The Potential Impact of Artificial Intelligence on Medical Malpractice Claims from Diagnostic Errors in Radiology in New York

September 2021

The Potential Impact of Artificial Intelligence on Medical Malpractice Claims from Diagnostic Errors in Radiology in New York

Authors Maggie Lee, FIAA CERA
Senior Lecturer, Department of Actuarial Studies and Business Analytics
Macquarie University, Australia

Simon Guthrie, FIAA
Senior Lecturer, Department of Actuarial Studies and Business Analytics
Macquarie University, Australia

Steven Girvan,
Director
Actuarial Services
Ernst & Young, Australia

SPONSOR Society of Actuaries Research Institute
General Insurance Research Committee

Give us your feedback!
Take a short survey on this report.
Click Here

Caveat and Disclaimer
The opinions expressed and conclusions reached by the authors are their own and do not represent any official position or opinion of the Society of Actuaries Research Institute, the Society of Actuaries or its members. The Society of Actuaries Research Institute makes no representation or warranty to the accuracy of the information.

Copyright © 2021 by the Society of Actuaries Research Institute. All rights reserved.

CONTENTS

Executive Summary **4**

Section 1: Introduction **5**

1.1 Medical Errors and Medical Malpractice 5

1.1.1 The relationship between medical errors and medical malpractice 5

1.1.2 Diagnostic errors and the Diagnostic Process 6

1.2 AI Technology in Diagnostic Errors and its legal considerations 8

1.2.1 AI Technology in the Diagnostic Process and Radiology 8

1.2.2 The Legal Landscape for AI in Health Care 10

1.3 Reliances and Limitations 12

1.4 Conclusion 12

Section 2: Potential Impact of AI technology on medical malpractice claims caused by diagnostic errors **13**

2.1 Phase 1 – Past medical malpractice experience 13

2.1.1 Data 13

2.1.2 Past paid medical malpractice claim numbers and payments 15

2.1.3 Rate of Paid claims 18

2.1.4 Average Indemnity Size 21

2.2 Phase 2 – Investigation of potential key changes from Impact of AI 24

2.2.1 Potential key changes on rate of paid claims from AI 24

2.2.2 Potential key changes in average indemnity amount from AI 28

Section 3: Legal Scenarios and Sensitivities **29**

3.1 Legal Scenarios 29

3.2 Sensivity Testing of Assumptions 32

3.2.1 Sensitivities on assumptions for the baseline projection 32

3.2.2 Sensitivities on assumptions for the potential impact of AI on rate of paid claims 33

Section 4: Conclusion **36**

Section 5: Acknowledgments **37**

Appendix A: Data and Analysis **38**

A.1 DATA validation 38

A.2 Analysis 38

References **42**

About The Society of Actuaries Research Institute **45**

Copyright © 2021 Society of Actuaries Research Institute

The Potential Impact of Artificial Intelligence on Medical Malpractice Claims from Diagnostic Errors in Radiology in New York

Executive Summary

This report provides an exploration of the potential impact of Artificial Intelligence (AI) on medical malpractice claims, with a focus on one specialty (radiology) and one state (New York). The discussions in this report are targeted to actuaries who are interested in developing a method to quantify the potential impact of AI to their own medical malpractice portfolio or to their own work.

The research aims to:

- extend existing research by exploring the potential quantitative impact of AI on medical malpractice claims caused by diagnostic errors in radiology in New York (NY), taking into account possible legal scenarios and sensitivities.
- Provide interested actuaries (particularly those practicing in medical malpractice or in health) an example of a framework for evaluating the potential impact of AI on medical malpractice claims that may help with the development of their own scenario tests related to this digital disruption. These scenario tests could then help inform how future claims and premium rates might change.

The research is focused on one specialty i.e., radiology and one jurisdiction of the United States (US) i.e. New York. New York was chosen as the state that had the highest average annual per-capita medical malpractice cost for all practitioners from 2012-2016 (Belk, 2019). The research is limited to this focus because exploration of all the different intended uses of AI across the healthcare industry and legal considerations across all jurisdictions will likely be limitless and far beyond the intended scope of this study.

Past medical malpractice data was obtained from the National Practitioner Data Bank (NPDB) public use database and the Westlaw legal research database. Past exposure data including the number of active radiologists and imaging procedures were obtained from the Harvey L. Neiman Health Policy Institute (NHPI). Information from these datasets were used together to form an understanding of past rates of paid claims and average indemnity amounts of medical malpractice claims from diagnostic errors by radiologists. The rate of paid claims was then projected to 2030, with a baseline projection based on historic experience with no impact from AI. Three legal scenarios were also developed to create three additional projections to 2030, allowing for the impact of AI and the change in the legal framework as AI is adopted. The resulting percentage reduction in the future rate of paid claims across the three scenarios compared to the baseline projection ranged between 5% to 36% across future years to 2030 depending on the legal scenario. A review of past jury verdicts and settled claims also showed that focused future developments of AI in mammography and MRI modalities and in the breast and brain anatomy could reduce the number of large claims and therefore reduce average indemnity size. The introduction of AI could however also have the potential to increase average indemnity amounts if the number of megaverdicts increases which may be driven by juror sentiment when AI is involved in a medical malpractice case.

Give us your feedback!
Take a short survey on this report.
Click Here
SOA Research INSTITUTE

Copyright © 2021 Society of Actuaries Research Institute

Section 1: Introduction

Artificial Intelligence (AI) technology is gradually being applied within the healthcare system as healthcare data increases in volume and complexity. This technology has the potential to transform the future of public health, community health and health care delivery from a personal level to a system level in the next 10 years. First demonstrations of AI have already emerged showing that methods such as deep neural networks can perform as well, if not better, than the best human clinicians in the context of diagnostic tasks. Many of these initial demonstrations of AI have been in field of radiology, the medical discipline that uses medical imaging to diagnose and treat diseases. Therefore, it has been suggested that the adoption of AI in radiology may improve the accuracy of medical imaging diagnosis and reduce the number of diagnostic errors in radiology.

The increasing adoption of AI technology in clinical practice is also likely to lead to an evolving legal and regulatory framework in healthcare to ensure the safety and effectiveness of this technology when implemented. The legal and regulatory landscape for healthcare AI is complex because of its need to consider: firstly, that AI systems across the different health and medical disciplines have different intended uses and audiences; secondly, that AI use environments will likely face different requirements at the state, federal and international levels and thirdly, that responsibility and liability for medical errors need to be determined if the AI system gets a diagnosis wrong even if its methodology for diagnosis may not be completely transparent because of its inherent 'black-box' nature. It is therefore important to understand how the legal landscape may change as the adoption of AI technology in healthcare increases.

The integration of AI in healthcare such as in diagnosis combined with the potential changes in the legal framework in healthcare over time to incorporate AI is likely to have a flow on effect on future medical malpractice claims. The qualitative exploration of AI in radiology and how use of AI in medicine may shape medical tort law and medical malpractice has been well established in other studies. This report aims to:

- Extend the research by exploring the potential quantitative impact of AI on medical malpractice claims caused by diagnostic errors in radiology in New York (NY) taking into account possible legal scenarios and sensitivities.
- Provide interested actuaries (particularly those practicing in medical malpractice or in health) an example or framework of evaluating the potential impact of AI on medical malpractice claims that may help with the development of their own scenario tests related to this digital disruption. These scenario tests could then help inform how future claims and premium rates might change.

The research is focused on one specialty i.e., radiology and one jurisdiction of the United States (US) i.e., New York. New York was chosen as the state that had the highest average annual per-capita medical malpractice cost for all practitioners from 2012-2016 (Belk, 2019). The research is limited to this focus because exploration of all the different intended uses of AI across the healthcare industry and legal considerations across all jurisdictions will likely be limitless and far beyond the intended scope of this study.

1.1 MEDICAL ERRORS AND MEDICAL MALPRACTICE

1.1.1 THE RELATIONSHIP BETWEEN MEDICAL ERRORS AND MEDICAL MALPRACTICE

A medical error is defined as "an unintended act (either of omission or commission) or one that does not achieve its intended outcome, the failure of a planned action to be completed as intended (an error of execution), the use of a wrong plan to achieve an aim (an error of planning), or a deviation from the

Copyright © 2021 Society of Actuaries Research Institute

process of care that may or may not cause harm to the patient" (Makary et al., 2016). Examples of medical errors include adverse drug events, surgical injuries, misdiagnosis, and equipment and lab report errors. These errors can occur anywhere in the healthcare system such as hospitals, clinics, surgery centers, doctors' offices, nursing homes, pharmacies, and patient homes.

Medical errors have been suggested by a John Hopkins University study to be the third leading cause of death in the United States (US). The study's aim was to identify the contribution of medical errors to US deaths in relation to the list of common causes of death compiled by the Centers for Disease Control and Prevention (CDC). This is because causes of death associated with human and system factors are generally not associated with an International Classification of Diseases (ICD) code and therefore not captured by CDC's list of common causes. The study analyzed medical death rates among hospital patients over an eight-year period and found that more than 250,000 deaths per year in the US were due to medical errors (Makary et al., 2016). As a result, there is now a growing focus in the health industry to introduce strategies that enhance care coordination and communication to prevent future medical errors and improve outcomes. An example of such strategies is the proposal of integrating AI in clinical care which may help physicians and other medical professionals reduce the number of medical errors in their practice (Parades, 2018).

Medical malpractice is defined as "professional negligence by any act or omission by a physician during treatment of a patient that deviates from the accepted norms of practice in the medical community and causes an injury to the patient" (Bal, 2009). Most medical malpractice cases involve a medical error. However, it is important to note that not all medical errors by physicians are medical malpractice. Human error is inevitable and therefore, if a physician meets the expected standard of care when treating a patient for a specific condition, then they may not be deemed negligent or liable when a medical error occurs. In fact, four elements of the tort of negligence are generally required for a successful medical malpractice claim (Bal, 2009). These elements are:

- The existence of a legal duty on the part of the doctor to provide care or treatment to the patient;
- A breach of this duty by a failure of the treating doctor to adhere to the standards of the profession;
- A causal relationship between such breach of duty and injury to the patient; and
- The existence of damages that flow from the injury such that the legal system can provide redress

While medical errors and medical malpractice are not the same, they are closely linked. Consequently, any initiatives that aim to reduce the likelihood or severity of medical errors (such as the initiative to introduce AI as noted above) will also likely lead to a reduction in the likelihood or severity of medical malpractice claims.

1.1.2 DIAGNOSTIC ERRORS AND THE DIAGNOSTIC PROCESS

Diagnostic errors are one of the most commonly occurring medical errors (Carver et al., 2021). The National Academy of Medicine defines diagnostic error in healthcare as the failure to (a) establish an accurate and timely explanation of the patient's health problem(s) or (b) communicate that explanation to the patient (Khullar et al., 2015). In line with this definition, diagnostic errors are often categorized as:

- **A delayed diagnosis** – This is where the diagnosis should have been made earlier. Common cases include the diagnosis of cancer where in some cases detection is difficult until the end stages.

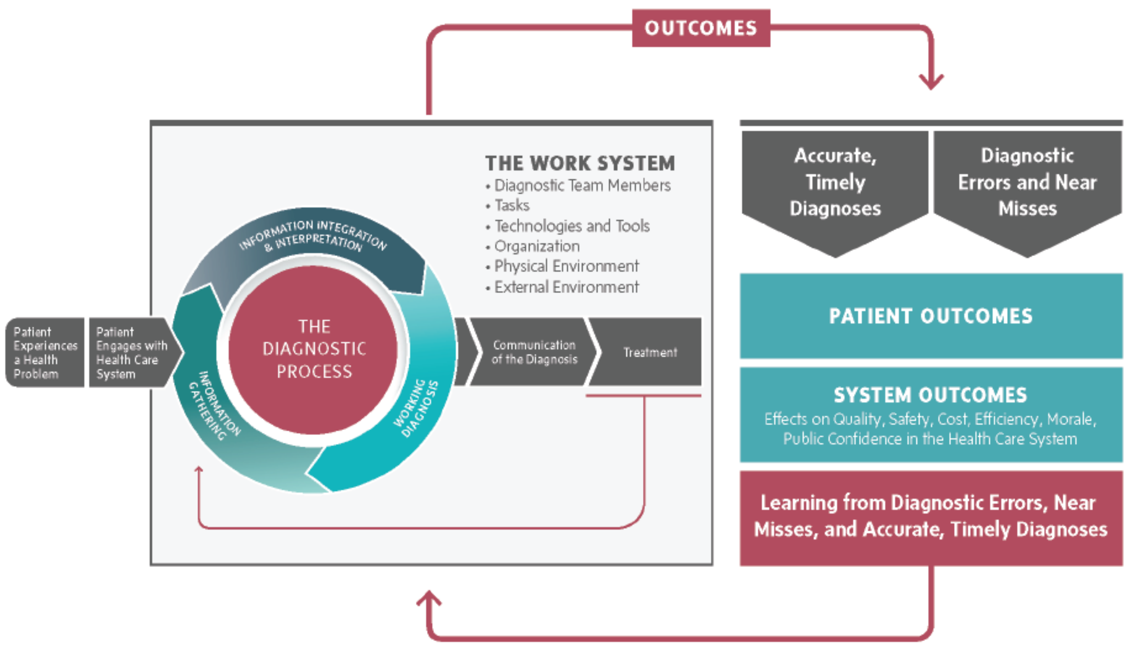
Copyright © 2021 Society of Actuaries Research Institute

- Misdiagnosis – This is where a wrong diagnosis occurs. For example, a patient whose pain is from a heart attack is told that it was from acid indigestion instead. In this case, the original diagnosis is incorrect because the true cause is discovered later.
- No diagnosis – This is where the diagnosis is missed altogether. Common cases include patients with chronic fatigue or chronic pain. It could also include where patients make specific complaints, but no diagnosis was formed.

It should be noted that there is no fixed definition for diagnostic error but several existing definitions and definitional frameworks (National Academies of Sciences, Engineering, and Medicine, 2015). For example, some definitions may include unavoidable errors whereas some may not. This is likely because of the complexity, multidisciplinary and iterative nature of the diagnostic process in which diagnostic errors occur.

The report "Improving Diagnosis in Health Care" by the National Academies of Sciences, Engineering and Medicine (National Academies of Sciences, Engineering, and Medicine, 2015) has developed a conceptual framework of the diagnostic process as summarized in the figure below.

Figure 1: The diagnostic process and the outcomes from the diagnostic process



Source: National Academies of Sciences, Engineering, and Medicine. (2015). Improving diagnosis in health care. National Academies Press. Figure. Reprinted with permission from National Academies Press.

This diagnostic process is defined within the context of a work system that is composed of:

- Diagnostic team members – This includes patients, their families, and all healthcare professionals (for example, the primary physician, radiologists, pathologists and so on) involved in their care. The diagnostic process requires intra and interpersonal collaborations amongst these members. The radiology and pathology fields are critical to diagnosis but also work with others to form a diagnostic team.
- Tasks –This includes goal-oriented actions that occur within the diagnostic process
- Technologies and tools – This includes health information technology used in the diagnostic process
- Organizational factors – This includes culture, rules and procedures, and leadership and management considerations

Copyright © 2021 Society of Actuaries Research Institute

- **The physical environment** – This includes elements such as layout, distractions, lighting, and noise
- **The external environment –** This includes factors such as the payment and care delivery system, the legal environment, and the reporting environment.

Within the work system, the diagnostic process unfolds over time and consists of information gathering, information integration and interpretation, and working diagnoses. As information is gathered, the goal is to reduce diagnostic uncertainty and to develop a deep understanding of a patient's health problem. Clinicians do not need to obtain diagnostic certainty prior to initiating treatment as the goal of information gathering is to reduce diagnostic uncertainty enough to make optimal decisions. The four types of information gathered include the clinical history and interview of the patient, a physical exam of the patient, diagnostic testing and sending a patient for referrals or consultations. Examples of diagnostic testing include medical imaging tests by radiologists and blood and pathology tests by pathologists.

Medical imaging, as part of the diagnostic testing phase of information gathering, plays a critical role in establishing the diagnoses of various conditions and is used in nearly every branch of medicine. Medical imaging can also be described in the form of a process. This process includes a pre-pre-analytic phase, which includes the selection and ordering of medical imaging; a pre-analytic phase, which includes preparing the patient for imaging; an analytic phase, which includes image acquisition and analysis; a post-analytic phase which includes interpreting and reporting the imaging results to the ordering clinician or the patient; and a post-post analytic phase which includes the integration of results into the patient context and further action. There are several medical imaging techniques and modalities including X-rays, Ultrasound, Computed Tomography (CT), Magnetic Resonance Imaging (MRI), Mammography and Positron Emission Tomography (PET) to diagnose diseases and health conditions. The interpretation of the image is typically performed by radiologists or for selected tests by nuclear medicine physicians. With the growing body of data, medical knowledge and imaging options, there are increased challenges for radiologists to achieve full competency in all imaging modalities and to fully digest all information presented. Consequently, perceptual and cognitive errors made by radiologists can contribute to diagnostic errors.

Diagnostic errors have been identified as the most common and most costly alleged error contributing to medical malpractice claims (Tehrani et al., 2013), with radiologists as one of the main physician specialties contributing to such claims (Schaffer et al., 2017). This likely reflects the large involvement of radiologists throughout the diagnostic process. Furthermore, research has shown diagnostic errors impact about 5.08% of outpatients in the US, which is approximately 12 million adults a year (Singh et al., 2014), account for between 6 and 17 percent of hospital complications and contribute to approximately 10 percent of patient deaths (Emerj, 2019). It is believed that without a dedicated focus on improving diagnosis, diagnostic errors are likely to worsen. To address these challenges and high costs, researchers and organizations have started exploring AI as a particular tool to improve medical diagnosis and thus reduce diagnostic errors. Much of the exploration has been in the utilization of AI in the area of medical imaging for diagnosis in radiology (Neri et al., 2019).

1.2 AI TECHNOLOGY IN DIAGNOSTIC ERRORS AND ITS LEGAL CONSIDERATIONS

1.2.1 AI TECHNOLOGY IN THE DIAGNOSTIC PROCESS AND RADIOLOGY

AI means different things to different people and therefore has several definitions. A typical definition of AI is "any technique which enables computers to mimic human behavior" (KDnuggets, 2017). Examples of these techniques include machine learning, natural language understanding, language synthesis, computer vision, robots, sensor analysis, optimization and simulation.

Copyright © 2021 Society of Actuaries Research Institute

AI has recently started making vast waves in healthcare because of the increasing availability of healthcare data and rapid development of big data analytic methods. AI systems across the different health and medical disciplines have different intended uses and audiences. For example, AI can be used to detect patterns that are useful for diagnostic insights, to automate diagnostic processes, to power remote-controlled robotic surgery and to assist clinicians in providing a personalized patient protocol to track patient drug dosages and the risks associated with cumulative doses. In the context of diagnostic medical imaging in radiology, which is an area with a lot of AI focused activity to date, AI also has varying intended uses. It can help with triaging critical findings in medical imaging, predicting outcomes for diseases such as strokes, or diagnose health problems and flag acute abnormalities. The focus for our study will be for the intended use of AI to assist radiologists to diagnose health problems and flag abnormalities.

The application of AI to assist radiologists in diagnosis typically involves deep learning, a subset of machine learning techniques. In particular, a convolutional neural network (CNN), which is a specialized deep learning method for analyzing images, is usually involved. Deep neural networks consist of a stack of multiple layers of artificial neuronal links that loosely simulates the brain's neuronal connections. For CNNs, the deep neural network conceptually mimics the visual pathway. Advances in deep learning has led to the potential of localizing and recognizing complex patterns in different medical imaging modalities and for different anatomies: chest, abdomen, pelvis and so on; and have been comparable to or exceeding the performance of expert radiologists in their decision-making (Jin et al., 2021).

The most relevant AI algorithm for medical imaging diagnosis and prediction are binary classification algorithms built with deep neural networks by means of supervised learning. That is, the computer is presented with example inputs (say x) and their desired outputs (say y) from a training dataset and the goal of the algorithm is to find a general rule that maps inputs to outputs as $y = f(x)$. The prediction is then compared to the test dataset. For example, the goal of the algorithm could be to find a rule that helps with differentiating lung cancer from benign diseases in lung nodules seen on plain chest radiographs. A general rule is generated by the computer from a training dataset of chest radiographs. The binary output from the algorithm of 1 or 0 signifies whether the prediction is lung cancer or not. This general rule is then applied to an independent test dataset of chest radiographs for diagnostic testing of the performance of the algorithm. The performance and therefore accuracy of such algorithms are typically measured in terms of sensitivity, specificity, Receiver Operating Characteristic (ROC) curve or area under the ROC curve (AUC). These are explained in Section 2.2.1. As noted above, besides binary classification of diagnosis, AI can be applied in radiology in many ways such as in the segmentation of medical images, estimation of continuous measurements and workflow automation. These applications are beyond the scope of this report.

AI is already in clinical use in the US. The AI technology, Arterys, a medical imaging platform with machine learning applications in radiology was the first to receive clearance by the Food and Drugs Administration (FDA) in 2017. By 2020, the FDA has cleared or approved approximately 64 AI-based medical devices, with 29 classified as AI-machine learning (ML) based. Most of the clinical use is in diagnostics and imaging, particularly in radiology (21 of the 29 approved AI-ML based devices were in radiology) (Benjamins et al., 2020). Our study therefore focuses on AI-ML based technology that provides classification for medical imaging diagnosis and prediction.

With diagnostic errors by radiologists a major source of medical malpractice claims and AI technology bringing a paradigm shift to healthcare particularly in imaging diagnosis by radiologists, it would be beneficial for actuaries specializing in medical malpractice to understand the potential impact of AI technology on such claims.

Copyright © 2021 Society of Actuaries Research Institute

1.2.2 THE LEGAL LANDSCAPE FOR AI IN HEALTH CARE

The legal landscape for healthcare is complex. This is because rules and regulations are in place for different intended uses and audiences and vary at state, federal and international levels. As such, rather than provide a full account of all legal requirements, this section aims to provide an overview of some legal considerations as AI becomes increasingly prevalent in healthcare and how this might impact medical liability.

With the growing potential for AI technology to transform numerous aspects of the healthcare sector, there is an increased focus on the risks of unintended and negative consequences associated with such technology. These risks are considered high, especially when scaled to the level of state-wide or national health systems and because of this, the current use of AI in everyday health care is still limited. Before AI can be used in healthcare at a large scale, questions concerning the robustness, interpretability and accountability of AI need to be answered.

The concept of 'responsible AI' has attracted considerable attention worldwide in recent years to ensure that best practices are in place for AI development, adoption, and maintenance, both in the healthcare sector as well as other industries. For example, the G20 AI Principles (OECD, 2020), which was drawn from the OECD recommendation on AI, was welcomed by G20 leaders in 2019. These values-based principles seek to foster public trust and confidence in AI technologies and realize their potential through the promotion of responsible stewardship of trustworthy AI. The five complementary values-based principles identified include:

- Inclusive growth, sustainable development and well-being,
- Human-centered values and fairness,
- Transparency and explainability,
- Robustness, security and safety,
- Accountability

Of these five values-based principles, the accountability and transparency and explainability principles are particularly relevant to the identification of where liability falls when a medical error caused by AI technology occurs. Legal clarification and responsibility for the AI model outputs is an important but also challenging area that is still being developed. Several questions need to be answered, such as "What is the best course of action when an AI algorithm provides a recommendation that is different to what the physician believes, based on their own substantial experience?" or "Who is responsible when the AI provides a recommendation that the physician adopts but the AI was incorrect, causing injury to the patient?". These questions are challenging due to the complexity in understanding how the AI makes its decisions because of its inherent 'black box' nature. If the reasoning of how an AI makes its recommendations cannot be understood, it is likely to be very difficult to assign liability under traditional tort paradigms. In the next section, we will explore the legal landscape in the US context. Based on this legal landscape, we will then explore the legal considerations when diagnostic errors occur in radiology due to AI.

<u>Legal landscape for AI in healthcare in the US</u>

In the US, users and developers of health care AI technology and systems are subject to many different legal regimes including federal statutes, federal regulations and state tort law. In the context of medical liability, two important laws and regulations in the US are:

Copyright © 2021 Society of Actuaries Research Institute

- **Federal Food, Drug and Cosmetic Act (FDCA)** – The Food and Drugs Administration (FDA) is the regulator that enforces the FDCA. It regulates the safety and effectiveness of drugs and medical devices, including certain forms of medical software such as clinical AI systems
- **State tort law** – Tort law allows injured individuals to recover damages under medical malpractice liability if the injury was caused by another individual/entity such as physicians, developers, providers, hospitals, or other health care actors who performed below the standard of care. Product liability may also be imposed on developers if there are defects in the design or manufacturing of AI systems or for failure to adequately warn users of the risks of a particular AI system. State law determines the applicable standard of care and when tort liability will exist.

Other significant laws and regulations in the US include the Health Insurance Portability and Accountability Act (HIPAA), Common Rule, Federal Trade Commission Act (FTCA) and the FTC Health Breach Notification Rule.

Clinical AI, which includes AI that provides diagnoses or treatment of disease without physician interpretation, such as the AI-ML based technology in radiology that provides classification of diagnosis from images, faces the highest scrutiny by the US FDA and by other similar regulatory agencies internationally. This includes regulatory oversight to ensure that false-negative and false-positive errors and misinterpretations of clinical AI algorithms' outputs, actions and recommendations to clinicians are minimized. Clinical AI systems must therefore demonstrate safety and efficacy under the FDA and may also generate liability (medical malpractice liability or product liability) under state tort law, which performs its own regulatory role. For example, tort law may exert pressure on the developers or users of AI by imposing liability for injuries caused by AI which could have been reasonably avoided either through more careful development or more careful use. Therefore, it should be no surprise that state tort law and FDA approval are closely intertwined in determining liability and accountability.

There is still uncertainty on how exactly tort law will deal with clinical AI systems. This is because court decisions are retrospective and clinical AI systems are still too new to have been challenged in medical malpractice lawsuits, making it unclear as to how courts will determine responsibility and what kind of transparency should be required. However, it is worth considering the factors that are likely to have impact on tort liability in this area. Examples of such factors include:

- the interaction of FDA approval and state tort law which may vary by states where some states may establish statutory requirements that are different from or in addition to FDA requirements and some may not
- the role of products liability and the learned intermediary doctrine in liability insurance. Products liability is a common law doctrine that allows patients to be entitled to recovery when they are injured by products that are not reasonably safe due to defective design, manufacture or warning. As AI may have many developers it may be difficult to assign fault to a certain developer or manufacturer. Furthermore, the learned intermediary doctrine prevents plaintiffs from suing medical device manufacturers directly as they have no duty to patients directly, particularly if they disclosed risks to intermediaries. Rather the physician is considered the end consumer of medical devices because they are in the best position to weigh the risks against the possible benefits of using the device
- the role of vicarious liability where one individual may be held legally responsible for the acts of another, such as vicarious liability being placed on hospitals for the negligent acts of their physicians as employees of the hospital. There is a question as to whether an AI can be deemed an employee of a hospital which would then lead to the hospital taking responsibility for the liability. This is difficult because there needs to be an established link that the hospital has control

Copyright © 2021 Society of Actuaries Research Institute

over the AI which may be difficult to determine, particularly due to the potential autonomous nature of the AI.

- the role of informed consent liability where physicians have a legal obligation to inform patients of material information about a proposed course of treatment. In the context of AI, determining what 'informed' means may be difficult because there is a potential that no-one knows exactly how a recommendation by the AI was derived and the risks associated with this recommendation.
- the impact of potential opacity of clinical AI systems on tort liability which may complicate the ability of injured patients, lawyers, providers or health systems to determine the precise cause of injury.

New York and Medical Liability

As noted above, careful consideration of state doctrines is also important when determining liability. For example, New York has a doctrine that applies to cases where there may be more than one responsible party, called comparative negligence. This doctrine may not be applicable in other states. In New York, the law involves determining the degree to which the party involved was negligent. If the plaintiff also contributed to the negligence, New York still allows the plaintiff to collect damages, whereas in other states this may not be possible. This also may have implications on how the liability of damages caused by AI is shared in the state of New York.

As a result of the likely change in legal landscape on an international, national and state level as AI becomes increasingly adopted in a clinical setting, there is still uncertainty regarding where the liability might fall if an injury is caused by AI. However, it is worth considering some of the possible legal solutions to address AI liability and how this might impact medical malpractice claims caused by diagnostic errors. This is considered in Section 3.

1.3 RELIANCES AND LIMITATIONS

This report has been produced for the SOA with the aim to advance the knowledge of its actuarial members through research and education.

While all duty and care has been taken to review the data for reasonableness, the results in this report rely heavily on the data provided. There are several limitations with the data which are outlined in Section 2.1.1.

In the event that this advice is reproduced, this report should be provided in its entirety. If the user of this report wishes to reproduce any part of this report not in the entirety, then the authors should be contacted to ensure the work is presented correctly.

1.4 CONCLUSION

With AI becoming increasingly introduced in medical imaging and with diagnostic errors being a key medical error driving medical malpractice claims, it would be beneficial for actuaries practicing in the medical malpractice space to consider the potential impact of AI on medical malpractice claims caused by diagnostic errors. The level of impact of AI would also depend on how the legal landscapes change to allow for AI in healthcare on state, federal and international levels. That is, the impact is likely to vary according to different legal scenarios. The next two sections seek to provide quantitative and qualitative discussions of the potential impact of AI on medical malpractice claims caused by diagnostic errors across a few different potential legal scenarios.

Copyright © 2021 Society of Actuaries Research Institute

Section 2: Potential Impact of AI technology on medical malpractice claims caused by diagnostic errors

The research explores the potential impact of AI technology on medical malpractice claims caused by diagnostic errors in radiology in New York in two phases.

- **Phase 1** – This phase includes an analysis of past paid medical malpractice claims cost (by rate of paid claims and average indemnity sizes) from diagnostic errors in radiology in New York to understand historic trends of these claims
- **Phase 2** – This phase includes the determination of how claims cost may change for medical malpractice claims due to the increased use of AI technology in healthcare. This section includes an initial exploration of key factors that may change the rate of paid claims and average indemnity amounts. How the factors for the rate of paid claims may change under different legal scenarios and the sensitivities of these factors will then be explored in Section 3.

2.1 PHASE 1 – PAST MEDICAL MALPRACTICE EXPERIENCE

Phase 1 of our analysis focuses on the past paid medical claims cost arising from diagnostic errors in radiology in NY.

2.1.1 DATA

To analyze past paid medical claims cost, information was extracted from the following sources:

1. The **National Practitioner Data Bank (NPDB) public use data file** – The NPDB serves as a federal information clearinghouse on malpractice payments for and disciplinary sanctions against health care practitioners. It has a NPDB public use data file which is a publicly available dataset containing retrospective cross-sectional analysis of de-identified physician data published by the US Department of Health and Human Services (HHS). This data file contains selected variables from medical malpractice payment and adverse licensure, clinical privileges, professional society membership and Drug Enforcement Administration (DEA) reports received by the NPDB concerning physicians, dentists and other licensed health care practitioners. It also includes reports of Medicare and Medicaid exclusion actions taken by the Department of HHS Office of Inspector General. This file is updated 4 times a year to include data as of March 31, June 30, September 30 and December 31.

   The version used for analysis in this report contains reports received from September 1, 1990 to December 31, 2020.

   This dataset has been previously used in other research (including Schaffer et al. (2017), Mello et al. (2014), Tehrani et al. (2013) and AbuDagga et al. (2016)) to analyze various characteristics of past US medical malpractice claims. This public use data file is used to draw insights on both past rate of paid claims and average indemnity amounts and to have a quantitative discussion on the potential impact of AI on the rate of paid claims of medical malpractice claims caused by diagnostic errors in radiology in NY. Data validation on this dataset can be found in Appendix A.

   The NPDB dataset has the following limitations:

   - This dataset only captures information on medical malpractice claims that result in a settlement payment, which account for about 32% of all US medical malpractice claims resolved (QBE, 2020). Given it is only on closed claims, there is no information on claims still open.
   - Defense costs and some medical malpractice events are not reportable to the NPDB and therefore not captured in the dataset. This includes events where: the physician resolves the claim early and such claims have not been set forth in writing, settlement payments have been paid out-of-

pocket by an individual practitioner, settlements occur with a high-low agreement resolution, the physician's medical corporation (such as a hospital) is named as the party in the lawsuit rather than the physician (this is known as corporate shield practice) and settlement payments follow receipt of an 'intent to sue' notice.

- Some of the records in this dataset might be from insurers settling a claim to avoid costly litigation or a contentious discovery period even when the practitioner's care may have been within the standard of care. Therefore, some of these payments may not necessarily reflect negligence.
- This dataset does not allow us to split the data by physician specialty. We are therefore unable to identify medical malpractice claims from diagnostic errors by radiologists. An assumption on the proportion that radiologists might contribute to medical malpractice claims from diagnostic errors is made using trend analysis conducted by CRICO from their Comparative Benchmarking System (CBS), which includes medical malpractice information from all Harvard medical institutions and their affiliations, representing approximately 30% of US medical malpractice claims (CRICO, 2021).

Despite these limitations, NPDB has been noted as the most complete source of malpractice payment data available in the US (Tehrani et al., 2013).

2. **Harvey L. Neiman Health Policy Institute (NHPI) Neiman Almanac data** – Neiman Almanac data by NHPI provides the number of radiologists in the physician workforce and the number of imaging procedures across the US and by states, including New York. The NHPI notes the source of the data for the radiology workforce is from the US Health Resources and Services Administration (HRSA) Area Health Resources File and the AMA Physician Characteristics and Distribution in the US Books, which are considered the authoritative sources for information on the US physician workforce (Hughes, 2015). The data is used in our analysis as the exposure component of our rate of paid claims analysis. The NHPI notes that the source of the imaging procedures data is from the Centers for Medicare and Medicaid Services (CMS) Physician/Supplier procedure summary.

   The NHPI Radiology Access and Workforce dataset has the following limitations:

   - This dataset only provides the number of radiologists from 1995 to 2013, and therefore historic values for 2014 to 2020 are missing. We investigated other potential sources of radiology workforce data and, to the best of our knowledge, this dataset is the most complete dataset on the radiology workforce that is publicly available. Our investigation of other sources of data revealed that the number of active radiologists each year can vary between sources and this is likely due to lags in the data collection processes and how physician's self-reporting of their own primary and secondary specialties are treated across the different sources. We decided to impute the missing values for 2014 to 2020, which is discussed in further detail in Section 2.1.3.

3. The **Westlaw legal research database (Thomson Reuters)** – This database is an online legal database which has information on State and Federal jury verdicts and settlements related to medical malpractice claims, including those specifically from diagnostic errors by radiologists in New York. This database is used to draw insights on average indemnity amounts to qualitatively explore the potential impact of AI on the average indemnity size of medical malpractice claims caused by diagnostic errors in radiology in NY.

   The Westlaw dataset has the following limitations:

   - The information in this dataset is all text-based and dependent on a search criterion. Relevant information was then extracted manually from the case files.
   - It does not include cases dismissed prior to trial or cases settled outside of court. Approximately only 7% of medical malpractice lawsuits make it to a jury trial (Canady, 2018).
   - Some cases only record the amount awarded while some cases also provide the amount settled. In situations where both were provided, the amount settled was taken. There were only a few cases where only the amount awarded was provided. This is provided in a separate analysis.

Copyright © 2021 Society of Actuaries Research Institute

- In cases where there was shared liability between the radiologists and the clinic or hospital, some were clear on the proportion that was shared whereas some only provided the total amount. In situations where the proportion was clear, only the radiologist's amount was taken. In cases where only the total amount was provided, we adopted the total amount with no apportionment.

While these publicly available sources of data have limitations, the focus of our research is to provide the audience a possible case study on analyzing the potential impact of artificial intelligence on medical malpractice claims. One of our aims is to allow actuaries practicing in medical malpractice insurance to be able to adapt the information and the analysis provided in this report to their own medical malpractice datasets.

2.1.2 PAST PAID MEDICAL MALPRACTICE CLAIM NUMBERS AND PAYMENTS

This research is focused on medical malpractice claims arising from diagnostic errors in radiology in NY. The NPDB public use datafile provides medical malpractice claim reports for all of the US across 11 main allegation groups (diagnosis related, anesthesia related, surgery related, medication related, IV and blood products related, obstetrics related, treatment related, monitoring related, equipment/product related, behavioral health related and other miscellaneous). Our initial analysis consisted of observing trends from 1995 to 2020 in:

- Past paid medical malpractice claims against Doctors of Medicine (MDs) and Doctors of Osteopathic medicine (DOs) across all of US **i.e., in aggregate**
- Past paid medical malpractice claims against MDs and DOs for **diagnostic errors only** across all of US
- Past paid medical malpractice claims against MDs and DOs for diagnostic errors only i**n NY**

The 1995 to 2020 period was selected to capture most of the period over which the NPDB has collected data on closed claims (Mello et al., 2014). All payments have been converted to 2020 dollars using historical consumer price index for all urban consumers from the Bureau of Labor Statistics (Bureau of Labor Statistics, 2020) as advised by the NPDB Public Use Data File.

Figure 2 below shows the number of claims and average indemnity amounts by physicians in each settlement year while Figure 3 shows the same information but for claims from diagnostic errors only.

Figure 2: Number of claims and Average indemnity amount by physicians (Doctor of Medicine - MDs and Doctor of osteopathic medicine - DOs) in each settlement year from NPDB

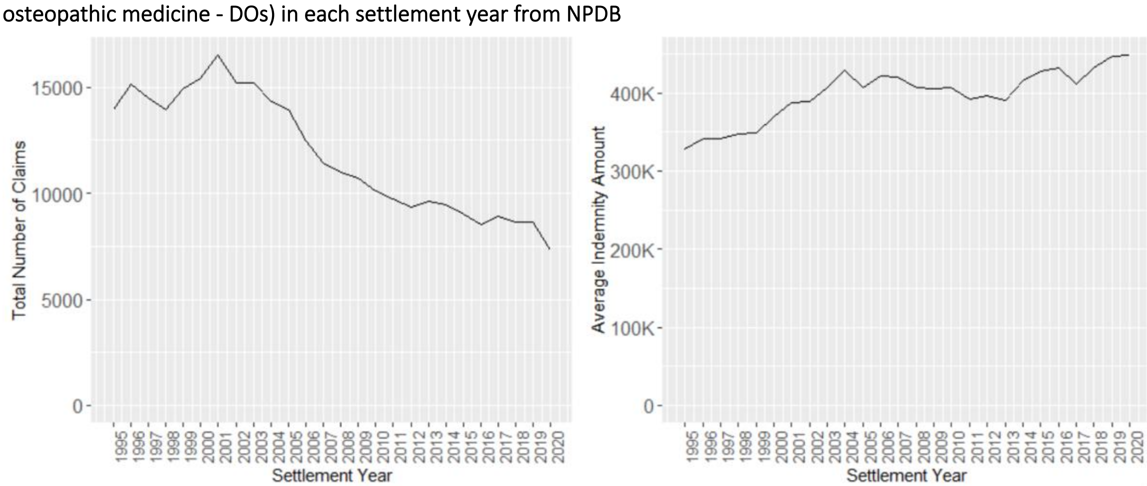
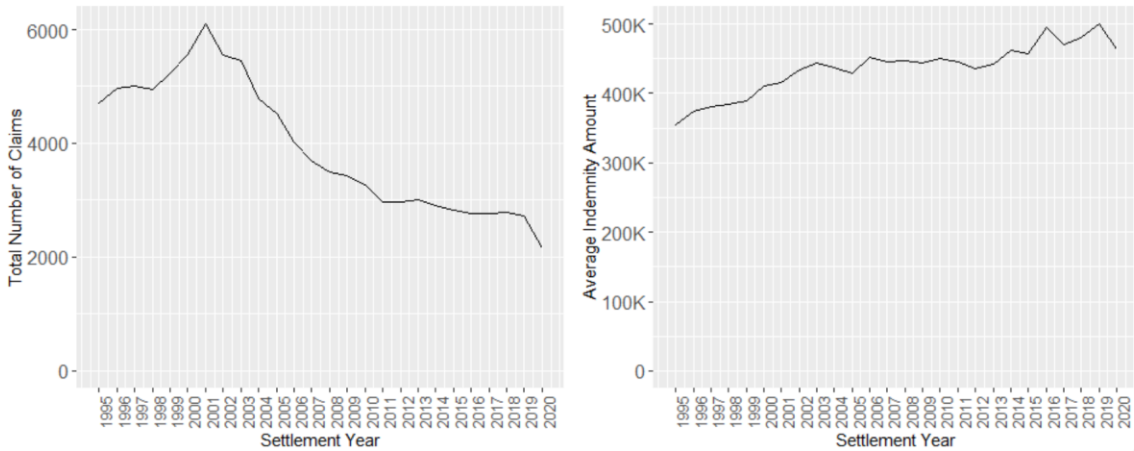


Figure 3: Number of claims and Average indemnity amount from diagnostic errors by physicians (MDs and DOs) in each settlement year from NPDB



Similar trends in the number of claims and average indemnity sizes can be observed in Figure 2 and Figure 3. That is, a downward trend in claim numbers and a slight increasing trend in average indemnity amounts. Prior analyses from other studies (Schaffer et al., 2017, Mello et al., 2014, Tehrani et al., 2013) demonstrate similar trends in the number of claims and average indemnity amounts. While the NPDB dataset alone cannot definitively distinguish drivers of the downward trend in the number of claims, Schaffer et al., 2017 outlines some factors that may help to explain why numbers of claims paid on behalf of physicians is declining. This includes:

- Increasing prevalence of communication and resolution (sometimes referred to as disclosure, apology and offer) programs, through which compensation for an injury due to negligence may be provided without requiring a written claim from a patient. As adoption of these programs increases, their effectiveness will likely grow over time.
- The manner by which institutions and insurers resolve claims – As noted in the NPDB data limitations, only written claims paid on behalf of physicians are required by Federal law to be reported to the NDPB. A growing number of claims may be settled on behalf of institutions alone, instead of individual physicians, thereby not triggering the NPDB reporting requirements. This practice has been referred to as corporate shielding and there has been concerns expressed by critics that this practice blunts the ability to detect and track physicians with an excessive number of negligent events.
- Improvements in patient safety; however, findings cannot establish this fact. Certain measures such as use of checklists, patient handoff protocols have been shown to enhance patient safety. Recent studies however have found that patient safety is still lacking in the US health care system.
- The impact of traditional tort reforms during this period such as damage caps and statutes of limitation – Critics have questioned whether such reforms have successfully lowered the cost of malpractice liability with several studies producing mixed results. For example, some studies found that the adoption of liability reform lowered the probability of physicians' experiencing a malpractice claim while others have found no effect (Seabury et al., 2014).

The reasons for the slight increase in the average indemnity amounts is also unclear from the NPDB dataset. However, other studies have noted the increase could be because of plaintiff attorneys not taking cases with a smaller potential payout or settling these cases outside the written claims process because of risk of loss or administrative costs associated with these cases when they are settled within the written claims process (Schaffer et al., 2017).

Copyright © 2021 Society of Actuaries Research Institute

Figure 4 and Figure 5 below highlight the split of medical malpractice claim payments (inflated to 2020 dollars) from diagnostic errors by each US state. Figure 4 shows this split for all payments aggregated across 1995 to 2020 whereas Figure 5 shows this split for the 2020 year only.

Figure 4: Total Medical Malpractice claim payments (dollars) from diagnostic errors between 1995 and 2020 by State

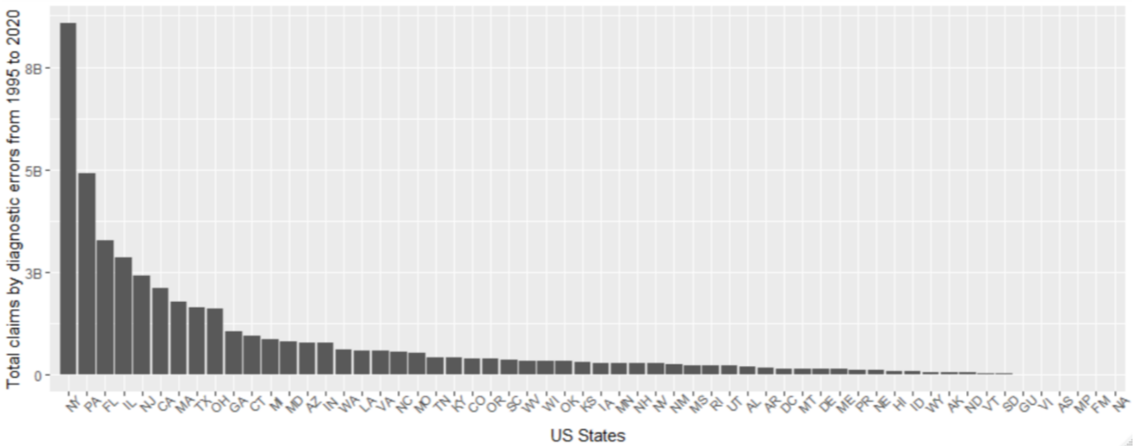
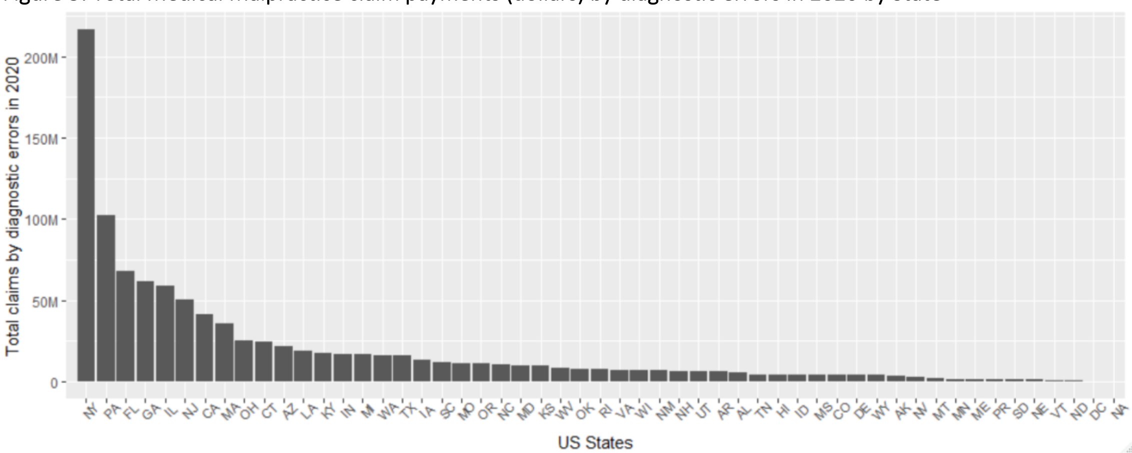


Figure 5: Total Medical Malpractice claim payments (dollars) by diagnostic errors in 2020 by State

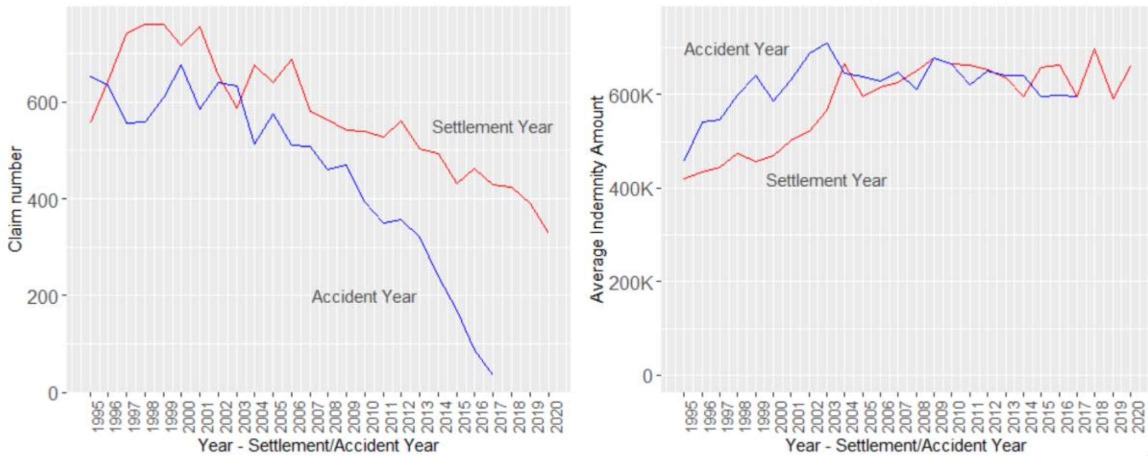


From the figures above, it is evident that New York had the highest total claim payments in both the most recent year, 2020 and across the entire study period from 1995 to 2020. This is probably not surprising given that, as the most populous state in the US, it is likely to have the largest patient and physician populations. The above therefore can be seen as a reflection of the large proportion of medical care delivered in New York. This analysis from our data further confirms the reasonableness of choosing New York as the main state of focus because of its considerably high proportion of medical malpractice claim payments from diagnostic errors compared to the other states.

Copyright © 2021 Society of Actuaries Research Institute

Figure 6 below shows similar graphs as those in Figure 2 and Figure 3 but for claims caused by diagnostic errors in New York only.

Figure 6: Number of claims and Average Indemnity amount caused by diagnostic errors by physicians (MDs and DOs) in each settlement year and accident year in NY from NPDB



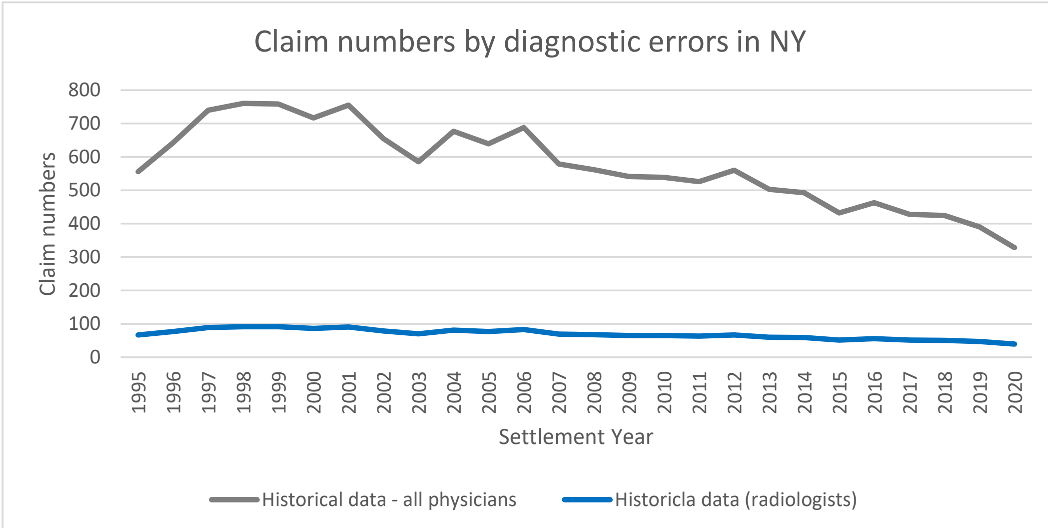
Medical malpractice claims from diagnostic errors in New York have similar trends to those observed for medical malpractice claims in general and medical malpractice claims from diagnostic errors in all of US (as shown in Figure 2 and Figure 3). That is, a decreasing trend for the number of claims and slight increasing trend in average indemnity. These trends are comparable by settlement year. In the figure above, claim numbers and average indemnity amounts are also shown by accident years. There is a clear lag between the date a medical malpractice claim is settled compared to when it occurred, which is on average about 5 years and is consistent with observations in other studies (Tehrani et al., 2014). Our study will focus on effects by settlement year instead of accident year because of this lag and because the NPDB dataset is on closed claims only.

2.1.3 RATE OF PAID CLAIMS

To obtain historic rate of paid claims from diagnostic errors by radiologists for our study, historic claim numbers from diagnostic errors from the NPDB dataset was firstly adjusted to only account for claims caused by radiologists. This number was then divided by the number of active radiologists.

As the NPDB dataset does not provide physician specialty information, the number of claims from diagnostic errors in New York by radiologists was derived by assuming that radiologists contribute to 12% of medical malpractice claims from diagnostic errors. This is based on data from CRICO's diagnosis-related cases report which indicated that radiologists accounted for 12% of medical malpractice diagnostic error cases from 2014-2018 (CRICO, 2021). Figure 7 below shows claim numbers from diagnostic errors by all physicians in New York (grey line) and specifically for those caused by radiologists in New York (blue line).

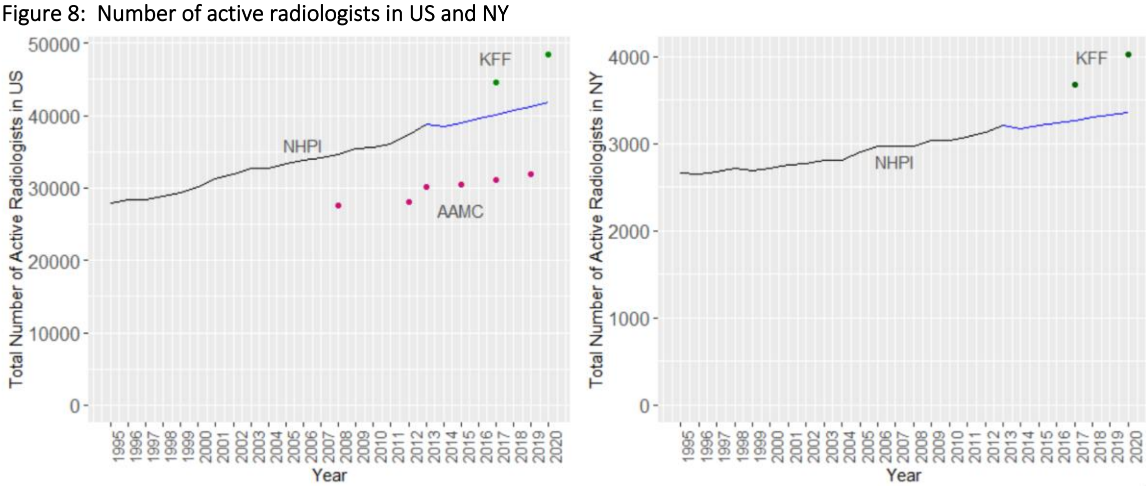
Figure 7: Number of claims per year based on historical data (1995 to 2020) from all diagnostic errors and from diagnostic errors by radiologists only in NY



Schaffer et al., 2017 also notes that radiologists are one of the top specialists that contribute to medical malpractice claims caused by diagnostic errors and are 1 of 9 specialists (or 11% of total specialists) who contribute to diagnostic errors.

Figure 8 below shows the number of active radiologists in US and NY in each year of the study period.

Figure 8: Number of active radiologists in US and NY



From 1995 to 2013, the number of active radiologists has been obtained from NHPI. Given the missing data from 2014 to 2020, other sources of active radiologists in the US and NY were obtained from the Kaiser Family Foundation (KFF) and Association of American Medical Colleges (AAMC). The number of active radiologists from AAMC (in red in Figure 8) was consistently lower than NHPI numbers whereas KFF numbers (in green in Figure 8) appeared to be higher.

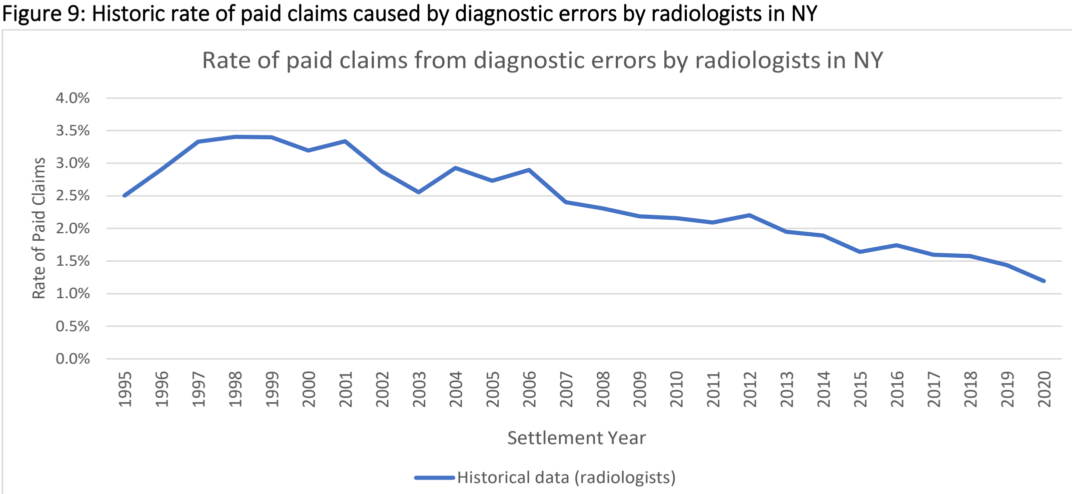
All sources (NHPI, AAMC and KFF) note that their information was based on the American Medical Association (AMA) Physician Masterfile. NHPI and AAMC directly referenced the AMA Physician Masterfile while KFF referenced Redi-Data which is an authorized AMA database licensee. While the AMA Physician Masterfile is considered the gold-standard source of physician workforce data, it does have limitations as noted in Section 2.1.1. Looking at one specific comparison in 2012/2013, the AMA fact book reported

37,300 radiologists, NHPI's independent examination of the raw 2012 AMA Physician Masterfile data reported 38,722 radiologists whereas the AAMC's examination of the 2013 AMA Physician Masterfile data reported substantially fewer radiologists at 27,570.

Given that NHPI's number of active radiologists were in the middle of AAMC's and KFF's and the AMA Physician Masterfile is considered the gold-standard source of physician workforce data, the number of active radiologists from NHPI from 1995 to 2013 has been used to reflect the number of active radiologists in that period. A linear regression was applied to impute the missing values between 2014 and 2020 to maintain the increasing trend observed across all three sources (blue line in Figure 8). These imputed values were used in the analysis.

The resulting historic rate of paid claims from diagnostic errors by radiologists in New York is shown in Figure 9. This was obtained by dividing the claim numbers from diagnostic errors by radiologists in New York by the 5-year rolling average of active radiologists. A 5-year average was chosen to reflect the average lag between the accident and settlement year of these claims. The results are slightly higher than those noted by Schaffer et al., 2017, which noted a 1.9% rate of paid claims in the period 2003-2008 and 1.4% in the period 2009-2014 for radiology.

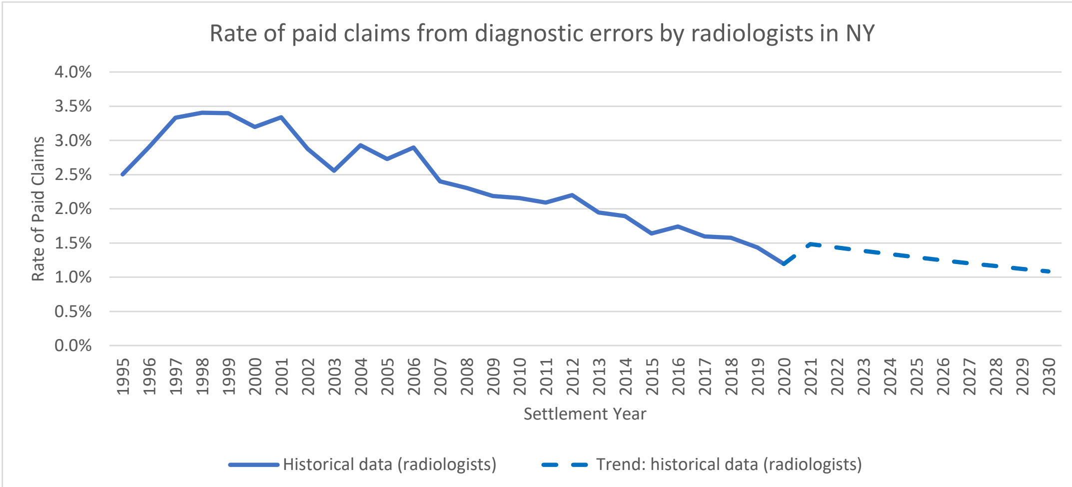
Figure 9: Historic rate of paid claims caused by diagnostic errors by radiologists in NY



The historic rate of paid claims was then projected to 2030 to obtain a baseline projection where we assume no AI has been used in radiology. As the first AI technology, called Arterys (a deep learning platform in radiology to help doctors diagnose heart problems), was approved in a clinical setting by the FDA at the beginning of 2017, our study projects to 2030 based on historic data to 2016 rather than to 2020 to avoid any potential AI impacts in the data between 2017 and 2020. This projection was obtained by fitting an exponential curve to the data between 1996 and 2016, as shown in Figure 10.

Copyright © 2021 Society of Actuaries Research Institute

Figure 10: Rate of paid claims (projected to 2030 from historic data between 1996 and 2016 by fitting an exponential curve. R^2 fit of 0.8886).



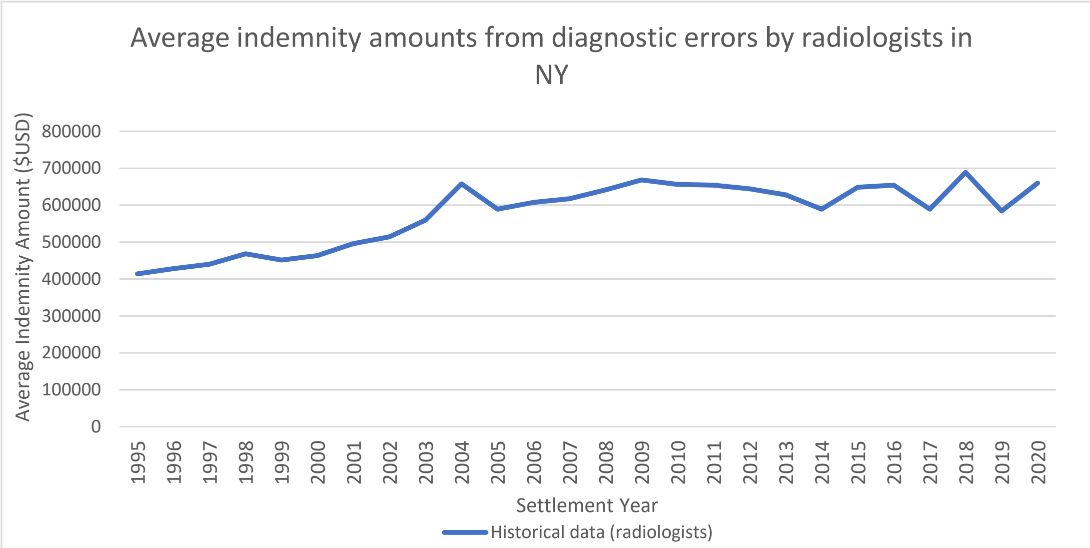
The projection above shows that the rate of paid claims is reducing at a reducing rate. We believe the decline in the rate of paid claims at a reducing rate is reasonable as this reflects the long-term downward trend in experience and because it is unlikely that negligence by physicians and therefore, reductions in rate of paid claims, will ever reduce to zero. This projection was compared to a deterministic projection using an assumed average reduction in historic rate of paid claims based on historic experience. The deterministic projection, as shown in Section 3.2.1 and Appendix A.2.1, produced similar results.

2.1.4 AVERAGE INDEMNITY SIZE

To obtain historic average indemnity size of claims from diagnostic errors by radiologists for our study, historic average indemnity size from diagnostic errors from the NPDB dataset was split for claims caused by radiologists. Similar to the rate of paid claims, as the NPDB dataset does not provide physician specialty information, the average indemnity size of claims from diagnostic errors in New York by radiologists was derived by assuming that radiologists contribute to 12% of medical malpractice claim payments from diagnostic errors. This is in line with data from CRICO's CBS report which indicated that radiologists accounted for 13% of medical malpractice diagnostic error payments from 2014-2019. This assumption (combined with the assumption that 12% of the number of claims from diagnostic errors is caused by radiologists) allows for the average indemnity size of claims from diagnostic errors by radiologists in NY per year to remain unchanged from the average indemnity size of claims from diagnostic errors for all physicians in NY. The historic average indemnity sizes are shown in Figure 11 and shows that the average indemnity size of claims from diagnostic errors by radiologists in NY in the last 5 years was about $650K.

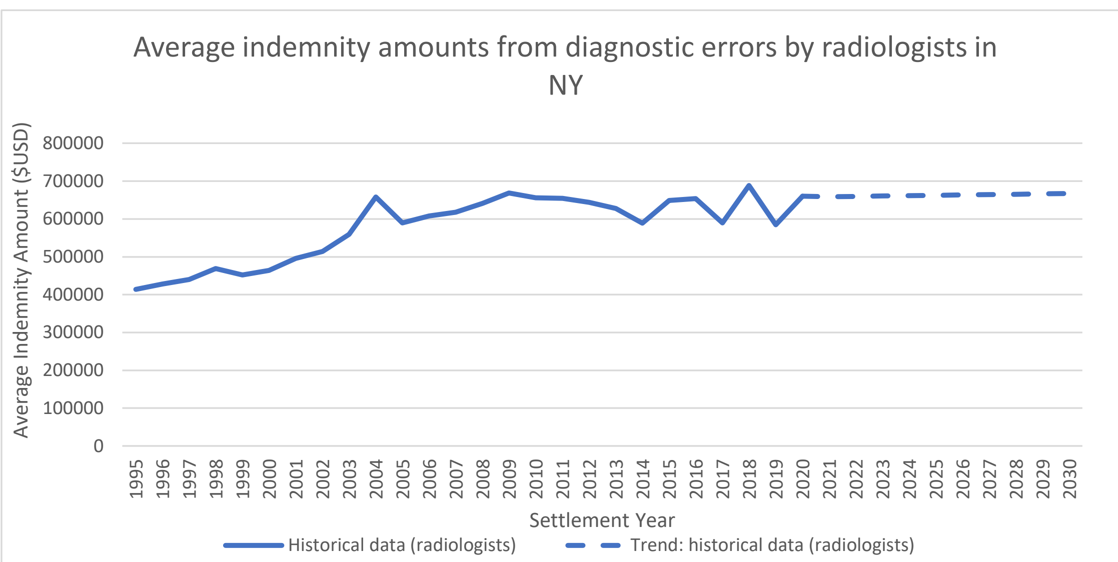
Copyright © 2021 Society of Actuaries Research Institute

Figure 11: Historic average indemnity amounts caused by diagnostic errors by radiologists in NY



The historic average indemnity size was then projected to 2030 to obtain a baseline projection where we assume no AI has been used in radiology. Similar to the rate of paid claims projection, our study projects to 2030 based on historic data to 2016 rather than to 2020, to avoid any potential AI impacts in the data between 2017 and 2020. This projection was obtained by analyzing observed average increases in average indemnity amounts per year. An assumption that average indemnity amounts increase 0.1% per year was adopted. The projection is shown in Figure 12 below.

Figure 12: Average indemnity amounts by radiologists based on historical data (1995 to 2016) and future projections to 2030 (projected from historic data to 2016).



Copyright © 2021 Society of Actuaries Research Institute

Data from the Westlaw database was also obtained to help gain a deeper understanding of the characteristics of average indemnity sizes. The figures below show the number of settled and awarded claims to the plaintiff and against radiologists due to a diagnostic error caused by negligence in NY. The number of settled and awarded claims have been categorized by settlement and award amount ranges (increments of $0.5 million for settlements and $1m for awards; inflated to 2020 using CPI). Within each amount range, the figures also show the type of modality used for diagnostic medical imaging and the anatomy area the radiologist focused on for the diagnosis.

Figure 13: Amounts that have been agreed as settlements to be paid by a radiologist as a result of a diagnostic error caused by negligence.

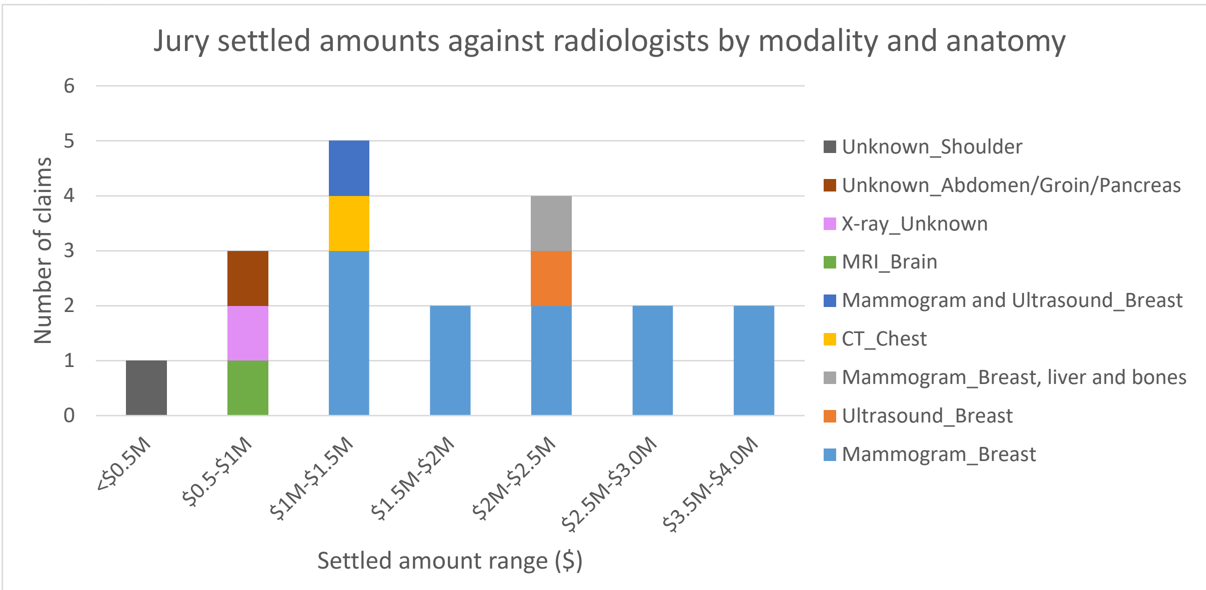
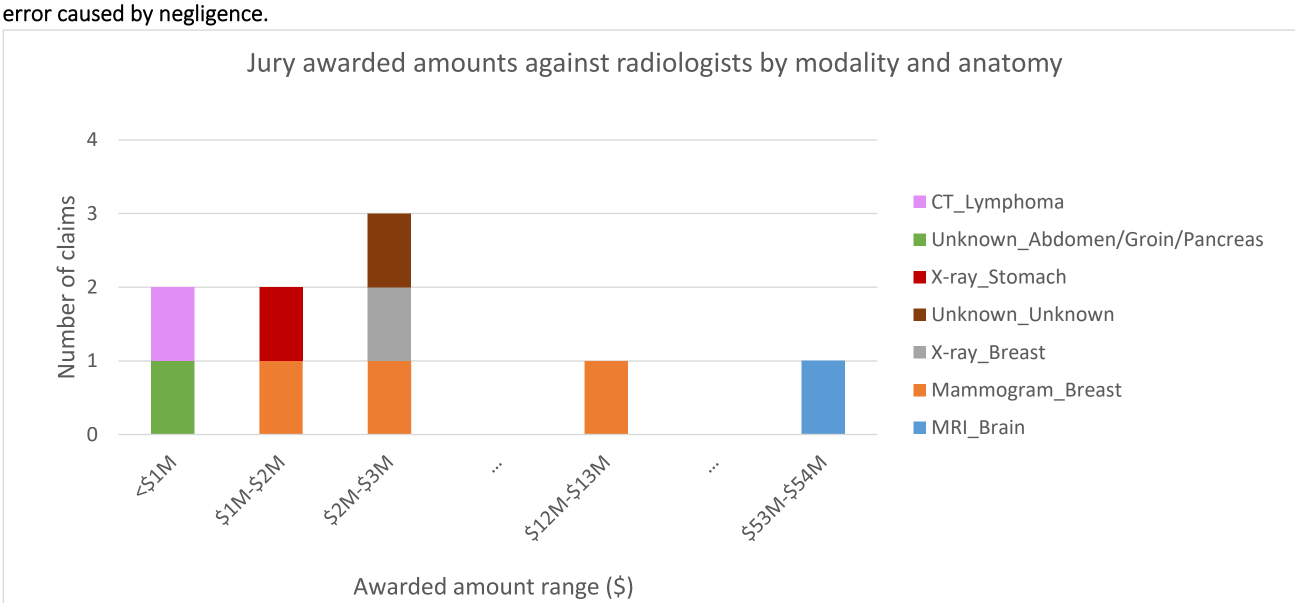


Figure 14: Amounts that are awarded by the jury to the plaintiff and against the radiologist as a result of a diagnostic error caused by negligence.



Copyright © 2021 Society of Actuaries Research Institute

Our observations from the two figures above are that:

- There were 19 cases that involved a negligent diagnostic error caused by a radiologist with a settled amount and 9 cases with a jury award amount.
- For the 19 cases with a settled amount, the settled amounts ranged between $225K and $4m (inflated to 2020 dollars). The average settled amount was $1.8m. The jury awarded amounts ranged between $293K and $54m (inflated to 2020 dollars). Of the 19 cases, we were not able to identify the share of the liability to the radiologist for 3 cases (2 cases concerning a mammogram of the breast of $1.2m and $2.4m and 1 case concerning a brain MRI of $0.6m). The majority of the 19 cases were cases related to mammograms of the breast. A study on the causes of medical malpractice suits against radiologists across 47 states in the US also found that errors in diagnosis were the most common generic cause of malpractice suits against radiologists, with breast cancer being the most frequently missed or incorrect diagnosis (Whang et al., 2013).
- For the 9 cases with a jury award amount, the average awarded amount was $7.6m. This was driven by two significantly large awards of $12m (a case concerned with a mammogram of the breast) and $53.7m (a case concerned with a brain MRI). Of the 9 cases, we were not able to identify the share of the liability to the radiologist for 1 case (the lymphoma CT case of $0.3m)

The above give us some insights of the mix of claims settled by a jury verdict. Compared to the average indemnity amount from NPDB of about $650K as seen in Figure 12, this shows that jury settled and awarded claims (which tend to be related to the breast and brain) are likely to be the larger medical malpractice claims from diagnostic errors by radiologists in NY.

2.2 PHASE 2 – INVESTIGATION OF POTENTIAL KEY CHANGES FROM IMPACT OF AI

After developing the baseline projection of future rate of paid claims with no AI influences, we consider how AI may impact the rate of paid claims and average indemnity size. For the rate of paid claims, we explore how AI may impact the frequency of diagnostic errors and subsequently medical malpractice claims. This is discussed in Section 2.2.1. For average indemnity sizes, we explore how AI may impact the mix of the claims, such as its impact on the larger claims compared to the smaller claims. This is discussed in Section 2.2.2.

2.2.1 POTENTIAL KEY CHANGES ON RATE OF PAID CLAIMS FROM AI

In this study, the potential impact of AI on rate of paid claims from diagnostic errors in radiology has been explored in terms of:

- The impact of AI on the accuracy of diagnosis
- The proportion of imaging modalities/procedures that are likely to be impacted by AI
- The proportion of claims from diagnostic errors in radiology likely to be impacted by AI
- The adoption rate of AI by radiologists

<u>The impact of AI on accuracy of diagnosis</u>

To assess the impact of AI on the accuracy rate of diagnosis, a literature review on the impact of AI in radiology was conducted. Common measures to assess the clinical performance of AI for diagnosis as noted by the Radiological Society of North America (RSNA, 2018) and in academic literature include:

- **Sensitivity** – metric that evaluates a model's ability to predict true positives of each available category (True Positives/True Positives + False Negatives)

Copyright © 2021 Society of Actuaries Research Institute

- **Specificity** – metric that evaluates a model's ability to predict true negatives of each available category (True Negatives/True Negatives + False Positives)
- **Area under the ROC Curve (AUC)** – represents the probability that a model's predictions are correct in terms of the positives

The focus for this study was on the sensitivity metric where *1 – sensitivity* represents the false negative error rate. In general, false negative errors have been found to be 5 times more common than false positive errors in radiological errors (Radiology Key, 2016). The potential reduction in the false negative error rate from AI compared to radiologists has therefore been treated as a proxy for the reduction in diagnostic errors in radiology. We then assume that medical malpractice claims reduce by the same proportion as the reduction in diagnostic errors. We note that medical malpractice claims arise from diagnostic errors that have been deemed to be negligent and so it is possible that the reduction of medical malpractice claims is different to the reduction of diagnostic errors from AI technology.

Our review of literature included a review of research which compared the diagnostic performance of AI technology compared to radiologists from medical images. This review considered academic journals as well as a review of existing FDA approved AI technology in radiology based on FDA summary reports of this technology. Across the sources, it was suggested that the rate of false negative errors may reduce between a range of 13% to 60%, with the average reduction in rate of false negative errors across all sources being 32%. It should be noted that some of the AI in academic journals may never be approved by the FDA and those that have been approved by the FDA may improve over time. The measures we have adopted however assume that the accuracy rate is static. As our sources include a large proportion of AI technology that is already approved by the FDA, the observed reduction in false negative errors may be conservative. We note the need to be aware of emerging research on the improvements in the accuracy rate of AI technology as it trains more data and becomes more widely accepted in practice. More information of our review can be found in Appendix A.2.2.

Assumption on reduction in rate of paid claims from improved accuracy on diagnosis

We therefore propose a scenario that assumes a potential reduction of **30%** for modalities where AI technology is having more of an impact now i.e., CT, MRI and X-Rays.

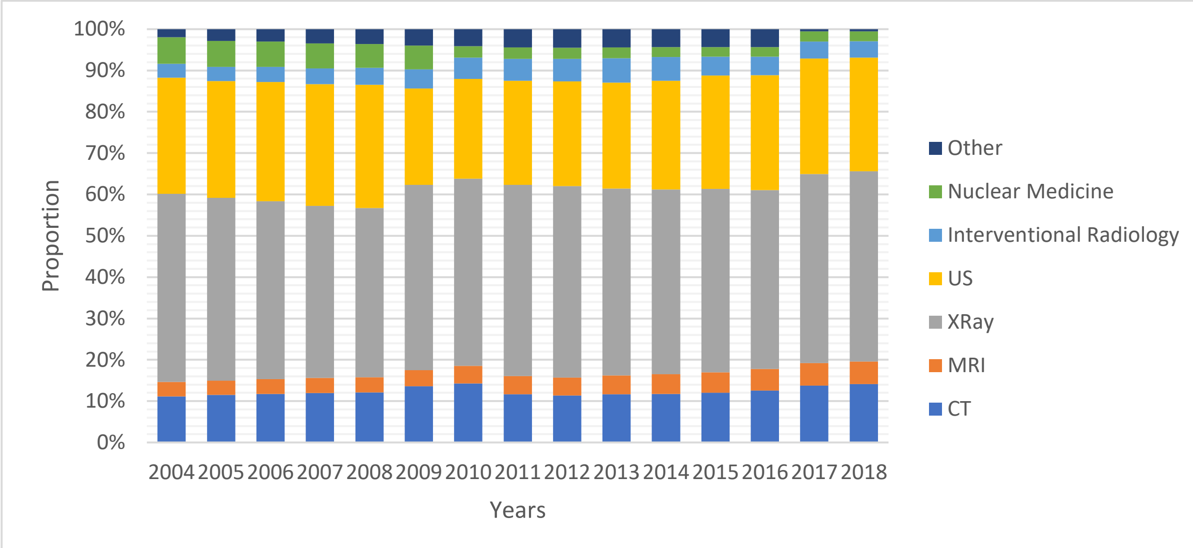
We define this assumption as b in Section 3.1.

The proportion of imaging modalities/procedures that are likely to be impacted by AI

The figure below shows the proportion of imaging procedures by modality from 2004 to 2018 (NHPI Volume of Services Performed data, 2018).

Copyright © 2021 Society of Actuaries Research Institute

Figure 15: Proportion of Imaging Procedures by Modality in New York in each year from 2004 to 2018



For example, in 2018, of all imaging procedures undertaken, approximately 14% were CT, 5.4% MRI, 28% Ultrasound (US), 4% Interventional Radiologist, 46% X-Ray and 2.4% Nuclear according to NHPI's Volume of Services Performed data, sourced from Centers for Medicare and Medicaid Services (CMS) 5% Research Identifiable Files. Adding the CT, MRI and X-Ray proportions, the total proportion is 65.4%. According to a technography study of applications of AI in diagnostic radiology (Mehrizi et. al 2020), most current AI applications target CT, MRI and X-Ray modalities with very few applications in ultrasound and mammography modalities.

Assumption (proportion of imaging procedures to be impacted)

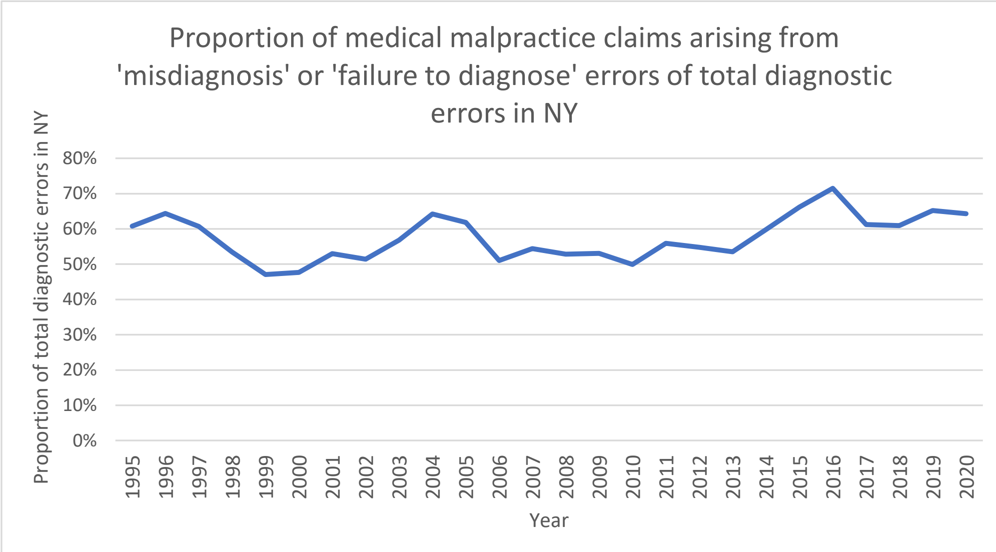
We therefore propose a scenario that assumes **65%** of imaging procedures have the potential to be impacted (based on the proportion of AI-impacted procedures – CT, MRI and X-Ray add up to 65.4%)

We define this assumption as c in Section 3.1.

The proportion of claims from diagnostic errors in radiology likely to be impacted by AI-ML technology

The figure below shows medical malpractice claims caused by a 'misdiagnosis' or a 'failure to diagnose' error as a proportion of all diagnostic errors in NY from 1995 to 2020. We assume that medical malpractice claims caused by 'misdiagnosis' or 'failure to diagnose' errors are more likely to be impacted by AI-ML technology focused on accurate diagnosis than claims from 'delay in diagnosis' errors which are more likely to be impacted by AI technology that focuses on triaging medical images.

Figure 16: Medical malpractice claims arising from misdiagnosis or failure to diagnose errors as a proportion of total diagnostic errors in NY from 1995 to 2020.



In the last 5 years, the proportion of medical malpractice claims caused by diagnostic errors that was a misdiagnosis or failure to diagnosis error was on average 65% of all diagnostic errors (based on NPDB data from 2016 to 2020)

Assumption (proportion of diagnostic errors that are misdiagnosis/failure to diagnose)

We therefore propose a scenario that assumes **65%** of diagnostic errors are due to misdiagnosis/failure to diagnosis where improved accuracy rates from AI-ML based technology will potentially have an impact.

We define this assumption as m in Section 3.1.

The adoption rate of AI by radiologists

Definitive Healthcare surveyed imaging leaders and radiologists from US and imaging centers from October to December 2019. According to this survey, approximately 30% of imaging centers and hospitals are using AI technology in radiology. Nearly one third said they plan to use the technology within the next two years (i.e., by end of 2021).

Another survey by Reaction Data in 2018, shows the percentage of application that radiologists plan to use AI in: Breast (68%), Lung (61%), Chest (58%), Bone (38%), Liver (34%), Cardiovascular (36%), Neural Aneurysm (26%) and Pulmonary hypertension (16%). The survey results showed that:

- 14% have been using ML for a while (i.e., prior to 2018)
- 7% just adopted (i.e., 2018)
- 11% are planning on adopting ML in the next 12 months (i.e., 2019)
- 27% are 1-2 years away from adopting ML (i.e., 2020)
- 25% are 3+ years away (i.e., 2021 and after)
- 16% do not think they will ever utilize ML

The two surveys were fairly consistent in noting that approximately 30% of radiologists will likely be using AI by 2020. The survey from Reaction Data (2018) further suggests a gradual increase in the adoption rate of AI by radiologists after 2020.

Copyright © 2021 Society of Actuaries Research Institute

Assumption (adoption rate of AI by radiologists)
We therefore propose a scenario that assumes the adoption rate of AI (based on the two survey results) to increase over time in line with the table below.

2020	2021	2022	2023	2024	2025	2026 and thereafter
30%	40%	50%	60%	70%	80%	85%

We define this assumption as a_i for each year i, where i =2021, 2022, 2023, 2024, 2025, 2026 and thereafter in Section 3.1.

In summary, when combining all these factors together, it is likely that AI will decrease the rate of paid claims from diagnostic errors in radiology due to the improved accuracy rate and therefore reduced rate of diagnostic errors from this technology compared to radiologists.

2.2.2 POTENTIAL KEY CHANGES IN AVERAGE INDEMNITY AMOUNT FROM AI

There is more uncertainty on how AI may impact the average indemnity amount compared to the rate of paid claims. We therefore seek to explore the potential impact qualitatively.

According to a technography study of applications of AI in diagnostic radiology (Mehrizi et. al 2020), most AI applications target CT, MRI and X-Ray modalities with very few applications in ultrasound and mammography modalities. The most popular anatomical region is the brain, followed by lung, chest, breast and cardiovascular. While current AI applications approved by the FDA for the mammography modality on the breast has been limited (Mehrizi et al., 2020), a survey conducted among members of the European Society of Radiology indicated that a large proportion of responders believe future AI developments will have the most impact on the breast, oncologic, thoracic and neuroimaging mainly involving mammography, computed topography (CT) and magnetic resonance (MRI) (Codari et al., 2019). This is consistent with another report by Reaction Data in 2018 which showed the percentage of application that radiologists plan to use AI include: Breast (68%), Lung (61%), Chest (58%), Bone (38%), Liver (34%), Cardiovascular (36%), Neural Aneurysm (26%) and Pulmonary hypertension (16%).

The above suggests that medical malpractice claims associated with the breast and the mammography modality or with the brain and the MRI modality could reduce with the increased focus of AI to improve diagnoses in those areas in the future. As these types of claims tend to be the larger claims, as seen from our Westlaw analysis in Section 2.1.4, this could in turn potentially reduce the average indemnity amounts of medical malpractice claims caused by diagnostic errors by radiologists when AI is adopted.

It should be noted that AI could also have the potential to increase average indemnity amounts. For example, it could have the potential to increase the number of megaverdicts (verdicts that award damages of millions of dollars or more) which may be driven by juror sentiment when AI is involved in a medical malpractice case. To date, there have been no major cases that have evidence of negligence by AI technology in radiology (D’Arcy, 2021) and therefore, it is difficult to quantify whether a decrease or an increase in average indemnity amounts is more likely when AI is involved.

Copyright © 2021 Society of Actuaries Research Institute

Section 3: Legal Scenarios and Sensitivities

In Section 2, the potential impact of AI on medical malpractice claims from diagnostic errors was explored in terms of the rate of paid claims and the average indemnity size. In this section, we continue the exploration of the potential impact of AI under different legal scenarios. Sensitivity testing of the assumptions used in the analysis are also discussed.

3.1 LEGAL SCENARIOS

With the introduction of AI in healthcare, situations where the AI technology may be trained in inappropriate environments, use imperfect techniques or incomplete data may arise. Even when the AI is trained as well as possible, diagnostic errors may still occur. For example, a tumor may still be missed or misdiagnosed by the AI technology when it is used for diagnosis from a medical image. When these errors occur, patients may be injured. Therefore, it is important to consider the parties that may be liable when AI is involved. This in turn affects when a physician may be liable which will have an impact on medical malpractice claims.

Under current tort law, one landmark of a lawsuit is the demonstration of whether the 'standard of care' was met by the physician. A legal definition of standard of care is "the watchfulness, attention, caution and prudence that a reasonable person in the circumstances could exercise; failure to meet the standard of care is negligence". Therefore, if it is deemed that the physician did not perform their duties within the expected standard of care, then they are likely be held liable for patient injuries resulting from their negligence. For example, if a radiologist used AI for diagnosis which resulted in a diagnostic error but was deemed negligent because they did not meet the expected 'standard of care', then this would likely result in a medical malpractice claim.

As AI is increasingly adopted by radiologists, the situation of defensive medicine may arise. This is the practice of recommending a diagnostic test or medical treatment that is not necessarily the best option for the patient but an option that mainly serves the function to protect the physician against the patient as potential plaintiff. For example, the new norm in diagnostic processes, as dictated by a new 'standard of care', may be that radiologists include the use of AI for diagnosis regardless of whether it is necessary or not. If a diagnostic error occurs in this situation, the radiologist may be able to prove that they performed under the expected 'standard of care' and may not be deemed negligent. Therefore, the potential for the definition of standard of care to change as AI is increasingly adopted by radiologists or other physicians in healthcare results in an endless possibility of scenarios.

There is also a high degree of complexity when considering the legal environment of AI in healthcare because of not only how the standard of care may change but also the interaction of rules and regulations (including the interactions of FDA approval and state tort liability, the role of products liability and the learned intermediary doctrine, the role of vicarious liability and the role of informed consent liability) as well as the difficulties of AI being a 'black box' that impacts transparency, interpretability, and ethics. If AI is involved in the provision of healthcare services, stakeholders including the developer, provider, and physician may have liability under a variety of tort law principles. For example, a developer may be held liable if defects in the AI are unreasonably dangerous to users, a physician may be held liable for negligent use of the AI in making clinical decisions and a provider may be held liable for not disclosing the use of AI to their patients as part of the informed consent process. In terms of where the liability will fall, that is, under professional liability or product liability, is likely to depend on the functions the AI is performing.

Based on the above considerations, we propose three simple legal scenarios that, although are simplifications of the complex legal environment that is likely to evolve as AI is increasingly adopted in

Copyright © 2021 Society of Actuaries Research Institute

healthcare, provide a high-level overview of important considerations from the legal perspective. We define the variables that impact the new projected rate of claims paid as follows:

$r =$ the base projected rate of claims paid

$r^* =$ new projected rate of claims paid

$b =$ reduction in rate of paid claims for improved diagnostic accuracy due to AI

$c =$ proportion of imaging procedures impacted by AI

$m =$ proportion of diagnostic errors that are misdiagnoses/failure to diagnose

$a_i =$ adoption rate for AI technology for year i

The three proposed legal scenarios are:

- **Scenario 1**: current general tort law principles still apply where the current legal standard of care is key to liability for medical AI i.e., healthcare professionals are responsible for harm if they do not take adequate measures in properly evaluating the AI technology used in care for the patient.

  This scenario assumes that past medical malpractice claims from diagnostic errors would remain medical malpractice claims because the standard of care definitions are unchanged. However, these claims would likely reduce in frequency if AI was introduced because of the improved accuracy of AI technology in diagnosis.

  Mathematically, this results in the **new projected rate of claims paid for each year for medical malpractice claims from diagnostic errors by radiologists** to be:

  $$r^* = r \times a_i \times c \times m \times (1-b) + r \times (1 - a_i \times c \times m)$$

- **Scenario 2:** current law is modified such that "personhood" is conferred on the AI machine i.e., the AI machine is viewed as an independent person. We assume claims made against an AI machine will go either to medical malpractice liability or products liability where the AI machine is covered instead of the radiologist.

  Mathematically, this results in the **new projected rate of claims paid for each year for medical malpractice claims from diagnostic errors by radiologists** to be:

  $$r^* = r \times (1 - a_i \times c \times m)$$

- **Scenario 3:** current law is modified for common enterprise liability i.e., if some injury is caused by the AI technology, then all groups involved in use and implementation of the AI system should jointly bear some responsibility. We assume the shared responsibility to be 50% of the liability to be borne by the AI machine and/or hospital and 50% of the liability to be borne by the physician.

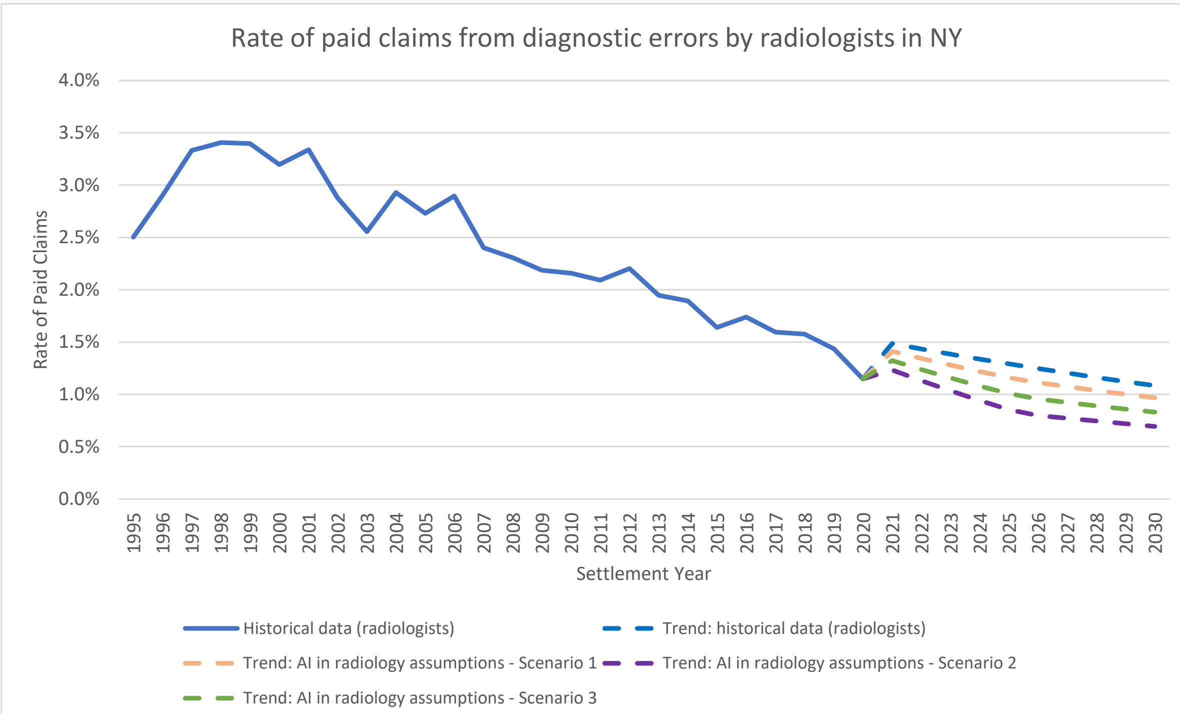
Copyright © 2021 Society of Actuaries Research Institute

Mathematically, this results in the **new projected rate of claims paid for each year for medical malpractice claims from diagnostic errors by radiologists** to be:

$$r^* = r \times a_i \times c \times m \times (1-b) \times 0.5 + r \times (1 - a_i \times c \times m)$$

Putting the three scenarios together, we observe the following changes in future rate of paid claims for radiologists.

Figure 17: Future rate of paid claims from diagnostic errors by radiologists in NY by three legal scenarios



The resulting percentage reduction in rate of paid claims for these three scenarios from the baseline projection for each year from 2021 to 2030 are shown in the table below.

Table 1: Percentage reduction in future rate of paid claims from diagnostic errors by radiologists in NY for the three legal scenarios compared to the baseline projection of future rate of paid claims from 2021 to 2030

	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030
Scenario 1	-5.1%	-6.3%	-7.6%	-8.9%	-10.1%	-10.8%	-10.8%	-10.8%	-10.8%	-10.8%
Scenario 2	-16.9%	-21.1%	-25.4%	-29.6%	-33.8%	-35.9%	-35.9%	-35.9%	-35.9%	-35.9%
Scenario 3	-11.0%	-13.7%	-16.5%	-19.2%	-22.0%	-23.3%	-23.3%	-23.3%	-23.3%	-23.3%

Figure 17 and Table 1 show that for all three legal scenarios, as AI is gradually adopted by radiologists in diagnosis between 2021 to 2030, the percentage reduction in the rate of paid claims from the baseline projection of rate of paid claims increases. The percentage reduction in the rate of paid claims is largest in Scenario 2, followed by Scenario 3 and then Scenario 1. This is not surprising as many of the claims in Scenario 2 become product liability claims or medical malpractice claims by the AI rather than the radiologists which results in the larger reduction. It is also reasonable for Scenario 3 to have a larger

Copyright © 2021 Society of Actuaries Research Institute

reduction in rate of paid claims than Scenario 1 because shared liability means that some of the liability moves to other parties other than the radiologist.

It should be noted that regardless of the change in assumptions used to derive the baseline projection (such as the changes in assumptions shown in the sensitivity testing in Section 3.2.1), the percentage reductions in rate of paid claims from the baseline projection in Table 1 will remain unchanged for each legal scenario, as they are relative to the baseline projection and only dependent on the assumptions (i.e. b, c, m and a_i in Section 2.2.1) used to derive the potential impact of AI on future rate of paid claims.

3.2 SENSIVITY TESTING OF ASSUMPTIONS

This section summarizes the sensitivity testing of assumptions to project future rate of paid claims based on historic experience (i.e., the baseline projection) and of assumptions that consider the impact of AI to project future rate of paid claims (including the assumption on reduction in rate of paid claims due to reduced false negative diagnostic errors and on the adoption rate of AI by radiologists).

3.2.1 SENSITIVITIES ON ASSUMPTIONS FOR THE BASELINE PROJECTION

Sensitivity on the assumed baseline future rate of paid claims reduction

One of the assumptions for the baseline projection is the future level of reduction in the rate of paid claims. The figure below shows the changes in the baseline projection (blue line) if the projection assumed a fixed 4% per annum reduction in the rate of paid claims (green line) or if the projection followed a linear regression fit (red line).

Figure 18: Sensitivity testing of the future reduction of rate of paid claims for the baseline projection

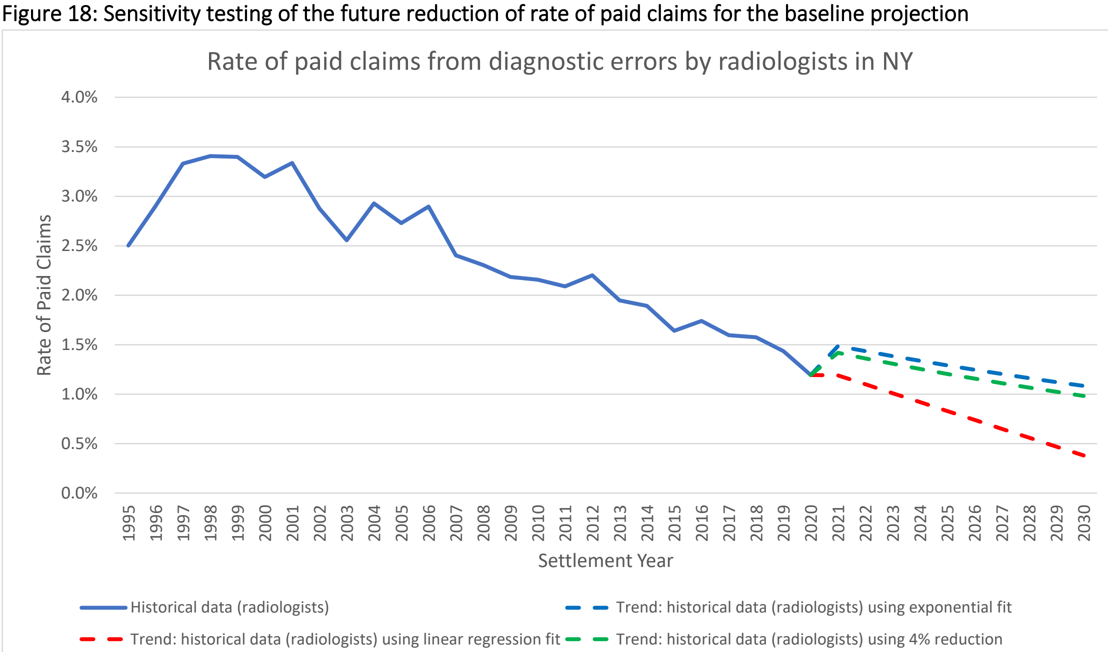


Figure 18 shows how the adopted baseline projection using an exponential fit is similar to the projected baseline projection if a fixed 4% per annum reduction in line with historic experience was deterministically applied. If a linear regression fit was applied to obtain the baseline projection, the rate of paid claims would decline to below 0.5% by 2030. This rate of reduction appears unreasonable as it is unlikely that the number of claims will reduce to close to zero. Although the analysis on the potential impact of AI on medical malpractice claims is not impacted by this assumption, the projection using the exponential fit was adopted for demonstration of the study's framework. This is because its shape in the decline in the rate of paid claims i.e., reducing at a reducing rate appeared most reasonable.

Sensitivity on the assumed proportion of medical malpractice claims from diagnostic errors that is caused by radiologists

Another assumption required for the baseline projection is the proportion of medical malpractice claims from diagnostic errors that were caused by radiologists. This is because a limitation of the NPDB public use datafile is that it does not provide physician specialty information. A 12% assumption based on CRICO's Diagnosis-Related Cases report was adopted, as noted in Section 2.1.3. The figure below shows the sensitivity of this assumption if 20% or 5% was adopted instead.

Figure 19: Sensitivity testing of the assumed proportion of medical malpractice claims that is caused by radiologists for the baseline projection

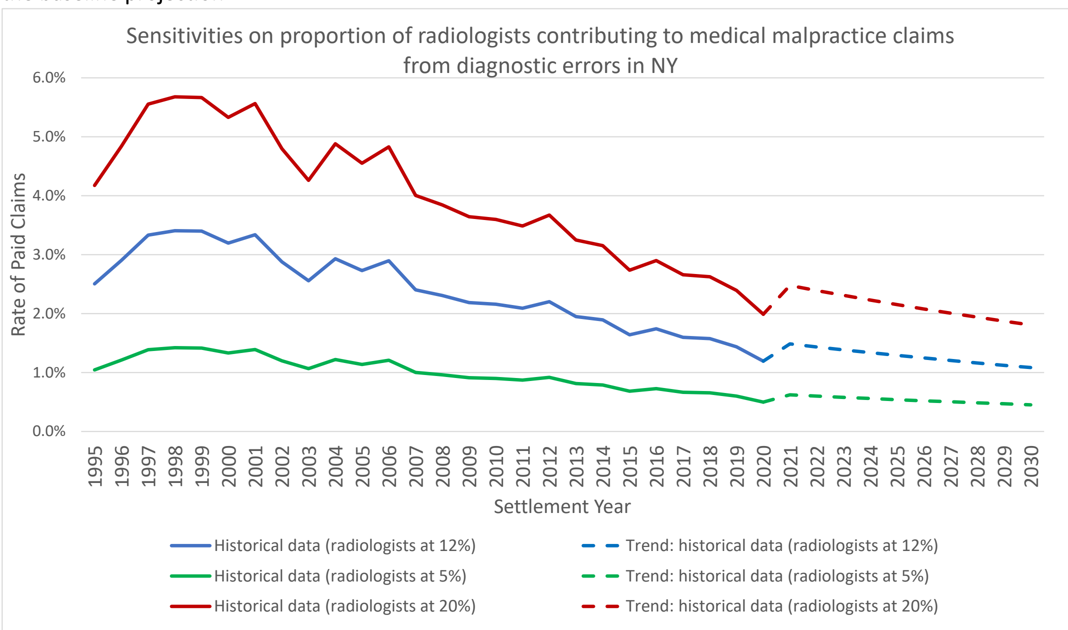


Figure 19 shows that if the proportion of medical malpractice claims from diagnostic errors by radiologists was 20%, the rate of paid claims would reduce from approximately 5.8% in the late 1990s to approximately 1.9% in 2030. If it was 12% (as assumed), the rate of paid claims would reduce from approximately 3.4% in the late 1990s to approximately 1.1% in 2030. If it was 5%, the rate of paid claims would reduce from approximately 1.5% in the late 1990s to approximately 0.5% in 2030. Out of these three sensitivities, the adopted assumption of 12% resulted in rates of paid claims that were most in line with Schaffer et al. (2017) of 1.9% in the period 2003-2008 and 1.4% in the period 2009-2014.

3.2.2 SENSITIVITIES ON ASSUMPTIONS FOR THE POTENTIAL IMPACT OF AI ON RATE OF PAID CLAIMS

The two key assumptions in the legal scenarios with the greatest uncertainty are the reduction in rate of paid claims due to reduced false negative diagnostic errors when AI is introduced and the adoption rate of AI by radiologists between 2021 to 2030. The sensitivity testing of these assumptions is demonstrated below. These assumptions are sensitivity tested from the legal scenario 1 projection (the current standard of care definition does not change), leaving other assumptions in Phase 2 as outlined in Section 2.2.1 unchanged.

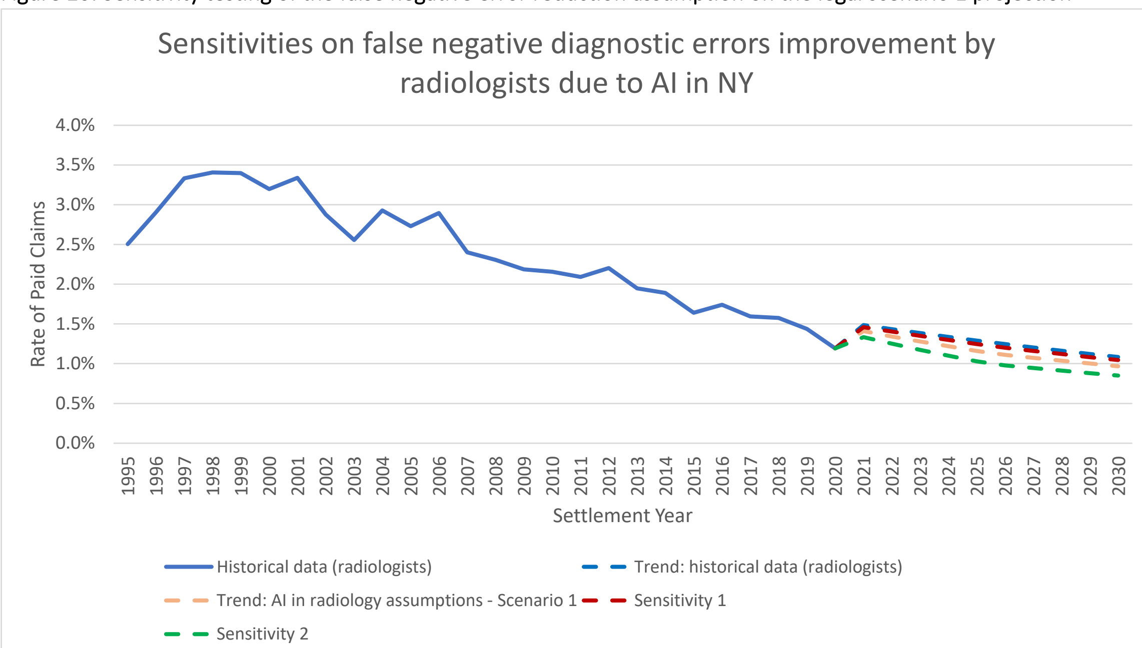
Sensitivities on the assumed reduction in the rate of paid claims due to reduced false negative diagnostic errors:

To explore the sensitivity testing of the assumed reduction in rate of paid claims due to reduced false negative diagnostic errors from AI, the following changes to the assumption was adopted.

- Sensitivity 1: If false negative error reduced by 10% instead of 30%
- Sensitivity 2: If false negative error reduced by 60% instead of 30%

The figure below shows the resulting changes in the projected rate of paid claims, where the blue line is the baseline projection based on historic experience, the yellow line is the projection using legal scenario 1, and the red and green lines are the projections using the changes noted in Sensitivity 1 and Sensitivity 2 respectively.

Figure 20: Sensitivity testing of the false negative error reduction assumption on the legal scenario 1 projection



The resulting percentage reduction in rate of paid claims from the baseline projection for each year from 2021 to 2030 are summarized in the table below.

Table 2: Percentage reduction in future rate of paid claims from diagnostic errors by radiologists in NY for sensitivity 1 and 2 compared to the baseline projection of future rate of paid claims from 2021 to 2030

	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030
Scenario 1	-5.1%	-6.3%	-7.6%	-8.9%	-10.1%	-10.8%	-10.8%	-10.8%	-10.8%	-10.8%
Sensitivity 1	-1.7%	-2.1%	-2.5%	-3.0%	-3.4%	-3.6%	-3.6%	-3.6%	-3.6%	-3.6%
Sensitivity 2	-10.1%	-12.7%	-15.2%	-17.7%	-20.3%	-21.5%	-21.5%	-21.5%	-21.5%	-21.5%

Copyright © 2021 Society of Actuaries Research Institute

Sensitivities on the assumed adoption rate of AI by radiologists:

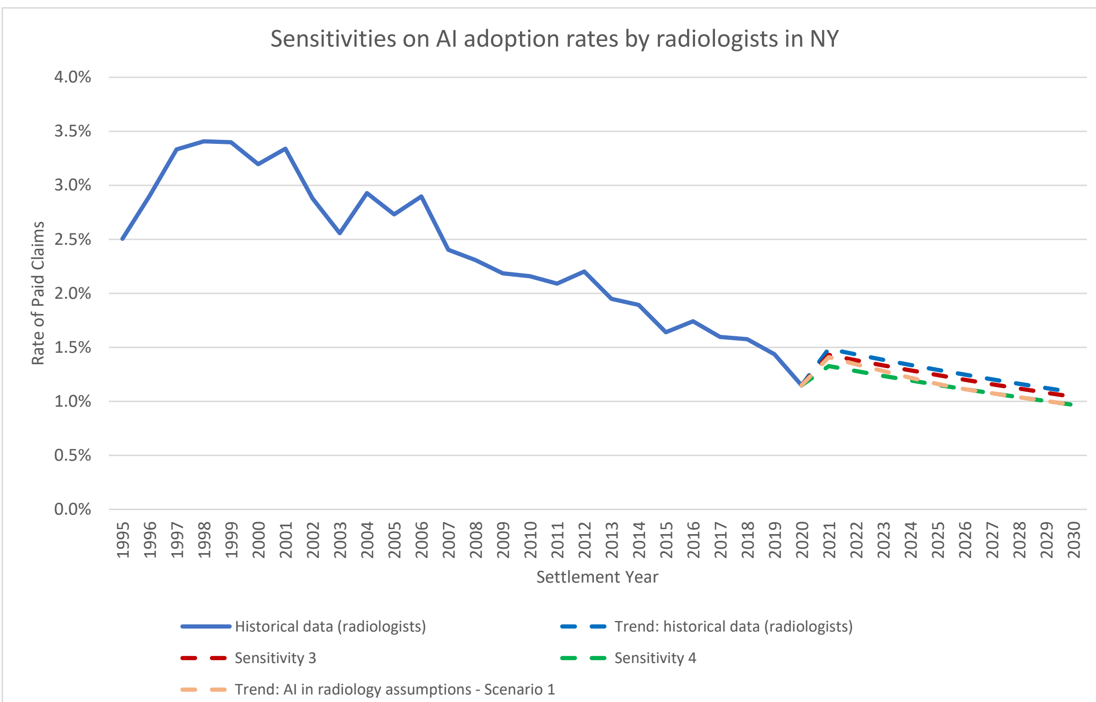
To explore the sensitivity testing of the assumed adoption rate of AI by radiologists, the following changes to the assumption were adopted. In Sensitivity 3, we retain the adoption rate of AI by radiologists in 2020 of 30% for all years from 2021 to 2030. In Sensitivity 4, the long-term adoption rate of AI of 85% is adopted for all years from 2021 to 2030.

Table 3: Changes in the assumed adoption rate of AI by radiologists for Sensitivity 3 and 4.

	2021	2022	2023	2024	2025	2026 and thereafter
Original	40%	50%	60%	70%	80%	85%
Sensitivity 3	30%	30%	30%	30%	30%	30%
Sensitivity 4	85%	85%	85%	85%	85%	85%

The figure below shows the resulting changes in the projected rate of paid claims, where the blue line is the baseline projection based on historic experience, the yellow line is the projection using legal scenario 1, and the red and green lines are the projections using the changes noted in Sensitivity 3 and Sensitivity 4 respectively.

Figure 21: Sensitivity testing of the AI adoption rate assumption on the legal scenario 1 projection



The resulting percentage reduction in rate of paid claims from the baseline projection for each year from 2021 to 2030 are summarized in the table below.

Table 4: Percentage reduction in future rate of paid claims from diagnostic errors by radiologists in NY for sensitivity 3 and 4 compared to the baseline projection of future rate of paid claims from 2021 to 2030

	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030
Scenario 1	-5.1%	-6.3%	-7.6%	-8.9%	-10.1%	-10.8%	-10.8%	-10.8%	-10.8%	-10.8%
Sensitivity 3	-3.8%	-3.8%	-3.8%	-3.8%	-3.8%	-3.8%	-3.8%	-3.8%	-3.8%	-3.8%
Sensitivity 4	-10.8%	-10.8%	-10.8%	-10.8%	-10.8%	-10.8%	-10.8%	-10.8%	-10.8%	-10.8%

Copyright © 2021 Society of Actuaries Research Institute

Section 4: Conclusion

The purpose of this study was to provide a possible framework for interested actuaries to analyze the potential impact of AI on medical malpractice claims. The study focused on one specialty, radiology and one jurisdiction of the US, New York as a useful case study. Exploration of all different intended uses of AI across the healthcare industry and legal considerations across all jurisdictions would likely be limitless and far beyond the intended scope of this study.

The key uncertainty related to the study was from the historic data used to project future experience. The NPDB public use dataset was only on closed claims and did not have physician specialty information. Assumptions on the proportion of medical malpractice claims from diagnostic errors by radiologists and the number of active radiologists were made based on other external sources. Despite the data limitations, this report was created with the intention to help actuaries understand the considerations addressed and perform similar analysis for their work using their own data. Therefore, the available data were used for demonstration purposes of a possible framework.

The results of this study show that AI may reduce the rate of paid medical malpractice claims from diagnostic errors by radiologists. The resulting potential percentage reduction in the future rate of paid claims across three simple legal scenarios compared to the baseline projection ranged between 5% to 36% for each future year to 2030 depending on the legal scenario. There is greater uncertainty about the impact of AI on the average indemnity size of these claims. There appears to be a greater focus on AI development for imaging modalities and anatomy areas that tend to have larger jury verdicts and settlements related to medical malpractice claims. Reduced claim rates in these contexts could result in a reduction in the average indemnity amounts of medical malpractice claims caused by diagnostic errors by radiologists. However, to date, there have been no jury cases that have evidence of negligence by AI in radiology. Therefore, there could also be an increase in the number of megaverdicts when AI is involved in a medical malpractice case, which may be driven by juror sentiment.

Actuaries can use the framework of this report to develop scenario tests in relation to this digital disruption to their own portfolio of medical malpractice claims. Similar work can be applied to the impact of AI on medical malpractice claims on other medical specialties such as pathology, surgery, ophthalmology and cardiology (Ahuja 2019, Bohr et al., 2020).

Copyright © 2021 Society of Actuaries Research Institute

Section 5: Acknowledgments

The researchers wish to extend their gratitude to collaborators including the Project Oversight Group and SOA members for overseeing the project and for reviewing the analysis and report for accuracy and relevance.

We also thank the research assistance of Macquarie University's PhD student, Kyu Hyung Park who helped us in data investigation and literature review related to this project.

Project Oversight Group Members:

Ken Avner FSA, MAAA
Joan Barrett FSA, MAAA
Yi Chen
Tyler Doiron
Janet Duncan FSA, MAAA, FCAS
Sophie Feng ASA, CERA, ACIA
Shasha Huang ASA
Julie Meadows FSA, MAAA, CERA
Tamara Wilt ASA, MAAA

At the Society of Actuaries:

Rob Montgomery, ASA, MAAA, FLMI, Consultant – Research Project Manager

Copyright © 2021 Society of Actuaries Research Institute

Appendix A: Data and Analysis

A.1 DATA VALIDATION

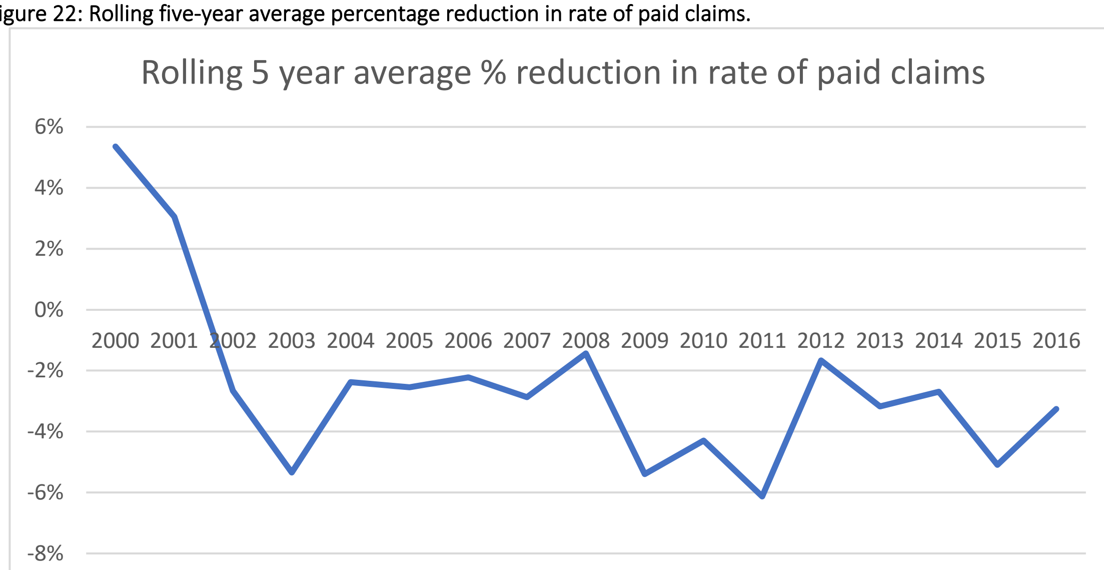
As the NPDB public data file has been a commonly used dataset, we performed data reconciliations of the historic data to other sources.

- Tehrani et al., 2013 – Compared to this study, we obtained similar number of claims and mean, and total claims paid by allegation description using the NPDB public use dataset. Our total number of claims was 1.5% higher than the study and the mean and total claims paid by allegation description ranged between 4.8% to 15% lower. The differences were likely due to the slight differences in cut off dates in year 2012 and different CPI adopted. We adopted the "Historical consumer price index for all urban consumers" as advised by NPDB and also used by Mello et al., 2014. Tehrani et al., 2013 used "Medical Care Consumer Price Index"
- Schaffer et al., 2017 – We obtained slightly higher total number of claims and payments compared to this study for this same period at 2.4% and 7.5% respectively. However, we note this study was conducted using data up to 2014 so differences are likely due to differences in CPI inflation assumption and any changes the NPDB data has made to historic data between 2014 and 2020.
- Mello et al., 2014 – Claim numbers and claims paid were graphed using the NPDB public use July 2020 dataset which showed a similar trend of decreasing claim numbers and slight increases in claims paid in this study.

A.2 ANALYSIS

A.2.1 Past percentage reduction in rate of paid claims

Figure 22: Rolling five-year average percentage reduction in rate of paid claims.



The above figure shows that a 4% p.a. reduction in rate of paid claims is not unreasonable for future rate of paid claims. This was shown as a sensitivity in Section 3.2.1.

Copyright © 2021 Society of Actuaries Research Institute

A.2.2 Literature Review to derive potential reduction in diagnostic errors from AI

This section shows the literature review conducted to derive the potential reduction in rate of paid claims from observed reductions in false negative errors.

Table 5 below summarizes the observations from a review of literature which compares the performance of AI technology and radiologists in diagnosis from medical radiology imaging. A search of the following on PubMed: (artificial intelligence [MeSH Terms]) AND (radiology) with filters including "systematic review" and "recent 5 years" returned 31 studies as at 21 April 2021 which were reviewed. A search on google scholar with the search term: "systematic review and meta-analysis accuracy of AI technology in radiology" was also conducted. Many of the studies produced either only the accuracy measures of the AI technology itself and not against practicing radiologists or provided a qualitative view of the accuracy. We summarize the studies that noted the accuracy measures of the AI technology that had been compared to radiologists. Table 6 below summarizes the observations from a review of existing FDA approved AI technology in radiology based on FDA summary reports of these technologies. We considered 80 FDA-approved AI technologies. Similar to the review in Table 5, Table 6 only lists the FDA-approved AI technologies that had a comparison of performance to radiologists available in the FDA summary reports.

Table 5: Summary table of review of literature

	Source	Summary
1	"Artificial intelligence performance in detecting tumor metastasis from medical radiology imaging: A systematic review and meta-analysis" (Zheng et al., 2020)	With the studies that compared performance between AI algorithms and health-care professionals, the pooled sensitivity was 89% for AI Algorithms and 72% for health-care professionals. The pooled specificity was 85% for AI algorithms and 72% for health-care professionals. This was a combination of MRI, CT, PET and US. An improvement of pooled sensitivity from 72% to 89% suggests that false negative error improves from 28% to 11% which is **a 60% reduction**
2	"Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXneXt algorithm to practicing radiologists" (Rajpurkar et al., 2018)	This study shows where the algorithm performed better than radiologists and did not perform better than radiologists. • There were 3 pathologies (cardiomegaly, emphysema and hiatal hernia) where radiologists performed better: AUCs of 0.888, 0.911 and 0.985 compared to 0.831, 0.704 and 0.851 by the AI respectively • In the other pathologies, it performed better for 1; AUC of 0.808 by radiologist compared to 0.863 by the AI This was a suggested AI algorithm and not one that has yet been approved. False negative to reduce from 20.1% to 13.7% i.e., **a 32% reduction**
3	"Artificial Intelligence versus Clinicians in Disease Diagnosis: Systematic Review" (Shen et al., 2019)	Identified two studies related to radiology: Study 1: AUC: 0.91 versus 0.885 Sensitivity: 80.7% versus 70.4% Specificity: No report Therefore, False negative error is suggested to reduce from 29.6% to 19.3% **i.e., a 35% reduction** Study 2: This was the source identified in source 2.

Copyright © 2021 Society of Actuaries Research Institute

4	"The use and performance of artificial intelligence applications in dental and maxillofacial radiology: A systematic review" (Hung et al., 2020)	Table 4 notes some sensitivity and specificity for diagnosis of osteoporosis Sensitivity ranged from 76.9% to 99.1% Specificity ranged from 43.8% to 98.4% In the recent studies, sensitivity and specificity all reached 95% and above
5	"Towards clinical application of image mining: a systematic review on artificial intelligence and radiomics" (Sollina et al., 2019)	There were two studies here that had an outcome of diagnosis in the Oncology (Breast and Ovary) field. The Oncology/Breast was on Ultrasound and recorded an AUC of 0.93 (i.e., 93% of outcomes likely to be correct). The Oncology/Ovary study was on Ultrasound and showed an 88% specificity and 98% sensitivity.
6	"Artificial Intelligence for Breast MRI in 2008 – 2018: A Systematic Mapping review" (Codari et al., 2019)	MRI breast cancer – using AI for lesion classification. Median AUC value was 0.87, IQR was 0.84 to 0.91
7	"Artificial Intelligence in radiology" (Hosny et. al 2018)	Noted that for classification tasks of lymph node metastasis in PET-CT, deep learning had higher sensitivities but lower specificities than radiologists.
8	"A systematic review of diagnostic accuracy of artificial intelligence-based computer programs to analyze chest x-rays for pulmonary tuberculosis" (Harris et. al 2019)	Development studies: median AUC 0.88 (0.82 – 0.90) versus Clinical studies median AUC 0.75 (0.66 to 0.87)

Note: the above studies may include AI algorithms that have not been approved by FDA yet. Below is a summary of FDA approved AI-ML technology in radiology.

Copyright © 2021 Society of Actuaries Research Institute

Table 6: Summary table of review of FDA approved AI-ML based technology in radiology

	AI technology	FDA approval date	Subspecialty	Body Area	Modality	Summary
9	QuantX	19/07/2017	Women's Imaging	Breast	MRI	AUC: mean AUC performance improved from 0.71 to 0.75. This suggests false negative errors to reduce from 29% to 25% i.e., **14% reduction**
10	OsteoDetect	24/05/2018	Musculoskeletal Imaging	Wrist	XRay	Sensitivity of 0.803 versus 0.747 and a Specificity of 0.914 versus 0.889. Looking at sensitivity, this suggests that false negative error improves from 25.3% to 19.7% i.e., a **22% reduction**
11	Koios DS for Breast	3/07/2019	Women's Imaging	Breast	MRI	AUC: mean AUC performance improved from 0.836 to 0.873. This suggests false negative errors to reduce from 16.4% to 12.7% i.e., a **23% reduction**
12	Transpara	5/03/2020	Women's Imaging	Breast	Mammogram	AUC: mean AUC performance improved from 0.833 to 0.863. This suggests false negative errors to reduce from 16.7% to 13.7% i.e., a **18% reduction**
13	MammoScreen	25/03/2020	Women's Imaging	Breast	Mammogram	AUC: mean AUC performance improved from 0.77 to 0.8. This suggests false negative errors to reduce from 23% to 20% i.e., a **13% reduction**
14	FractureDetect (FX)	30/07/2020	Musculoskeletal Imaging	Upper Extremity, Lower Extremity	X-ray	Sensitivity of 0.900 versus 0.819 and a specificity of 0.918 versus 0.890. Looking at sensitivity, this suggests that false negative error improves from 18.1% to 10% i.e., **a 45% reduction**

Source: Data Science Institute – American College of Radiology https://models.acrdsi.org/. Technologies have been reviewed up to 26 August 2020.

Note: the nature of AI-ML is that it can keep improving over time but, at the moment these accuracy metrics are static.

Copyright © 2021 Society of Actuaries Research Institute

References

AAMC, Physician Specialty Data Report, https://www.aamc.org/data-reports/workforce/interactive-data/active-physicians-age-and-specialty-2019

AbuDagga, A., Wolfe, S. M., Carome, M., & Oshel, R. E. (2016). Cross-sectional analysis of the 1039 US physicians reported to the National Practitioner Data Bank for sexual misconduct, 2003–2013. PloS one, 11(2), e0147800.

Ahuja, A. S. (2019). The impact of artificial intelligence in medicine on the future role of the physician. PeerJ, 7, e7702.

Bal B. S. (2009). An introduction to medical malpractice in the United States. Clinical orthopaedics and related research, 467(2), 339–347. https://doi.org/10.1007/s11999-008-0636-2

Belk (2019). Malpractice Statistics. See http://truecostofhealthcare.org/malpractice_statistics/

Benjamens, S., Dhunnoo, P., & Meskó, B. (2020). The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database. NPJ digital medicine, 3(1), 1-8.

Bohr, A., & Memarzadeh, K. (2020). The rise of artificial intelligence in healthcare applications. In Artificial Intelligence in healthcare (pp. 25-60). Academic Press.

Bureau of Labor Statistics (2020). Historical Consumer Price Index for All Urban Consumers (CPI-U): https://www.bls.gov/cpi/tables/supplemental-files/historical-cpi-u-202101.pdf

Canady, M. (2018). The Verdict Is In: Surviving a Medical Malpractice Trial. https://www.physicianleaders.org/news/the-verdict-surviving-medical-malpractice-trials

Carver, N., Gupta, V., & Hipskind, J. E. (2020). Medical error. StatPearls [Internet].

Codari, M., Melazzini, L. et al. (2019) Impact of artificial intelligence on radiology: a EuroAIM survey among members of the European Society of Radiology. Insights Imaging 10, 105. https://doi.org/10.1186/s13244-019-0798-3

Codari, M., Schiaffino, S., Sardanelli, F., & Trimboli, R. M. (2019). Artificial intelligence for breast MRI in 2008–2018: a systematic mapping review. American Journal of Roentgenology, 212(2), 280-292.

CRICO (2021). Diagnosis-Related Cases: National Overview. https://www.rmf.harvard.edu/Malpractice-Data/CBS-Data-Request/Available-CBS-Data

D’Arcy, A. (2021). Who is Liable If Artificial Intelligence Causes Medical Malpractice? https://djdlawyers.legalexaminer.com/health/who-is-liable-if-artificial-intelligence-causes-medical-malpractice/

Data Science Institute (2020). FDA Cleared AI Algorithms. See https://models.acrdsi.org/

Definitive Healthcare (2019). The Future of the AI Market: 2019 Study Results. See https://blog.definitivehc.com/2019-artificial-intelligence-study

Emerj (2019). Machine Learning For Medical Diagnostics – 4 Current Applications. https://emerj.com/ai-sector-overviews/machine-learning-medical-diagnostics-4-current-applications/

Copyright © 2021 Society of Actuaries Research Institute

Harris, M., Qi, A., Jeagal, L., Torabi, N., Menzies, D., Korobitsyn, A., ... & Ahmad Khan, F. (2019). A systematic review of the diagnostic accuracy of artificial intelligence-based computer programs to analyze chest x-rays for pulmonary tuberculosis. PloS one, 14(9), e0221339.

Hosny, A., Parmar, C., Quackenbush, J., Schwartz, L. H., & Aerts, H. J. (2018). Artificial intelligence in radiology. Nature Reviews Cancer, 18(8), 500-510.

Hughes (2015). How many radiologists? It depends on who to ask? https://www.neimanhpi.org/commentary/how-many-radiologists-it-depends-on-who-you-ask/

Hung, K., Montalvao, C., Tanaka, R., Kawai, T., & Bornstein, M. M. (2020). The use and performance of artificial intelligence applications in dental and maxillofacial radiology: A systematic review. Dentomaxillofacial Radiology, 49(1), 20190107.

Jin, D., Harrison, A. P., Zhang, L., Yan, K., Wang, Y., Cai, J., Miao, S., & Lu, L. (2021). Artificial intelligence in radiology. Artificial Intelligence in Medicine, 265–289. https://doi.org/10.1016/B978-0-12-821259-2.00014-4

KDnuggets (2017). What Are Artificial Intelligence, Machine Learning and Deep Learning. https://www.kdnuggets.com/2017/07/rapidminer-ai-machine-learning-deep-learning.html

Khullar, D., Jha, A. K., & Jena, A. B. (2015). Reducing Diagnostic Errors--Why Now?. The New England journal of medicine, 373(26), 2491–2493. https://doi.org/10.1056/NEJMp1508044

KFF, Professionally Active Specialist Physicians by Field. https://www.kff.org/other/state-indicator/physicians-by-specialty-area/?currentTimeframe=0&sortModel=%7B%22colId%22:%22Location%22,%22sort%22:%22asc%22%7D

Makary, M. A., & Daniel, M. (2016). Medical error—the third leading cause of death in the US. Bmj, 353.

Mello, M. M., Studdert, D. M., & Kachalia, A. (2014). The medical liability climate and prospects for reform. Jama, 312(20), 2146-2155

Mehrizi, M. H. R., van Ooijen, P., & Homan, M. (2020). Applications of artificial intelligence (AI) in diagnostic radiology: a technography study. European Radiology, 1-7.

National Academies of Sciences, Engineering, and Medicine. (2015). Improving diagnosis in health care. National Academies Press.

Neri, E., de Souza, N., Brady, A., Bayarri, A. A., Becker, C. D., Coppola, F., & Visser, J. (2019). What the radiologist should know about artificial intelligence-an ESR white paper.

NPDB. Public Use Data File. https://www.npdb.hrsa.gov/resources/publicData.jsp

NHPI (2021). Neiman Almanac data. https://www.neimanhpi.org/almanac/

NHPI (2018). Volume of Services Performed data. https://www.neimanhpi.org/almanac/series-select/#category-25

OECD (2020). Trustworthy AI in Health. https://www.oecd.org/health/trustworthy-artificial-intelligence-in-health.pdf

Paredes, M. (2018). Can Artificial Intelligence help reduce human medical errors? Two examples from ICUs in the US and Peru.

QBE (2020). The Elephant in the Room – The NPDB and Dispute Resolution. https://www.qbe.com/us/conversations/elephant-in-the-room-npdb-dispute

Copyright © 2021 Society of Actuaries Research Institute

Radiology Key (2016). Approach to Characterising Radiological Errors. See https://radiologykey.com/approach-to-characterising-radiological-errors/

Rajpurkar, P., Irvin, J., Ball, R. L., Zhu, K., Yang, B., Mehta, H., ... & Lungren, M. P. (2018). Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists. PLoS medicine, 15(11), e1002686.

Reaction Data (2018). Machine Learning in Medical Imaging. See https://www.reactiondata.com/report/machine-learning-medical-imaging/

RSNA (2018). Methodologic guide for evaluating clinical performance and effect of artificial intelligence technology for medical diagnosis and prediction. Radiology, 286(3), 800-809.

Schaffer, A. C., Jena, A. B., Seabury, S. A., Singh, H., Chalasani, V., & Kachalia, A. (2017). Rates and characteristics of paid malpractice claims among US physicians by specialty, 1992-2014. JAMA internal medicine, 177(5), 710-718.

Seabury, S. A., Helland, E., & Jena, A. B. (2014). Medical malpractice reform: noneconomic damages caps reduced payments 15 percent, with varied effects by specialty. Health Affairs, 33(11), 2048-2056.

Shen, J., Zhang, C. J., Jiang, B., Chen, J., Song, J., Liu, Z., ... & Ming, W. K. (2019). Artificial intelligence versus clinicians in disease diagnosis: systematic review. JMIR medical informatics, 7(3), e10010.

Singh, H., Meyer, A. N., & Thomas, E. J. (2014). The frequency of diagnostic errors in outpatient care: estimations from three large observational studies involving US adult populations. BMJ Qual Saf, 23(9), 727-731.

Sollini, M., Antunovic, L., Chiti, A., & Kirienko, M. (2019). Towards clinical application of image mining: a systematic review on artificial intelligence and radiomics. European journal of nuclear medicine and molecular imaging, 46(13), 2656-2672.

Tehrani, A. S. S., Lee, H., Mathews, S. C., Shore, A., Makary, M. A., Pronovost, P. J., & Newman-Toker, D. E. (2013). 25-Year summary of US malpractice claims for diagnostic errors 1986–2010: an analysis from the National Practitioner Data Bank. BMJ quality & safety, 22(8), 672-680.

Whang, J. S., Baker, S. R., Patel, R., Luk, L., & Castro III, A. (2013). The causes of medical malpractice suits against radiologists in the United States. Radiology, 266(2), 548-554.

Westlaw Classic, Thomson Reuters (2021). https://next.westlaw.com

Zheng, Q., Yang, L., Zeng, B., Li, J., Guo, K., Liang, Y., & Liao, G. (2021). Artificial intelligence performance in detecting tumor metastasis from medical radiology imaging: A systematic review and meta-analysis. EClinicalMedicine, 31, 100669.

Give us your feedback!
Take a short survey on this report.
Click Here
SOA Research INSTITUTE

Copyright © 2021 Society of Actuaries Research Institute

About The Society of Actuaries Research Institute

Serving as the research arm of the Society of Actuaries (SOA), the SOA Research Institute provides objective, data-driven research bringing together tried and true practices and future-focused approaches to address societal challenges and your business needs. The Institute provides trusted knowledge, extensive experience and new technologies to help effectively identify, predict and manage risks.

Representing the thousands of actuaries who help conduct critical research, the SOA Research Institute provides clarity and solutions on risks and societal challenges. The Institute connects actuaries, academics, employers, the insurance industry, regulators, research partners, foundations and research institutions, sponsors and non-governmental organizations, building an effective network which provides support, knowledge and expertise regarding the management of risk to benefit the industry and the public.

Managed by experienced actuaries and research experts from a broad range of industries, the SOA Research Institute creates, funds, develops and distributes research to elevate actuaries as leaders in measuring and managing risk. These efforts include studies, essay collections, webcasts, research papers, survey reports, and original research on topics impacting society.

Harnessing its peer-reviewed research, leading-edge technologies, new data tools and innovative practices, the Institute seeks to understand the underlying causes of risk and the possible outcomes. The Institute develops objective research spanning a variety of topics with its strategic research programs: aging and retirement; actuarial innovation and technology; mortality and longevity; diversity, equity and inclusion; health care cost trends; and catastrophe and climate risk. The Institute has a large volume of topical research available, including an expanding collection of international and market-specific research, experience studies, models and timely research.

Society of Actuaries Research Institute
475 N. Martingale Road, Suite 600
Schaumburg, Illinois 60173
www.SOA.org

Copyright © 2021 Society of Actuaries Research Institute