Original Investigation

# Big Data and the Future of Radiology Informatics

Akash P. Kansagra, MD, MS, John-Paul J. Yu, MD, PhD[1], Arindam R. Chatterjee, MD, Leon Lenchik, MD, Daniel S. Chow, MD, Adam B. Prater, MD, MPH, Jean Yeh, MD, Ankur M. Doshi, MD, C. Matthew Hawkins, MD, Marta E. Heilbrun, MD, Stacy E. Smith, MD, Martin Oselkin, MD, Pushpender Gupta, MBBS, Sayed Ali, MD

**Abbreviations and Acronyms**

| | |
|---|---|
| **EMR** | electronic medical record |
| **MCI** | mild cognitive impairment |
| **ADNI** | Alzheimer's Disease Neuroimaging Initiative |
| **PE** | pulmonary embolism |
| **CTA** | computed tomography angiography |
| **RIS** | radiology information system |
| **PACS** | picture archiving and communication system |
| **ACR** | American College of Radiology |
| **HIPAA** | Health Insurance Portability and Accountability Act |

Rapid growth in the amount of data that is electronically recorded as part of routine clinical operations has generated great interest in the use of Big Data methodologies to address clinical and research questions. These methods can efficiently analyze and deliver insights from high-volume, high-variety, and high-growth rate datasets generated across the continuum of care, thereby forgoing the time, cost, and effort of more focused and controlled hypothesis-driven research. By virtue of an existing robust information technology infrastructure and years of archived digital data, radiology departments are particularly well positioned to take advantage of emerging Big Data techniques. In this review, we describe four areas in which Big Data is poised to have an immediate impact on radiology practice, research, and operations. In addition, we provide an overview of the Big Data adoption cycle and describe how academic radiology departments can promote Big Data development.

**Key Words:** Radiology; informatics; personalized medicine; workflow; value.

## INTRODUCTION

Advances in medicine have traditionally been the result of hypothesis-driven research, often in the form of controlled clinical trials. In this approach, a clinical variable believed to influence outcome is identified a priori, and great effort is made—through patient selection and predefined research protocols—to control confounding clinical variables and isolate the effect of the variable of interest. Although this approach is effective, it may be impractical, time-consuming, and costly to run such controlled trials for each of the countless variations in patient demographics, pathophysiology, and clinical decision-making that define each case. As a result, many investigators see promise in a *data-driven* approach in which care is allowed to proceed as it does in the real world, and naturally occurring variations in care

---

**Acad Radiol 2015; ■:■■–■■**

From the Mallinckrodt Institute of Radiology, Washington University School of Medicine, 510 South Kingshighway Blvd., Campus Box 8131, Saint Louis, MO 63110 (A.P.K., A.R.C.); Department of Radiology and Biomedical Imaging, University of California, San Francisco, San Francisco, California (J.-P.J.Y., J.Y.); Department of Radiology, Wake Forest School of Medicine, Winston-Salem, North Carolina (L.L., P.G.); Department of Radiology, Columbia University Medical Center, New York, New York (D.S.C.); Department of Radiology and Imaging Sciences, Emory University School of Medicine, Atlanta, Georgia (A.B.P., C.M.H.); Department of Radiology, NYU Langone Medical Center, New York, New York (A.M.D.); Department of Radiology, University of Utah School of Medicine, Salt Lake City, Utah (M.E.H.); Department of Radiology, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts (S.E.S.); Department of Radiology, Maimonides Medical Center, Brooklyn, New York (M.O.); Department of Radiology, Temple University School of Medicine, Philadelphia, Pennsylvania (S.A.). Received June 29, 2015; revised October 8, 2015; accepted October 14, 2015. [1]Present address: Department of Radiology, University of Wisconsin, Madison, Madison, Wisconsin 53792. **Address correspondence to:** A.P.K. e-mail: apkansagra@gmail.com

http://dx.doi.org/10.1016/j.acra.2015.10.004

delivery from patient to patient are studied in aggregate to determine the effect of each on overall outcome (1,2).

This type of research relies on analytical methods from the emerging science of "Big Data" informatics. Big Data refers to extremely complex datasets characterized by the four Vs: *volume*, which refers to the sheer number of data elements within these extremely large datasets; *variety*, which describes the aggregation of data from multiple sources; *velocity*, which refers to the high speed at which data is generated; and *veracity*, which describes the inherent uncertainty in some data elements (3,4). These sources of complexity exceed the capabilities of conventional data analysis techniques, but Big Data methods are specifically designed to overcome these challenges.

This approach is inspired in part by the successes of Big Data methods in leveraging the immense data collected by mobile and internet-enabled technologies over the last decade. These data have been successfully used as the basis for targeted advertising, personalized consumer recommendations, and real-time traffic maps, among countless other applications. As electronic medical records (EMRs) and other clinical databases make patient data more readily accessible in the healthcare enterprise (1,5), there is hope that Big Data analytics may yield important insights in medicine. This vision of the future has been formalized in the concept of a Learning Healthcare System proposed by the Institute of Medicine (6). Indeed, early applications of Big Data to health care—such as an informatics platform to integrate neonatal physiological monitoring to predict the onset of nosocomial infections prior to the onset of clinical symptoms (7)—have produced promising results.

The promise of Big Data is particularly strong within radiology. Nearly two decades ago, the specialty became an early adopter of digital workflows and electronic integration of healthcare information and now enjoys a mature information technology (IT) infrastructure that has virtually eradicated the use of nondigitized data (8). As a result, information has become the currency of radiology, and electronically accessible information—the key ingredient needed to power Big Data analytics—is available in immense quantities within the information systems at the center of every modern radiology department. Despite the rich troves of digital data available in radiology, most of the methods needed to analyze these data need to be studied and developed before the impact of Big Data on clinical radiology can be fully appreciated.

In this paper, we review potential applications of Big Data in modern radiology practice through the lens of four big questions facing our specialty. Specifically, we consider how emerging Big Data methods can enable personalized image interpretation, facilitate discovery of new imaging markers, quantify the value of radiology services to patient health, and characterize and optimize radiology workflows. We then review the four stages of Big Data adoption and use these insights as a guide for academic radiology departments that wish to encourage Big Data research, development, and utilization. In so doing, we hope to provide both inspiration and a blueprint for departmental decision-makers as the specialty of radiology steps into the next era of informatics and data science.

## CAN IMAGE INTERPRETATION AND MANAGEMENT RECOMMENDATIONS BE PERSONALIZED FOR INDIVIDUAL PATIENTS?

### Background

Radiologists routinely rely upon pattern recognition and morphological features of visually apparent abnormalities to arrive at diagnoses and generate recommendations for management, including follow-up imaging. However, imaging features alone are not sufficient to completely determine diagnosis and management, as similar imaging findings in two different patients may have vastly different significance. For example, management of a simple ovarian cyst detected in a premenopausal woman with no risk factors for cancer is markedly different from an identical cyst discovered in a postmenopausal woman with multiple risk factors for malignancy (9).

Accordingly, radiologists must look beyond imaging findings alone and consider patient-specific and population-level factors—personal and family history, physical exam findings, laboratory data, genetic profiles, and local disease incidence, among others—to render the most accurate diagnoses and recommend appropriate management. Much of this information is recorded in the EMR, and radiologists already know to look to this resource for specific information that would affect diagnosis or alter their approach to a patient—for example, directing high-risk carriers of the BRCA mutation to begin early screening for breast cancer (10).

### Why This Needs Big Data?

Although radiologists already account for some relevant patient-specific factors (e.g., patient age, history of cancer) when interpreting studies, the vast quantity of lesion-, patient-, and population-specific data contained in the EMR exceeds the ability of a radiologist to meaningfully incorporate into interpretation. For example, biopsy samples from a wide range of tumor types now routinely undergo detailed genetic analysis to extract information about DNA, RNA, and protein expression in tumor cells that profoundly influence prognosis and subsequent management. For instance, the presence of 1p/19q codeletion confers improved survival in patients with oligodendroglioma (11), whereas EGFR and HER2/neu expression in lung and breast cancer, respectively, determine the effectiveness of certain targeted chemotherapies (12,13). But the number of genetic mutations present in any particular tumor is enormous.

To adequately synthesize these and other factors that can influence diagnosis and management requires a Big Data approach capable of mining a large volume and variety of data. Furthermore, these data have variable veracity in that some clinical data elements are inherently uncertain or unreliable. With such an approach, Big Data methods may allow morphological features of a lesion to be supplemented by a larger range of nonradiological factors that can enable more precise and individualized diagnosis and management (Fig 1).