

Evaluation of Health Information Systems

2: Evaluating anything: detective work and methods

Hamish S. F. Fraser, MBChB, MSc,
FACMI



*Associate Professor of Medical Science,
Brown Center for Biomedical Informatics*

Outline

- Health information systems (HIS) are a critical part of health care delivery world wide, with a key role in improving in quality of care
- Many systems work well and are well accepted e.g.
 - primary care EHRs in Europe
 - Some commercial US EHRs
 - The VA VistA EHR
 - Radiology Picture archive and communication systems (PACs)
 - Systems supporting HIV care
 - Many Apps – but not a big proportion of the 300,000+ available
- But many problems with current HIS
- A good evidence base from rigorous evaluation studies can prevent most of these problems

Artificial intelligence system 'as good as world-leading experts' at detecting eye disease

Programme identified patients in need of urgent treatment in 94 per cent of cases during tests

Tom Barnes | @thomas_barnes | Tuesday 14 August 2018 01:37 | 0 comments



Like Click to follow
The Independent



Clinically applicable deep learning for diagnosis and referral in retinal disease

[Jeffrey De Fauw](#), et al,
August 13th, 2018,
Nature Medicine



BROWN

Center for
Biomedical Informatics

<https://www.independent.co.uk/news/health/ai-eye-disease-detection-good-as-experts-moorfields-hospital-glaucoma-macular-degeneration-a8490486.html>



“Babylon’s study does not offer convincing evidence that its Babylon Diagnostic and Triage System can perform better than doctors in any realistic situation, and there is a possibility that it might perform significantly worse.”

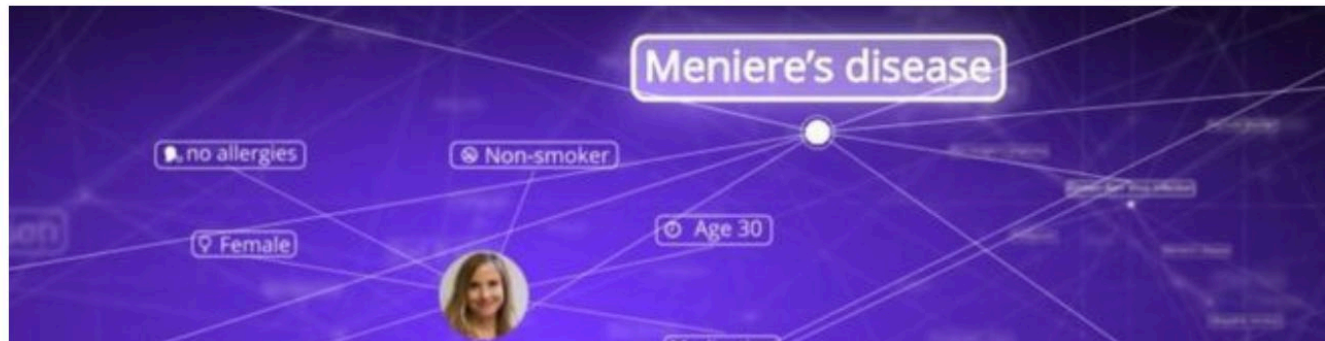
Fraser, Coiera, Wong, Lancet 2018

Worklife | Travel | Future | N

ch | Science | Stories | Enter

beats GPs at

    Share



Diagnostic Inaccuracy of Smartphone Applications for Melanoma Detection

Joel A. Wolf, BA; Jacqueline F. Moreau, BA; Oleg Akilov, MD; Timothy Patton, DO;
Joseph C. English III, MD; Jonhan Ho, MD; Laura K. Ferris, MD, PhD

Conclusions: The performance of smartphone applications in assessing melanoma risk is highly variable, and 3 of 4 smartphone applications incorrectly classified 30% or more of melanomas as unconcerning. Reliance on these applications, which are not subject to regulatory oversight, in lieu of medical consultation can delay the diagnosis of melanoma and harm users.

removal as a part of routine care.

Main Outcome Measures: Sensitivity, specificity, and positive and negative predictive values of 4 smartphone applications designed to aid nonclinician users in determining whether their skin lesion is benign or malignant.

sign, in lieu of medical consultation can delay the diagnosis of melanoma and harm users.

JAMA Dermatol. 2013;149(4):422-426.
Published online January 16, 2013.
[doi:10.1001/jamadermatol.2013.2382](https://doi.org/10.1001/jamadermatol.2013.2382)



Accuracy of a Popular Online Symptom Checker for Ophthalmic Diagnoses

Carl Shen, MD; Michael Nguyen, MD; Alexander Gregor, MD; Gloria Isaza, MD; Anne Beattie, MD

IMPORTANCE Because more patients are presenting with self-guided research of symptoms, it is important to assess the capabilities and limitations of these available health information tools

 [Invited Commentary](#)
page 693

 [Supplemental content](#)

“The correct diagnosis was not included in the symptom checker’s list in 18 of 42 (43%; 95%CI, 32%-63%) cases.

Triage urgency based on the top diagnosis was appropriate in 7 of 18 (39%; 95%CI, 14%-64%) emergent cases”

The primary diagnosis by the symptom checker was correct in 11 of 42 (26%; 95% CI, 12%-40%) cases. The correct diagnosis was included in the online symptom checker’s top 3 diagnoses in 16 of 42 (38%; 95% CI, 25%-56%) cases. The correct diagnosis was not included in the symptom checker’s list in 18 of 42 (43%; 95% CI, 32%-63%) cases. Triage urgency based on the top diagnosis was appropriate in 7 of 18 (39%; 95% CI, 14%-64%) emergent cases and 21 of 24 (88%; 95% CI, 73%-100%) nonemergent cases. Interuser variability for the correct diagnosis being in the top 3 listed was at least moderate (Cohen κ = 0.74; 95% CI, 0.54-0.95).

CONCLUSIONS AND RELEVANCE The most popular online symptom checker may arrive at the correct clinical diagnosis for ophthalmic conditions, but a substantial proportion of diagnoses may not be captured. These findings suggest that further research to reflect the real-life application of internet diagnostic resources is required.

Why we can always improve what we know about an organization, or system

Questions and detective work

- Asking interesting questions about our world and not taking things for granted
- Bringing your own perspective and understanding
- Finding design documents, project plans and data
- Surveying, observing, interviewing, measuring, doing experiments
- In health informatics problems and their causes are usually obvious – we want to change that!
- Simple observations by users can be very valuable
- Often the harder the question the more interesting or useful the answer

Weighing up evidence

- Use of multiple sources of evidence
- Each source can reduce your level of uncertainty
- Even when you have limited knowledge about an issue you wish to evaluate
 - E.g. role of peer support in patients recovery

Principles of evaluation of complex
health interventions:
Evaluation frameworks

Dimensions of evaluation

- Stage of system development and implementation
- Functional area
- Stakeholders
- Type of benefit/impact

5 Levels of Evaluation

1. Problem definition
2. Bench testing
3. Field trials: observational
4. Field trials: interventional
5. Field trials: scale up
6. Long term follow-up

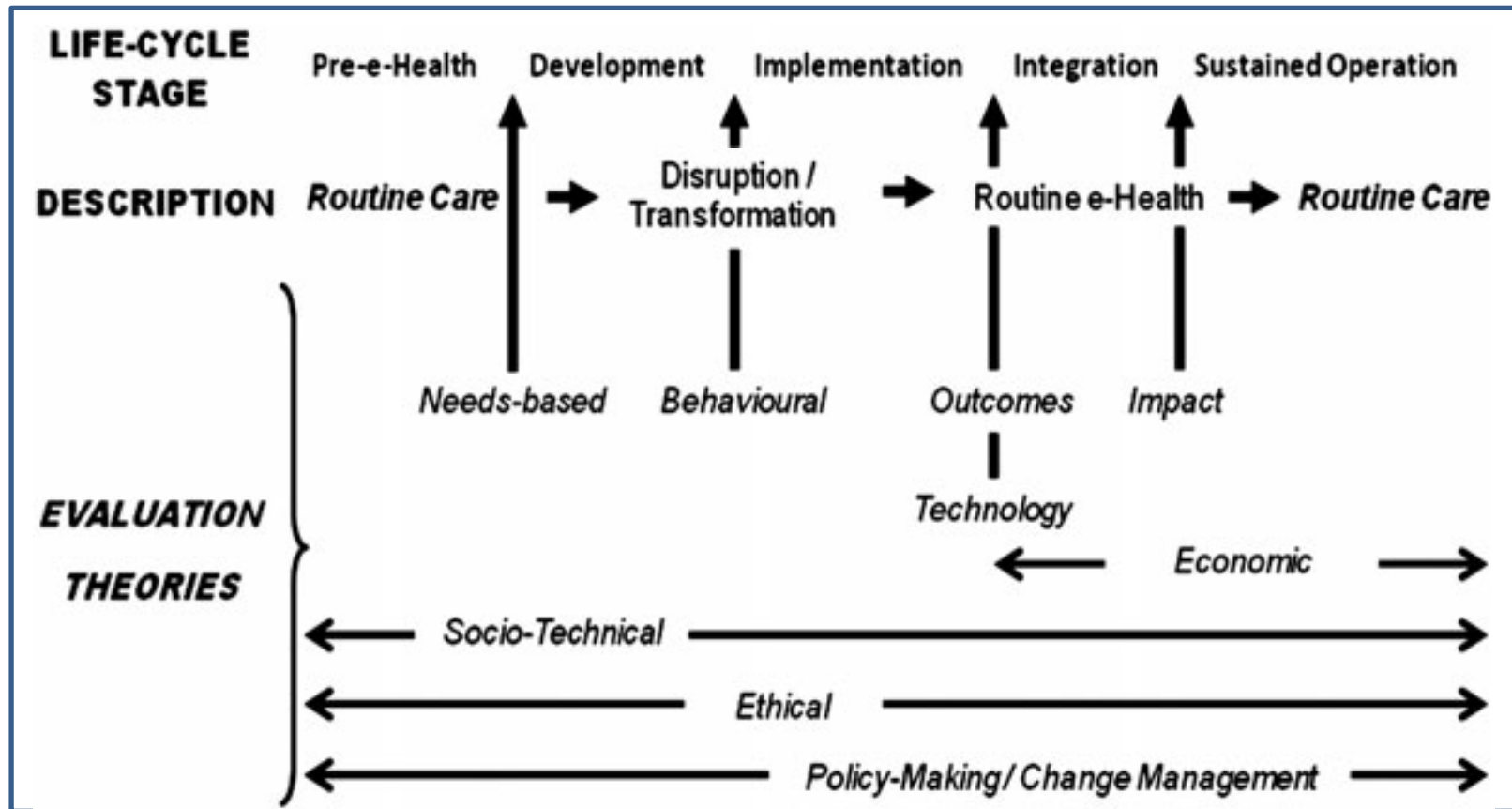
Stakeholders

Stakeholders vary by application and environment

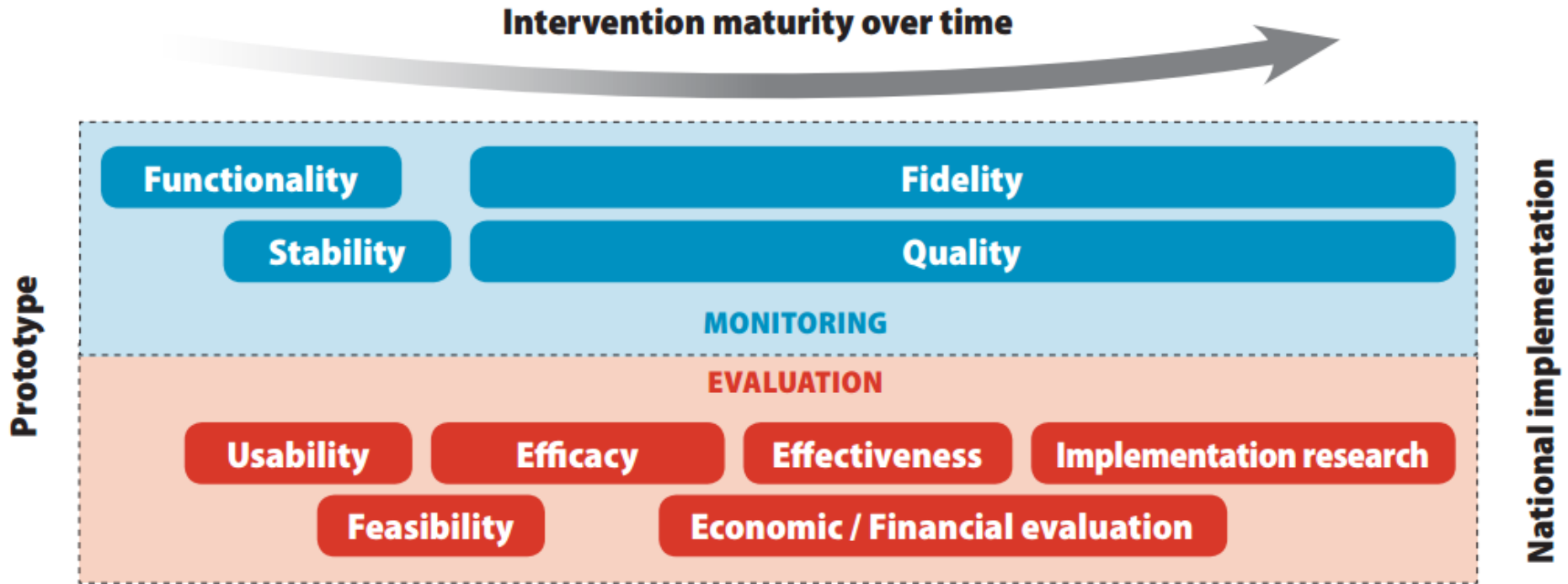
- Users:
 - Patients
 - Clinicians
 - Healthcare staff
- Managers
- Funders
- MOH
- Researchers
- IT team
- Software developers

Conceptual Framework for Development of Comprehensive e-Health Evaluation Tool

Shariq Khoja, Hammad Durrani, Richard E. Scott, Afroz Sajwani, Usha Piryani



Resources to support HIS evaluations

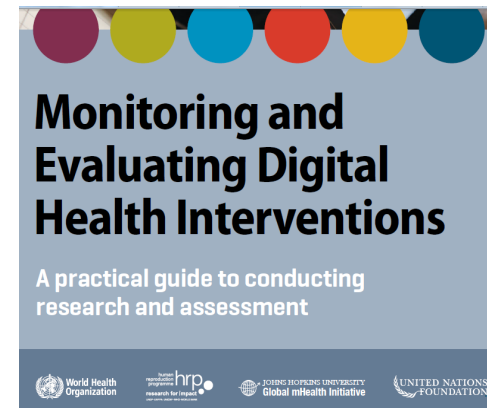


Monitoring and Evaluating Digital Health Interventions, 2016, WHO

<http://apps.who.int/iris/bitstream/10665/252183/1/9789241511766-eng.pdf>



BROWN
Center for
Biomedical Informatics



HIV care in Africa

- OpenMRS open source EHR system used to support HIV care in 12+ countries in Africa
- Rolled out to over 400 clinics in Rwanda, 700 in Kenya, 700 in Mozambique and 1000 in Uganda and Nigeria
- Several published studies show benefits for patient care:
 - Improved detection of HIV+ children in Rwanda[1]
 - 4x increase in compliance with treatment guidelines in pediatric HIV clinic in Kenya[2]
- Much still to learn!



1) Amoroso, ... Fraser, Stud Health Technol Inform. 2010;160:337-41

2) Were et al, Pediatrics 2013



Process evaluation of OpenMRS EHR in Rwanda

- System Monitoring Tool –uptime, usage, data completeness (quant)
- Survey of EHR users (quant)
- Key informant interviews (qual)
- User observation (qual)
- Formal data quality assessment (quant)
(e.g. Lot Quality Assurance Sampling)

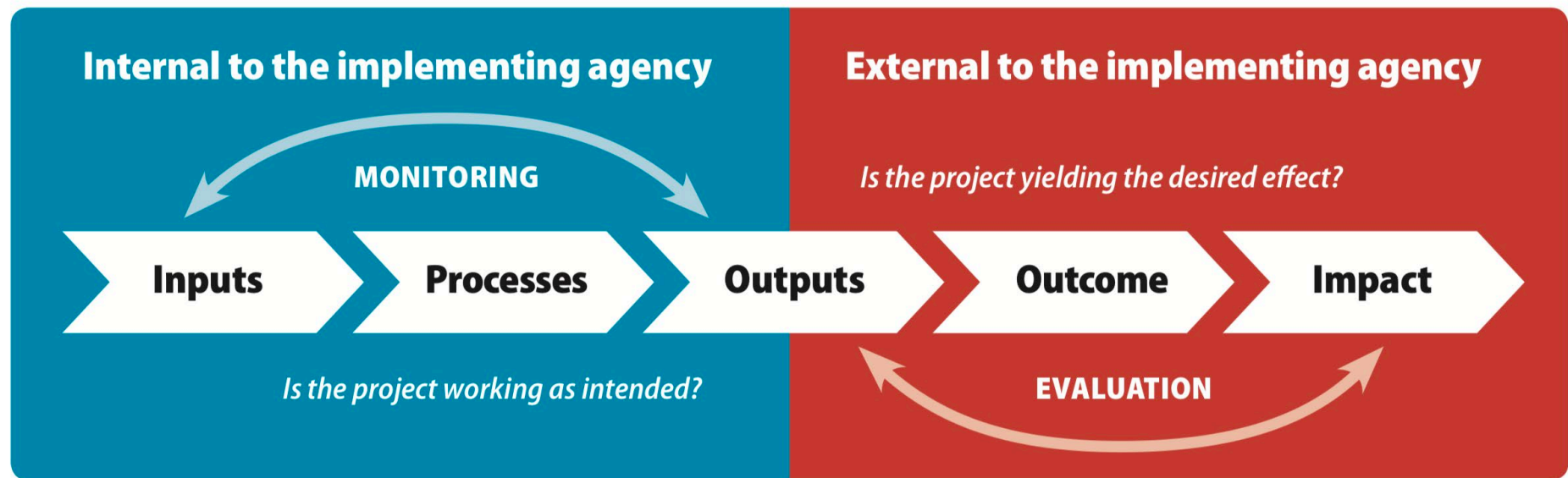


Log Frame

- Useful long established tool for analyzing the impact of an intervention in a complex system e.g. a health system
- Categorizes activities in context of the goal of actual measured impacts
- Useful for determining the appropriate metrics for the system you are studying

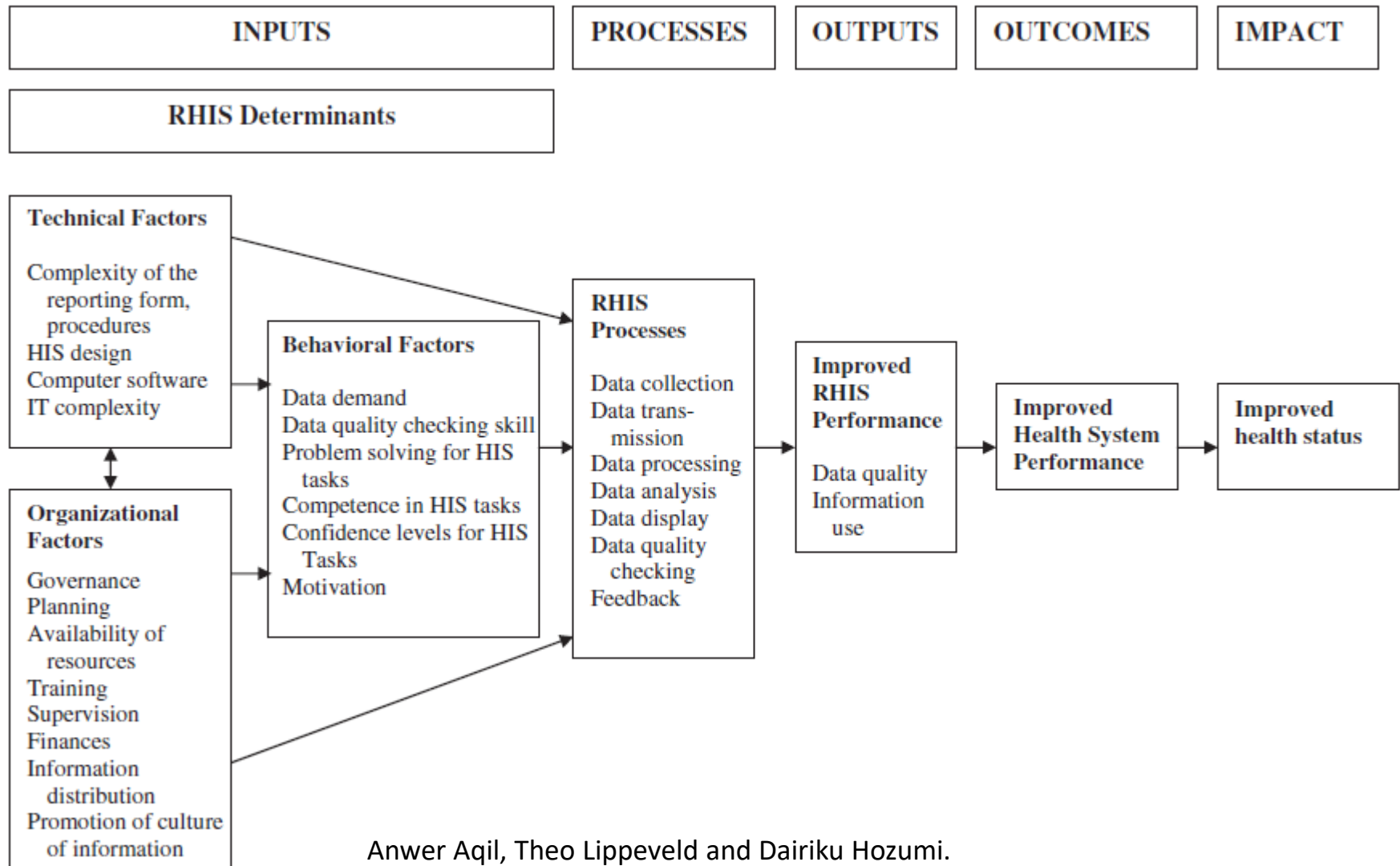
Log Frame

Figure 1.3. Schematic diagram of the interplay between monitoring and evaluation activities



Source: adapted from Pritchett et al. 2013 (4).

The PRISM Framework, example of a “log frame”



Anwer Aqil, Theo Lippeveld and Dairiku Hozumi.
Health Policy and Planning 2009;24:217–228

Review of common evaluation methods

- Document analysis
- Surveys
- Interviews
- Observations
- Monitoring and analysis of system data
- Measuring changes in clinical process/care
- Measuring changes in patient outcomes
- Measuring effects on costs

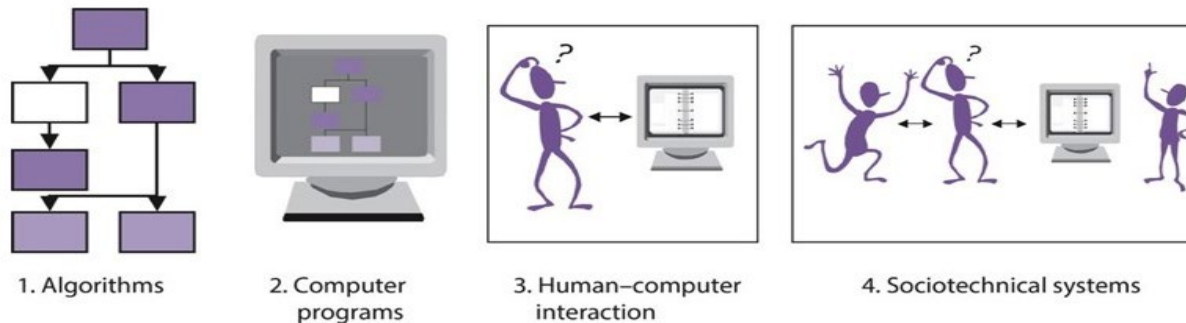
Typical metrics

- System performance against design metrics e.g. creation of reports, patient summaries
- Usability, time to complete tasks
- Data quality and completeness
- User acceptance, confidence
- User compliance with guidelines with and without system
- Quality of care, e.g. time to start treatment
- Measurement of clinical improvement
- Survival rates

Harder

What can be evaluated in HIS

Information and communication systems are sociotechnical systems made up of four different classes of system



1. **Algorithm** – how particular information tasks are to be executed. E.g. how to calculate a specific drug dosage based on renal function and patient's weight) or the type and order of tasks to be executed in a patient pathway
2. **Computer program** – how to carry out each step of the algorithm, data inputs and outputs, functions to manipulate the program
3. **Human-computer interaction** – how acceptable to users, design of interface, task flow etc
4. **Sociotechnical system** – how well a system fits organisational structure and workflow, how it interacts with humans , willingness to adopt and understand systems

Collier 2015,
Chapter 11

Evaluation Types

Formative Evaluation

1. Determine important functionality of and improve system
2. More qualitative methods
3. Usually performed by implementers

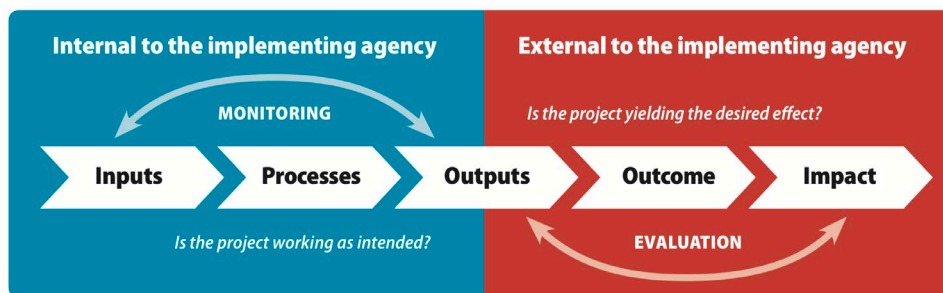
Summative Evaluation

1. Determine benefits and sustainability of system
2. More quantitative methods
3. Usually performed by outside researchers

Example End Points

1. Use of the system
 - Inputs/Processes
2. Confidence in system
 - Outputs
3. *Effect of system* on physician behavior, diagnoses, prescribing etc.
 - Outcomes
4. Changes in patient outcomes
 - Impact

Figure 1.3. Schematic diagram of the interplay between monitoring and evaluation activities



Source: adapted from Pritchett et al. 2013 (4).

Guidelines on higher risk devices

Key questions:

- What are the risks to patients?
- How much are devices under control of health care workers?
- How closely monitored are patients using the device?
- How many patients are users/at risk?

Calculating risk to population

1. %age risk to patient of use of device +ve
2. Close role of clinician -ve
3. High degree of monitoring in environment -ve
4. Larger number of patients using device +ve

Rate the population risk of these examples:

- Infusion pump in ICU setting
- Patient diagnostic aid used in the home

Remote robotic surgery ?



Study Sampling

- Patient characteristics, enrolment criteria
- Study location
 - type of institution
 - stage in clinical process
 - geographical
- Type of sampling
 - consecutive, convenience, selection
- Generalizability?

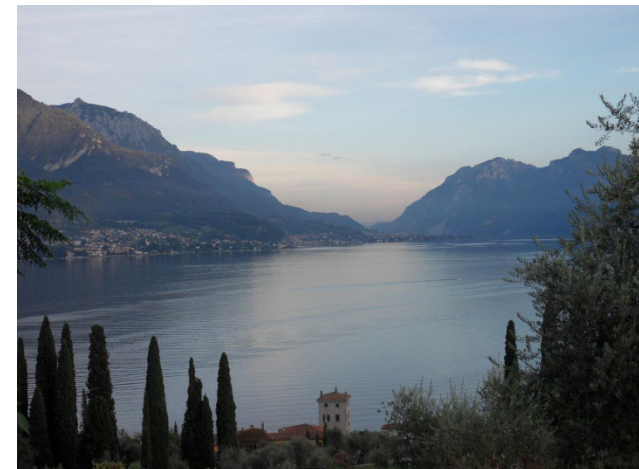
Bellagio Call to Action on Global eHealth Evaluation

Consensus Statement of the WHO Global eHealth Evaluation Meeting,
Bellagio, September 2011

“To improve health and reduce health inequalities, rigorous evaluation of eHealth is necessary to generate evidence and promote the appropriate integration and use of technologies.”



e health
valuation
vidence



Impact of eHealth

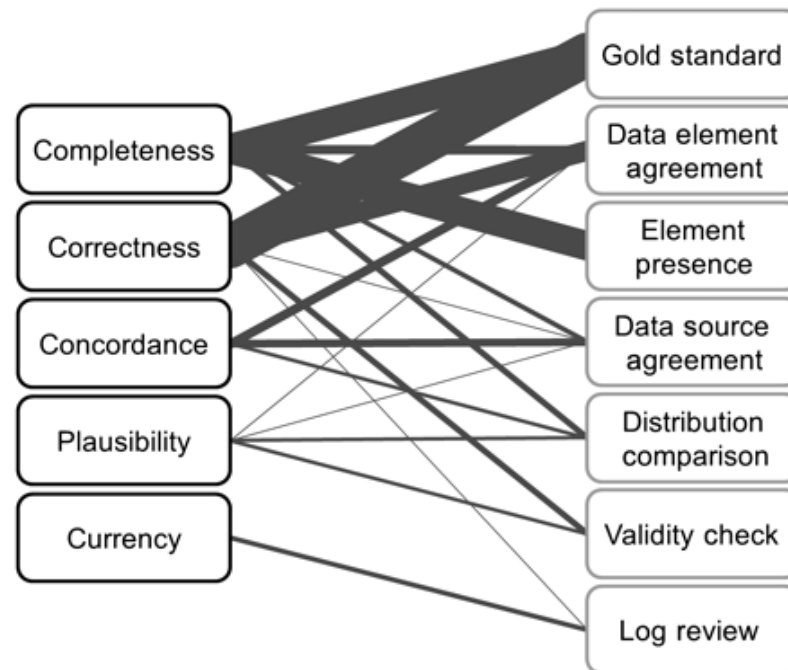
Used appropriately, eHealth has the potential to catalyze, support and monitor health improvements at scale, and to accelerate achievement of national and global development goals, including the United Nations Millennium Development Goals

If used improperly, eHealth may divert valuable resources and even cause harm

1) Core principles underlie the structure, content, and delivery of an eHealth system *independent of the rapidly changing technology used.*



2) High quality data collection, communication and use are central to the benefits of eHealth systems.



Weiskopf, Weng, J Am Med Inform Assoc doi:10.1136/amiajnl-2011-000681

3) Evaluating eHealth both demonstrates its impact and fosters a culture that values evidence and uses it to inform improvements in eHealth deployments.

Utilization- Focused Evaluation



Michael Quinn Patton



4) To ensure the greatest benefit from eHealth and enhance sustainability and scale, eHealth evaluations should recognize and address the needs of all key constituencies.

- Clinician
- Lab technicians
- Patients
- Data entry staff
- Managers
- Researchers

5) Evidence is needed to demonstrate costs and benefits of eHealth implementations, and maximize eHealth's beneficial impact on health system performance and population health.

Decision makers should be encouraged and empowered to utilize evaluation findings to identify and implement solutions with the greatest potential impact on health, and implement them in the most cost-effective, sustainable and practical ways, at scale.

Potential economic benefits of EMR systems in LMICs



OPEN ACCESS

Modeling return on investment for an electronic medical record system in Lilongwe, Malawi

Julia Driessen,¹ Marco Cioffi,² Noor Alide,³ Zach Landis-Lewis,⁴ Gervase Gamadzi,³ Oliver Jintha Gadabu,² Gerald Douglas⁴

¹Department of Health Policy & Management, University of Pittsburgh, Pittsburgh, Pennsylvania, USA

²Baobab Health Trust, Lilongwe, Malawi

³Kamuzu Central Hospital, Lilongwe, Malawi

⁴Center for Health Informatics for the Underserved, Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh, Pennsylvania, USA

Correspondence to

Julia Driessen,
Crabtree Hall A614, Graduate
School of Public Health,
University of Pittsburgh,
130 De Soto Street, Pittsburgh,
PA 15261, USA;
driessen@pitt.edu

ABSTRACT

Objective To model the financial effects of implementing a hospital-wide electronic medical record (EMR) system in a tertiary facility in Malawi.

Materials and Methods We evaluated three areas of impact: length of stay, transcription time, and laboratory use. We collected data on expenditures in these categories under the paper-based (pre-EMR) system, and then estimated reductions in each category based on findings from EMR systems in the USA and backed by ambulatory data from low-income settings. We compared these potential savings accrued over a period of 5 years with the costs of implementing the touchscreen point-of-care EMR system at that site.

Results Estimated cost savings in length of stay, transcription time, and laboratory use totaled US \$284 395 annually. When compared with the costs of installing and sustaining the EMR system, there is a net financial gain by the third year of operation. Over 5 years the estimated net benefit was US\$612 681.

primary barrier to adoption of EMR in the USA, and represent a more significant obstacle in low-income settings.² So far, there has been little consideration of the financial benefits of EMR in the literature, much less their potential cost savings in low-income settings. In this paper we present a prospective analysis of some of the main categories in which a hospital-wide EMR system could generate financial benefits in a low-income setting, modeled on an EMR system in development in a hospital in Lilongwe, Malawi.

While there is interplay between clinical and financial outcomes, analyses of EMR in low-income settings have almost exclusively examined the former. EMR in these settings have been shown to reduce wait times and medication order errors.³ Financial analyses of EMR systems in high-income countries have identified cost savings due to reductions in transcription and chart pulling,



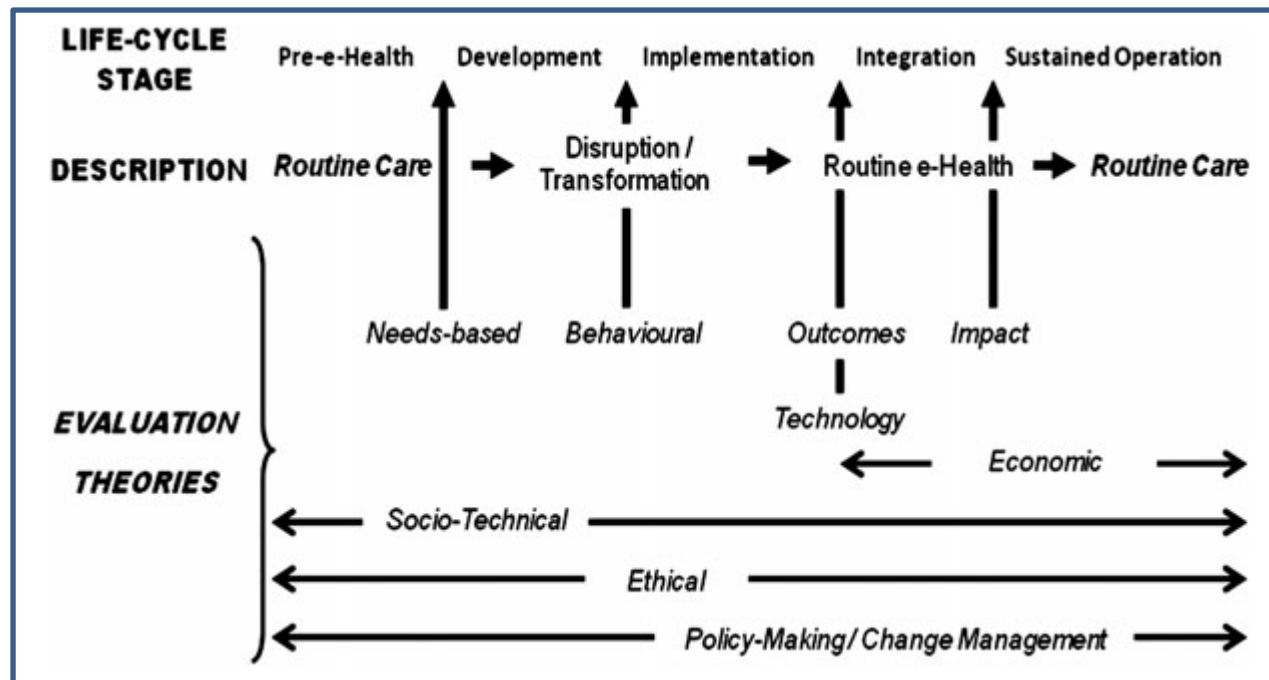
BROWN

Center for
Biomedical Informatics

Advance

Clinical
Translational
Research

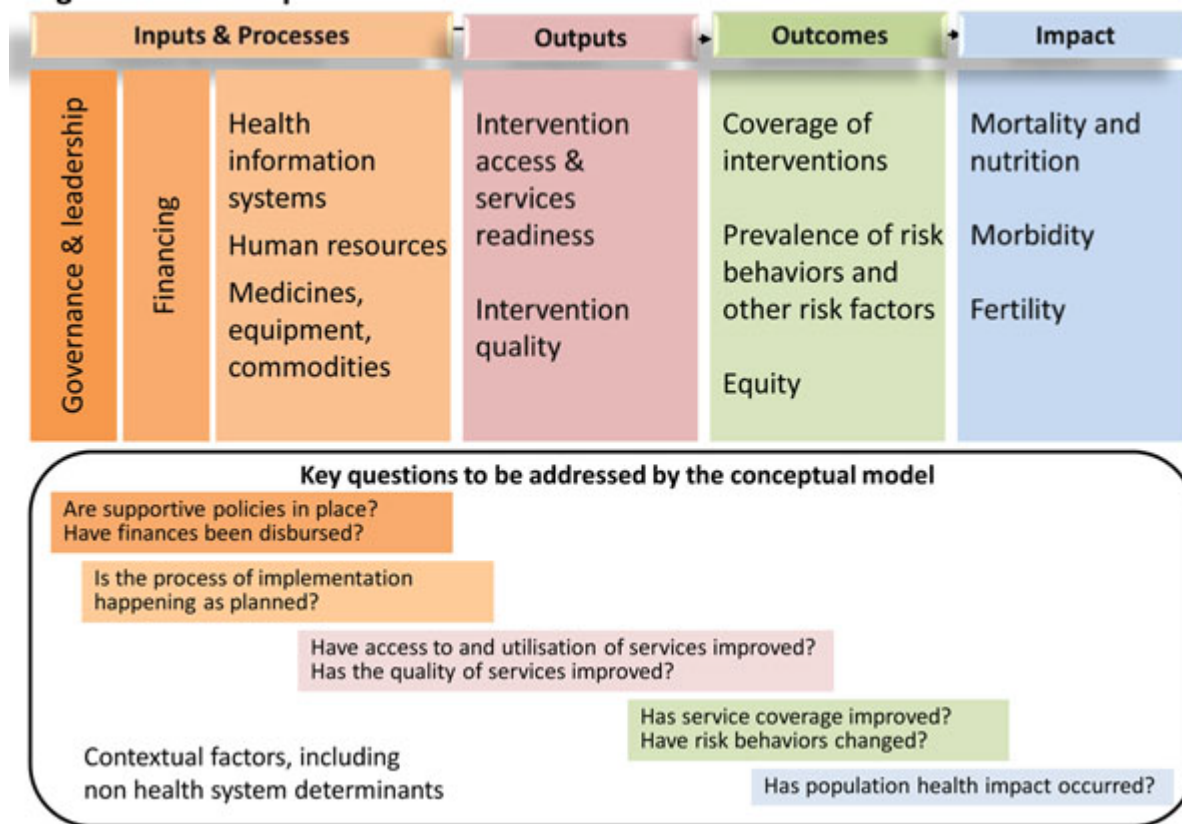
6) The value of a complete evaluation program is enhanced through research that is attuned to the differing requirements throughout the life-course of the project, whether at needs assessment, pilot-, facility level-, regional and national scale-up stages.



7) Independent and objective outcome-focused evaluation represents the ideal of **impact evaluation**.

- *Randomized, controlled trials*
- *Interrupted time series analyses*
- *Natural experiments*

8) Country alignment and commitment to a clear eHealth vision, plan, and evaluation strategy is essential.



A conceptual model for the AHl Collaborative. Adapted from the CHeSS framework [12].
Bryce *et al.* *BMC Health Services Research* 2013 **13**(Suppl 2):S10

9) Improving the eHealth evidence base requires more than increased numbers of studies but also improved quality of eHealth research studies.

OPEN ACCESS Freely available online

PLOS MEDICINE

The Effectiveness of Mobile-Health Technologies to Improve Health Care Service Delivery Processes: A Systematic Review and Meta-Analysis

Caroline Free^{1*}, Gemma Phillips², Louise Watson³, Leandro Galli⁴, Lambert Felix⁵, Phil Edwards³, Vikram Patel³, Andy Haines³

1 Clinical Trials Unit, London School of Hygiene & Tropical Medicine, London, United Kingdom, **2** Department of Health Services Research and Policy, London School of Hygiene & Tropical Medicine, London, United Kingdom, **3** Department of Population Health, London School of Hygiene & Tropical Medicine, London, United Kingdom, **4** Warwick University, Coventry, United Kingdom, **5** Department of Primary Care and Public Health, Imperial College, London, United Kingdom

Abstract

Background: Mobile health interventions could have beneficial effects on health care delivery processes. We aimed to conduct a systematic review of controlled trials of mobile technology interventions to improve health care delivery processes.



BROWN
Center for
Biomedical Informatics



Example End Points

1. Successful generation of reports, alerts etc.
2. Use of the system
3. Confidence in system

Outputs

4. Effect of system on physician behavior, diagnoses, prescribing etc.
5. Effect of system on patient behavior

Outcomes

6. Changes in patient outcomes
7. Changes in costs for health system

Impact

Some questions we still need to answer

- How to get started in an evaluation study
- What evidence exists and does it apply to a particular problem
- What evaluation techniques are available
- How to get accurate and useful results
- Results for whom? Will they create change?
- How to minimize biases

TOM SIMONITE

BUSINESS 06.21.2021 11:00 AM

An Algorithm That Predicts Deadly Infections Is Often Flawed

A study found that a system used to identify cases of sepsis missed most instances and frequently issued false alarms.



Class exercise

Please read the [Wired](#) article and in groups of 2-3 identify answers to the following questions.

- Why are prediction models for sepsis like this one potentially valuable?
- How can they help medical staff diagnose and manage patients?
- What are the concerns about this algorithm and how it is being used?
- How did the researchers identify these problems?
- Would an accurate algorithm work well in the NHS? Why or Why not?
- What other research techniques could help determine if the algorithm embedded in an EHR would improve care?

External Validation of a Widely Implemented Proprietary Sepsis Prediction Model in Hospitalized Patients

Andrew Wong, MD; Erkin Otles, MEng; John P. Donnelly, PhD; Andrew Krumm, PhD; Jeffrey McCullough, PhD; Olivia DeTroyer-Cooley, BSE; Justin Pestruie, MEcon; Marie Phillips, BA; Judy Konye, MSN, RN; Carleen Penzoza, MHSA, RN; Muhammad Ghous, MBBS; Karandeep Singh, MD, MMSc

IMPORTANCE The Epic Sepsis Model (ESM), a proprietary sepsis prediction model, is implemented at hundreds of US hospitals. The ESM's ability to identify patients with sepsis has not been adequately evaluated despite widespread use.

OBJECTIVE To externally validate the ESM in the prediction of sepsis and evaluate its potential clinical value compared with usual care.

DESIGN, SETTING, AND PARTICIPANTS This retrospective cohort study was conducted among 27 697 patients aged 18 years or older admitted to Michigan Medicine, the academic health system of the University of Michigan, Ann Arbor, with 38 455 hospitalizations between December 6, 2018, and October 20, 2019.

EXPOSURE The ESM score, calculated every 15 minutes.

MAIN OUTCOMES AND MEASURES Sepsis, as defined by a composite of (1) the Centers for Disease Control and Prevention surveillance criteria and (2) *International Statistical Classification of Diseases and Related Health Problems, Tenth Revision* diagnostic codes accompanied by 2 systemic inflammatory response syndrome criteria and 1 organ dysfunction criterion within 6 hours of one another. Model discrimination was assessed using the area under the receiver operating characteristic curve at the hospitalization level and with prediction horizons of 4, 8, 12, and 24 hours. Model calibration was evaluated with calibration plots. The potential clinical benefit associated with the ESM was assessed by evaluating the added benefit of the ESM score compared with contemporary clinical practice (based on timely administration of antibiotics). Alert fatigue was evaluated by comparing the clinical value of different alerting strategies.

RESULTS We identified 27 697 patients who had 38 455 hospitalizations (21 904 women [57%]; median age, 56 years [interquartile range, 35-69 years]) meeting inclusion criteria, of whom sepsis occurred in 2552 (7%). The ESM had a hospitalization-level area under the receiver operating characteristic curve of 0.63 (95% CI, 0.62-0.64). The ESM identified 183 of 2552 patients with sepsis (7%) who did not receive timely administration of antibiotics, highlighting the low sensitivity of the ESM in comparison with contemporary clinical practice. The ESM also did not identify 1709 patients with sepsis (67%) despite generating alerts for an ESM score of 6 or higher for 6971 of all 38 455 hospitalized patients (18%), thus creating a large burden of alert fatigue.

CONCLUSIONS AND RELEVANCE This external validation cohort study suggests that the ESM has poor discrimination and calibration in predicting the onset of sepsis. The widespread adoption of the ESM despite its poor performance raises fundamental concerns about sepsis management on a national level.

← Editorial page 1040

+ Multimedia

+ Supplemental content

+ CME Quiz at
jamacmelookup.com and
CME Questions page 1148



Results Wong et al

We identified 27 697 patients who had 38 455 hospitalizations (21 904 women [57%]; median age, 56 years [interquartile range, 35-69 years]) who met inclusion criteria for our study cohort ([Table 1](#)). Sepsis occurred in 2552 of the hospitalizations (7%).

ESM Performance

The ESM had a hospitalization-level AUC of 0.63 (95% CI, 0.62-0.64) ([Table 2](#)). The AUC was between 0.72 (95% CI, 0.72-0.72) and 0.76 (95% CI, 0.75-0.76) when calculated at varying time horizons. At our selected score threshold of 6, the ESM had a hospitalization-level sensitivity of 33%, specificity of 83%, positive predictive value of 12%, and negative predictive value of 95% ([Figure 1](#)). The median lead time between when a patient first exceeded an ESM score of 6 and the onset of sepsis was 2.5 hours (interquartile range, 0.5-15.6 hours) ([Figure 2](#)). The calibration was poor at all time horizons possibly considered by the developer.



Results Wong et al

Evaluation of Potential Clinical Benefit and Alert Fatigue

Of the 2552 hospitalizations with sepsis, 183 (7%) were identified by an ESM score of 6 or higher, but the patient did not receive timely antibiotics (ie, prior to or within 3 hours after sepsis). The ESM did not identify 1709 patients with sepsis (67%), of whom 1030 (60%) still received timely antibiotics.

An ESM score of 6 or higher occurred in 18% of hospitalizations (6971 of 38 455) even when not considering repeated alerts. If the ESM were to generate an alert only once per patient when the score threshold first exceeded 6—a strategy to minimize alerts—then clinicians would still need to evaluate 8 patients to identify a single patient with eventual sepsis ([Table 2](#)). If clinicians were willing to reevaluate patients each time the ESM score exceeded 6 to find patients developing sepsis in the next 4 hours, they would need to evaluate 109 patients to find a single patient with sepsis.

Sensitivity Analysis

When ESM scores up to 3 hours after the onset of sepsis were included, the hospitalization-level AUC improved to 0.80 (95% CI, 0.79-0.81).

It's mainly about people and processes, not technology

The hard stuff is the soft stuff

Dennis Protti 2007

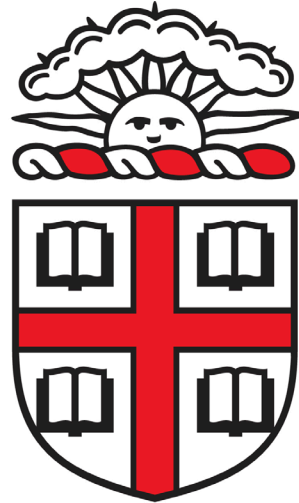
Why is e-Health so hard?

*We need to respect the basic rules of informatics
and invest in e-health experts*

Enrico Coiera 2013

Summary

- Information systems are widely used in health care in many settings
- Good systems can improve care process and quality
- Many system are not optimal in design implementation or evidence base
- Evaluations can be carried out in a wide range of ways – open mind and detective work is key
- A range of evaluation methods exist for different stages in development and scale up
- High quality data is key



BROWN

Center for Biomedical Informatics



BROWN
Center for
Biomedical Informatics

