression	0.66	Beta	se=0.03	0.59	0.72	[12]

Table S1: *Continued.*

Parameter	Base case	Distribution	Parameters	Lower	Upper	Source
Costs						
Costs of next generation sequencing as diagnostic tool	€7439 per case (€3719 per genome)	Gamma	Se=€1960	€100	€11,502	[13]
Watch & wait health state cycle 1 to 6	€45 Based on a consult with a hematologist and laboratory tests during the consult 6 months after diagnosis and a complete blood count every 2 months.	Gamma	Se=€11	€26	€69	[27,28, expert opinion]
Watch & wait health state remaining cycles	€22 Based on a consult with a hematologist and laboratory tests during the consult every 15 months and a complete blood count every 2 months.	Gamma	Se=€5	€13	€34	[27,28, expert opinion]
Fludarabine	€271	Gamma	Se=€68	€155	€418	[5]
Cyclophosphamide	€8	Gamma	Se=€2	€4	€12	[5]
Rituxumab	€773	Gamma	Se=€193	€442	€1195	[5]
Chlorambucil	€66	Gamma	Se=€16	€38	€102	[5]
Obinituzumab	€2393	Gamma	Se=€598	€1368	€3700	[5]
Ibrutinib	€3,813	Gamma	Se=953	€2180	€5,897	[5]
Venetoclax	€4096	Gamma	Se=1024	€2341	€6334	[5]
Adverse Events FCR	€643	Gamma	se=€161	€367	€994	[5]
Adverse Events ChlO	€304	Gamma	Se=€76	€174	€470	[5]
Adverse Events VO	€733	Gamma	Se=€183	€419	€1133	[5]

Adverse Events Ibrutinib	€804	Gamma	Se=€201	€459	€1243	[5]
Administration costs	€386	Gamma	Se=€97	€221	€598	[5]
TLS prevention	€2009	Gamma	Se=€502	€1149	€3107	[5]
Parameter	**Base case**	**Distribution**	**Parameters**	**Lower**	**Upper**	**Source**
Progression Current Care	€2,579 Based on 50% of patients treated with ibrutinib, 30% of patients treated with venetoclax and 20% of patients treated with FCR and follow-up costs.	Gamma	Se=€645	€1,474	€3,988	[10,23,24,25,29]
Progression Intervention	€ 1,572 Based on all patients receiving monotherapy with Rituxumab and follow-up costs.	Gamma	Se=393	€899	€2,431	[29,30]
Follow-up first-line	€145	Gamma	Se=€36	€83	€224	[29]
Follow-up second-line	€780	Gamma	Se=195	€446	€1207	[29,30]

HR= Hazard Ration, TP= Transition Probability, W&W= Watch and Wait, FCR= Fludarabine + Cyclophosphamide+ Rituximab, ChlO= Chlorambucil + Obinutuzumab, VO= Venetoclax + Obinutuzumab, VR= Venetoclax + Rituximab, TLS= Tumor Lysis Syndrome

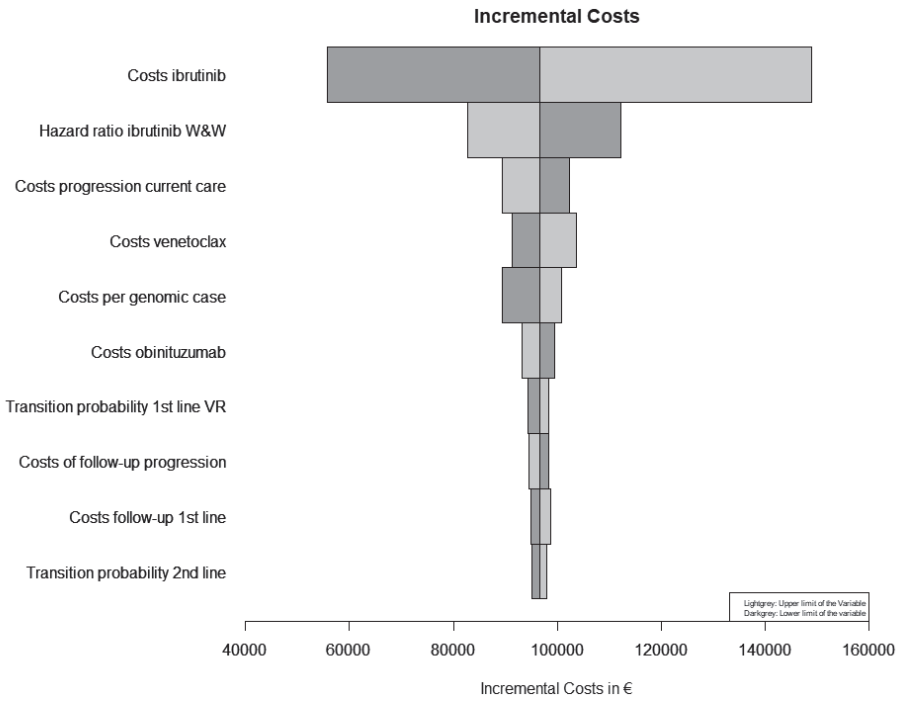


Figure S2: Tornado diagram for incremental costs when using analytics to estimate the risk of progression of watch and wait patients with chronic lymphocytic leukemia compared to current care. W&W= watch & wait patients, VR= venetoclax-rituximab.

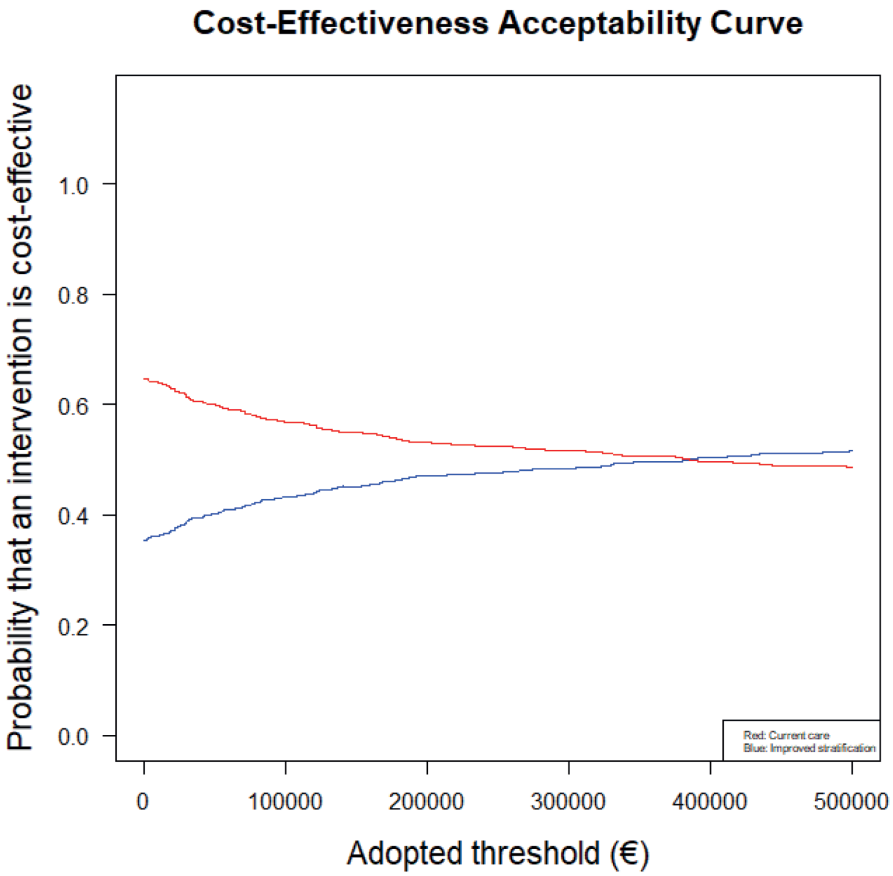


Figure S3: Cost-effectiveness acceptability curve with on the X-axis a range of willingness-to-pay thresholds and on the Y-axis the probability that the analytics and subsequent treatment with Ibrutinib would be cost-effective.

Model & Input parameters case study 2: Catheter related bloodstream infection

Infections are a recurring issue in the intensive care unit and can result in sepsis and septic shock. A common cause of these infections in Greece is placement of a central venous catheter [31]. Catheter related bloodstream infections (CRBSI) have been associated with increased mortality, prolonged intensive care unit (ICU) and hospital stay and prolonged mechanical ventilation [31,32]. Furthermore, CRBSIs are relatively easy to avoid by timely identification and infection control [33].

A decision tree model was combined with a four state Markov model to estimate the health and economic benefits of using analytics for earlier detection of CRBSI compared to current care in patients admitted to the ICU (Figure S4). In the decision tree, patients could receive care where CRBSI is diagnosed early with novel analytics (Arms 1- Arm7) or current care in which CRBSI was diagnosed according to clinical symptoms (Arm 8 -Arm 14). The diagnosis of CRBSI in current care is primarily based on clinical symptoms whereas when using analytics, CRBSI would be

diagnosed in real-time. After diagnosis, the interventions for current care and care with analytics are identical; the catheter is replaced, and antibiotics are administered. In care with analytics, this intervention would be administered earlier due to a timely diagnosis. Patients falsely classified as having CRBSI would have their catheter replaced and receive antibiotics unnecessarily. False negatives would progress to having clear clinical symptoms resulting in a delayed diagnosis for which outcomes were assumed to be identical to patients in current care. When replacing the catheter, a subset of patients experienced complications such as pneumothorax, hematoma, and arterial puncture.

The decision tree ended in a Markov model in which all patients transitioned from ICU care to the general ward and were then discharged from the hospital. Cycle length for both of these states was identical to the median length of stay. For the post discharge state, a yearly cycle length was adopted. The possibility to return from discharge to the hospital ward state was not included given that there is no conclusive evidence that 30-day readmission rates are higher in patients with CRBSI [34]. Relevant decision-makers were hospital employees such as clinicians and budget managers in Greece. With these decision-makers in mind relevant health outcomes modelled were ICU length of stay, mortality, life years gained, and quality-adjusted life years gained (QALYs). Furthermore, a healthcare payer perspective was adopted including only direct medical costs. A discount rate of 3.5% was used for both costs and effects since national guidelines for performing economic evaluations in Greece are lacking. All analyses were performed using R v3.6.3.

Probabilities

All input parameters can be found in Table S2. The prevalence of CRBSI was based on an earlier report for the collaborating ICU in Greece (7.6%) [35]. Uncertainty surrounding these estimates was based on the large variation reported in the literature (0.5%-29%) [31,32]. In current care, the diagnosis is made according to clinical symptoms and the accuracy of the diagnosis in current care is uncertain. Sensitivity was high given that the diagnosis was made at a late stage at which clinical symptoms were clearly present whereas specificity was much lower. Sensitivity and specificity of the analytics were considered to be at least as good as algorithms already available in the literature (sensitivity= 85%, specificity= 83% [36,37]).

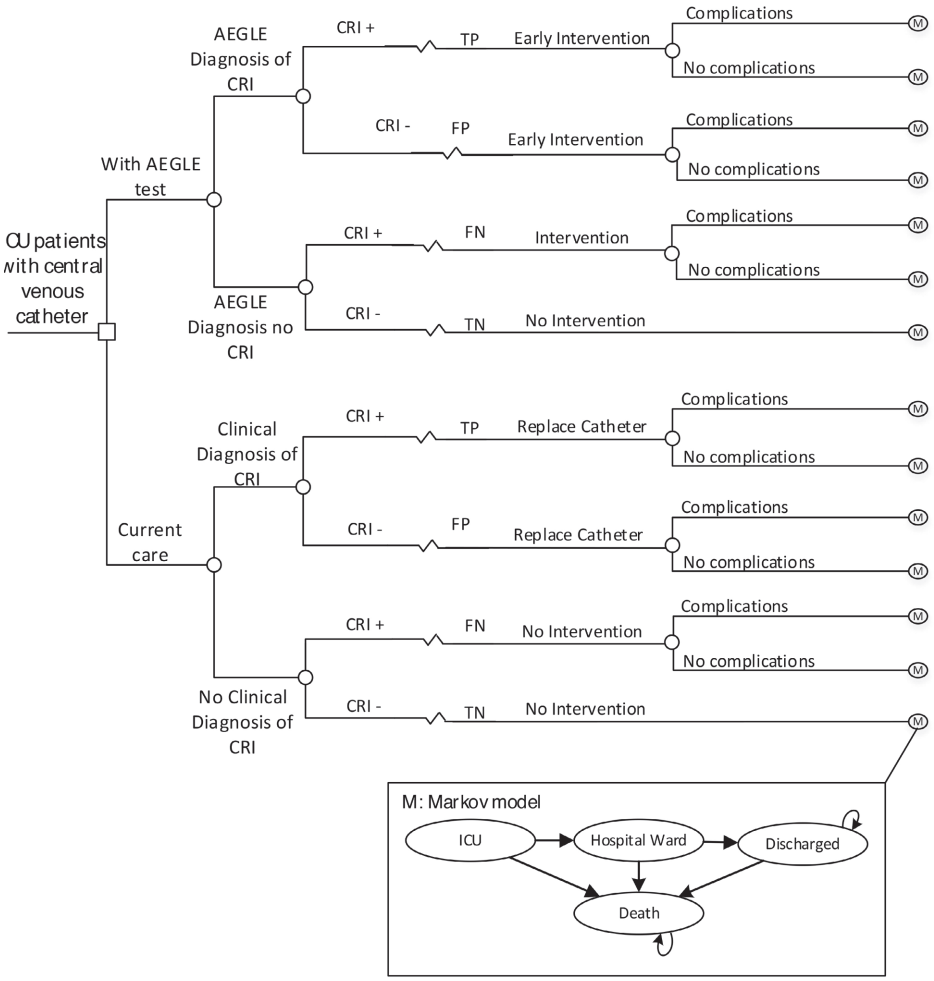


Figure S4: Decision tree and Markov model used to estimate costs and effects of using analytics for earlier identification of catheter related bloodstream infection compared to current care. CRI=Catheter related Infection, TP= True Positive, FP= False Positive, FN=False Negative, TN= True Negative, ICU= Intensive Care Unit.

The probabilities of ICU and hospital mortality were obtained from a large multicenter study following patients with CRBSI (28.5%) and without CRBSI (19.6%) [38]. We adopted a conservative approach in which we assumed that the analytics would result in earlier diagnosis and thus less severe outcomes instead of avoiding the CRBSI event altogether. We assumed that mortality with timely antibiotics resulted in a relative risk reduction in mortality (0.74) similar to the impact of early administration of antibiotics reported by Ferrer et al [39]. For survival after hospital discharge, the hazard ratio of dying after an ICU stay without CRBSI [40] and with CRBSI [41] were combined with survival in the Greek population [42,43]. The average incidence of

complications across puncture sites was used from a Greek study [44] since all sites were used in the hospital in question.

Utilities

Utility estimates were obtained from the literature. No research is available reporting quality of life of patients during their ICU stay. Therefore, we adopted a utility estimate of 0.30 which corresponds to an EQ-5D state of extreme problems with selfcare, mobility and usual activities but no pain or discomfort assuming sedation was adequate. Quality of life after discharge was varied according to the time since hospital discharge and the mean age of the Greek patient population [45].

Unit costs and resource use

Length of stay in the ICU for patients with CRBSI (13) and without CRBSI (3) were obtained from Vught et al [38]. We assumed that ICU length of stay reduced with 24% with an early intervention, similar to the effect of early administration of antibiotics reported by Ferrer et al [39]. Length of stay after ICU discharge (15 days) was obtained from Vught et al. [38] and assumed to be identical for the intervention and current care. Patients with CRBSI in the model received treatment with antibiotics for 10.5 days [46]. This estimate was varied from 7-14 days. Costs of catheter replacement were based on a duration of change of 10 minutes. For the base case analysis, annual licensing costs per bed were included for the analytics (€959) [47]. Daily costs of antibiotics were derived from the literature [46] and unit costs of an ICU and hospital day were derived from Greek micro-costing studies [48,49]. Costs were reported in 2019 euros.

Analyses

For the base case estimate, incremental costs, length of stay, mortality, life years gained, QALYs and the incremental costs-effectiveness ratio were reported. Base case estimates were varied in univariate and probabilistic sensitivity analyses. In the probabilistic sensitivity analyses, all parameters were varied simultaneously except for the costs of the analytics. Underlying distributions adopted for probabilities were the beta and beta pert distribution. For costs and resource use, the gamma and beta pert distribution were used. We also estimated the headroom according to the following formula:

$$Headroom = N + \lambda * Q$$

Here N refers to the potential savings where the costs of the technology are set to zero, λ is the willingness-to-pay threshold and Q are the health effects gained [50]. Willingness-to-pay thresholds used were €4,946, €7,758 [51] and €30,000 [52,53]. We assumed patients occupied the bed for 7.4 days on average [35] and that the analytics should be functional for at least three years. At least 49 patients would be using the analytics each year. Costs of implementation were obtained from a systematic review that reported cost estimates for developing and implementing clinical decision support systems in EHRs for diabetes [54]. Costs of validation were based on recommendations reported by Calster et al. [55] for an ICU with 13 beds and a validation study including 100 patients.

Table S2: Input parameters used to estimate costs and effects of using analytics for earlier identification of catheter related bloodstream infection compared to current care. The values for the input parameters were obtained from the literature and if no evidence was available through discussions with experts and assumptions.

Parameter	Base Case estimate	Distribution	PSA	Lowest estimate	Highest estimate	Source
Probabilities						
Prevalence of CRBSI	7.6%	Beta pert	Min=0.5%, Max=29%, Mode=4.025%	0.5%	29%	[31,32,35]
Sensitivity CRBSI diagnosis Current Care	100%					Assumption
Specificity CRBSI diagnosis Current Care	60%	Beta pert	Min=40%, Max=100%, Mode=55%	40%	100%	Assumption
Sensitivity analytics	85%	Beta pert	Min=75%, Max=100%, Mode=84%	75%	100%	[36,37]
Specificity analytics	83%	Beta pert	Min=63%, Max=100%, Mode=84%	63%	100%	[36,37]
30-day mortality without CRBSI	19.6%	Beta	s.e.=1.96	15.9%	23.6%	[38]
30-day mortality with CRBSI	28.5%	Beta	s.e.=2.85	23.1%	34.2%	[38]
Relative risk of mortality with early intervention	0.74	Normal	s.e.=0.007	60%	100%	[38,39]
Hazard ratio of mortality after discharge sepsis vs. no sepsis	1.39	Normal	s.e.=0.07	1.26	1.52	[41]
Hazard ratio for survival after ICU discharge	2.01	Normal	s.e.=0.1	1.64	2.46	[40]
Incidence of arterial puncture	6%	Beta pert	Min=4.95% Max=7.75% Mode=6%	4.95%	7.75%	[44]
Incidence of hematoma	2%	Beta pert	Min=1.28% Max=2.73% Mode=2%	1.28%	2.73%	[44]
Incidence of pneumothorax	0.5%	Beta pert	Min=0.15% Max=0.85% Mode=0.5%	0.15%	0.85%	[44]

Table S2: *Continued.*

Parameter	Base Case estimate	Distribution	PSA	Lowest estimate	Highest estimate	Source
Utilities						
Quality of Life ICU	0.30	Beta	s.e.=0.03	0.24	0.36	Assumption
Quality of Life hospital	0.60	Beta	s.e.=0.06	0.48	0.71	[56]
Quality of Life First 5 years after discharge	0.67	Beta	s.e.=0.023	0.62	0.71	[45]
Quality of Life 5-10 years after discharge	0.70	Beta	s.e.=0.025	0.65	0.75	[45]
Quality of Life >10 years after discharge	0.68	Beta	s.e.=0.031	0.62	0.74	[45]
Unit costs (2019 Euros)						
Analytics (annual)	€959			€100	€20.000	[47]
ICU day	€670.4	Gamma	s.e.=335.2	€565.9	€1,469.5	[48,57,58]
Hospital day	€297.6	Gamma	s.e.=148.8	€81.1	€652.2	[49]
Antibiotics for CRBSI per day	€114.4	Beta pert	Min=€85.2 Max=€137.4 Mode=€114.4	€85.2	€137.4	[46]
Catheter replacement	€17.7	Gamma	s.e.=€8.9	4.8	38.9	[58]
Catheter	€12. 6	Gamma	s.e.=€6.3	3.4	27.6	[59]
Treatment of arterial puncture	€10.1	Gamma	s.e.=€5.0	2.7	22.1	[60]
Treatment of hematoma	€0	Beta pert	Min= €0 Max=€50 Mode=€0	0	50	[61]
Treatment of pneumothorax	€96.1	Gamma	s.e.=€48.0	26.2	210.6	[61]
Resource Use						
Duration of infection (days)	10.5	Beta pert	Min=7 Max=14 Mode=10.5	7	14	[46]
LOS ICU without CRBSI	3	Gamma	s.e.=0.6	1.9	4.3	[38]
LOS ICU with CRBSI	13	Gamma	s.e.=2.6	8.4	18.6	[38]
Relative change in ICU LOS with intervention	0.76	Normal	Se=0.08	61%	100%	[39]
LOS hospital after ICU discharge	15	Gamma	s.e.=3	9.7	21.43	[35,38]

REFERENCES

1 Condoluci A, Terzi di Bergamo L, Langerbeins P, et al. International prognostic score for asymptomatic early-stage chronic lymphocytic leukemia. Blood. 2020 May 21;135(21):1859-69.
2 Pflug N, Bahlo J, Shanafelt TD, et al. Development of a comprehensive prognostic index for patients with chronic lymphocytic leukemia. Blood. 2014 Jul 3;124(1):49-62.
3 International CLL-IPI Working Group. An international prognostic index for patients with chronic lymphocytic leukaemia (CLL-IPI): a meta-analysis of individual patient data. Lancet Oncol. 2016 Jun 1;17(6):779-90.
4 Langerbeins P, Bahlo J, Rhein C, et al. Ibrutinib versus placebo in patients with asymptomatic, treatment-naïve early stage CLL: primary endpoint results of the phase 3 double-blind randomized CLL12 trial. Hematol Oncol. 2019 Jun;37:38-40.
5 TLV The Dental and Pharmaceuticals Benefits Agency. 2020 [Internet] https://www.tlv.se/download/18.2b305b2817457f407a2a598b/1599732770817/bes200827_underlag_venclyxto.pdf Accessed January 26, 2020
6 Guyot P, Ades AE, Ouwens MJ, Welton NJ. Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan-Meier survival curves. BMC Med Res Methodol. 2012 Dec;12(1):1-3.
7 WHO mortality database: Query the online database. [Internet] http://apps.who.int/healthinfo/statistics/mortality/causeofdeath_query/start.php. Updated 20172016. Accessed February 21, 2017.
8 Population pyramids of the world from 1950 to 2100. [Internet] https://www.populationpyramid.net/sweden/2016/. Accessed February 21, 2017.
9 Sylvan SE, Asklid A, Johansson H, et al. First-line therapy in chronic lymphocytic leukemia: a Swedish nation-wide real-world study on 1053 consecutive patients treated between 2007 and 2013. Haematologica. 2019 Apr;104(4):797.
10 Svenska KLL gruppen: 2020 [Internet] https://kunskapsbanken.cancercentrum.se/diagnoser/kll/ Accessed January 12, 2020
11 Holtzer-Goor KM, Schaafsma MR, Joosten P, et al. Quality of life of patients with chronic lymphocytic leukaemia in the Netherlands: results of a longitudinal multicentre study. Qual Life Res. 2015 Dec;24(12):2895-906.
12 Kosmas CE, Shingler SL, Samanta K, et al. Health state utilities for chronic lymphocytic leukemia: importance of prolonging progression-free survival. Leuk Lymphoma. 2015 May 4;56(5):1320-6.
13 Schwarze K, Buchanan J, Fermont JM, et al. The complete costs of genome sequencing: a microcosting study in cancer and rare diseases from a single center in the United Kingdom. Genet Med. 2020 Jan;22(1):85-94.
14 Krijkamp EM, Alarid-Escudero F, Enns EA, et al. A multidimensional array representation of state-transition model dynamics. Med Decis Making. 2020 Feb;40(2):242-8.
15 Kutsch N, Bahlo J, Robrecht S, et al. Long Term Follow-up Data and Health-Related Quality of Life in Frontline Therapy of Fit Patients Treated With FCR Versus BR (CLL10 Trial of the GCLLSG). HemaSphere. 2020 Feb;4(1).
16 Jeyakumaran D, Kempel A, Côté S. An assessment of the number of chronic lymphocytic leukemia (CLL) patients eligible for front-line treatment but unsuitable for full-dose fludarabine across the European Union. Value Health. 2016 Nov 1;19(7):A574-5.
17 Goede V, Fischer K, Busch R, et al. Obinutuzumab plus chlorambucil in patients with CLL and coexisting conditions. N Engl J Med. 2014 Mar 20;370(12):1101-10.
18 Fischer K, Al-Sawaf O, Bahlo J, et al. Venetoclax and obinutuzumab in patients with CLL and coexisting conditions. N Engl J Med. 2019 Jun 6;380(23):2225-36.
19 Emond B, Sundaram M, Romdhani H, Lefebvre P, Wang S, Mato A. Comparison of Time to Next Treatment, Health Care Resource Utilization, and Costs in Patients with Chronic Lymphocytic Leukemia Initiated on Front-line Ibrutinib or Chemoimmunotherapy. Clin Lymphoma Myeloma Leuk. 2019 Dec 1;19(12):763-75.
20 Campo E, Cymbalista F, Ghia P, et al. TP53 aberrations in chronic lymphocytic leukemia: an overview of the clinical implications of improved diagnostics. Haematologica. 2018 Dec;103(12):1956.
21 Seymour JF, Kipps TJ, Eichhorst B, et al. Venetoclax–rituximab in relapsed or refractory chronic lymphocytic leukemia. N Engl J Med. 2018 Mar 22;378(12):1107-20

22 Eichhorst B, Robak T, Montserrat E, et al. Chronic lymphocytic leukemia: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol. 2020. In Press.
23 Munir T, Brown JR, O'Brien S, et al. Final analysis from RESONATE: Up to six years of follow-up on ibrutinib in patients with previously treated chronic lymphocytic leukemia or small lymphocytic lymphoma. Am J Hematol. 2019 Dec;94(12):1353-63.
24 Hallek M, Fischer K, Fingerle-Rowson G, et al. Addition of rituximab to fludarabine and cyclophosphamide in patients with chronic lymphocytic leukaemia: a randomised, open-label, phase 3 trial. Lancet. 2010 Oct 2;376(9747):1164-74.
25 Jones J, Mato A, Coutre S, et al. Evaluation of 230 patients with relapsed/refractory deletion 17p chronic lymphocytic leukaemia treated with ibrutinib from 3 clinical trials. Br J Haematol. 2018 Aug;182(4):504-12.
26 Buchanan J, Wordsworth S, Clifford R, et al. Using genomic information to guide ibrutinib treatment decisions in chronic lymphocytic leukaemia: a cost-effectiveness analysis. Pharmacoeconomics. 2017 Aug;35(8):845-58.
27 Ghatnekar O, Hjalte F, Taylor M. Cost-effectiveness of dasatinib versus high-dose imatinib in patients with Chronic Myeloid Leukemia (CML), resistant to standard dose imatinib–a Swedish model application. Acta Oncol. 2010 Aug 1;49(6):851-8.
28 Glenngård AH, Persson U, Söderman C. Costs associated with blood transfusions in Sweden–the societal cost of autologous, allogeneic and perioperative RBC transfusion. Transfus Med. 2005 Aug;15(4):295-306.
29 Sorensen SV, Peng S, Dorman E, et al. The cost-effectiveness of ibrutinib in treatment of relapsed or refractory chronic lymphocytic leukemia. Health Econ Outcome Res. 2016;2(4):1-9.
30 Furman RR, Sharman JP, Coutre SE et al. Idelalisib and rituximab in relapsed chronic lymphocytic leukemia. N Engl J Med. 2014 Mar 13;370(11):997-1007.
31 Apostolopoulou E, Raftopoulos V, Filntisis G, et al. Surveillance of device-associated infection rates and mortality in 3 greek intensive care units. Am J Crit Care. 2013;22(3):e12-20.
32 Blot SI, Depuydt P, Annemans L, et al. Clinical and economic outcomes in critically ill patients with nosocomial catheter-related bloodstream infections. Clin Infect Dis. 2005;41(11):1591-1598.
33 Blot S, Poulakou G, Timsit JF. Catheter-associated bloodstream infection rates: how low can you go? Intensive Care Med. 2019 Jun;45(6):896-7.
34 Stevens V, Geiger K, Concannon C, Nelson R, Brown J, Dumyati G. Inpatient costs, mortality and 30-day re-admission in patients with central-line-associated bloodstream infections. Clin Microbiol Infect. 2014;20(5).
35 GiViT, Gruppo Italiano per la Valutazione degli Interventi In Terapia Intensiva. Report PROSAFE project; 2014. Report nr Centre GR001.
36 Henry KE, Hager DN, Pronovost PJ, Saria S. A targeted real-time early warning score (TREWScore) for septic shock. Sci Transl Med. 2015;7(299):299ra122.
37 Gardner-Thorpe J, Love N, Wrightson J, Walsh S, Keeling N. The value of modified early warning score (MEWS) in surgical in-patients: A prospective observational study. Ann R Coll Surg Engl. 2006;88(6):571-575.
38 van Vught LA, Klouwenberg PMK, Spitoni C, et al. Incidence, risk factors, and attributable mortality of secondary infections in the intensive care unit after admission for sepsis. JAMA. 2016;315(14):1469-1479.
39 Ferrer R, Martin-Loeches I, Phillips G, et al. Empiric antibiotic treatment reduces mortality in severe sepsis and septic shock from the first hour: Results from a guideline-based performance improvement program. Crit Care Med. 2014;42(8):1749-1755.
40 Williams TA, Dobb GJ, Finn JC, et al. Determinants of long-term survival after intensive care. Crit Care Med 2008; 36:1523-30.
41 Dick A, Liu H, Zwanziger J, et al. Long-term survival and healthcare utilization outcomes attributable to sepsis and pneumonia. BMC Health Serv Res. 2012;12(1):1.
42 World Health Organization: WHO Mortality Database: Query the Online Database. [Internet] http://apps.who.int/healthinfo/statistics/mortality/causeofdeath_query/start.php. Accessed June 12, 2017.
43 Index Mundi: Greece Age structure. [Internet] http://www.indexmundi.com/greece/age_structure.html. Accessed June 12, 2017
44 Tsotsolis N, Tsirgogianni K, Kioumis I, et al. Pneumothorax as a complication of central venous catheter insertion. Ann Transl Med. 2015;3(3):40-5839.2015.02.11.

45 Cuthbertson BH, Elders A, Hall S, Taylor J, MacLennan G, Mackirdy F, et al. Mortality and quality of life in the five years after severe sepsis. Crit Care 2013; 17:1.

46 Vandijck DM, Depaemelaere M, Labeau SO, et al. Daily cost of antimicrobial therapy in patients with intensive care unit-acquired, laboratory-confirmed bloodstream infection. Int J Antimicrob Agents. 2008;31(2):161-165.

47 Datamed: EKG Format Translators. Product Price List: Effective 4/1/2017. [Internet] http://www.datamed.com/docs/datamed_product_price_list.pdf. Accessed June 12, 2017

48 Leftakis A, Geitona M. Cost analysis and estimation of thoracic surgical patients with lung cancer in Greece: The case of Sotiria ICU. Intensive Crit Care Nurs 2001; 17:322-30.

49 Geitona M, Hatzikou M, Steiropoulos P, Alexopoulos EC, Bouros D. The cost of COPD exacerbations: A university hospital–based study in Greece. Respir Med 2011; 105:402-9.

50 Girling A, Lilford R, Cole A, Young T. Headroom approach to device development: Current and future directions. International journal of technology assessment in health care. 2015;31(5):331-8.

51 Woods B, Revill P, Sculpher M, Claxton K. Country-level cost-effectiveness thresholds: initial estimates and the need for further research. Value Health 2016;19(8):929-935.

52 Makras P, Athanasakis K, Boubouchairopoulou N, et al. Cost-effective osteoporosis treatment thresholds in Greece. Osteoporosis Int 2015;26(7):1949-1957.

53 Geitona M, Kousoulakou H, Kalogeropoulou M, Mitsiki E, Panitti E, Steiropoulos P. Cost-effectiveness analysis of the fixed combination indacaterol/glycopyrronium vs. tiotropium and salmeterol/fluticasone in the management of COPD in Greece. Value Health 2015;18(7):A500.

54 Jacob V, Thota AB, Chattopadhyay SK, et al. Cost and economic benefit of clinical decision support systems for cardiovascular disease 55 prevention: a community guide systematic review. J Am Med Inform Assoc. 2017 May 1;24(3):669-76.

55 Van Calster B, Wynants L, Timmerman D, Steyerberg EW, Collins GS. Predictive analytics in health care: how can we know it works?. J Am Med Inform Assoc. 2019 Dec;26(12):1651-4

56 Cox CE, Carson SS, Govert JA, Chelluri L, Sanders GD. An economic evaluation of prolonged mechanical ventilation. Crit Care Med 2007; 35:1918-27.

57 Karabatsou D, Tsironi M, Tsigou E, Boutzouka E, Katsoulas T, Baltopoulos G. Variable cost of ICU care, a micro-costing analysis. Intensive Crit Care Nurs. 2016;35:66-73.

58 Hakkaart-van Roijen L, van der Linden N, Bouwmans C, Kanters T, Tan SS. Bijlage 1. Kostenhandleiding: methodologie van kostenonderzoek en referentieprijzen voor economische evaluaties in de gezondheidszorg. Institute for Medical Technology Assessment, Erasmus Universiteit Rotterdam, Rotterdam; 2015: 1-120 (in Dutch).

59 Lorente L, Lecuona M, Jiménez A, et al. Cost/benefit analysis of chlorhexidine-silver sulfadiazine-impregnated venous catheters for femoral access. Am J Infect Control. 2014;42(10):1130-1132.

60 Moretti R, Moretti F. Ultrasound-guided internal jugular access: Systematic review and economic evaluation of cost-effectiveness. Evid Based Med. 2013;1(2):11.

61 Boland A, Haycox A, Bagust A, Fitzsimmons L. A randomised controlled trial to evaluate the clinical and cost-effectiveness of hickman line insertions in adult cancer patients by nurses. NCCHTA; 2003.

Extrapolating empirical long-term survival data: the impact of updated follow-up data and parametric extrapolation methods on survival estimates in multiple myeloma.

Bakker L, Thielen F, Redekop W, Uyl-de Groot C, Blommestein H.

Submitted.

Paper under embargo

Paper under embargo

Paper under embargo

Paper under embargo

Paper under embargo

Paper under embargo

Paper under embargo

Paper under embargo

Paper under embargo

Paper under embargo

Paper under embargo

Paper under embargo

Paper under embargo

Paper under embargo

Paper under embargo

Paper under embargo

Paper under embargo

Paper under embargo

Paper under embargo

Paper under embargo

Paper under embargo

Paper under embargo

Paper under embargo

Paper under embargo

Paper under embargo

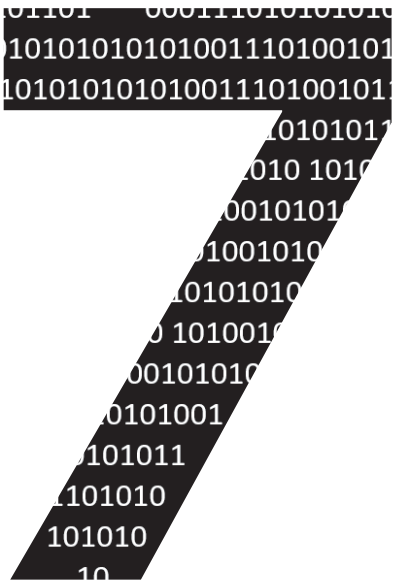
Paper under embargo

Paper under embargo

Paper under embargo

Paper under embargo

Paper under embargo


7

Discussion.

"*So in health care, it turns out maybe we* [IBM] *were too optimistic*" [1]

For many clinical areas, diseases and data types, papers are available that discuss the 'reaping' of health benefits and savings from technologies such as big data analytics and AI [2-14]. Following these high expectations, companies, hospitals, and governmental bodies invested billions to realize the benefits [15-18]. Unfortunately, success stories in this area are rare and few big data analytics and AI technologies have been implemented so far [19-21]. Better yet, many are familiar with the failure of IBM's supercomputer Watson. Ten years after winning Jeopardy, stories of failed development, a lack of clinical success and a lack of revenue for IBM have been widely reported [22-24].

In this dissertation, I explored how economic evaluations may assist decision making of developers of healthcare analytics during the exciting process of development aiming to increase the likelihood of success. In the following paragraphs, I will first discuss how economic evaluations are currently used to evaluate analytics. Hereafter, I discuss how the use of economic evaluations might assist developers to increase the likelihood of feasible development and adoption. Economic evaluations can assist developers by estimating clinically relevant benefits, identify cost components that impact cost-effectiveness and enable developers to identify variation in between practices in current care (Figure 1).

In the final section, I offer recommendations for those conducting economic evaluations since several challenges could influence the timing and methods of economic evaluations. As can be seen in Figure 1, I argue that an assessment of feasibility should precede any economic evaluation and that any further assessments and evaluations should be conducted iteratively. Moreover, when collecting evidence on survival for these economic evaluations reducing the amount of censoring is crucial and both standard parametric and spline models should be fitted for extrapolating long term survival.

HOW ARE ECONOMIC EVALUATIONS USED TO EVALUATE HEALTHCARE ANALYTICS?

Despite expectations that analytics can improve care, my findings in Chapter 2 demonstrated that economic evaluations corroborating these claims are scarce [25]. These findings align with those of Voets et al., Mehta et al., and Wolff et al [2,26,27]. The economic analyses performed were often of poor quality and did not compare alternative strategies or excluded costs or consequences [25]. Cost calculations were frequently incomplete and especially costs of the technology and implementation were generally missing [25]. In prior research, the costs of validation and deployment are considered important barriers that jeopardize widespread adoption of analytics [21,28] and thus including them in estimates that facilitate investment and implementation decisions is essential. Possibly, developers are simply not aware of these costs during development and therefore do not include them. Alternatively, their exclusion is a conscious choice when economic analyses are used to attract investors. Demonstrating the potential value of analytics in the most favorable light would allow developers to pique interests of those willing to invest. However, the lack of studies that include costs of analytics makes it difficult to verify claims that data technologies can lead to savings for healthcare payers.

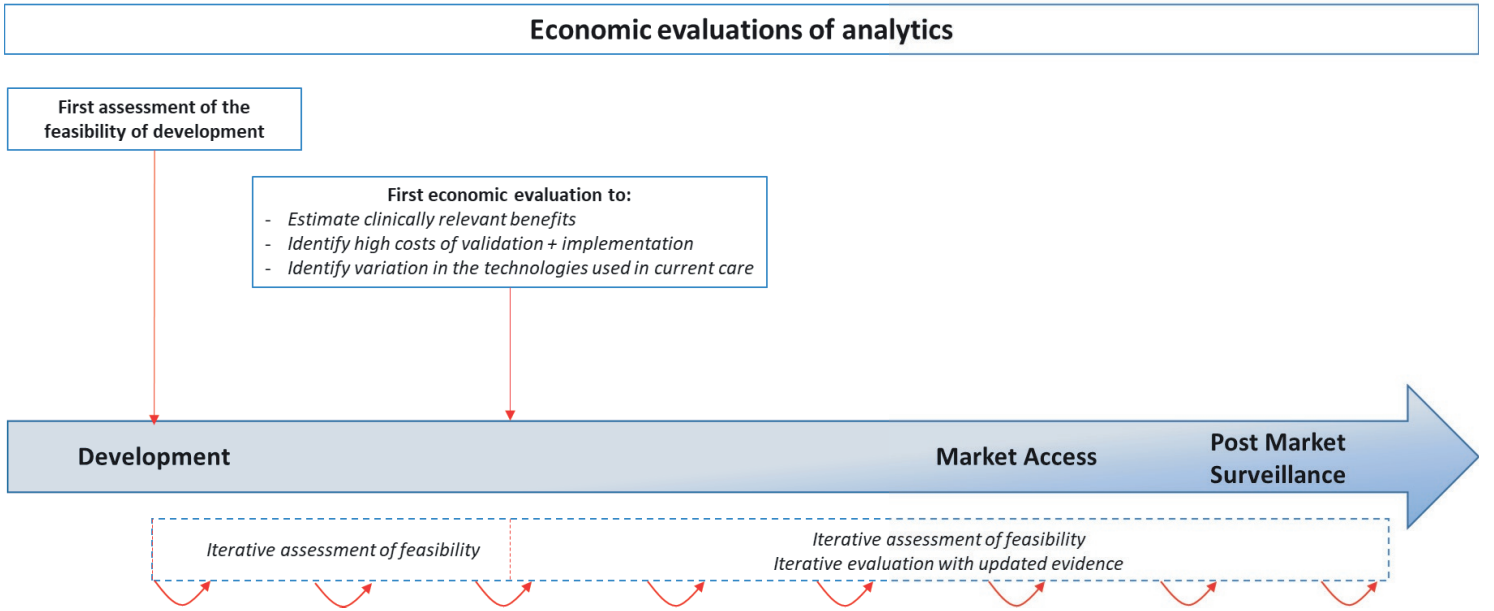


Figure 1: Procedural overview of how and when economic evaluations can assist developers throughout development and market access.

The outcome most often reported, besides costs, was technical model performance (e.g., area under the curve, accuracy, sensitivity, specificity). Estimates on clinical utilities or quality adjusted life years (QALYs) were reported in roughly half of the papers included [25].

Fifty percent of the analyses were performed alongside development, a frequency much higher than of other health technologies such as pharmaceuticals and diagnostic tests [25]. A possible explanation could be that the costs of additional validation are an important financial barrier which might require developers to seek investors or other means of novel funding [28]. Evidence from an economic analysis might stimulate third parties to provide the necessary additional investments needed to continue with validation. Moreover, it is likely that many earlier economic evaluations are performed but not published, sometimes to ensure a competitive advantage and sometimes because the results are unfavorable. The question arises whether the shortcomings of published economic evaluations included in this review also apply to unpublished evaluations. Although the impact of any existing publication bias is uncertain, I expect that this is the case since these challenges have also been emphasized by other researchers.

Even though economic evaluations (such as cost-effectiveness analyses) of big data analytics were rare, other methods that aim to translate model accuracy into clinically relevant and economic outcomes are increasingly reported. These methods are considerably less complex than more traditional methods, such as decision analytic models, to estimate benefits [29]. Methods, in order of increasing complexity, include the number needed to benefit [30], decision curve analysis [31,32], and a third method suggested by Katki et al [33]. The number needed to benefit, for instance, estimates the potential economic benefits, is easily applicable since it utilizes readily available input, and is understandable for a wide variety of stakeholders involved in analytics development [30]. Decision curve analysis has been previously suggested to evaluate predictive algorithms beyond model performance [31,32]. Katki et al. suggested the use of a third method to estimate the incremental net benefit which was slightly more detailed than decision curve analysis but slightly less complex than a full decision analytic model [33]. To my knowledge, no study has ever compared these three methods and economic evaluations in terms of the outcomes and recommendations they provide. I think such insights could be relevant for developers but also for health economists performing these analyses.

HOW CAN ECONOMIC EVALUATIONS ASSIST DEVELOPERS OF ANALYTICS?

For a variety of application domains, I demonstrated how economic evaluations can be used to estimate outcomes relevant to users. Moreover, early economic evaluations provide relevant insights for developers regarding the potentially high costs of novel analytics. Relevant outcomes were presented, such as the thresholds for realizing benefits, the headroom for development, and the uncertainty in the impact of healthcare analytics (Chapter 3 & 5). Moreover, recommendations are made concerning the research needed to gain more insight into the costs of purchasing and using these analytics.

Present evidence relevant to users and buyers

Economic evaluations stimulate developers of analytics to calculate and present the effectiveness of their analytics in outcomes relevant to users and healthcare payers. It has been stressed that evidence on the effectiveness of analytics is often limited to technical model performance. However, model performance does not clearly illustrate the value for end users of the technology. Because of this, it has been previously stressed that developers and researchers should evaluate and present outcomes relevant to these users [19,34,35]**.**

In Chapter 3 and 5, outcomes were reported beyond model performance, estimating outcomes relevant to users such as costs and QALYs. In Chapter 3, I described an economic evaluation to estimate the potential impact of analytics to improve the interaction between patients and their mechanical ventilator. The analysis was performed during the early stages of development. In this early stage of development, the potential of the technology to yield healthcare and economic benefits was high (i.e., + 0.21 QALYs and -€264 on average per patient). However, the effectiveness of intervening as well as the costs of the technology and the monitor that measured suboptimal interaction were highly uncertain. This can be a challenge since the actual costs are unclear, thus increasing the risk for developers.

In Chapter 5, I estimated the potential benefits of identifying catheter related bloodstream infection. Although initial results demonstrated that the average savings per patient could be high (€886), the expected savings depended on the prevalence of the disease. The prevalence of catheter related bloodstream infection was relatively high in the hospital involved in development. Further research is needed to determine if the prevalence in other settings is sufficiently high for further development to be worthwhile.

In this chapter, the potential impact on QALYs and costs of better stratification of newly diagnosed chronic lymphocytic leukemia (CLL) patients was also assessed. Results demonstrated that better stratification and subsequent treatment of patients with a poor prognosis resulted in high incremental costs and a marginal QALY gain (incremental cost effectiveness ratio (ICER) of €166,879 per patient), making it unlikely that further development would be worthwhile given the current costs of treatment. Therefore, for stratification and treatment of watch & wait patients to be considered a cost-effective alternative to current care, the treatment would need to be more effective and/or considerably cheaper compared to estimates currently reported for ibrutinib.

Moreover, developers should go beyond the evidence relevant to users and consider what kind of evidence is needed to facilitate reimbursement and purchasing decisions of analytics. Hartz & John have previously emphasized that for all health technologies, the outcomes to evaluate depend on where results will eventually be presented [36]. I recommend that developers conduct further research into the market access procedures for analytics in different countries. In this process, developers should consider who the users and purchasers will be and thus who they need to convince with their results. Then, when evaluating their own technologies, developers can apply the rules and techniques that are applied by their prospective clients early on alongside development. This might facilitate alignment with outcomes relevant to buyers but perhaps also influence the target market selected.

Although it is currently not always apparent which rules are applied by prospective clients, the initiatives to develop guidelines and recommendations for developers and purchasers of healthcare analytics are increasing. The National Health Service (NHS) in the United Kingdom (UK) has published guidelines for both developers and potential users and buyers of data-driven technologies (e.g., NHS Buyers Guide, NICE's DHT Evaluation Framework [37,38]). For developers, guidelines are available that provide recommendations to assist them in the steps to consider when developing data-driven technologies for healthcare [38,39]. An important outcome according to these guidelines is cost-effectiveness. For buyers, the guideline emphasizes that they should be aware of the interoperability, data compatibility and the costs associated with all those elements needed to use a novel AI algorithm [37].

Analyze whether the benefits of analytics justify their potentially high costs

The high costs of analytics, and especially the costs of validation and implementation, are considered important barriers to the success of healthcare analytics. From a developer's perspective, these high future costs might be acceptable if the benefits for potential clients are also high. Thus, if there is great potential for healthcare analytics to save money for a potential client, the developer can ask a steep price and ensure a positive return on investment for both the developer and the client. In my economic evaluations for the ICU (Chapter 3 & 5), I demonstrated that the costs of the technology are an important parameter influencing incremental costs. However, there also seems to be considerable headroom for development. So, although economic evaluations might not lead to a reduction in the high costs of analytics, they could enable developers to estimate whether the price they need to ask for their technology to cover their costs is reasonable for the client given the potential savings from using the technology.

Nonetheless, literature on the costs of developing, validating, and implementing analytics is grossly lacking and further research is essential. Prices and costs of simple analytics offered by commercial parties are available and economic evaluations of analytics often include a fee-for-use per patient (e.g., Rossi et al [40]). However, to my knowledge, the costs of implementing complex analytics in for instance hospitals have not been estimated. Future research should study whether the total costs of validating, purchasing, and implementing analytics allow sufficient room for developers to generate a positive return on investment. A cost estimation of analytics from the perspective of the developer should contain a wide variety of elements relevant for a potential buyer, such as the costs of data storage, computing power, validation, and evaluation [28,37]. Moreover, they should consider the infrastructure needed in the future to host many different pipelines and the requirements that enable use of, for instance, siloed data.

Ideally, a future study that assesses the costs of analytics should cover multiple sites to understand inter-site variation. In the EU project AICCELERATE the line of inquiry outlined in this thesis will be continued in a novel project in which multiple hospitals collaborate with several small and large companies developing technologies that enable the use of AI in healthcare [41]. Here, the costs of adopting AI technologies can be explored at these five different hospitals across Europe. To my knowledge other costs of development are not mentioned explicitly as a barrier. However, one can imagine that if development which is doomed to fail due to poor data is not abandoned 'quickly' enough, the price of any analytics that eventually reach the market will also have to cover the development costs of these failed attempts. Hartz & John have previously stressed the essence of failing early and focusing on those treatments which

will most likely generate positive returns and thus cover the expenses of failed development in the pharmaceutical industry [36].

Identify variation in technologies used in current care

The costs of implementing data analytics are high, partly because of the variability of technologies used in clinical practice. For instance, the electronic health records, the monitoring devices, and the ways in which electronic health data are stored differ between sites. Narrowing down the scope is the first step in any economic evaluation and could be an important tool for developers and clinical stakeholders to better understand what happens in current care and how care would change with their analytics. In the scope, the PICO is defined, requiring developers to critically consider their target (patient) population, current care, the intervention (i.e., any technologies and supporting infrastructures but also any subsequent treatment), and relevant outcomes. Results of current care form the benchmark for performance of the novel intervention and the poorer results are in current care, the more room for improvement there is [36]. Thus, a good understanding of current care at an early stage of technology development enables developers to estimate whether additional investments are worthwhile. Only then can developers include the costs of technologies that enable use of analytics by addressing poor interoperability when estimating their potential benefits for future customers.

For analytics, describing current care during the scoping phase requires a detailed discussion of technologies currently in place, including their limitations and barriers such as lack of interoperability. Interoperability refers to the ability to exchange and use information from two or more different systems [42]. Interoperability can be a challenge, but it is essential when developing data-driven technologies [39,43]. Lehne *et al* go so far as to say that "*Digital health depends on interoperability*" [42]. In recent years, progress has been made with electronic health record (EHR) vendors such as Epic and Cerner investing in natural language processing algorithms to be used within their EHRs, thereby facilitating access to the data stored within them [44]. Moreover, initiatives such as the Observational Health Data Sciences and Informatics initiative [45,46] facilitate better use of vast amounts of EHR data [44]. Nonetheless, analytics developed using datasets from a certain hospital might be complex, time-consuming, and costly to implement in other settings due to limited interoperability and therefore need extensive data cleaning and pre-processing [42]. Identifying relevant barriers such as limited interoperability during development is worthwhile [39] and the procedures performed to narrow down the scope in an early economic evaluation might help developers in this regard.

The challenges of variation between clinical sites also relate to the 'technologies' part of the intervention. Descriptions should go beyond merely stating the analytics, and developers should determine what technologies should be in place to enable use of their solution in the target market of interest. This may include certain devices (e.g., monitors that register ventilation interaction, remote monitoring devices) but also software (e.g., EHRs). Differences between sites can be expected regarding the technologies already available in current care and what needs to be purchased. These costs for future users should be considered when estimating results such as the potential headroom for analytics. For instance, to use the analytics to improve patient ventilator interaction evaluated in this dissertation, it is essential to have a monitor that collects

data on ineffective efforts. However, the need to purchase this monitor reduces the price a developer can ask for the analytics (Chapter 5).

These considerations will also likely influence whether the economic evaluation should focus on the initially selected target market or whether a broader approach should be adopted. Depending on the use case, it might be essential to consider a wider target market, including other countries at an early stage. This is because the technology infrastructures will likely vary considerably between countries. Where EHRs from one vendor available at site A enables users to easily extract data, the EHR used at site B may not, because of different labels and siloed data. Moreover, the results from validation could differ due to differences in patient populations. As shown in Chapter 5, the prevalence of a disease may also vary, which could impact the cost-effectiveness of the analytic. Therefore, developers should decide early on what other markets they will aim to reach in the future.

WHEN AND HOW TO PERFORM ECONOMIC EVALUTIONS OF HEALTHCARE ANALYTICS?

When initiating this dissertation, the aim was to conduct economic evaluations alongside development for the use cases presented in the introduction. However, as discussed in Chapters 4 and 5, development was challenging for many of these use cases. In the next paragraphs, I argue that economic evaluations should only be initiated *after* an assessment of the feasibility of development, for instance given the data quality. Moreover, as discussed in Chapter 6, the robustness of the data used in economic evaluations is decisive in the uncertainty surrounding estimates. After generating an initial estimate of the potential impact of the analytics, the evidence should be updated frequently using economic evaluations iteratively prior, during, and after market access. Moreover, when collecting evidence on survival as input for an economic evaluation, reducing the amount of censoring is crucial and both standard parametric and spline models should be fitted when extrapolating long term survival.

Initiate an economic evaluation after assessing whether development is feasible

Many challenges can arise during development relating to the quality of the ('big') data sets used for development. These challenges include, but are not limited to, small sample size [28,47], shorter duration of follow-up [48], confounding bias [19,49], patient heterogeneity and selection bias [28,35,48], and bias due to missing data [8,49,50]. In Chapter 4, some of the challenges that arose when using EHR data to optimize treatment response for patients with type 2 diabetes were presented. Moreover, I discussed how target trial emulation might increase awareness of these challenges during development. Target trial emulation can be used to identify differences between an ideal RCT, and the study actually performed, using observational data in a systematic way [51]. This allows researchers and developers to identify and understand potential sources of bias.

In the field of epidemiology, target trial emulation has been proposed to reduce the risk of bias when analyzing observational data. The risk of bias can be an important problem for use and development of novel analytics since observational data sources are often used. Bias could lead to invalid conclusions, with the risk of realizing lower benefits than anticipated, and can sometimes even cause harm to patients. Many examples of results from machine learning

algorithms subject to for instance confounding bias exist [19] and there is increasing emphasis that methods to address confounding in deep-learning models and others are needed [52]. Moreover, recent Food and Drug Administration (FDA) recommendations have emphasized that developers should be aware of bias and adopt methods to identify it [53].

Target trial emulation might enable developers and clinicians to critically review the data available, prior to starting projects to develop healthcare analytics using observational data. As discussed in Chapter 4, heterogeneity in patients and treatment can limit the value of EHR data, while the amount of missing data can limit the ability to adjust for confounders. Target trial emulation might stimulate developers to anticipate sources of bias when analyzing observational data. For instance, the variables needed to adjust for bias can be identified at an early stage to avoid disappointment of developers. Moreover, it could be used to determine the plausibility of their results during interpretation.

Challenges relating to the data used for development also resulted in the flowchart presented in Chapter 5, which aims to stimulate efficient use of economic evaluations. This flowchart positions the use of economic evaluations after verifying that development is likely to be feasible given the data available. Ensuring feasibility of development *before* performing economic evaluations facilitates efficient *use* of economic evaluations. This enables developers to allocate resources for development and evaluation to those analytics most likely to succeed, thereby avoiding wasted investment. The question of efficiency is essential for healthcare analytics because development of multiple application domains is often anticipated with the data available (e.g., EHR data). Economic evaluations can help select those application domains with the highest potential health and economic benefits and the ones that maximize profits. Analyses such as the headroom method enable developers to determine the maximum price, they could ask for their technology given a certain willingness to pay threshold. If such a price is deemed insufficient to cover costs, development could of course be halted, and another application could be selected. However, selection of a new application should begin with an assessment of feasibility and not estimation of potential benefits and profits (Figure 1).

I have demonstrated that the variety in application domains of analytics could be a reason for developers to fail fast for those analytics unlikely to succeed. This enables them to invest resources in those applications most likely to succeed based on data quality, ease of access to this data, and the expected impact of the application. For instance, we found that continuing development was challenging for several application domains (e.g., diabetes, CLL). The application of analytics to prevent catheter related bloodstream infection could potentially have a larger impact than analytics to reduce the prevalence of ineffective effort events (Chapter 3 and 5). However, there is also much uncertainty in the input parameters of the models that must be reduced to select the analytics worthy of further development.

A variety of use cases is examined in this dissertation and further research into the relevance of the flowchart for different clinical scenarios is needed. For instance, the expectations are that AI based on imaging (e.g., radiology, dermatology) is the field that will make the most progress in the coming years [18]. Studies that validate the flowchart for use alongside analytics development, for instance for applications using imaging data, would be highly relevant. These studies should adopt a multidisciplinary approach including clinical experts that have access

to data and the developers with the technical know-how. Examples of such projects are those funded by for instance the European Union in the H2020 framework programme. However, such collaborations could also simply be initiated by a developer that is exploring the setting and research questions for which to develop analytics.

Challenges relating to data quality might differ across countries. Issues relating to sample size might be less prevalent in larger countries (US) and countries with a developed infrastructure for data sharing and secondary use of data (e.g., Estonia). Initiatives such as the Clinical Practice Research Datalink [54], the National Patient-Centered Clinical Research Network (PCORNET) [55], and organizations such as Kaiser Permanente [56], enable access to databases that contain millions of individuals. This greatly broadens the potential number of questions that can be answered and perhaps makes some challenges such as a small sample size less relevant. However, for these databases other issues (i.e., lack of ethnic variation, limited information on patient characteristics and treatment response) could be very relevant and should therefore be addressed in the first steps of the flowchart. For instance, ethnic variation in datasets used for training is limited, which can result in discrimination based on race, sex, and socioeconomic status [19].

Carefully considering the data available and narrowing down the scope (patient, intervention, the comparator, and the outcome) at an early stage in development could perhaps make development and validation considerably easier (Chapter 5). Patient heterogeneity could result in the need for costly, site-specific validation and customization and limit the sample size considerably. For instance, we saw considerable variation in the treatment of diabetes and CLL patients, limiting development of analytics but also making it essential that analytics developed in one location are validated in another site (Chapter 4 and 5). Here prior treatments can differ from those used in patients at the development site, potentially influencing the accuracy of novel analytics. The availability (or rather absence) of large, *representative* data sets is a major challenge for development and validation [43]. Narrowly defining the scope and carefully considering the data available for development and validation might identify these potential risks early on, avoiding downstream disappointment.

Use evidence from robust clinical studies that inform iterative forms of evaluation

In Chapter 6, the possibility to estimate long-term survival of patients while varying duration of follow-up was explored. I assessed how the accuracy of extrapolations varied across different data cut offs. Variations in the length of follow-up coincided with variations in the percentage censored and the absolute number of events that occurred. The accuracy was measured using the absolute error between the extrapolated restricted mean survival time (RMST) and empirical RMST. I found that the error was large when the follow-up was short and thus the percentage censored was high. The reduction in RMST error was especially high when increasing the maximum follow-up from three to six years where all percentages censored were <60%. Moreover, there was no clear benefit of using for instance spline models over standard parametric models. However, two of the standard parametric models (Weibull and Gamma) seemed to perform slightly better (i.e., < RMST error and good visual fit) when the follow-up was short and the percentage censored was high.

The results from Chapter 6 underline the relevance of clinical studies with long term follow-up to evaluate analytics. There is a trend towards a more flexible approach of evaluation by governmental bodies such as the FDA and European Drug Association (EMA) to speed up evaluation and better align with these technologies [18]. Moreover, Hendrix et al. also emphasized that the experience of health economists provides them with opportunities to model the clinical impact from the limited evidence available [57]. As I have previously emphasized, I agree that there is a role for health economic modelling to enable developers to translate their technical accuracy into clinically relevant outcomes. Even good data has its limits (e.g., follow-up duration), so good modelling techniques are needed to make valid estimates of long-term health and economic outcomes. However, the results in Chapter 4 and 6 clearly demonstrate the essence of using data with large sample sizes and low percentages censored which often requires long-term follow up.

In Chapter 5, I recommend a flexible and iterative approach to economic evaluations alongside the development of novel healthcare analytics. Development of analytics is an iterative process [18,28]. A prototype of a technology is developed, tested, and then adjusted accordingly. This process is repeated until the technology is considered to function well enough for it to be of value for end-users. Hereafter, site-specific training and/or validation is often required prior to implementation of analytics [19]. After their initial release, they are frequently updated and improved when novel data become available [18,19]. As Stevens emphasized: *"…data-driven technologies need constant adaptation to the healthcare practices in which they are used and vice versa"* [58]. Whereas methods for evaluating and granting market access for novel healthcare technologies are often somewhat static, evaluation of analytics would benefit from a more flexible approach that aligns with the iterative nature of development. Such an approach where updates of the technology coincide with updates in results from evaluation would perhaps allow a more efficient use of resources for evaluating analytics and is therefore highly recommended.

An iterative approach to evaluation is in fact distinctive of early economic evaluations, since they are used iteratively to inform development and assist investment decisions [59-61]. For other healthcare technologies, the use of iterative economic evaluations *during* development has been recommended for many years [36,59-61]. However, contrary to other technologies, analytics are also likely to change *after* implementation. For instance, dataset shift implied that after adopting a new technology in clinical practice, the population in clinical practice may change [19].

Future research should explore whether the efficient use of resources for evaluation can be stimulated not only through use of the flowchart but also by automating the process of evaluation. A flexible approach to evaluation allows developers to update results based on site-specific performance depending on the most recent, validated version of an algorithm. Better yet, adaptive algorithms automatically update input over time, leading to changes in accuracy and therefore outcomes. If continuous collection of data is anticipated after implementation, developers could explore whether they can automate the process of (economic) evaluation after implementation. Such an automated process could enhance the efficiency of performing these evaluations by reducing future costs of evaluation.

Using early economic evaluations aligns with recommendations by other authors. He *et al*. emphasized that national and international recommendations promote the use of an iterative

form of evaluation for analytics [18]. Moreover, the relevance of early assessment of technologies such as discussed in this dissertation has also been demonstrated by the development of the Medtech Early Technical Assessment (META) tool by NICE [62]. The META tool tries to increase the understanding of developers of the consequences of any limitations in the evidence that they have and allows them to better understand their potential customers. It also offers research recommendations concerning the evidence still required to gain insights into the potential benefits of the product. For instance, the META tool can assist developers to develop value claims needed for a health technology assessment for NICE or another organization. Moreover, the need for iterative evaluation has also been emphasized in recommendations from governmental institutes such as the FDA [53]. The FDA have also acknowledged the need to revise the review processes for market entry of analytics since the current review processes are not suitable for the iterative development of analytics [18]. Here, developers that adhere to certain standards of excellence can apply for precertification and faster review [18].

To conclude, the use of RCTs and robust methods of evaluation are needed to assess the potential benefits of analytics. The use of RCTs to evaluate AI for instance is gradually increasing [63]. However, it has been emphasized that the external validity is an issue because RCTs of analytics are often performed in a similar geographic region as the one where they have been developed [63]. Policymakers and developers should jointly develop strategies and guidelines that can ensure that the quality of evidence is robust while assessments for market access keep up with development of analytics. Authors have previously recommended the use of dynamic modelling to assess the health economic impact of AI [26,57]. However, these complex models will also require robust evidence to estimate parameters. We can increase model complexity, but the quality of the evidence used to populate the model will need to improve.

CHALLENGES FOR DEVELOPING AND ADOPTING HEALTHCARE ANALYTICS

Despite the potential of healthcare analytics, many studies have reported challenges that need to be addressed for big data analytics and AI to revolutionize healthcare as promised [8,9,19,34,35,42,47,48]. As I have demonstrated in this discussion, economic evaluations can assist developers in identifying and addressing some of these challenges (Challenge 1-3 Table 1), whereas other challenges should be considered when performing economic evaluations (Challenge 4-5 Table 1). For instance, a timely estimation of the potential cost-effectiveness enables developers to better understand whether there is sufficient headroom for development despite the high costs of analytics or supporting technologies (Challenge 2). However, there are also other challenges (e.g., Challenge 4 & 5), which should be considered by those *conducting* an economic evaluation since I think they will influence *when* and *how* to perform an economic evaluation. For instance, an economic evaluation should be initiated after the feasibility of development with the data available has been critically assessed. Moreover, depending on these challenges, sometimes uncertainty analyses should be more detailed (e.g., Challenge 3-5).

Table 1: Several challenges relating to development and/or implementation of analytics that economic evaluations can address, and which should be considered when conducting an economic evaluation.

How can economic evaluations assist developers of analytics?			
	Challenge	**How can economic evaluations address this challenge?**	**Relevant chapter in this dissertation**
Present evidence relevant to users and buyers	**Challenge 1:** Few studies report clinically relevant outcomes	In an economic evaluation multiple outcomes relevant to end users can be identified using the PICO method. Hereafter, the potential impact of the analytics on these outcomes can be estimated and presented to users, buyers, and investors.	3
Analyze whether the benefits of analytics justify their potentially high costs	**Challenge 2:** The costs of analytics and supporting technologies are high	In an economic evaluation, the intervention and current care are defined using the PICO method. Hereafter, analyses in economic evaluations can explore the potential ROI while varying costs of analytics. In uncertainty analyses, a variety of cost components can be included and excluded, and unit costs can be decreased and increased to estimate the impact on benefits.	3, 5
Identify variation in technologies used in current care	**Challenge 3:** There is considerable variation in technologies[a] used between prospective users	In an economic evaluation, current care is defined using the PICO method. This should be done in sufficient detail and any anticipated variation between sites and/or target markets can be modelled and its impact on outcomes estimated.	3, 5

Table 1: *Continued.*

When and how to perform economic evaluations of healthcare analytics?			
	Challenge	**How should this challenge be considered when conducting economic evaluations of analytics?**	**Relevant chapter in this dissertation**
Initiate an economic evaluation after assessing whether development is feasible.	**Challenge 4:** Feasibility of development can be threatened for instance due to poor data quality and quantity	The quality and quantity [b] of the data available for development can eventually limit the feasibility of development. An economic evaluation should only be initiated after an assessment of the feasibility for development given the data available. Here, a good definition of the PICO might enable a better assessment of the data quality.	4, 5
Use evidence from robust clinical studies that inform iterative forms of evaluation.	**Challenge 5:** Lack of robust trials	Few robust clinical trials report the long-term impact of analytics. Uncertainty surrounding outcomes can have a profound impact on extrapolated long-term survival estimates. Here, model parameters need to be varied in uncertainty analyses and should be updated once new evidence is available.	6

PICO= Patient Intervention Comparator Outcome, ROI= Return on Investment, [a] Includes differences in the way data is stored, terminology, types of tests performed etc., [b] Can include issues such as confounding, small sample size, missing data, lack of follow-up.

Health Technology Assessment: diversity in challenges and thus diversity in assessment

Although table 1 provides a list of critical challenges, I do not offer an exhaustive list of the different types of challenges that can arise during analytics development and adoption in this dissertation. Challenges relating to explainability of results [19,35,48], privacy, ethical and legal concerns [8, 35], but also the lack of policy and regulation frameworks for many countries [19] are important concerns for which conducting an economic evaluation may or may not offer some assistance. Moreover, some of these challenges might need to be considered when performing an economic evaluation.

There are many factors that may influence whether a technology reaches the market and could influence decision-makers when developing their technology. Health technology assessment (HTA) adopts a much wider scope than just an assessment of health economic benefits by including for instance, organizational and legal aspects, ethical considerations, and patient preferences [64]. Economic evaluations should be considered as a single step, when performing an early Health Technology Assessment thus widening the number of challenges addressed

when assessing the technology. The order in which these steps are addressed will probably depend on the technology. However, one can imagine that for all steps, the processes are likely to be iterative. The feasibility assessment presented in the flowchart in Chapter 5 in this dissertation could be supplemented with issues relating to these other HTA components, such as privacy, safety, organization, and ethics, since they can lead to termination of (or changes in) development. Eventually, the true benefit of the flowchart will really depend on whether developers critically address challenges and barriers relating to all these elements in the early phases of development, prior to performing an economic evaluation.

The use of early HTA might further stimulate the success of analytics. Here, challenges which are mentioned in the literature but were not addressed in this dissertation such as interpretability and usability could be addressed. Trocin *et al*. reported criteria for 'responsible' AI, defined as the field that ensures design, implementation and use of AI technologies that are ethical, transparent, and accountable to reduce the potential risk associated with use of AI [65]. Social factors, such as the way AI is explained to users, are considered essential [65]. Acceptance of users has also been mentioned by the NHS as an important outcome to be assessed [38], to which explainability greatly contributes, while the FDA emphasizes the necessity of explainable AI to facilitate trust of users and enhance adoption [53]. A wide variety of elicitation methods are available that could be of use here, including interviews but also methods that quantify preferences such as discrete choice experiments. Hendrix et al for instance elicited preferences of patients and providers regarding explainable AI [66].

CONCLUDING REMARKS

In this dissertation, I argue that early economic evaluations could assist developers of healthcare analytics in their decision-making during development. Thus far, evidence from economic evaluations is scarce. National guidelines and recommendations on assessing analytics are not yet widely available, which increases uncertainty for developers as to which outcomes would be relevant to prospective purchasers. The lack of clarity about relevant outcomes means that developers will not know which outcomes to analyze and optimize and efficiency in development could be enhanced by clarifying assessment criteria of purchasers and users.

Moreover, good-quality data is essential for developing and evaluating analytics. The flowchart presented in my dissertation can hopefully increase the efficiency of development by selecting those application domains of healthcare analytics with limited barriers to implementation and the highest impact. However, it is essential to perform a feasibility check before estimating the potential benefits and profits. Moreover, more research is needed as to the way analytics need to be evaluated and the evidence required.

The lack of insights into total costs of adopting analytics for the prospective customers of these developers and the lack of purchasing guidance makes it difficult for developers to realize widespread adoption of their analytics. I have emphasized that further research is needed to estimate the costs of analytics and to further validate the flowchart in other settings and countries.

The potential benefits of healthcare analytics have often been mentioned and it is this faith in their ability to improve care that has resulted in billions of Euros and USD being invested in their development. However, as a society we should also be cautious to ensure that scarce resources are invested wisely. Early economic evaluations can assist development by initiating the discussion of their potential in the early phases of development. However, they are also merely one piece of a very complex puzzle which will take more than just the availability of computing power, neural networks and EHRs to solve any time soon.

REFERENCES

1 Krishna A. 2021. [Internet] https://twitter.com/axios/status/1460027961957101577 Accessed September 29, 2022
2 Mehta N, Pandit A. Concurrence of big data analytics and healthcare: a systematic review. Int J Med Inform. 2018;114:57–65.
3 Bates DW, Saria S, Ohno-Machado L, et al. Big data in health care: Using analytics to identify and manage high-risk and high-cost patients. Health Aff (Millwood). 2014;33:1123–31.
4 Celi LA, Mark RG, Stone DJ, et al. "Big data" in the intensive care unit. Closing the data loop. Am J Respir Crit Care Med. 2013;187(11):1157.
5 Sanchez-Pinto LN, Luo Y, Churpek MM. Big data and data science in critical care. Chest. 2018;154(5):1239–48.
6 Gutierrez G. Artificial intelligence in the intensive care unit. Crit Care (Fullerton). 2020 Dec;24:1-9
7 Janke AT, Overbeek DL, Kocher KE, et al. Exploring the Potential of Predictive Analytics and Big Data in Emergency Care. Ann Emerg Med. 2016;67:227–36.
8 Hemingway H, Asselbergs FW, Danesh J, et al. Big data from electronic health records for early and late translational cardiovascular research: challenges and potential. Eur Heart J. 2017;39(16):1481–95.
9 Rumsfeld JS, Joynt KE, Maddox TM. Big data analytics to improve cardiovascular care: promise and challenges. Nat Rev Cardiol. 2016;13(6):350.
10 Hofmann-Apitius M. Is dementia research ready for big data approaches? BMC Med. 2015;13:145.
11 Phillips KA, Trosman JR, Kelley RK, et al. Genomic sequencing: assessing the health care system, policy, and big-data implications. Health Aff (Millwood). 2014;33(7):1246-1253.
12 Roski J, Bo-Linn GW, & Andrews TA. Creating value in health care through big data: opportunities and policy implications. Health Aff (Millwood). 2014;33(7):1115-1122.
13 Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential Review. Health Inf Sci Syst. 2014;2:3.
14 Yang C, Kong G, Wang L, et al. Big data in nephrology: Are we ready for the change? Nephrology. 2019;24:1097–102.
15 Morello L, Guglielmi G. US science agencies set to win big in budget deal. Nature. 2018;555(7698).
16 Banks, MA. Sizing up big data. Nat Med 2020;26:5–6
17 Kisner J. Creating Shareholder Value with AI? Not so Elementary, My Dear Watson. 2017.
18 He J, Baxter SL, Xu J, et al. The practical implementation of artificial intelligence technologies in medicine. Nat Med. 2019 Jan;25(1):30-6.
19 Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. BMC Med; 2019 Dec 1;17(1):195.
20 Magrabi F, Ammenwerth E, McNair JB, et al. Artificial intelligence in clinical decision support: Challenges for evaluating AI and practical implications: A position paper from the IMIA Technology Assessment & Quality Development in Health Informatics Working Group and the EFMI Working Group for Assessment of Health Information Systems. Yearb Med Inform; 2019 Aug;28(1):128.
21 Morse, KE, Bagely, SC, Shah, NH. Estimate the hidden deployment cost of predictive models to improve patient care. Nat Med; 2020:26;18–19.
22 Advisory Board 2021. 10 years ago, IBM's Watson threatened to disrupt health care. What happened? [Internet] https://www.advisory.com/daily-briefing/2021/07/21/ibm-watson Accessed September 29, 2022
23 LPB. 2018. Overpromising and Underdelivering at Sloan Kettering: is AI Still 'Human' After All? [Internet] https://d3.harvard.edu/platform-rctom/submission/overpromising-and-underdelivering-at-sloan-kettering-is-ai-still-human-after-all/ Accessed September 29, 2022.
24 Pringle S. Scoop: IBM tries to sell Watson Health again, [Internet] Accessed from: https://www.axios.com/ibm-tries-to-sell-watson-health-again-82f691a4-ab81-4b2b-a5bb-13a7556c8ef1.html Accessed on March 23, 2022
25 Bakker L, Aarts J, Uyl-de Groot C, Redekop W. Economic evaluations of big data analytics for clinical decision-making: a scoping review. J Am Med Inform Assoc. 2020 Sep;27(9):1466-75.

26 Voets MM, Veltman J, Slump CH, Siesling S, Koffijberg H. Systematic Review of Health Economic Evaluations Focused on Artificial Intelligence in Healthcare: The Tortoise and the Cheetah. Value Health. 2021 Dec 16.
27 Wolff J, Pauling J, Keck A, Baumbach J. The economic impact of artificial intelligence in health care: systematic review. J Med Internet Res. 2020;22(2):e16866.
28 Fröhlich H, Balling R, Beerenwinkel N, et al. From hype to reality: Data science enabling personalized medicine. BMC Med. 2018;16.
29 Vickers AJ, van Calster B, Steyerberg EW. A simple, step-by-step guide to interpreting decision curve analysis. Diagn Progn Res. 2019 Dec;3(1):1-8.
30 Liu VX, Bates DW, Wiens J, Shah NH. The number needed to benefit: estimating the value of predictive analytics in healthcare. J Am Med Inform Assoc; 2019 Dec;26(12):1655-9.
31 Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Making; 2006 Nov;26(6):565-74.
32 Vickers AJ, Van Calster B, Steyerberg EW. Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. BMJ; 2016 Jan 25;352:i6.
33 A.Katki H, Bebu I. A simple framework to identify optimal cost-effective risk thresholds for a single screen: Comparison to Decision Curve Analysis. J R Stat Soc Ser A Stat Soc. 2021 Jul;184(3):887-903.
34 Shah NH, Milstein A, Bagley SC. Making machine learning models clinically useful. JAMA. 2019 Oct 8;322(14):1351-2.
35 Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med; 2019 Jan;25(1):44-56
36 Hartz S, John J. Contribution of economic evaluation to decision making in early phases of product development: a methodological and empirical review. Int J Technol Assess Health Care. 2008 Oct;24(4):465-72.
37 Joshi I, Cushnan D. 2020. A buyer's guide to AI in health and care. [Internet] https://transform.england.nhs.uk/ai-lab/explore-all-resources/adopt-ai/a-buyers-guide-to-ai-in-health-and-care/a-buyers-guide-to-ai-in-health-and-care/ Accessed September 30, 2022.
38 NICE. 2019. Evidence standards framework for digital health technologies: user guide. [Internet] https://www.nice.org.uk/corporate/ecd7/resources/evidence-standards-framework-for-digital-health-technologies-user-guide-pdf-11696158815685 Accessed September 30, 2022.
39 Department of Health & Social Care. 2021. A guide to good practice for digital and data-driven health technologies. [Internet] https://www.gov.uk/government/publications/code-of-conduct-for-data-driven-health-and-care-technology/initial-code-of-conduct-for-data-driven-health-and-care-technology#section-10 Accessed September 30, 2022.
40 Rossi JG, Rojas-Perilla N, Krois J, Schwendicke F. Cost-effectiveness of Artificial Intelligence as a Decision-Support System Applied to the Detection and Grading of Melanoma, Dental Caries, and Diabetic Retinopathy. JAMA Netw Open. 2022 Mar 1;5(3):e220269-.
41 Aiccelerate. [Internet] https://aiccelerate.eu/ Accessed September 30, 2022.
42 Lehne M, Sass J, Essenwanger A, Schepers J, Thun S. Why digital medicine depends on interoperability. NPJ Digit Med. 2019 Aug 20;2(1):1-5.
43 Tang A, Tam R, Cadrin-Chênevert A, et al. Canadian Association of Radiologists white paper on artificial intelligence in radiology. Can Assoc Radiol J. 2018 May;69(2):120-35.
44 Bohr A, Memarzadeh K. The rise of artificial intelligence in healthcare applications. In Artificial Intelligence in healthcare 2020 Jan 1 (pp. 25-60). Academic Press.
45 Observational Health Data Sciences and Informatics. [Internet] https://www.ohdsi.org/ Accessed January 8, 2023.
46 Hripcsak G, Duke JD, Shah NH, Reich CG, Huser V, Schuemie MJ, Suchard MA, Park RW, Wong IC, Rijnbeek PR, Van Der Lei J. Observational Health Data Sciences and Informatics (OHDSI): opportunities for observational researchers. InMEDINFO 2015: eHealth-enabled Health 2015 (pp. 574-578). IOS Press.
47 Prosperi M, Min JS, Bian J, et al. Big data hurdles in precision medicine and precision public health. BMC Med Inform Decis Mak. 2018;18(1):139.
48 Shilo S, Rossman H, Segal E. Axes of a revolution: challenges and promises of big data in healthcare. Nat Med. 2020;26(1):29-38.

49 Farmer R, Mathur R, Bhaskaran K, Eastwood SV, Chaturvedi N, Smeeth L. Promises and pitfalls of electronic health record analysis. Diabetologia. 2018 Jun 1;61(6):1241-8.

50 Goldstein BA, Navar AM, Pencina MJ, Ioannidis J. Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review. J Am Med Inform Assoc; 2017 Jan 1;24(1):198-208.

51 Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. Am J Epidemiol. 2016 Apr 15;183(8):758-64.

52 Zhao Q, Adeli E, Pohl KM. Training confounder-free deep learning models for medical applications. Nature Commun. 2020 Nov 26;11(1):1-9.

53 Food & Drug Administration. Executive Summary for the Patient Engagement Advisory Committee Meeting. Artificial Intelligence (AI) and Machine Learning (ML) in Medical Devices. [Internet] https://www.fda.gov/media/142998/download Accessed September 29 2022.

54 Clinical Practice Research Datalink. [Internet] https://cprd.com/ Accessed September 30, 2022.

55 Klonoff DC. The expanding role of real-world evidence trials in health care decision making. J Diabetes Sci Technol. 2020 Jan;14(1):174-9.

56 Boland MV. Big data, big challenges. Ophthalmology. 2016 Jan 1;123(1):7-8.

57 Hendrix N, Veenstra DL, Cheng M, Anderson NC, Verguet S. Assessing the Economic Value of Clinical Artificial Intelligence: Challenges and Opportunities. Value Health. 2022 Mar 1;25(3):331-9.

58 Stevens M. Dreaming with Data: Assembling responsible knowledge practices in data-driven healthcare. Netherlands, Erasmus University Rotterdam, 2021, pg 178

59 Buisman LR, Rutten-van Mölken MPMH, Postmus D, et al. The early bird catches the worm: early cost-effectiveness analysis of new medical tests. Int J Technol Assess Health Care. 2016;32(1–2):46–53.

60 Pietzsch JB, Paté-Cornell ME. Early technology assessment of new medical devices. Int J Technol Assess Health Care. 2008;24(1):36–44.

61 Sculpher M, Drummond M, Buxton M. The iterative use of economic evaluation as part of the process of health technology assessment. J Health Serv Res Policy; 1997 Jan;2(1):26-30.

62 META. Medtech Early Technical Assessment (META) tool. [Internet] https://meta.nice.org.uk/ Accessed September 30, 2022.

63 Siontis GC, Sweda R, Noseworthy PA, Friedman PA, Siontis KC, Patel CJ. Development and validation pathways of artificial intelligence tools evaluated in randomised clinical trials. BMJ Health Care inform. 2021;28(1).

64 Eunethta. HTA Core Model. [Internet] http://www.corehta.info/model/HTACoreModel3.0.pdf Accessed September 30, 2022

65 Trocin C, et al. Responsible AI for digital health: a synthesis and a research agenda. Inf Syst Front, 2021, 1-19.

66 Hendrix N, et al. Artificial intelligence in breast cancer screening: Primary care provider preferences. J Am Med Inform Assoc, 2021, 28.6: 1117-1124.


A
101 00
01010101
01010101010
10101010101
0011100 010110
1010100 011010
0001010 010100
0000010 101010
0101010 1010010
10010101000001010
101011010101001

Appendices.

SUMMARY

Healthcare systems are under increasing pressure to reduce costs while maintaining the quality of care. The use of healthcare analytics has been suggested as one of the possible solutions to realize this ambitious aim. In the past ten years, many have emphasized the potential of (big) data analytics and artificial intelligence to improve health and reduce healthcare costs. However, when initiating the research for my dissertation, it was unclear if, and how, economic evaluations were used by developers to corroborate these claims.

In this dissertation, I explored how economic evaluations may assist decision making of developers of healthcare analytics during the process of development aiming to increase the likelihood of success. I first reviewed how economic evaluations are currently used to evaluate analytics. Hereafter, I explored how the use of economic evaluations might assist developers to increase the likelihood of feasible development and adoption. In the final chapters I provided recommendations for those conducting economic evaluations since several challenges could influence the timing and methods adopted for economic evaluations.

HOW ARE ECONOMIC EVALUATIONS USED TO EVALUATE HEALTHCARE ANALYTICS?

In Chapter 2, I present results from a scoping review in which I assessed whether economic evaluations of big data analytics were available, whether these studies adhered to best practice guidelines and if they informed decision-making during development or during the market access phase after development. Most of the studies that reported an economic analysis, did not compare alternative strategies, or excluded consequences and/or costs. I found few full economic evaluations (N=22) and roughly half of the studies were performed alongside development. Cost calculations were often incomplete and only 20% included the costs of the technology whereas effects were often reported in terms of model performance. Only 7 studies reported cost savings and better outcomes and could be classified as 'big' data analytics. Economic evaluations that evaluate (big) data analytics that adhere to best practice guidelines are lacking but the analyses performed are often performed alongside development.

HOW CAN ECONOMIC EVALUATIONS ASSIST DEVELOPERS OF ANALYTICS?

In this dissertation, I demonstrate that economic evaluations can be used to assist decision-making of developers alongside development. Economic evaluations can assist developers by estimating clinically relevant benefits, identify cost components that impact cost-effectiveness, and enable developers to identify variation in between practices in current care.

In Chapter 3, I performed an economic evaluation to estimate the potential impact of analytics that aim to improve the interaction between patients and their mechanical ventilator. The analysis was performed during the early stages of development. In this early phase of development, the technology demonstrated the potential to lead to benefits (i.e., + 0.21 quality adjusted life years (QALY) and -€264 on average per patient). However, the effectiveness of intervening was highly uncertain, and the costs of the technology and the monitor that measured suboptimal interaction strongly influenced results.

In Chapter 5, I estimated the potential benefits of identifying catheter related bloodstream infection. Although initial results demonstrated that the average savings per patient could be high (0.06 QALYs, €886), results fluctuated depending on the prevalence of the disease. The prevalence of catheter related bloodstream infection was relatively high in the hospital involved in development. Further research is needed to determine if the prevalence in other settings is sufficiently high for further development to be a worthwhile investment. In this chapter, I also assessed the potential impact on QALYs and costs of better stratification of newly diagnosed chronic lymphocytic leukemia patients. Results demonstrated that better stratification and subsequent treatment of those with a poor prognosis resulted in high incremental costs and a slight improvement in QALYs (ICER of €166,879). The high costs of treatment were the main cause of the high ICER making it unlikely that further development would be worthwhile.

In my economic evaluations for the intensive care unit (ICU) (Chapter 3 & 5), I demonstrate that the costs are an important parameter influencing incremental costs but that there seems to be sufficient headroom for development. In these estimates, I also considered the costs of validation and implementation in the price of the analytics. From a developer's perspective, these future high costs might be acceptable if the benefits for potential clients are also high, enabling them to ask a steep price ensuring a positive return on investment. However, further research into the costs of developing, validating, and implementing analytics and the willingness to pay of purchasers is needed.

In Chapter 5, I also demonstrate how the first step in an economic evaluation in which the scope of the problem is clarified could assist developers in identifying challenges that can be expected when implementing the technology. The patient, intervention, comparator, and outcome method could stimulate developers to take some of the future challenges (e.g., lack of interoperability) into account at an early stage in development. For instance, the infrastructures in place could vary across sites making it excessively difficult to gain access to the data required as input for the analytics. A good understanding of current care at an early stage of technology development enables developers to estimate whether additional investments are worthwhile. Only then can developers include the costs of technologies that enable use of analytics by addressing poor interoperability when estimating their potential benefits for future customers.

WHEN AND HOW TO PERFORM ECONOMIC EVALUATIONS OF HEALTHCARE ANALYTICS?

In this dissertation I offer several recommendations. I argue that an assessment of feasibility should precede any economic evaluation and that any further assessments and evaluations should be conducted iteratively. Moreover, when gathering clinical evidence to estimate long-term survival benefits of the intervention, developers and researchers should take care to reduce the percentages censored to increase the likelihood that long-term benefits can be estimated accurately.

First, I discuss that economic evaluations should only be conducted after an initial assessment of the feasibility of development. Feasibility issues can relate for instance to the quality of the data (e.g., the sample size, missing data, length of follow-up), variation in treatments prescribed and data collected, and the risk of bias. In Chapter 4, I present some of the challenges that arose when using electronic health record (EHR) data to optimize treatment response for patients with type 2 diabetes. Heterogeneity in patients and treatments limited the value of the EHR data,

whereas the amount of missing data limited the ability to adjust for confounders. I discussed how target trial emulation could increase awareness of these challenges during development. Target trial emulation enables stakeholders involved in development to identify potential sources of bias by comparing an ideal randomized controlled trial (RCT) with the study performed. The method might enable developers and clinicians to critically review the data available, prior to engaging in projects that aim to develop data analytics using observational data.

These challenges relating to the data used for development, also resulted in the flowchart presented in Chapter 5. This flowchart positions the use of economic evaluations after verifying that development is likely to be feasible given the data available. Prior to selecting analytics based on potential benefits, a selection based on feasibility given the data available is recommended. Moreover, in this flowchart I also explicitly emphasize the first step of any economic evaluation in which the scope is clearly defined. As demonstrated for several use cases in Chapter 5, a clear definition of the scope including the population, the intervention, a description of current care and the outcomes desired, could help developers identify critical barriers that they could face during future development and implementation.

I also stress the use of an iterative approach to evaluating analytics where results are continuously updated until eventually robust clinical studies are available for evaluation. In Chapter 5, I recommend a flexible and iterative approach to economic evaluations alongside the development of novel healthcare analytics. Whereas methods for evaluating and granting market access for novel healthcare technologies are often somewhat static, evaluation of analytics would benefit from a more flexible approach that aligns with the iterative nature of development. Such an approach where updates of the technology coincide with updates in results from evaluation would perhaps allow a more efficient use of resources for evaluating analytics.

For the evidence used in these economic evaluations, developers and researchers should take care to reduce the percentages censored in survival data to increase the likelihood that long-term benefits can be estimated accurately. In Chapter 6, the possibility to estimate long-term survival of patients while varying duration of follow-up was explored. I assessed how the accuracy of extrapolations varied across different data cut offs. Variations in the length of follow-up coincided with variations in the percentage censored and the absolute number of events that occurred. In the results I found that the error was large when the follow-up was short and thus the percentage censored was high. The reduction in RMST error was especially high when increasing the maximum follow-up from three to six years where all percentages censored were <60%. However, two of the standard parametric (Weibull and Gamma) models seemed to perform slightly better (i.e., < RMST error and better fit) when the follow-up was short and the percentage censored was high.

In this dissertation, I argue that economic evaluations alongside development could assist developers of healthcare analytics in their decision-making during development. The flowchart presented in my dissertation can hopefully increase the efficiency of development by selecting those application domains of healthcare analytics with limited barriers to implementation and the highest impact. However, it is essential to perform a feasibility check before estimating the potential benefits and profits. Moreover, more research is needed as to the way analytics need to be evaluated and the evidence required.

The lack of insights into total costs of adopting analytics for the prospective customers of these developers and the lack of purchasing guidance makes it difficult for developers to realize widespread adoption of their analytics. I have emphasized that further research is needed to estimate these costs of analytics and to further validate the flowchart in other settings and countries.

The potential benefits of healthcare analytics have often been mentioned and it is this faith in their ability to improve care that has resulted in billions of Euros and USD being invested in their development. However, as a society we should also be cautious to ensure that scarce resources are invested wisely. Early economic evaluations can assist development by initiating the discussion of their potential in the early phases of development. However, they are also merely one piece of a very complex puzzle which will take more than just the availability of computing power, neural networks and EHRs to solve any time soon.

SAMENVATTING

Zorgstelsels staan onder toenemende druk om zorg te leveren van dezelfde of betere kwaliteit voor minder geld. In de afgelopen tien jaar hebben velen benadrukt dat nieuwe technologieën die gebruik maken van data in de zorg zoals (big) data-analytics en kunstmatige intelligentie kunnen bijdragen aan het realiseren van dit ambitieuze doel. De potentiële kosten en effecten van deze nieuwe technieken kunnen worden berekend en vergeleken met de huidige zorg in economische evaluaties. Ten tijde van de start van het onderzoek voor dit proefschrift, was het echter onduidelijk of, en hoe, economische evaluaties werden gebruikt door ontwikkelaars om de beweringen ten aanzien van betere zorg en kostenbesparingen te onderbouwen en te realiseren.

In dit proefschrift heb ik onderzocht hoe economische evaluaties kunnen helpen bij het nemen van beslissingen van ontwikkelaars van analytics in de zorg tijdens het ontwikkelingsproces, met als doel de kans op succes te vergroten. Ik ben eerst nagegaan hoe economische evaluaties momenteel worden gebruikt om analytics in de zorg te evalueren. Hierna heb ik onderzocht hoe het gebruik van economische evaluaties ontwikkelaars zou kunnen helpen om de kans op haalbare ontwikkeling en adoptie te vergroten. In de laatste hoofdstukken geef ik aanbevelingen voor hen die economische evaluaties uitvoeren, aangezien verschillende uitdagingen de timing en methoden voor economische evaluaties van analytics in de zorg kunnen beïnvloeden.

HOE WORDEN ECONOMISCHE EVALUATIES GEBRUIKT OM ANALYTICS IN DE ZORG TE EVALUEREN?

In Hoofdstuk 2 presenteer ik de resultaten van een scoping review waarin ik heb onderzocht of economische evaluaties werden gebruikt om big data-analytics te evalueren, of deze studies conform de richtlijnen werden uitgevoerd en of ze de besluitvorming informeerden tijdens ontwikkeling of tijdens de fase waarin toegang wordt verkregen tot markt, na ontwikkeling. De meerderheid van de economische analyses, vergeleken geen alternatieve strategieën of hadden consequenties en/of kosten niet geïncludeerd. Het aantal volledige economische evaluaties was beperkt (N=22) en ongeveer de helft van de onderzoeken werd uitgevoerd naast ontwikkeling. Kostenberekeningen waren vaak onvolledig en slechts 20% includeerde de kosten van de technologie, terwijl effecten vaak werden gerapporteerd in termen van de technische eigenschappen van de analytics. Slechts 7 studies rapporteerden kostenbesparingen en betere resultaten en konden worden geclassificeerd als ‘big’ data-analytics. Dus, economische evaluaties van (big) data-analytics die zijn uitgevoerd in overeenstemming met de richtlijnen zijn schaars, maar de uitgevoerde analyses worden vaak naast ontwikkeling uitgevoerd.

HOE KUNNEN ECONOMISCHE EVALUATIES ONTWIKKELAARS VAN ANALYTICS IN DE ZORG ASSISTEREN?

In dit proefschrift laat ik zien dat economische evaluaties gedurende ontwikkeling kunnen worden gebruikt om de besluitvorming van ontwikkelaars te ondersteunen. Economische evaluaties kunnen ontwikkelaars helpen door het effect van hun technologie op klinisch relevante uitkomsten en kosten te onderzoeken, kostencomponenten te identificeren die van invloed zijn op de kosteneffectiviteit, en hen in staat te stellen variatie tussen praktijken in de huidige zorg te identificeren.

In Hoofdstuk 3 heb ik een economische evaluatie uitgevoerd om de potentiële impact te onderzoeken van analytics in de zorg die gericht zijn op het verbeteren van de interactie tussen patiënten en hun mechanische ventilator. De analyse werd uitgevoerd tijdens de vroege stadia van ontwikkeling. In deze vroege ontwikkelingsfase blijkt dat de technologie mogelijk in betere zorg en lagere kosten kan resulteren (d.w.z. + 0,21 voor kwaliteit gecorrigeerde levensjaren (QALY) en gemiddeld -€ 264 per patiënt). Het effect van tijdig ingrijpen met behulp van deze analytics was echter zeer onzeker, en de kosten van de analytics en de monitor waarmee suboptimale interactie tussen de patiënt en de ventilator werd gemeten, waren bepalend voor de resultaten.

In Hoofdstuk 5 is het effect van analytics voor de vroegtijdige detectie van katheter gerelateerde bloedbaan-infecties onderzocht. Hoewel de eerste resultaten aantoonden dat de gemiddelde besparing per patiënt hoog kunnen zijn (0,06 QALY's, € 886), fluctueerden de resultaten afhankelijk van de prevalentie van de ziekte. De prevalentie van katheter gerelateerde bloedbaan-infecties was relatief hoog in het ziekenhuis dat bij de ontwikkeling betrokken was. Verder onderzoek is nodig om te bepalen of de prevalentie in andere landen en ziekenhuizen voldoende hoog is om verdere ontwikkeling een waardevolle investering te maken. In dit hoofdstuk heb ik ook de mogelijke impact op QALY's en kosten onderzocht van een betere stratificatie van net gediagnosticeerde patiënten met chronische lymfatische leukemie. Resultaten toonden aan dat een betere stratificatie en daaropvolgende behandeling van degenen met een slechte prognose resulteerde in hoge incrementele kosten en een kleine verbetering van QALY's (ICER van € 166.879). De hoge kosten van behandeling waren de belangrijkste oorzaak van de hoge ICER, waardoor het onwaarschijnlijk was dat verdere ontwikkeling de moeite waard zou zijn.

In de economische evaluaties voor de intensive care (IC) (Hoofdstuk 3 & 5) zijn de kosten een belangrijke parameter die de incrementele kosten beïnvloeden. Er blijkt echter voldoende financiële ruimte lijkt te zijn voor ontwikkeling, zelfs indien rekening wordt gehouden met de hoge kosten van validatie en implementatie in de prijs van de analytics. Vanuit het perspectief van een ontwikkelaar kunnen deze toekomstige hoge kosten acceptabel zijn als de voordelen voor potentiële klanten ook hoog zijn, waardoor ze een hoge prijs kunnen vragen die een positief rendement op de investering garandeert. Er is echter meer onderzoek nodig naar de kosten van het ontwikkelen, valideren en implementeren van analytics in de zorg en de betalingsbereidheid van toekomstige klanten.

In Hoofdstuk 5, laat ik ook zien hoe de eerste stap in een economische evaluatie, ontwikkelaars kan helpen bij het identificeren van uitdagingen die kunnen worden verwacht bij het implementeren van de technologie. De patiënt-, interventie-, comparator- en uitkomstmethode in deze eerste stap zou ontwikkelaars kunnen stimuleren om in een vroeg stadium van ontwikkeling rekening te houden met enkele van de toekomstige uitdagingen (bijv. gebrek aan interoperabiliteit). De aanwezige technologische infrastructuren kunnen bijvoorbeeld per locatie verschillen, waardoor het buitengewoon moeilijk wordt om toegang te krijgen tot de data die nodig zijn als input voor de analyses. Een goed begrip van de huidige zorg in een vroeg stadium van technologieontwikkeling stelt ontwikkelaars in staat om in te schatten of extra investeringen de moeite waard zijn. Alleen dan kunnen ontwikkelaars de kosten van technologieën die het gebruik van de analytics mogelijk maken, meenemen bij het inschatten van de potentiële waarde van hun technologie voor toekomstige klanten.

WANNEER EN HOE MOETEN ECONOMISCHE EVALUATIES VAN ANALYTICS IN DE ZORG WORDEN UITGEVOERD?

In dit proefschrift doe ik tevens een aantal aanbevelingen. Ik pleit ervoor dat een beoordeling van de haalbaarheid van ontwikkeling voorafgaat aan elke economische evaluatie en dat eventuele verdere beoordelingen en evaluaties iteratief moeten worden uitgevoerd. Bovendien moeten ontwikkelaars en onderzoekers bij het verzamelen van klinisch bewijs, ervoor zorgen dat het percentage gecensureerde patiënten wordt verlaagd om de kans te vergroten dat lange termijn overleving nauwkeurig kan worden geschat.

Allereerst dienen economische evaluaties pas worden uitgevoerd na een eerste beoordeling van de haalbaarheid van ontwikkeling. Uitdagingen met betrekking tot de haalbaarheid van ontwikkeling zijn bijvoorbeeld gerelateerd aan de kwaliteit van de data beschikbaar voor ontwikkeling (bijvoorbeeld de omvang van de dataset, ontbrekende gegevens, duur van de follow-up), variatie in voorgeschreven behandelingen en de variabelen verzameld, en het risico op bias. In Hoofdstuk 4, presenteer ik enkele van de uitdagingen die zich voordeden bij het gebruik van elektronische patiëntendossier (EPD)-gegevens om de behandelrespons voor patiënten met type 2-diabetes te optimaliseren. Heterogeniteit in patiënten en behandelingen beperkte de bruikbaarheid van de EPD-gegevens, terwijl de hoeveelheid ontbrekende gegevens de mogelijkheid om voor confounders aan te passen beperkte. Het gebruik van target trial-emulatie zou het bewustzijn van onderzoekers en ontwikkelaars kunnen vergroten met betrekking tot het risico dat deze uitdagingen kunnen voorkomen tijdens ontwikkeling. Target trial-emulatie stelt belanghebbenden die betrokken zijn bij de ontwikkeling in staat om mogelijke bronnen van bias te identificeren door een ideale RCT te vergelijken met de toegepaste methode. Target trial-emulatie kan ontwikkelaars en clinici in staat stellen om de beschikbare gegevens kritisch te beoordelen, voordat ze zich bezighouden met projecten die gericht zijn op het ontwikkelen van analytics met behulp van observationele data.

Deze uitdagingen met betrekking tot de gegevens die voor ontwikkeling worden gebruikt, hebben ook geleid tot het ontwikkelen van het stroomdiagram in Hoofdstuk 5. Deze stroomdiagram positioneert het gebruik van economische evaluaties nadat is geverifieerd dat ontwikkeling waarschijnlijk haalbaar is gezien de beschikbare data. Voorafgaande aan het selecteren van analytics op basis van hun potentiële impact of gezondheid en kosten, wordt aanbevolen een selectie te maken op basis van haalbaarheid gegeven de beschikbare data. Bovendien wordt in deze stroomdiagram ook expliciet benadrukt dat in de eerste stap van elke economische evaluatie de scope moet worden gedefinieerd. Zoals aangetoond voor verschillende casussen in Hoofdstuk 5, zou een duidelijke definitie van de scope, inclusief de populatie, de interventie, een beschrijving van de huidige zorg en de gewenste uitkomsten, ontwikkelaars kunnen helpen bij het identificeren van kritieke barrières waarmee ze te maken kunnen krijgen tijdens toekomstige ontwikkeling en implementatie van analytics in de zorg.

De essentie van een iteratieve benadering voor het evalueren van analytics wordt in dit hoofdstuk eveneens benaderd. Hier worden de resultaten continu bijgewerkt indien nieuwe informatie (bijvoorbeeld afkomstig uit klinische onderzoeken) beschikbaar is. In Hoofdstuk 5 beveel ik een flexibele en iteratieve benadering van economische evaluaties aan naast de ontwikkeling van nieuwe analytics in de zorg. Terwijl methoden voor het evalueren en verlenen van toegang tot de markt voor nieuwe zorgtechnologieën vaak enigszins statisch zijn, zou de evaluatie van

analytics gebaat zijn bij een flexibelere aanpak die aansluit bij de iteratieve aard van ontwikkeling. Een dergelijke benadering waarbij updates van de technologie samenvallen met updates in de resultaten van evaluatie, zou het evalueren van analytics aanzienlijker efficiënter maken.

Voor het bewijs dat in deze economische evaluaties wordt gebruikt, moeten ontwikkelaars en onderzoekers ervoor zorgen dat de percentages patiënten wiens overleving is gecensureerd worden beperkt. Hogere percentages censoring maakt het lastig voor onderzoekers om de lange termijn overleving te modelleren welke noodzakelijk voor het berekenen van de kosteneffectiviteit op basis van deze korte(re) termijn patiëntgegevens. In Hoofdstuk 6 werd de mogelijkheid onderzocht om de overleving van patiënten op lange termijn te schatten terwijl de duur van de follow-up varieerde. De nauwkeurigheid van extrapolaties varieerde aanzienlijk aan de hand van de verschillende data cut-offs. Variaties in de duur van de follow-up vielen samen met variaties in het percentage gecensureerd en het absolute aantal gebeurtenissen dat plaatsvond. In de resultaten vond ik dat de discrepantie tussen gemodelleerde en daadwerkelijke gemiddelde overleving groot was indien de follow-up kort was en dus het gecensureerde percentage hoog was. De vermindering van deze discrepantie in gemiddelde overleving was vooral hoog bij het verhogen van de maximale follow-up van drie naar zes jaar, waarbij alle gecensureerde percentages <60% waren. Twee van de standaard parametrische modellen (Weibull en Gamma) leken echter iets beter te presteren (d.w.z. < discrepantie in gemiddelde overleving en betere fit) wanneer de follow-up kort was en het percentage gecensureerd hoog was.

In dit proefschrift beargumenteer ik dat economische evaluaties naast ontwikkeling ontwikkelaars van analytics in de zorg kunnen helpen bij hun besluitvorming gedurende ontwikkeling. Het stroomdiagram dat in mijn proefschrift wordt gepresenteerd, kan hopelijk de efficiëntie van ontwikkeling vergroten door die applicaties van analytics te selecteren met beperkte barrières voor implementatie en de grootste impact. Het is echter essentieel om de haalbaarheid van ontwikkeling te onderzoeken voorafgaande aan het schatten van de potentiële gezondheidseffecten en kostenbesparingen. Bovendien is er meer onderzoek nodig naar de manier waarop analytics moeten worden geëvalueerd en het benodigde bewijs.

Het gebrek aan inzicht in de totale kosten van het adopteren van analytics voor de potentiële klanten van deze ontwikkelaars en het gebrek aan richtlijnen voor toekomstige kopers maakt het voor ontwikkelaars moeilijk om brede acceptatie van hun analytics te realiseren. Verder onderzoek is dan ook nodig om de kosten van analytics in kaart te brengen en om het stroomdiagram in andere instellingen en landen verder te valideren.

De potentiële voordelen van analytics in de zorg zijn vaak genoemd en het is dit vertrouwen in hun vermogen om de zorg te verbeteren dat ertoe heeft geleid dat miljarden euro's en dollars zijn geïnvesteerd in hun ontwikkeling. Als samenleving moeten wij er echter ook op toezien dat de schaarse middelen beschikbaar optimaal worden geïnvesteerd. Vroege economische evaluaties kunnen de ontwikkeling van analytics in de zorg ondersteunen door vroegtijdig een inschatting te maken van hun potentieel tijdens ontwikkeling. Ze zijn echter ook slechts een stukje van een zeer complexe puzzel waarvoor meer nodig is dan de beschikbaarheid van computing power, neurale netwerken en EPD's om hem op te lossen.

LIST OF ABBREVIATIONS

abs.=absolute
AI=Artificial Intelligence
AIC= Akaike Information Criterion
AUC= Area under the curve
BDA =Big data analytics
BIC= Bayesian Information Criterion
Bort= Bortezomib
CI= Confidence Interval
CLL= Chronic lymphocytic leukemia
CPRD= Clinical Practice Research Datalink
CRBSI= Catheter related bloodstream infection
DAG= Directed Acyclic Graph
DCO= data cut off
DPP4= dipeptidyl-peptidase 4 inhibitor
eGFR= glomerular filtration rate
EHR= Electronic health record
EMA= European Medicines Agency
GLPs= glucagon-like peptide-1 agonists
HbA1C= Hemoglobin A1C
HOVON= Dutch Haemato-oncology Foundation for Adults in the Netherlands
ICER= Incremental Cost Effectiveness Ratio
ICU= Intensive care unit
IEEVs= Ineffective effort events
IPD= individual patient data
IPW= Inverse probability weighting
IQR= interquartile range
KM= Kaplan-Meier
LCI=Lower Confidence Interval
LOS= Length of Stay
LYG= life years gained
META= Medtech Early Technical Assessment
MI= Multiple Imputation
MM= Multiple Myeloma
MMSE= Mini-Mental State Examination
MP = Melphalan + Prednisone
MRI= Magnetic Resonance Images
MV = Mechanical Ventilation
NCR= Netherlands Cancer Registry
NGS= Next generation sequencing
NHS=National Health Service
NKR+= Dutch National Cancer Registry
OAD= Oral antidiabetic drug
PCORNET = National Patient-Centered Clinical Research Network
PHAROS= Population based HAematological Registry for Observational Studies

PICO= Patient Intervention Comparator Outcome
PRISMA-ScR = Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews
PSA= Probabilistic sensitivity analysis
QALYs= Quality-adjusted life years
RCT=Randomized Controlled Trial
RMST= Restricted Mean Survival Time
ROI= Return on Investment
sd= standard deviation
SES= social economic status
SGLTs= sodium glucose transporter-2 inhibitors
Thal= Thalidomide
UCI= Upper Confidence Interval
UK=United Kingdom
US= United States
W&W= Watch and wait

PHD PORTFOLIO

Peer- Reviewed Publications

Bakker L, Aarts J, Uyl-de Groot C, Redekop W. Economic evaluations of big data analytics for clinical decision-making: a scoping review. Journal of the American Medical Informatics Association. 2020 Sep;27(9):1466-75.

Bakker L, Vaporidi K, Aarts J, Redekop W. The potential of real-time analytics to improve care for mechanically ventilated patients in the intensive care unit: an early economic evaluation. Cost Effectiveness and Resource Allocation. 2020 Dec;18(1):1-1.

Bakker LJ, Goossens LM, O'Kane MJ, Uyl-de Groot CA, Redekop WK. Analysing electronic health records: The benefits of target trial emulation. Health Policy and Technology. 2021 Sep 1;10(3):100545.

Bakker L, Aarts J, Uyl-de Groot C, Redekop K. How can we discover the most valuable types of big data and artificial intelligence-based solutions? A methodology for the efficient development of the underlying analytics that improve care. BMC medical informatics and decision making. 2021 Dec;21(1):1-2.

Bakker L, Thielen F, Redekop W, Uyl-de Groot C, Blommestein H. Extrapolating empirical long-term survival data: the impact of updated follow-up data and parametric extrapolation methods on survival estimates in multiple myeloma- In submission

Gregoor A, Sangers T, Bakker L, Hollestein L, Uyl - de Groot C, Nijsten T, Wakkee M. The impact of an artificial intelligence (AI) based app for skin cancer detection: a first clinical practice evaluation in a population-based setting - In submission

Courses

EWP05 Diagnostic Research (2016)

EWP24 Survival Analysis (2016)

Scientific Integrity Workshop (2016)

Health-Exploring Complexity-Medical Informatics Europe (2016)

R Programming (online) (2016)

Advanced Decision Analytic Modelling for Economic Evaluation (2016)

Bayesian Analysis-Overview and Applications (2016)

Update in health Information Technology: Healthcare Data Analytics (2017)

Network meta-analysis (2018)

Erasmus MC - Next Generation Sequencing Data Analysis (2019)

Using R for Decision Analytic Modeling in Health Technology Assessment [CE16] (2021)

Conferences

E-Health week Europe HIMSS Speakers Corner (2016)

Panel participant- Workshop IMIE (2016)

ISPOR 19th Annual European Congress (2016)

Lola HESG (2017)

ISPOR 21st Annual European Congress (2018)

FPM-Centenary Conference 2018 (2018)

HPT European Cooperation (2019)

Lectures Provided

Lecture - Afscheidssymposium J. Aarts (2017)

Lecture- Takeda, CEAs for Big Data Analytics (2017)

Lecture- ZIN, Early Cost Effectiveness Analyses (2019)

Lecture Given- R for Health Technology Assessment, Estimating Restricted Means Survival Time (2022)

Teaching

Kwantitatief Leeronderzoek (KLO) (2016-2017)

Statistiek A (Premaster) (2017-2018)

Statistiek B (Premaster) (2017-1018)

Inleiding Methode & Technieken (Premaster & Bachelor) (2018-2019)

Technologie & Innovatie- Vakontwerp (2018)

Lecture series- Technologie & Innovatie- KEAs (2019-2022)

Technologie & Innovatie- Working & Project Groups (2019-2022)

Master Thesis supervision - (2018-2022)

Reading Committee- (2021-2022)

Grants Awarded

Research Grant Awarded- AICCELERATE (€745.764) (2020)

Research Grant Awarded- Open Mind (€4.294) (2021)

Managerial & Other

Organising Committee Lola HESG (2017)

Reviewer scientific publications (2017-2018)

Commissioning Editor-Health Policy & Technology (2019)

Under embargo

Under embargo

Under embargo

Under embargo

ABOUT THE AUTHOR

Lytske Jantien Bakker was born on September 19th, 1988, in Muscat, Oman. After an elementary education in the Netherlands, Nigeria, and Australia, she graduated high school in 2006 in the Netherlands. After not being drawn for dentistry several times, she became a registered dental hygienist in 2012. However, it was quickly apparent that this did not fulfil all her interests and thus she started the ESHPM premaster in 2013 followed by the Health Economics Policy and Law master program in 2014. After graduating cum laude in 2015 she started as a PhD candidate at ESHPM on the AEGLE project. Throughout her time working on the AEGLE project, her work focused on performing early cost-effectiveness analyses of the big data analytics developed throughout the project. In 2021, she was awarded a small grant as part of the Open Mind Convergence program estimating the cost impact of adopting radar for the neonatal intensive care unit. Throughout her time as a PhD candidate, Lytske has been actively involved in education including courses on elementary statistics, technology and innovation and master theses. Moreover, she has organized several conferences (i.e., LOLAHESG) and briefly worked for the journal Health Policy and Technology as a commissioning editor. In 2020, she received a large grant from the European Horizon 2020 program for the AICCELERATE project on which she is presently working with several colleagues. Throughout the project she will continue with her interests that followed from this dissertation focusing on the cost-effectiveness, and eliciting stakeholder preferences, of artificial intelligence.