

PERSPECTIVE

Tracking Progress in Improving Diagnosis: A Framework for Defining Undesirable Diagnostic Events

Andrew P. J. Olson, MD^{1,2}, Mark L. Graber, MD^{3,4}, and Hardeep Singh, MD, MPH⁵

¹Department of Medicine, University of Minnesota, Minneapolis, MN, USA; ²Department of Pediatrics, University of Minnesota, Minneapolis, MN, USA; ³RTI International, Raleigh, NC, USA; ⁴Society to Improve Diagnosis in Medicine, New York, NY, USA; ⁵Houston Veterans Affairs Center for Innovations in Quality, Effectiveness and Safety, Michael E. DeBakey Veterans Affairs Medical Center and the Section of Health Services Research, Department of Medicine, Baylor College of Medicine, Houston, TX, USA.

Diagnostic error is a prevalent, harmful, and costly phenomenon. Multiple national health care and governmental organizations have recently identified the need to improve diagnostic safety as a high priority. A major barrier, however, is the lack of standardized, reliable methods for measuring diagnostic safety. Given the absence of reliable and valid measures for diagnostic errors, we need methods to help establish some type of baseline diagnostic performance across health systems, as well as to enable researchers and health systems to determine the impact of interventions for improving the diagnostic process. Multiple approaches have been suggested but none widely adopted. We propose a new framework for identifying “undesirable diagnostic events” (UDEs) that health systems, professional organizations, and researchers could further define and develop to enable standardized measurement and reporting related to diagnostic safety. We propose an outline for UDEs that identifies both conditions prone to diagnostic error and the contexts of care in which these errors are likely to occur. Refinement and adoption of this framework across health systems can facilitate standardized measurement and reporting of diagnostic safety.

KEY WORDS: diagnosis; diagnostic error; patient safety; measurement.

J Gen Intern Med 33(7):1187–91

DOI: 10.1007/s11606-018-4304-2

© Society of General Internal Medicine (outside the USA) 2018

INTRODUCTION

Diagnostic error is a prevalent, harmful, costly phenomenon.^{1,2} The National Academy of Medicine (NAM) defines diagnostic error 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.”¹ More actionable definitions such as “a missed opportunity to make a diagnosis”^{3–6} or “any mistake or failure in the diagnostic process leading to a misdiagnosis, a missed diagnosis, or a delayed diagnosis”⁷ have been used in research. With mounting pressure to address diagnostic error, many health systems, providers, policymakers

(including the Centers for Medicare and Medicaid and Services [CMS]), and patient safety stakeholders (Centers for Disease Control and Prevention [CDC], Agency for Healthcare Research and Quality [AHRQ], and National Quality Forum [NQF]) are interested in developing and evaluating measurement strategies to improve the diagnostic process.^{8–10}

Often, a broad array of cognitive and systems-related contributing factors interact in complex ways to make the diagnostic process risk-prone.¹¹ Diagnostic errors include overlapping situations of missed, delayed, and/or incorrect diagnoses, and these three concepts often become hard to disentangle. Nevertheless, robust measurement is foundational to any improvement initiative. While it is generally agreed that multifaceted approaches are needed to reduce diagnostic error,¹² measuring the success of any intervention is challenging because of the absence of universally accepted measures for determining progress in diagnostic safety. The 2017 CMS Quality Measure Development Plan includes “diagnostic accuracy” as a priority for measure development.⁸ The science in this area is underdeveloped, and the NQF only recently convened a committee to develop a conceptual framework for measurement and to make recommendations for addressing these gaps.¹⁰ The NQF framework builds on the NAM framework, which considers diagnosis as a process: a complex and collaborative activity that unfolds over time and occurs within the context of a health care work system.¹⁰

To enable further progress in this area, we propose a pragmatic framework to identify “undesirable diagnostic events” (UDEs) that health systems, professional organizations, and researchers could further define and develop to enable standardized measurement and reporting related to diagnostic safety. The framework identifies clinical situations that denote potentially preventable breakdowns in the diagnostic process for which improved diagnostic processes would lead to improved health for patients. We propose the UDE framework as a starting point for further refinement and discussion among major stakeholders (health systems, professional organizations, and researchers) in diagnostic safety.

THE NEED FOR A NEW FRAMEWORK

Measuring diagnostic accuracy and safety in real-world clinical settings is difficult, labor-intensive, and subject to substantial inter-rater variability.⁶ Tools to improve the accuracy

Received July 6, 2017

Revised November 21, 2017

Accepted December 20, 2017

Published online January 29, 2018

and reliability of diagnostic error measurement are under development, but are still primarily research tools and may be difficult for a health system to implement.¹³ Given the wide scope, variability, and time scale of the diagnostic process, projects to improve diagnosis in one area (such as cancer) may have little effect on another area (such as sepsis or pulmonary embolus). Similarly, it is unknown whether interventions in one setting, such as primary care, would be useful in others, such as inpatient settings. Analytical tools have been developed to aid in determining whether a diagnostic error has occurred^{4,14} and in the classification of diagnostic errors,⁷ but are limited in their ability to standardize reporting and measurement, given the context specificity of many diagnostic errors.

So how do we move diagnostic safety forward in the absence of validated and reliable measures? Elsewhere in patient safety, the concept of “never events” for hospitalized patients has stimulated widespread efforts to monitor and reduce the occurrence of these unambiguous, serious, and usually preventable events.¹⁵ A major limitation in the study of diagnostic safety is the lack of standardized reporting metrics across health systems; in contrast, the advent of reporting “never events” has allowed national reporting of local practice patterns, thereby providing impetus for improvement and shared learning. Given the nascent state of scientific knowledge in diagnostic error, we are not yet able to define a comparable series of diagnostic never events. However, we propose a framework to identify “undesirable diagnostic events” (UDEs) as a building block to enable progress in measurement. In keeping with the “never event” concept, we propose that UDEs represent a reliable outcome measure (irrespective of the causes resulting in that outcome) that allows for standardization.

We define UDEs as specific, measurable, and actionable clinical situations likely to denote the presence of diagnostic error. Through the use of this framework, lists of UDEs can be developed across fields of medicine and health care systems. The concepts and criteria we propose are based on previous experience and prior research, and in consultation with experts in the field. Two well-known frameworks were also used to inform criteria development: “never events” and the well-known World Health Organization (WHO) screening test guidelines developed by Wilson and Jungner.¹⁶

UDEs are based on principles similar to those of trigger tools that “alert patient safety personnel to possible adverse events so they can review the medical record to determine if an actual or potential adverse event has occurred.”⁹ Once UDEs are identified, health systems could use existing tools such as the Diagnostic Error Evaluation and Research (DEER) taxonomy⁷ or the SaferDx¹⁴ instrument to analyze individual cases to identify contributing factors. Thus, UDEs could lead to improved detection of contributory factors and identification of solutions. Lessons may be learned and solutions shared and implemented across systems with appropriate context specificity.

CRITERIA FOR UNDESIRABLE DIAGNOSTIC EVENTS

UDEs must have a number of characteristics to enable standardization across health systems and interventions, ensure validity, and maximize generalizable lessons. UDEs should include both a condition (health problem) and a context (means, including timing and setting, by which that condition is diagnosed). Proposed criteria for defining the specific conditions and contexts that constitute UDEs are discussed below, and examples are given in Table 1. These criteria ensure that UDEs are strongly indicative of an opportunity to make the diagnosis earlier in the diagnostic process.

The Target Condition Should Be Readily Diagnosable in Routine Clinical Care

The condition for which suboptimal diagnosis would constitute a UDEs should be “readily diagnosable”; that is, the target condition should typically be diagnosed in an accurate and timely fashion when providers are evaluating a patient for that condition. Conditions that are generally diagnosed only *after* a thorough evaluation for other conditions (diagnoses of exclusion) or diagnosed during evaluation for another condition (diagnosis by serendipity) would be poor target conditions, as their discovery may be less reflective of the usual diagnostic process used to make the diagnosis. For example, proposing delayed diagnosis of adult-onset Still’s disease as a UDE would be inappropriate because it is generally a diagnosis of exclusion and typically diagnosed only after an exhaustive evaluation for other conditions. The heterogeneity of the processes used to arrive at such diagnoses is significant, making measurement and standardization difficult. Conditions should not be “zebras,” but instead conditions where usual medical care should be able to arrive at an accurate and timely diagnosis. Conditions where a prolonged period of time is typically required for their manifestations to become evident (e.g., amyloidosis or Alzheimer’s disease) are not good candidates due to the heterogeneity in their presentation and disease course.

The Condition Must Have an Objective, Time-Bound, Valid, and Nearly Universally Available Reference Standard for Verification of Diagnosis

In order for UDEs to be standardized and valid across patients and health systems, the process of defining the presence and absence of the condition must be unambiguous, through the use of a reference standard. Many studies on diagnostic accuracy use consensus clinical reference standards that are prone to bias and subject to varying (and often suboptimal) inter-rater reliability, making measurement difficult. For example, a study examining the effectiveness of an intervention to reduce misdiagnosis of depression in primary care would be most useful if there were an appropriate reference standard for the diagnosis of

Table 1 Example Candidate Undesirable Diagnostic Events (UDEs)

Condition	Context (setting and timing)	Readily clinically diagnosable	Objective, valid, and widely used reference standard	Prone to error	Timely diagnosis likely to have a positive effect on patients' health	Sufficiently common	Diagnostic process definable	Balance measure available
Rheumatoid arthritis	Outpatient >12 months	No (presentation varied, often over time)	Yes	Yes	Yes	Yes	No (given varied presentation, many different routes to diagnosis)	No
Sepsis	Inpatient >12 h	Yes	No (reference standard is usually clinical diagnosis of sepsis, subject to marked variation and false positive/negatives)	Yes	Yes	Yes	Maybe (many patients at risk for sepsis from many conditions, making diagnostic process heterogeneous)	Yes (blood culture rate, antibiotic days)
Bacterial meningitis	Emergency department/inpatient >1 visit	Yes	Yes	Yes	Yes	Yes	Yes	Yes (lumbar puncture rate, empiric antibiotic use)
Tuberculosis	Diagnosed on autopsy	Yes	Yes	Yes	Yes	Yes	Yes	Yes (patient isolation days)
Spinal epidural abscess	Emergency department/inpatient >2 visits	Yes	Yes	Yes	Yes	Yes	Yes	Yes (MRI rate)
Colorectal cancer	Outpatient >12 months after + screening test	Yes	Yes	Yes	Yes	Yes	Yes	No

A number of clinical conditions and contexts (timing and setting) are presented and evaluated using the defined criteria. Examples that meet all seven criteria are in bold font

depression available and used across all systems. Many conditions do have a widely used and objective reference standard, such as bacteremia (positive blood culture), spinal epidural abscess (MRI or operative findings), and pulmonary embolism (positive CT angiogram).

The Condition Should Be Prone to Diagnostic Error

While much of medicine would benefit from improved diagnosis, early interventions must focus on areas with potential for significant early gains, both for proof of concept and for strategic engagement with researchers, funders, and the public. Thus, UDEs should be chosen from conditions and contexts where errors are common and are associated with preventable harm. This is best demonstrated through research studies, local trends, and/or analysis of medical liability data. Evidence from such data sources should indicate that UDEs occur—in other words, in a given context, opportunities to diagnose a specific condition are missed relatively frequently and lead to

preventable harm. For instance, conditions such as colorectal and other cancers,⁹ pulmonary embolism, and infections such as pneumonia and spinal epidural abscess are known to be missed, based on previous reports^{7,17} and more recent autopsy studies.⁹ In contrast, conditions based largely on visual cues and/or pattern recognition, such as certain classic dermatologic or radiologic conditions, are likely much less subject to error.^{18,19} Interventions to improve the diagnosis of herpes zoster infection or pneumothorax would consequently have less impact.

Timely Diagnosis of the Condition Should Provide an Opportunity for a Positive Effect on Patients' Health

Diagnoses are generally a means to an end; that is, the diagnosis allows a patient and her health care team to make subsequent decisions about treatment and prognosis. As we aim to improve diagnostic safety, attention should first be focused on improving the diagnosis of conditions for which

improved (often earlier) diagnosis leads to improved treatment and measurably improved health. Thus, conditions considered as part of a UDE should be those for which more accurate and/or timely diagnosis is likely to have a positive effect on a patient's health. Treatments available for certain conditions, such as spinal metastases in a patient with already-advanced cancer, are unfortunately relatively unlikely to lead to a measurably positive effect on the health of a patient even if diagnosed earlier. In contrast, a diagnosis of colon cancer prior to metastasis allows for more effective treatment and better health. It should be noted, however, that this definition of health should be broad and should include multiple aspects of well-being for patients and families.

The Condition Must Be Common Enough to Warrant System-Level Interventions and for Interventions to Show Effect

UDE conditions must have sufficient prevalence to be realistically considered in the differential diagnosis for patients presenting to an institution. For example, in a large urgent care clinic, mitochondrial disorders would not be good candidates, given their rarity, while bacterial meningitis or cancer are likely considered frequently enough in the diagnostic process to justify interventions to improve their diagnosis. Conditions must thus be encountered with sufficient frequency to allow for meaningful measurement and for interventions to show effect.

Definable Steps, Pathways, or Processes Must Exist to Diagnose the Condition

Diagnosis-specific interventions will be difficult to define if steps taken to make a diagnosis are unclear or occur primarily through serendipity. Further, if there are highly variable diagnostic pathways used to diagnose a condition, then interventions focused on improving the diagnostic process would be difficult to design, implement, and measure. For example, many conditions are identified primarily after other conditions have been excluded (e.g., bacterial overgrowth syndromes, adrenal insufficiency) or primarily through provider opinion (some psychiatric conditions).⁹ Such processes are hard to measure and even harder to standardize, and interventions to improve them may prove elusive.

A Balance Measure Must Be Identifiable to Measure Unintended Consequences of Interventions to Improve Diagnosis

As with all quality improvement initiatives, projects to improve diagnostic safety may have unintended consequences, especially overutilization of tests and overdiagnosis and/or overtreatment. For example, if a health system were to focus on improving the diagnosis of spinal epidural abscess, the

percentage of patients presenting to the emergency department with back pain who undergo an MRI might increase. This would increase the cost of care and lead to additional unnecessary testing. Similarly, a focus on improving the diagnosis of meningitis might lead to more unnecessary lumbar punctures. Balance measures could help counter unintended consequences so that interventions to improve the diagnosis of appendicitis, for example, would also avoid substantially increasing the rate of CT scans performed in children. If such unintended consequences are not anticipated and identified as balance measures, well-meaning interventions to improve diagnostic safety could lead to increased costs of care, overdiagnosis, and overtreatment.

Using these criteria, candidate UDE examples are evaluated in Table 1. At this exploratory stage, we suggest that all seven criteria be met before the UDE is further considered for evaluation. It is likely that some factors in the course of further study will be deemed more important than others, while others will undergo further refinement after future research and implementation efforts. These candidate UDEs were selected based on the clinical experience of the authors, consultation with multiple experts in the field, and feedback after oral presentations, as well as previously published research.^{3,5,20–22}

Some measures of safety in health care have a denominator (e.g., events/1000 patient-days), while others, such as several serious safety events, do not. For selected UDEs, all patients evaluated for a certain condition or who are ultimately diagnosed with that condition could be used as a denominator, allowing for better benchmarking prior to intervention and comparison across systems. However, UDEs (like certain never events including wrong-site surgery) may also be used without a denominator if they are deemed to be nearly completely avoidable and severe.

HOW UDES MAY BE USED

As researchers and health systems consider interventions to improve diagnostic safety, defining UDEs could help not only measure the impact of specific interventions but also allow more standardized reporting of meaningful measures across health systems. Furthermore, it will allow comparison of the effectiveness of different interventions by measuring the effect on a certain UDE. As a next step, we suggest convening working groups including stakeholders from health systems, regulatory bodies, and professional societies to define high-priority UDEs, somewhat similar to the "Choosing Wisely" campaign.²³ The UDE implementation should then be piloted in several health systems, and concepts should be validated. Subsequently, research and quality improvement initiatives can target the diagnostic process related to each UDE.

CONCLUSIONS

Valid metrics of diagnostic safety must be developed to measure the effectiveness of interventions designed to improve the diagnostic process. Given the inherent problems with measurement of diagnostic error, we propose a new approach using UDEs, focusing on specific situations where an outcome likely reflects a breakdown in the diagnostic process. We propose robust criteria to ensure that UDEs focus on opportunities not only for earlier diagnosis, but also for timely diagnosis that leads to a positive effect on patients' health. UDEs may be followed prospectively over time by health systems and researchers and reported and/or compared across systems. Sharing across systems may occur in the context of research studies or within the purview of patient safety organizations to promote learning and improvement efforts. Further development and implementation of UDEs can stimulate large-scale research and improvement efforts related to diagnostic safety.

Acknowledgements: The authors thank Joe Grubenhoff, MD, MSCS, for his assistance with the refinement of some of the ideas presented in this paper.

Corresponding Author: Andrew P. J. Olson, MD; Department of Medicine University of Minnesota, Minneapolis, MN, USA (e-mail: olso5714@umn.edu).

Funding: Dr. Singh is supported by the VA Health Services Research and Development Service (CRE 12-033; Presidential Early Career Award for Scientists and Engineers USA 14-274), the VA National Center for Patient Safety, the Agency for Health Care Research and Quality (R01HS022087 and R21HS023602), and the Houston VA HSR&D Center for Innovations in Quality, Effectiveness and Safety (CIN 13-413).

Compliance with Ethical Standards:

Conflict of Interest: The authors declare that they do not have a conflict of interest.

REFERENCES

- National Academies of Sciences, Engineering, and Medicine. Improving Diagnosis in Health Care. Washington, DC: National Academies Press; 2015.
- Singh H, Graber ML. Improving diagnosis in health care—the next imperative for patient safety. *N Engl J Med*. 2015;373(26):2493–5. <https://doi.org/10.1056/NEJMp1512241>.
- Bhise V, Meyer AND, Singh H, et al. Errors in diagnosis of spinal epidural abscesses in the era of electronic health records. *Am J Med*. 2017;130(8):975–81. <https://doi.org/10.1016/j.amjmed.2017.03.009>.
- Davalos MC, Samuels K, Meyer AND, et al. Finding diagnostic errors in children admitted to the PICU. *Pediatr Crit Care Med*. 2017;18(3):265–71. <https://doi.org/10.1097/PCC.0000000000001059>.
- Singh H, Meyer AND, Thomas EJ. The frequency of diagnostic errors in outpatient care: estimations from three large observational studies involving US adult populations. *BMJ Qual Saf*. 2014;23(9):727–31. <https://doi.org/10.1136/bmjqs-2013-002627>.
- Zwaan L, Singh H. The challenges in defining and measuring diagnostic error. *Diagnosis*. 2015;2(2):97–103. <https://doi.org/10.1515/dx-2014-0069>.
- Schiff GD, Hasan O, Kim S, et al. Diagnostic error in medicine: analysis of 583 physician-reported errors. *Arch Intern Med*. 2009;169(20):1881–7. <https://doi.org/10.1001/archinternmed.2009.333>.
- CMS Quality Measure Development Plan: Supporting the Transition to The Quality Payment Program 2017 Annual Report. 2017. <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Value-Based-Programs/MACRA-MIPS-and-APMs/Draft-CMS-Quality-Measure-Development-Plan-MDP.pdf>. Accessed December 17, 2017.
- CLIA Home. <https://www.cdc.gov/cliac/>. Accessed December 17, 2017.
- National Quality Forum - Improving Diagnostic Quality and Safety. http://www.qualityforum.org/Improving_Diagnostic_Quality_and_Safety.aspx. Accessed December 17, 2017.
- Graber ML, Franklin N, Gordon R. Diagnostic error in internal medicine. *Arch Intern Med*. 2005;165(13):1493–9. <https://doi.org/10.1001/archinte.165.13.1493>.
- Singh H, Schiff GD, Graber ML, Onakpoya I, Thompson MJ. The global burden of diagnostic errors in primary care. *BMJ Qual Saf*. 2017;26(6):484–94. <https://doi.org/10.1136/bmjqs-2016-005401>.
- Singh H, Graber ML, Hofer TP. Measures to Improve Diagnostic Safety in Clinical Practice. *J Patient Saf*. 2016;1. <https://doi.org/10.1097/PTS.0000000000000338>.
- Al-Mutairi A, Meyer AND, Thomas EJ, et al. Accuracy of the Safer Dx instrument to identify diagnostic errors in primary care. *J Gen Intern Med*. 2016;31(6):602–8. <https://doi.org/10.1007/s11606-016-3601-x>.
- Patient Safety Primer: Never Events. <https://psnet.ahrq.gov/primers/primer/3/never-events>. Accessed December 17, 2017.
- Wilson JMG, Jungner G. Principles and practice of screening for disease. Public Health Paper No. 34. Geneva: World Health Organization; 1968.
- Combes A, Mokhtari M, Couvelard A, et al. Clinical and autopsy diagnoses in the intensive care unit. *Arch Intern Med*. 2004;164(4):389. <https://doi.org/10.1001/archinte.164.4.389>.
- Lowenstein EJ. Dermatology and its unique diagnostic heuristics. *J Am Acad Dermatol*. November 2017. <https://doi.org/10.1016/J.JAAD.2017.11.018>.
- Berner ES, Graber ML. Overconfidence as a cause of diagnostic error in medicine. *Am J Med*. 2008;121(5 Suppl):S2–23. <https://doi.org/10.1016/j.amjmed.2008.01.001>.
- Singh H, Daci K, Petersen LA, et al. Missed opportunities to initiate endoscopic evaluation for colorectal cancer diagnosis. *Am J Gastroenterol*. 2009;104(10):2543–54. <https://doi.org/10.1038/ajg.2009.324>.
- Singh H, Giardina TD, Petersen LA, et al. Exploring situational awareness in diagnostic errors in primary care. *BMJ Qual Saf*. 2012;21(1):30–8. <https://doi.org/10.1136/bmjqs-2011-000310>.
- Okafor N, Payne VL, Chathampally Y, Miller S, Doshi P, Singh H. Using voluntary reports from physicians to learn from diagnostic errors in emergency medicine. *Emerg Med J*. 2016;33(4):245–52. <https://doi.org/10.1136/emmermed-2014-204604>.
- Choosing Wisely. <http://www.choosingwisely.org/>. Accessed December 17, 2017.