THE HIGH COST OF LOW VALUE CARE

T. McGinn
Zucker School of Medicine at Hofstra/Northwell, tmcginn@northwell.edu

S. Cohen
S. Khan

S. Richardson
Zucker School of Medicine at Hofstra/Northwell, srichard12@northwell.edu

M. Oppenheim

See next page for additional authors

Follow this and additional works at: https://academicworks.medicine.hofstra.edu/articles

Part of the Internal Medicine Commons

Recommended Citation

This Article is brought to you for free and open access by Donald and Barbara Zucker School of Medicine Academic Works. It has been accepted for inclusion in Journal Articles by an authorized administrator of Donald and Barbara Zucker School of Medicine Academic Works. For more information, please contact academicworks@hofstra.edu.
THE HIGH COST OF LOW VALUE CARE

THOMAS MCGINN, MD, STUART COHEN, MD, SUNDAS KHAN, MD, SAFIYA RICHARDSON, MD, MPH, MICHAEL OPPENHEIM, MD, and JASON WANG, PHD

THOMAS MCGINN, NEW HYDE PARK, NEW YORK;

Correspondence and reprint requests: Thomas McGinn, MD, MPH, 2000 Marcus Avenue, 3rd Floor, New Hyde Park, New York 11042, Phone: 516-321-6049, Fax: 516-600-1756, tmcginn@northwell.edu

Copyright © 2019 The American Clinical and Climatological Association

Abstract

The main focus of this study is bridging the “evidence gap” between frontline decision-making in health care and the actual evidence, with the hope of reducing unnecessary diagnostic testing and treatments. From our work in pulmonary embolism (PE) and over ordering of computed tomography pulmonary angiography, we integrated the highly validated Wells' criteria into the electronic health record at two of our major academic tertiary hospitals. The Wells' clinical decision support tool triggered for patients being evaluated for PE and therefore determined a patients' pretest probability for having a PE. There were 12,759 patient visits representing 11,836 patients, 51% had no D-dimer, 41% had a negative D-dimer, and 9% had a positive D-dimer. Our study gave us an opportunity to determine which patients were very low probabilities for PE, with no need for further testing.

INTRODUCTION

A fundamental problem with health care delivery in the United States is that a large percentage of the care delivered is not indicated and often harmful. U.S. citizens receive only half of recommended medical care (1) and one-third of the care received is unnecessary (2). In addition, in many clinical situations there is readily available evidence to inform proper decision-making to help reduce unnecessary testing and treatments, but that evidence is not used. This juxtaposition of patients receiving tests or treatments not indicated when available evidence exists to inform the decision not to perform those tests or treatment we classify as the “evidence gap” (Figure 1).

Bridging the evidence gap between frontline decision-making in health care and the published evidence with the hope of reducing unnecessary diagnostic testing and treatments is the main focus of our Center for Predictive Medicine at Northwell Health. Our center's particular focus is on the integration of well-validated clinical prediction rules into the point-of-care through a process of usability testing. A clinical prediction rule (CPR) is a decision aid that brings together various components of the history, physical exam, and other easily obtainable lab data and quantifies them as to their ability to predict certain outcomes. CPRs are typically developed to help with quick frontline decisions in clinical care (3).

The process of deriving and validating CPRs is complex and includes multiple steps (Figure 2). The first step in the process is the derivation study where the rule is created often using a retrospective database. Once a CPR is derived, it must be both retrospectively and prospectively validated on multiple patient populations. Ideally, these patient populations will vary in important demographics, prevalence, and
severity of the condition being studied. Finally, once a CPR is validated, the third step is an impact analysis which is performed to determine the true impact of the CPR on care delivery, once deployed. Impact analysis can determine such things as cost avoidance, cost increase, as well as doctor and patient satisfaction with using the CPR.

After years of deriving and validating CPRs, it became clear that although these tools had tremendous potential to reduce gaps in care, few providers were actually adopting them at the point of care. This problem of adoption launched the research projects called “Integrated Clinical Prediction Rules (iCPR).” The goal of iCPR (4-6) studies is to take well-validated CPRs and integrate them into clinical workflow to enhance adoption and to then measure the impact of that process on test ordering or treatment selection. Our process for integrating CPRs at the point of care has multiple steps (Figure 3) (7). The first step includes a vigorous assessment of the quality and validity of the CPR to ensure it is of the highest validity and should therefore be broadly adopted into the decision-making process throughout health care. The second step involves looking for areas of potential overuse and underuse of both treatments and diagnostic tests where a CPR might enhance the evidence-based decision-making process. The third step in our iCPR process is assessing whether a CPR can be easily integrated into the workflow and electronic health record. Some CPRs are too complex for integration.

After applying these three basic parameters, a high level of evidence CPR, a clinical area of overuse or underuse of evidence-based health care, and finally the ease or difficulty of integration of the CPR into the flow of care, our group identified two clinical areas which required attention. The first area was antibiotic prescribing for upper respiratory infections in primary care settings and the second area was the use of computed tomography pulmonary angiography (CTPA) in patients suspected for pulmonary embolism (PE) in emergency room settings. In both instances there are well-validated CPRs, significant over prescribing of antibiotics, and over ordering of CTPAs, and we believed that with appropriate usability testing the CPR could be integrated into the point of care (8,9).

For our work in PE and over ordering of CTPAs, we integrated the highly validated Wells' criteria (10-12) (Table 1) into the electronic health record at two our major academic tertiary hospitals (13). Use of CTPA in the assessment for PE in the emergency department (ED) has significantly increased over the past 2 decades. Whereas this has improved the accuracy of radiologic testing for PE, CTPA also carries the risk of significant iatrogenic harm (14) including a 14% risk of contrast-induced nephropathy (15) and a lifetime malignancy risk of 2.76% in young female patients (16). In 2006, results from the landmark PIOPED II trial established CTPA as the first choice diagnostic imaging modality, with sensitivity over 90% for patients with high clinical suspicion of PE and specificity of 96% (17,18). Over the next 5 years, there was a 4-fold increase in CTPA use and a 33% decrease in ventilation/perfusion imaging (19).

As part of an ongoing study, our research team developed a Wells' criteria clinical decision support (CDS) tool with a user-centered design method by applying our previous experience in creating these tools and tailoring for the ED workflow and environment (9,13,20). Using adaptive principles in web and health information technology design, this tool was developed in three phases: 1) focus-group feedback on documented workflow and compliance barriers; 2) key-informant interviews on the tool's wireframes; and 3) iterative rounds of usability testing of the initial prototype. Following the three phases, the Wells' CDS tool was integrated into the order entry workflow and was triggered when a provider attempted to order tests that are used to evaluate a suspected PE (i.e., CTPA or D-dimer). With the launch of the CDS tool, a tracking tool was developed to monitor the Wells' CDS tool post-implementation. As the provider used the CDS tool to place orders for PE evaluation, the tracking tool monitored the number of times the CDS tool was completed which was defined as completion of the Wells' risk assessment score for PE, following recommendations, and accompanying orders.
CTPA is not recommended in patients who are deemed a low probability for PE. However, we began to realize that a large percentage of low-probability patients had a CTPA performed and were being recommended for follow-up studies based on incidental findings of their CTPA. In an attempt to assess the complete nature of this, we retrospectively characterized the rate of incidental findings on low-probability CTPA patients discharged home after the workup of PE. Highlighted below are the preliminary results of our ongoing study.

MATERIALS AND METHODS

As part of a larger ongoing cost-effectiveness study, we performed a retrospective analysis on all ED patients being evaluated for suspected pulmonary embolism seen at our two tertiary academic hospitals from August 1, 2015, to July 31, 2017. The Wells’ CPR CDS tool triggered for patients being evaluated for PE and therefore determined a patients’ pretest probability for having had a PE.

Each CDS visit, the first visit per patient was analyzed. Low-probability patients were determined based on a negative D-dimer during the visit (<500 ng/mL). Very-low-probability patients were determined based on CDS patients with a negative D-dimer who were discharged home. For those with negative D-dimer who were discharged home, all CTPAs within 24 hours of a negative D-dimer were evaluated for computed tomography findings. This was chosen to ensure that the physician had the negative D-dimer before ordering the test. The percent of patients with positive PE was determined for each group. For each PE visit, PE diagnosis was determined based on diagnosis code, ICD 9 415.xx, ICD 10 I26., as principal or secondary ED discharge diagnosis. Computed tomography angiography (CTA) chest was defined as any CTA chest without concurrent CTA neck, upper extremity, or abdomen. All CTA chest reports were manually confirmed to be CTPA and not CTA aorta by a cardiothoracic radiologist.

A fellowship-trained cardiothoracic radiologist performed manual review of the reports of all very-low-risk CDS patients with CTPA within 24 after D-dimer to account for incidental findings. CTPA follow-up recommendations were categorized into nine categories of follow-up (Table 2).

RESULTS

There were 12,759 patient visits representing 11,836 patients (Table 3). Of all the visits, 51% (6,456 of 12,759) had no D-dimer, 41% (5,195 of 12,759) had a negative D-dimer, and 9% (1,108 of 12,759) had a positive D-dimer. For the 11,836 patients, 5,068 were patients in the low-probability category and 3,645 were in the very-low-probability category. Of the very-low-probability group, 877 had a CTPA within 24 hours after a negative D-dimer.

Review of the follow-up recommendations for the 877 low-probability patients resulted in 79.8% (200 of 877) with no follow-up, 17.6% (154 of 877) had one item follow-up recommended, 2.5% (2 of 877) had two items recommended, and 0.1% (1 of 877) had three items recommended for follow-up. Test recommendation follow-ups are summarized in Table 4.

DISCUSSION

Avoiding unnecessary medical testing is important for several reasons. There are the more obvious reasons such as cost to the patient, health systems, and the country. In addition, there is the potential harm and inconvenience of the test to the patient. But a more concerning issue regarding expense and harm is the potential for unnecessary follow-up tests that may need to be performed but could have been avoided if the test had never been performed. Our study gave us a unique opportunity by having available pre-test probability estimates; therefore, we could determine which patients were very-low-probability for PE and not in need of further testing.
We defined unnecessary CTPA scans as patients receiving the scan who were known to be very-low-probability for having had a PE and therefore unlikely to benefit from testing with a CTPA. Beyond cost and inconvenience, there is potential for direct harm related to dye load and nephropathy, radiation exposure leading to increased cancer risks, and the potential for unnecessary biopsies for false-positive findings. In addition, there is the possibility for indirect harm such as emotional distress including anxiety and depression related to the false-positive study (21).

Typically, low-risk patients can be defined by the Wells' criteria, but in our retrospective review we defined it by a negative D-dimer. This was confirmed by the low PE diagnosis rate of 1.05% and that those patients discharged home had a PE diagnosis rate of 0.08%. Wells' criteria were not used because of lack of complete information. CDS patients were used because despite the lack of Wells' criteria data; it insured that patients were being evaluated for PE. Previous published work evaluating the effectiveness of the Wells' criteria found similar rates of 0% PE diagnosis rate in patients with a low Wells' score and a negative D-dimer (22).

Physicians are testing more for PE and seem to be finding and treating clinically insignificant PEs. In addition, CDS tools built to estimate pre-test probability of PE and discourage the CTPA use in low-risk patients have been shown to improve the CTPA yield (14). We have evidence to help reduce unnecessary testing and we must support studies to determine the best way to seamlessly integrate that evidence into care.

ACKNOWLEDGMENTS

This study was supported by the U.S. Department of Health and Human Services, Agency for Healthcare Research and Quality 1R24HS022061-01.

Footnotes

Potential Conflicts of Interest: None disclosed.

Contributor Information

THOMAS MCGINN, NEW HYDE PARK, NEW YORK.

STUART COHEN, NEW HYDE PARK, NEW YORK.

SUNDAS KHAN, NEW HYDE PARK, NEW YORK.

SAFIYA RICHARDSON, NEW HYDE PARK, NEW YORK.

MICHAEL OPPENHEIM, NEW HYDE PARK, NEW YORK.

JASON WANG, NEW HYDE PARK, NEW YORK.

DISCUSSION

Selker, Boston: That was a very astute example of what a learning health system should actually be — where you're actually learning stuff. You're applying science and hopefully improving care. At the rate we make clinical predictions, by the next millennium we'll have sufficient numbers to actually affect care. What is a way that we can actually get prediction support in a way that is timely (i.e., in our lifetime)?

McGinn, Manhasset: I'm a firm believer that just using large datasets and doing retrospective validation is not going to take the day. When I first arrived in Northwell we derived our prediction rule and we validated our prediction rule on clots and bleeds. It took me about a year and a half just to do the
derivation and retrospective study. Then, a couple of years later, I could do the same dataset in about 3 months. So, the speed in which we can derive and retrospectively validate I think is astronomically improved.

Golden, Baltimore: I was looking at your EMR and it looked like EPIC. You don't have to comment one way or the other but that's what we….

McGinn, Manhasset: I have one comment about EMRs. You know they all really suck? Some suck more than others.

Golden, Baltimore: So, I won't comment about that part specifically, but in the Johns Hopkins Health System we do use EPIC. We have developed some insulin prescribing clinical decision support tools. I was really impressed with the degree of uptake you had in usage. So, my question is how did you integrate it into EPIC so that it was actually easy for providers to use it?

McGinn, Manhasset: I have a whole talk on usability — the science of integration. The biggest problem we have right now in electronic health records is people don't think it's a science — to create methods by which the doctor and the patient in the computer can interface. The first year of our study was just the integration of EHR and doing it in a way that we could adopt. We had simulation lab usage. We had patients coming in and doctors testing it. These days, people just shove things into the computer and then people click through it. So, the key is time, effort and design.

Thibault, Boston: You've got a new medical school there doing innovative things. How are you and Larry Smith incorporating this into how you're teaching the next generation of doctors?

McGinn, Manhasset: When you start introducing some of these tools you do it the right way. They are sophisticated users. So, if you talk about design, they get really angry very quickly if you don't know how to design these things. But once they learn how to use them and they like them, it's actually hard to do randomized trials because they find the tools and use them. So, they are really some of the great partners in creating these things and designing them….and they love it. Also, this a way to get people to do a history and physical, because the prediction models are basically good solid approaches of history and physical exam findings and trying to take them away from just ordering diagnostic tests inappropriately. So, we use it as a great teaching tool.

Petersen, Houston: A lot of the reasons we physicians do things, or don't do things, are very complicated and have more to do with the field of implementation science than they do with evidence. Sometimes we know evidence, but there are other reasons, whether they're patient-driven or system-driven. So, what are your thoughts about other ways to overcome this?

McGinn, Manhasset: Another body of work that we look at is what we call doctor-patient evidence congruence. Take the Ottawa ankle rule for ankle fracture. Doctors wanted help to protect them from the demands of their patients for getting the x-rays. So, how do you help them. If there's not congruence between what the doctor really wants, and the evidence adoption is poor? They don't want to prescribe an antibiotic, so you can help them. But in cardiac work and chest pain workup they don't want to be told not to admit the patient. So, then you have to work on the doctor's side of the equation. It's a cultural phenomenon. That's the whole usability technique; to get into the mind of the physician and why they're resisting the adoption. One also has to examine the general work area. ERs are very difficult to work in; you've got PAs and nurses. If you don't take into account the cultural environment, it doesn't matter. The electronic health record is a small, tiny piece of the puzzle. We usually do environmental workflow analysis and cultural analysis and other things before we go in.

Del Rio, Atlanta: On your data, you say there's a lot of things that we discover. When we start to chasing something, maybe some of those chasing things are actually good. It'll be interesting to see what is a false-positive versus a true new incidental finding that actually changes outcomes on a patient. The other thing
is waste. We all quote there's 30% of waste. But the problem with waste is you don't know it's waste until later. No physician I know does a test and says, “Oh this is a waste of time.” The waste happens afterwards that you look back. How good is the prediction model in actually helping people in that decision-making?

**McGinn, Manhasset:** That's what a prediction model is supposed to do. The Wells' criteria with a negative D-dimer is an amazingly accurate predictor of not finding a PE. You don't need to do anything else after that. Now it doesn't know that you happen to have a small lung nodule that could end up being an early stage cancer. But you weren't ordering the test for that reason. That's the whole job of prediction model. It is to tell you with a high degree of confidence there's no ankle fracture, there's no head fracture, there's no pneumonia… and they do a very good job of it.

**Sherman, Houston:** Your talk started with a one-sided hypothesis of overutilization. Can you discuss the application of these prediction rules to areas of notorious underutilization, under-diagnosis and under-treatment, like cardiac disease in women?

**McGinn, Manhasset:** Rachel Burger at the University of Pittsburgh described the under-detection of infant child abuse. We created two prediction models that are now integrated in all of Pennsylvania and now rolling out in New York. So, it can work on increasing use of tests and of treatments. Some prediction models can highlight missed increased risk and show the need for a test or treatment that has not been ordered enough.

**REFERENCES**


Figures and Tables
Fig. 1.

Bridging the “EVIDENCE GAP.”
Fig. 2.

Development of a clinical prediction rule.
Fig. 3.

Steps to complex clinical decision support tool integration.

Open in a separate window
TABLE 1.
Wells' CPR for Pulmonary Embolism

<table>
<thead>
<tr>
<th>Clinical Characteristics of PE</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinical signs and symptoms of DVT (minimum of leg swelling and pain with palpation of the deep veins)</td>
<td>3</td>
</tr>
<tr>
<td>PE as or more likely than an alternative diagnosis</td>
<td>3</td>
</tr>
<tr>
<td>Heart rate greater than 100</td>
<td>1.5</td>
</tr>
<tr>
<td>Immobilization in the past 4 days or surgery in the previous 4 weeks</td>
<td>1.5</td>
</tr>
<tr>
<td>Previous DVT/PE</td>
<td>1.5</td>
</tr>
<tr>
<td>Hemoptysis</td>
<td>1</td>
</tr>
<tr>
<td>Malignancy</td>
<td>1</td>
</tr>
</tbody>
</table>

PE, pulmonary embolism; DVT, deep vein thrombosis.
### TABLE 2.

Categories of Recommended Follow-Up

<table>
<thead>
<tr>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lung nodule</td>
</tr>
<tr>
<td>“Other” chest finding</td>
</tr>
<tr>
<td>Upper abdominal finding</td>
</tr>
<tr>
<td>Thyroid finding</td>
</tr>
<tr>
<td>Breast, cardiac, musculoskeletal, vascular finding</td>
</tr>
</tbody>
</table>
TABLE 3.
Risk Categories and Percentage of PE Diagnosis

<table>
<thead>
<tr>
<th>Category</th>
<th>Number Patients</th>
<th>% PE Diagnosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDS patients</td>
<td>11,836</td>
<td>5.86%</td>
</tr>
<tr>
<td>Low prob (CDS and negative D-dimer)</td>
<td>5,068</td>
<td>1.05%</td>
</tr>
<tr>
<td>Very low prob (CDS and neg D-dimer and + discharged home)</td>
<td>3,645</td>
<td>0.08%</td>
</tr>
<tr>
<td>Very low prob with CTPA within 24 hours D-Dimer</td>
<td>877</td>
<td>0.34%</td>
</tr>
</tbody>
</table>

CDS, clinical decision support; PE, pulmonary embolism; CTPA, computed tomography pulmonary angiograph; prob, probability.
### TABLE 4.
Test Recommendation Follow-Ups

<table>
<thead>
<tr>
<th>Name of Test</th>
<th>Number of Tests</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chest CT scans for nodules</td>
<td>69</td>
</tr>
<tr>
<td>Thyroid follow-up</td>
<td>33</td>
</tr>
<tr>
<td>Non nodule chest follow-ups recommended (CT, PET-CT, CT versus MRI, lab values, or follow-up to resolution)</td>
<td>31</td>
</tr>
<tr>
<td>Abdominal finding (clinical correlation, abd MRI, abd CT, abd US, abd CT or MRI, dedicated liver imaging, or endoscopy)</td>
<td>30</td>
</tr>
<tr>
<td>Breast follow-up</td>
<td>19</td>
</tr>
<tr>
<td>Repeat CTPA</td>
<td>7</td>
</tr>
<tr>
<td>Vascular studies for follow-up (neck US, MRI or CTA, aneurysm follow-up, vascular follow-up, or clinical correlation)</td>
<td>5</td>
</tr>
<tr>
<td>Echocardiograms</td>
<td>4</td>
</tr>
<tr>
<td>Musculoskeletal follow-up (2 spine MRIs, 1 chest MRI)</td>
<td>3</td>
</tr>
</tbody>
</table>

abd, abdominal; CT, computed tomography; PET, positron-emission tomography; MRI, magnetic resonance imaging; US, ultrasound; CTPA, computed tomography pulmonary angiograph; CTA, computed tomography angiography.