The short title of our first article is “High-Risk Breast Lesions: [A] Machine Learning Model to Predict Pathologic Upgrade...” The authors are from Massachusetts General Hospital and MIT in Boston. The first author is Dr. Manisha Bahl; the senior author is Dr. Connie Lehman. Dr. Lehman is in particular one of the leading experts in breast imaging in the United States, and is a frequent contributor to Radiology. Quite a few years ago, Dr. Lehman and I were co-investigators in a large multi-center trial led by Dr. Mitchell Schnall during the early days of breast MRI. Those studies led to the acceptance of breast MRI for evaluation of women who were at high risk for breast cancer. I noticed this article by Dr. Lehman’s group has an extraordinarily high Altmetric score of more than 700. The Altmetric score assesses the immediate newsworthiness of the article, as assessed by social medial, news outlets and other appearances on the web. That score of more than 700 is really high. I guess the new term is “buzz worthy.” Let’s see what this is all about.

(Background) The topic is important. A woman has a suspicious mammogram requiring a biopsy. The biopsy does not show evidence of invasive cancer or DCIS. However, the biopsy reveals histology that is considered “high risk” due to its association with breast cancer. For example, there may be some adjacent tissue that show atypical ductal hyperplasia, ADH. ADH is a “high risk” finding on histology.

Even though the patient has a benign biopsy, the patient with ADH on the needle biopsy is very likely to have the lesion surgically removed. Why? The thinking is that when ADH is present, then cancer may be lurking nearby. The needle biopsy samples only a small amount of tissue. The needle biopsy simply could have missed the cancerous lesion.

If we consider all types of “high-risk” pathology, about 14% of patients, with benign biopsies but with high-risk pathology may be upgraded to cancer after surgical excision. The highest rate of cancer upgrade is for ADH. Other benign but “high-risk” lesions include (in order of highest risk) lobular carcinoma in situ, radial scar, papilloma, nonspecific atypia, and biphasic neoplasm.

So 14% of lesions may be upgraded to cancer, but that leaves 86% of women with unnecessary surgical treatment. At present, there is no way to tell if you are in the group of 14% who might have invasive cancer or DCIS, or in the majority 86% of women who do not have cancer.

Well – this is the age of Google, Facebook, Apple, and Amazon. Can a computer algorithm help us?

Purpose: The purpose of this paper was to see if machine learning can be used to sift through the results from the image-guided needle biopsy, to see if the computer can predict which patients really need surgical excision.

The authors’ institution had a strong track record of patients with high-risk lesions on needle biopsy who routinely undergo surgical excision. They looked back at their mammography database over about 10 years from 2006 to 2015 and correlated that database with their pathology database from both the image-guided biopsy and the eventual surgical excision. Besides image features on the digital mammogram or tomosynthesis and pathology, there was lots of other available information for the computer to analyze. The electronic medical record contained information on a tremendous number of patient traits, including age, age at first menses or menopause, ancestry, height, weight, smoking habits, etc. All of this data was fed into a machine learning computer algorithm. The algorithm was a random forest classifier.

You will see this term of a random forest algorithm in a lot of our latest research articles. What exactly is it? It’s possible that this specific computer algorithm will influence your medical treatments, or determine what products Amazon advertises to you. So let’s take just a few minutes to review this term.

If you do a Google search on the term “random forest model,” there are 8.4 million results that pop up. It is a hot topic. Why? A random forest classifier is a computer algorithm that is able to sift through a huge database starting in small pieces, and eventually trying to predict an outcome or response. In our study, we want to predict surgical upgrade to cancer, of a needle biopsy lesion that was benign but at high risk. How is this done?

The random forest classifier algorithm repeatedly selects a random set of a limited number of features from the dataset. That smaller set of features, such as age, gender and breast density, might be used to predict a high-risk lesion. But we can also use another set of randomly chosen features, maybe weight, family history and history of breast cancer. Each small group of factors tries to predict cancer. We take the average of all of the individual predictions, and that average of all the small groups turns out to be a very good and robust predictor of cancer.

One quick analogy on random forest prediction. Let’s say you are interviewing for a job. Instead of interviewing 1-on-1...
with each person, instead there are small, randomly selected interview groups each with five people. It stands to reason that each of those interview groups will probably ask you different questions. And they might differ, for example, as to whether you get the job or not. At the end of the day, you’ve interviewed with five different groups. Each group votes to see if you get the job. That is the concept of the random forest. Each small interview group gets a vote. Their vote is based on several different, relatively random features that they learned about you.

Back to breast cancer. In the current study, the authors had 1,006 high-risk lesions, such as ADH or lobular carcinoma in situ. To train the computer random forest algorithm, investigators used their first 671 cases (or about 2/3 of their data). Then, they use the remaining 1/3 of high-risk lesions to test the random forest algorithm to see if the results remain consistent.

There were about 20,000 data elements for the computer to sift through in the 1,006 patients. Those data elements included not only the mammographic findings, but also demographics from the health record as well as actual phrases from the pathology report. The authors ranked all of the 200 random decision trees. I guess that’s 200 separate job interview committees, and each committee had 12 different interviewers. That’s a lot of combinations. To decide on the best prediction algorithm, the authors set a threshold that the algorithm was supposed to identify essentially all of the patients who would be upgraded at surgery. They did not want to miss anyone.

So what are the results? Well, a disadvantage of the random forest method is that it is a black box. We do not know exactly how each committee votes. On the other hand, the computer algorithm ranks the most important features at the end of the analysis. The most important features included patient age, path results from the needle biopsy, and text from the pathology report, such as severely atypical findings. Not so surprising I suppose. But the key is how to combine all of those findings mathematically to make a prediction of surgical upgrade.

Next, the authors compared the new random forest computer algorithm to their current practice. Remember at MGH, essentially all of the high-risk lesions were excised, and 14% of patients were upgraded to cancer after that biopsy. On the other hand, if the computer algorithm made the decision for surgery or not, 30% fewer surgeries would have been performed on benign lesions. And for patients for whom the computer said should have surgery, 37 of 38 malignancies would have been diagnosed at surgery.

Yes, thank goodness, the computer missed one patient that her doctors would not have missed. That patient had Cowden syndrome. Cowden syndrome is characterized by multiple hamartomas on the skin and mucous membranes. But in Cowden syndrome, there is also an increased risk of breast cancer. The authors did not reveal to the computer that Cowden syndrome was associated with breast cancer. Why not? Well, there was probably only one person in the entire cohort with Cowden, so not much data was present. If the database were even larger than 1,000 patients, probably these genetic predispositions would be caught by the computer in the future.

Conclusion: Computer algorithms such as a random forest algorithm can help us sort through large datasets. They can help us identify predictors of disease that are very robust and that we might have otherwise overlooked. If we apply this to the entire medical record, on average, it seems like medical care would be improved. In the case of Dr. Lehman’s group, about 30% of women would not have to have biopsy compared to their current standard of care.

I also want to mention an excellent editorial on the Lehman article. It is written by Dr. Kitt Shaffer from Boston Medical Center. Dr. Shaffer is a professor of radiology who is very interested in information technology. She is the vice chair for education in her department. Her editorial is enlightening and I hope you have a chance to read it in the March 2018 issue of Radiology.

**Generalist versus Subspecialist Characteristics of the U.S. Radiologist Workforce**


Andrew B. Rosenkrantz, MD, MPA, • Wenyi Wang, MA, • Danny R. Hughes, PhD, • Richard Duszak, Jr, MD

Our second research article is a little bit different than our usual hard-core science. But I thought we needed a break from computer algorithms. The title is “Generalist versus Subspecialist Characteristics of the U.S. Radiologist Workforce.” The first author is Dr. Andrew Rosenkrantz from NYU. Andrew is extremely prolific and energetic and a very good author. The senior author is Dr. Richard Duszak from Emory. Dr. Duszak is nationally recognized for his work in healthcare policy. So let’s see what this talented group of authors has to say.

What is the background? We would like to know the composition of our specialty. It is useful for planning purposes, and tells each of us about our jobs, relative to others in our profession. For subspecialization, the U.S. government only classifies radiologists as either nuclear medicine, interventional radiology, or general radiology. That’s not very good information. And as you will see in the following discussion, a workforce survey tells us about the youngest members in our profession and their interests.

What was done? The authors evaluated claims data from all radiologists who participated in Medicare billing. Medicare is generally for patients who are 65 years or older. So this approach will unfortunately not give us information about the pediatric subspecialty. The authors used the RVU data that was submitted and they classified that billing data into previously validated subspecialty groups between 2012 and 2014.

The RVU data is interesting. A radiologist may have a particular interest in musculoskeletal imaging. But if he or she is reading only plain films in that area, it does not add up to a lot of RVU workload. So the definition of subspecialty in this study is not based on an individual’s own interests, but rather what he or she gets paid to do.

The results: In total, the authors analyzed the workload of more than 33,000 radiologists in the United States. There was previous work done to define a specialist in radiology. That definition is 50% of billed workload RVUs in a single specialty. So, I would be designated as a subspecialist in cardiothoracic imaging if 50% of my RVU workload is in the specialty area of cardiothoracic.

So, in the U.S., 55% of radiologists were found to be classified as generalists. You may think you are specializing, but in reality, your RVU workload is in more than one specialty. 45% of radiologists were classified as subspecialists.
In the U.S., about 10% of radiologists specialized in neuroradiology, 8% were breast imagers. Abdominal and interventional were both about 5%. Only 2% were in nuclear medicine.

Other facts and figures: For women, nearly a quarter, or 24% of all women subspecialize in breast imaging. I was not aware the number was so high. That amounts to about 2/3 of subspecialists in breast imaging are women.

On the other hand, only about 3% of women are in interventional radiology as specialists. That seems remarkably low.

What about the younger radiologists? Younger radiologists tend to specialize more, and one of the highest rates of subspecialization was in cardiothoracic imaging at 9%, which was similar to the number performing neuroradiology. For these younger physicians, almost none are subspecializing in nuclear medicine, less than 1% of the young physicians.

Otherwise, the results are what you expect. In academic practices, 76% of the radiologists were specialized, versus 36% of the non-academic radiologists. In the United States, the northeast and west coast areas had more specialists than, for example, the Midwest. Finally, for radiologists in large-group practices of more than 100 radiologists, the subspecialization rate was much higher. 63% of radiologists had a subspecialty compared to about 1/3 of radiologists in smaller group practices.

Conclusion: That’s probably enough numbers for now. Younger people are choosing cardiothoracic imaging over specialties such as nuclear medicine. We have a striking lack of gender diversity in fields such as interventional radiology.

For me, it is also interesting that the majority of radiologists in the U.S. are general radiologists based on their RVU workload. That means that you need to keep up with developments in multiple specialties. When I took this position, my impression was that the articles in *Radiology* were becoming very complex – perhaps understood by a small number of radiologists in the entire world. One result of this is these podcasts. We want to try to decipher the complex articles into discussion points to keep us all up to date.

A second consequence of this is subspecialty journals. The RSNA will publish subspecialty journals in cancer, artificial intelligence, and cardiothoracic imaging starting in less than a year. If I’m a specialist and need to know the details, I will probably go to those journals – but I also need to keep up with topics in radiology in general. Over the coming months to years, we hope to have articles in *Radiology* on topics and innovations that all of us in our profession should know. Yes, still specialists will be better at understanding the detail, but a breast imager should probably know about fractional anisotropy in the brain. A neuroradiologist should have some knowledge about national healthcare programs in lung cancer or breast cancer screening.

You can read the details of this article by Dr. Rosenkrantz and his colleagues in the March 2018 issue of *Radiology*.

---

**Technical Developments: Zero Echo Time Imaging of the Shoulder: Enhanced Osseous Detail by Using MR Imaging**


Ryan E. Breighner, PhD, • Yoshimi Endo, MD, • Gabrielle P. Konin, MD, • Lawrence V. Gulotta, MD, • Matthew F. Koff, PhD, • Hollis G. Potter, MD

The next article is about a technical development in MSK imaging that should be of general interest about MRI. The short title is, “Zero Echo Time Imaging of the Shoulder.” The study was done at the Hospital for Special Surgery in New York. The first author is Dr. Ryan Breighner. Ryan is a computational scientist. The senior author is Dr. Hollis Potter. It’s almost certain you have heard Dr. Potter lecture at RSNA. She is a prolific researcher and chair of her department at the Hospital for Special Surgery.

Background: The background and rationale for the study is straightforward. The authors wanted to see if there was a way to get a CT-like bone image on an MRI. As you know, on conventional MRI, there is no signal from cortical bone. Lack of cortical bone detail in musculoskeletal imaging often means that the patient needs both an MRI and a CT scan. Is there a way to get bone-like detail on the MRI?

One clue that you know of is about PET/ MRI. In that modality a major problem has been that bone density is not available. Bone density correction on PET is the main reason why PET/ CT exists. To overcome this, physicists have developed ways to simulate a bone image on the MRI. The solution is called zero TE image. So, what is a zero TE image? Typically, when we perform MRI, the TE time is relatively short, but it is still on the order of 5 or 10 milliseconds. In cardiac imaging, we use very short TE times, about 1 millisecond. But that is still too long for bone imaging.

Bone has a very short T2 time. A short T2 time means basically that the sample decays away before we have time to image it. The signal is black. With zero TE MRI, the TE is reduced from milliseconds to a few microseconds, or near zero. Thus, T2 signal can be visualized even in tissues like bone with a very short T2 time.

You may have heard about ultrashort TE imaging, or UTE MRI. UTE MRI is also useful for very short T2 tissues, such as the meniscus or cartilage. We perform lung imaging with a UTE pulse sequence and the images are quite good. UTE of the lung has been proposed as an alternative to CT for pediatric patients. But the TE time for UTE is still a little long: about 80 microseconds. To summarize, for conventional MRI, a short TE would be 8 milliseconds. For ultrashort TE MRI, the TE is 80 microseconds. But for zero TE MRI, the TE is only 8 microseconds.

What was done? The purpose of this study was to compare zero TE MRI to CT images, for evaluation of shoulder pathology. The authors evaluated 34 patients. All had both a CT scan of the shoulder and an MRI of the shoulder. The MRI included the conventional pulse sequences plus the zero TE images.

In order to make a zero TE image look like a bone image, there are several additional steps. Most importantly, the authors take the inverse of the image on a log scale. This causes the bone to appear white, much like on a CT scan. In addition, the authors segment some of the background tissue away. The result looks remarkably like a CT image. You can check this out for yourself. Images from this article appear on the front cover of the March issue of *Radiology*.

There are limitations, however. The spatial resolution of CT is about 0.6 mm. For MRI, the resolution was only about 1.2...
mm in plane. As a result, trabecular detail is poorly seen on zero TE MRI.

The results: The results were very good. First the authors compared the zero TE images to the conventional MRI sequences. Additional bone detail in the shoulder was seen in 21 out of 34 patients on the zero TE images compared to conventional MRI. These details included subchondral bone depression in a Bankart lesion. A Bankart lesion is an injury of the anterior glenoid labrum. In an anterior shoulder dislocation, the humerus is displaced anteriorly and can fracture or depress the anterior húmero labrum. In addition, lesions such as osteophytes and cysts were seen better on the zero TE images.

Compared to CT, the agreement of 2 MSK readers for diagnosing shoulder bone abnormalities was quite comparable for the MRI compared to CT. A major difference however was that CT could resolve the trabecula. The zero TE MRI could not identify the trabecular bone since spatial resolution was lower. But some features, such as bone marrow edema and small cysts were better seen on the MRI.

Conclusion: Dr. Potter and her colleagues at the Hospital for Special Surgery have introduced a new technique for bone imaging. Zero TE MRI is able to generate bone-like detail from an MRI scan. This is an extra pulse sequence lasting about 5 minutes. With MRI, it seems like there is always just one more pulse sequence to add, lengthening the study. Still, a zero TE sequence will be quicker than getting a CT. The images are impressive. Although I can describe these to you, you might want to look for yourself at the zero TE MR images in the March 2018 issue of Radiology.

Summary of March review articles

Free-Text Radiology Reports (Chen et al), • Focal Liver Lesions: Computer-aided Diagnosis by Using Contrast-enhanced US Cine Recordings (Ta et al), • Prevalence of Carotid Web in Patients with Acute Intracranial Stroke Due to Intracranial Large Vessel Occlusion (Campagne et al), • Cervical Internal Carotid Occlusion versus Pseudo-occlusion at CT Angiography in the Context of Acute Stroke: An Accuracy, Interobserver, and Intraobserver Agreement Study (Diouf et al), • and Assessment of Response to Transcatheter Arterial Chemoembolization with Doxorubicin-eluting Microspheres: Tumor Biology and Hepatocellular Carcinoma Recurrence in a 5-year Transplant Cohort (Sandow et al).

Finally, I’d like to try something different. Usually I give you a somewhat in-depth discussion of three articles per podcast. That’s useful, but you are missing a lot of content in each issue.

This week, I will also give you short summaries of five different articles. Again, the goal is to have you stay up to date with what is happening in the field.

First article: “Deep Learning to Classify Radiology Free-Text Reports,” from Stanford University. The authors use a deep learning convolutional network to read CT reports and identify if pulmonary embolus was diagnosed on that report. If a computer reads our reports, this would be useful to me. I would not have to call the clinician. Calling a clinician in the middle of the readout process takes a lot of time and interrupts the diagnostic process.

The authors started with thoracic CT scans at Stanford. They trained the computer on 2,500 reports that were annotated for the presence or absence of PE. A radiologist did the final determination of a PE being reported or not. The convolutional network was then tested on an additional 2,000 reports to evaluate performance. Finally, it’s important to test the computers on data that is completely external to your institution. In this case, the authors tested their algorithm on 839 CT reports from the University of Pittsburgh.

The computer accuracy for reading the CT reports for the presence of PE was 99% for the Stanford reports. For the external reports from Pittsburgh, the accuracy was 92%. The authors compared their neural network to a software program from the web, called PeFinder. PeFinder uses natural language processing to sort through reports, recognition of certain phrases. PeFinder was developed based on the University of Pittsburgh reports, so not surprisingly, it performed about the same as the neural network.

The authors conclude that the deep learning network performed as well or better than natural language processing for reading the thoracic CT reports. They also make the argument that deep learning networks require less time and energy to develop. Dr. Elizabeth Krupinski from Emory wrote a nice editorial about this result. She points out that we cannot have one neural network just for a PE, another one just for pneumothorax or infiltrate, and so on. But over time, it appears that we might get very reliable computer assistants to proofread our reports, indicate discrepancies in right or left sides of the body, and send alerts to referring physicians. This is only the first step.

Number two: Next a paper about computer diagnosis of focal liver lesions using contrast-enhanced ultrasound. The first author is Dr. Casey Ta; the senior author is Dr. Bob Mattrey at the University of California San Diego. The authors compared an artificial neural network, a support vector machine, and two radiologists for diagnosis of liver lesions. One radiologist had only a one-hour seminar in interpreting contrast enhanced ultrasound of the liver. The other radiologist had 20 years of experience in the field.

The support vector computer was better at diagnosis than the inexperienced radiologist. But both computer systems performed at a level on-par with the experienced radiologist. The most important result in my mind, however; is that the use of the computer algorithm improved the performance of both the inexperienced as well as the experienced radiologist. This article points out that computer algorithms can help all radiologists at all experience levels. Maybe we will all become super radiologists with computer assistance. That’s like Tony Stark in the Iron Man movies. Once you put on your computer suit of armor, everyone becomes an amazing diagnostician. I think that’s a few years off. But I am fairly confident it will eventually happen.

Number three: Next article, are you familiar with carotid artery webs? A multi-institutional study from the Netherlands discusses diagnosis of carotid webs on CT. The short title is “Prevalence of Carotid Web in Patients with Acute Intracranial Stroke…”

Carotid webs are thin, circumferential filling defects in the proximal internal carotid artery bulb. They are not due to atherosclerosis, but instead are the result of fibrous intimal hyperplasia.
Carotid webs are a possible cause of ischemic stroke, especially in young female patients.

The authors show six examples of carotid artery webs in their March 2018 research article. I personally do not recall seeing webs of this type, but I am not a specialist in neuro. But these webs need to be distinguished from small protrusions of the posterior carotid artery wall at the bifurcation. I have seen those protrusions quite frequently in asymptomatic individuals.

Out of 443 patients in the Netherlands presenting with ischemic stroke, carotid artery webs were seen in 11 or 2.5% of patients on the symptomatic side of the stroke. Ten of those 11 patients were female. So these carotid webs are quite interesting and have a suspicious association with stroke in young females. If you have such a patient, it seems important to consider a carotid web as a possible etiology. I think it’s an interesting finding. Please go ahead and take a look at the images in this article.

Number four: the next paper is also about CT angiography for diagnosis of acute stroke. The short title is, “Cervical Internal Carotid Occlusion versus Pseudo-occlusion at CTA...” The article is from radiology department at the Central University Hospital in Montreal, Canada. The challenge on CT is to distinguish two types of non-filling of the carotid artery. In the first case, there could be a true occlusion of the carotid artery with associated intracranial embolus. This is called a tandem occlusion.

The other possibility is an isolated intracranial clot that impedes ascending blood flow. There is no clot in the carotid artery itself at all. This is called a pseudo-occlusion. The authors looked to see if CTA could tell the difference. The main issue is that the treatment of the two lesions is quite different. Mis-diagnosis of pseudo-occlusion as a carotid artery occlusion will exclude patients from certain treatments.

The standard of reference for the study was digital subtraction angiography. In this study, non-attenuation of the internal carotid artery occurred in 37 of 166 stroke cases – that’s 22%.

For these 37 cases, seven radiologists had rather poor agreement as to whether the diagnosis was true, or pseudo-occlusion. CT was simply not reliable in these cases.

So this is something to be aware of for cases of acute stroke. The solution is probably to do multi-phase CTA. In one of the later CT phases, the iodine contrast may eventually opacify the vessel proximal to the true level of the intracranial obstruction. Still the authors caution this multi-phase CT approach works only in about 90% of cases, with 10% of cases still incorrectly diagnosed on CT. You might want to take a look at the findings of this paper out of Montreal by an expert stroke team.

Five: The final article is about treatment of hepatocellular carcinoma. The article is from the Ochsner Health System in New Orleans. The short title is “...Response to [TACE] with Doxorubicin-eluting Microspheres: ...Hepatocellular Carcinoma Recurrence in a... Transplant Cohort.” The underlying concept is that Transcatheter Arterial Chemoembolization, (or TACE,) is used as a bridge to liver transplantation. The hepatoma is treated, the liver is eventually transplanted, and all goes well. However, that does not always happen. Some tumors do not respond to TACE. The authors asked the question if those patients might ultimately have tumor recurrence after liver transplant.

The article included 93 patients who eventually underwent liver transplant over a five-year period. Low grade hepatocellular carcinoma had a favorable response to TACE in 87% of patients. But none of the patients with high grade tumors had a complete response, and only 1/3 had a partial response to TACE. Patients who did not respond to TACE had recurrence of hepatocellular carcinoma in about 1/3 of cases after liver transplantation. Only 1 of 75 patients with a complete or partial response to TACE had recurrent hepatocellular carcinoma. These high grade or poorly differentiated liver tumors tend to respond poorly to TACE, and suggests poor prognosis even after liver transplantation.

That concludes this week’s articles. I hope these podcasts were helpful to you. Until next time, this is Dr. David Bluemke for the journal of Radiology. I hope you have a good rest of your week.