

Ashwin Gupta*, Martha Quinn, M. Todd Greene, Karen E. Fowler and Vineet Chopra

Implementation of a bundle to improve diagnosis in hospitalized patients: lessons learned

<https://doi.org/10.1515/dx-2024-0099>

Received May 31, 2024; accepted September 30, 2024;

published online October 18, 2024

Abstract

Objectives: The inpatient setting is a challenging clinical environment where systems and situational factors predispose clinicians to making diagnostic errors. Environmental complexities limit trialing of interventions to improve diagnostic error in active inpatient clinical settings. Informed by prior work, we piloted a multi-component intervention designed to reduce diagnostic error to understand its feasibility and uptake.

Methods: From September 2018 to June 2019, we conducted a prospective, pre-test/post-test pilot study of hospital medicine physicians during admitting shifts at a tertiary-care, academic medical center. Optional intervention components included use of dedicated workspaces, privacy barriers, noise cancelling headphones, application-based breathing exercises, a differential diagnosis expander application, and a checklist to enable a diagnostic pause. Participants rated their confidence in patient diagnoses and completed a survey on intervention component use. Data on provider resource utilization and patient diagnoses were collected, and qualitative interviews were held with a subset of participants in order to better understand experience with the intervention.

Results: Data from 37 physicians and 160 patients were included. No intervention component was utilized by more than 50 % of providers, and no differences were noted in

diagnostic confidence or number of diagnoses documented pre- vs. post-intervention. Lab utilization increased, but there were no other differences in resource utilization during the intervention. Qualitative feedback highlighted workflow integration challenges, among others, for poor intervention uptake.

Conclusions: Our pilot study demonstrated poor feasibility and uptake of an intervention designed to reduce diagnostic error. This study highlights the unique challenges of implementing solutions within busy clinical environments.

Keywords: diagnostic error; implementation science; pilot study

Introduction

The inpatient setting is a clinical environment in which high complexity medical patients, often with myriad different illnesses of varying acuity, receive care. On top of challenges inherent in managing multiple medical processes (both within an individual patient and across a team of patients), systems and situational factors often make achieving an accurate and timely diagnosis – and rendering appropriate treatment – challenging [1–3]. Reflective of this, errors in diagnosis are the most costly and morbid of all types of medical error in hospitalized patients [4].

Owing to this complexity and the challenges associated with controlling for innumerable contextual factors, interventions to curb diagnostic error have generally occurred either in simulated or highly standardized clinical scenarios or environments (e.g., electrocardiogram or radiology interpretation) [5, 6]. Few, if any, interventions to address diagnostic error have focused on inpatient medical providers or been carried out in the inpatient clinical setting as care was being delivered [7, 8].

To address these challenges, we embedded clinicians and researchers within inpatient medical teams to better understand the challenges associated with diagnosis within the hospital setting. Informed by that work, we devised a multi-component intervention with the goal of reducing diagnostic errors. In this manuscript, we report the results of a pilot study aimed to assess feasibility and uptake of our intervention and the lessons learned.

***Corresponding author: Ashwin Gupta**, MD, Associate Professor, Internal Medicine VA Ann Arbor Healthcare System, 2215 Fuller Road, Mail Code 111, Ann Arbor, MI 48105, USA; and Division of Hospital Medicine, Department of Medicine, University of Michigan Medical School, Ann Arbor, MI, USA, E-mail: ashwing@med.umich.edu

Martha Quinn, School of Public Health, University of Michigan, Ann Arbor, MI, USA

M. Todd Greene, VA Ann Arbor Healthcare System, Ann Arbor, MI, USA; and Division of Hospital Medicine, Department of Medicine, University of Michigan Medical School, Ann Arbor, MI, USA

Karen E. Fowler, VA Ann Arbor Healthcare System, Ann Arbor, MI, USA

Vineet Chopra, Division of Hospital Medicine, Department of Medicine, University of Michigan Medical School, Ann Arbor, MI, USA; and Department of Medicine, University of Colorado School of Medicine, Aurora, CO, USA

Methods

Settings and participants

Between September 2018 and June 2019, we conducted a prospective, pre-test/post-test pilot study of a multi-component intervention focused on reducing diagnostic errors among hospital medicine physicians (hospitalists) at a large, academic, tertiary-care medical center. Hospitalist participants were recruited by a trained research assistant and were consented immediately prior to a clinical admitting shift, during which time they were responsible for patient admission, triage, documentation, and hand off to oncoming physicians. Admitting shifts were selected (rather than rounding shifts) as the diagnostic process is heavily weighted early within a hospitalization. Hospitalists were recruited consecutively, Monday through Friday, based on existing schedules. Participants on service between September and December 2018 provided data for the pre-intervention period, and those on service between January and June 2019 participated in the intervention.

Intervention

Informed by our prior work [1, 3, 9], we implemented a multi-component intervention aimed to address noted systems and environmental contributions to diagnostic error between September 2018 and June 2019. All intervention components were optional and could be used either individually or in their entirety by the provider. In order to limit distractions, dedicated workspace was identified, and privacy barriers were erected. Based on prior feedback, this location remained within the larger team room environment to facilitate groupthink, an often-utilized method of cognitive debiasing [10]. Noise cancelling headphones were available to eliminate ambient noise. A brief application-based guided breathing exercise was performed at the beginning of the shift and was available throughout the shift as a mechanism to promote mindfulness. To facilitate the diagnostic process, an electronic tablet that was preloaded with an application that expanded differential diagnoses (Diagnosaurus® DDX; Unbound Medicine Inc., Charlottesville, VA) was provided. Finally, to enable a diagnostic pause, a diagnostic checklist that could be imported into the electronic health record (EHR) through use of an electronic shortcut was created (Supplementary Material 1). This checklist encouraged specific actions to be performed before the patient encounter, during the patient encounter, and in follow-up. While the encouraged timing of use of the checklist was during drafting of the patient's documentation

(as many had previously identified that this is when reflection of patients occurs), a physical copy of the checklist was displayed within the privacy screens so that pre-encounter items could be considered. Hospitalists were educated to these intervention components during a group-wide conference in September 2018 and were again oriented to them at shift onset by one of two research assistants.

Data collection methods and processes

Data were collected by one of two research assistants, who were present beginning at shift start (1:00pm) and for either the duration of the shift or 9:00pm, whichever was earlier. At either shift completion or 9:00pm, participants in both the pre- and post-intervention period completed a survey in which they rated their diagnostic confidence in their leading diagnosis for each patient on a 10-point Likert scale (with one indicating least confident and 10 indicating most confident). In the post-intervention period, participants were also asked to indicate which components of the intervention (if any) were utilized.

Following shift conclusion, research staff reviewed the EHR chart of each patient admitted by the participating provider and documented the number and type of labs and imaging studies ordered in the ED, by consultants, and by the admitting provider; the number of consults requested by the ED and by the admitting provider; number of diagnoses listed in the admitting provider's initial differential diagnosis; and admission and discharge diagnoses. Patients admitted with known diagnoses (either via direct admission or transfer from another unit) were excluded from inclusion.

At the conclusion of the intervention, a trained qualitative methodologist (MQ) conducted in-person interviews, using a semi-structured interview guide (Supplementary Material 2), with a subset of participating hospitalists based on a convenience sample. The purpose of the interviews was to better understand their overall experience with the intervention and their views on specific intervention components. Interviews were conducted in a private office space located close to the hospital team rooms and scheduled at times most convenient for the hospitalists. A research assistant was also present during interviews to consent participants, assist with audio recording, and ask clarifying questions. All interviews were recorded and transcribed verbatim.

Outcomes

The primary outcome of interest was the feasibility and use of the intervention. Secondary outcomes included

differences in diagnostic confidence, number of diagnoses in admission documentation, and resource utilization (including lab studies, imaging studies, and consult orders) in the pre- and post-intervention period. We additionally assessed the use of the intervention (either use of any component or use of individual components) on the same outcomes.

Data analyses

Statistical analyses

Data were summarized using descriptive statistics (means for continuous variables and frequencies and percentages for categorical variables). Differences in key variables between pre-intervention and intervention periods were assessed via t-tests and chi-square tests as appropriate. To model the effect of intervention use and account for repeated measures at the physician level (as participants may have participated in multiple shifts and often admitted multiple patients) linear mixed effects models, using maximum likelihood estimation methods, were used to assess associations between use of intervention components (at least one component or each component individually) and diagnostic confidence, number of diagnoses listed on admitting documentation, and resource utilization.

For all models, random intercept components were included for participants to account for clustering at the physician-level. All models were also adjusted with fixed components for gender, race, and shift duration (measured in total minutes). Fixed effect coefficient estimates, 95 % confidence intervals, and p-values were estimated. An alpha-level of 0.05 was used to determine statistical significance of the fixed-effect coefficient of interest for all models. All analyses were conducted in Stata MP 14.1 (StataCorp, College Station, TX).

Qualitative analysis

A rapid analysis approach was used to analyze interview transcripts [11]. To begin, two team members trained in qualitative analysis methods (MQ, KF) constructed a template based on the interview guide to reflect the main domains of interest. These domains included participant views on specific intervention components (e.g., privacy screen, breathing exercises, headphones, diagnostic checklist). The two team members then read through each transcript independently. One took the lead summarizing data for each domain in the template, including supporting quotations, and the other conducted a review to ensure that all data was

captured accurately and consistently. In some cases, additional data were added to the templates by the second reviewer. These additions, along with any discrepancies between first and second reviews, were discussed until agreement was reached. Once all of the templates were complete, data under each domain were reviewed, synthesized, and summarized to better understand similarities and differences in views among and across participants.

Ethical and regulatory oversight

This study was reviewed and approved by the Institutional Review Board at the University of Michigan Health System (HUM00145793). All relevant ethical guidelines, including the Declaration of Helsinki, were followed in the conduct of this research.

Results

Quantitative results

A total of 53 physicians were approached for study participation; 47 agreed to participate and completed surveys. Of these 47 physicians, 3 did not admit any patients during their study shifts, and 3 others only admitted patients that were ineligible for the study (i.e., direct admits, planned treatment, or transfer patients) and were therefore removed from the study sample. An additional 4 physicians were excluded from analyses due to missing data elements. Complete data were available for 37 unique hospitalists who admitted 160 unique patients (Figure 1). A total of 10 hospitalists participated during the pre-intervention period only,

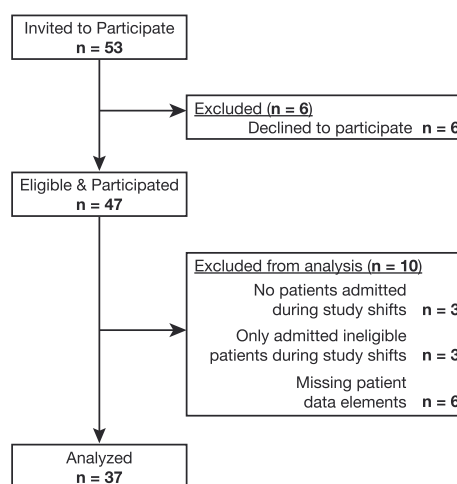


Figure 1: Consort flow diagram.

13 participated during the intervention period only, and 14 participated during both pre-intervention and intervention periods. Of the 24 hospitalists contributing data during the pre-intervention period, 11 (46 %) were men and 15 (63 %) were White. Of the 27 hospitalists contributing data during the intervention period, 13 (48 %) were men and 12 (44 %) were White. Hospitalists admitted a total of 75 and 85 patients during the pre-intervention and intervention periods, respectively.

The 27 hospitalists participating during the intervention admitted a total of 85 included patients across 44 unique shifts. The number of shifts per hospitalist during the intervention period ranged from 1–5 and the number of patients per shift ranged from 0–5. At least one intervention component was used during 38 (86 %) of the 44 shifts during the intervention period. Across the 44 shifts, the breathing exercise ($n=22$, 50 %), privacy screen ($n=19$, 43 %) and diagnostic checklist ($n=17$, 39 %) were the most frequently used interventions. The headphones ($n=8$, 18 %) and diagnostic application ($n=3$, 6.8 %) were the most infrequently used modalities.

No differences were noted in diagnostic confidence (mean confidence 7.7 vs. 7.8, $p=0.61$) or number of diagnoses documented in the initial differential (mean 1.8 vs. 2.0, $p=0.29$) between the pre- and post-intervention periods. There was an increase in overall lab testing in the post-intervention period (mean 4.7 vs. 6.1, $p=0.009$). No changes were noted in number of consults (mean 0.7 vs. 0.5, $p=0.14$) or imaging studies (mean 2.1 vs. 1.8, $p=0.21$) requested.

After adjusting for gender, race, shift duration and clustering of patients by provider, the use of at least one intervention component was not associated with diagnostic confidence ($p=0.76$), the total number of diagnoses listed in differential ($p=0.36$), imaging orders ($p=0.40$), or consults requested ($p=0.06$). However, use of at least one intervention component was associated with a statistically significant increase in the total number of labs ordered (intervention fixed coefficient=1.24, 95 % CI=0.17 to 2.31, $p=0.02$).

Qualitative results

Five hospitalists participated in interviews following the conclusion of the intervention. Review of interview data elicited several important findings. First, participants valued the “social” aspect of the intervention. To elaborate, during the design of the intervention, work locations were intentionally included within the team room environment and privacy screens were not completely obstructive. Nonetheless, most interviewed individuals found neither the privacy screen nor the headphones to be helpful.

It's not unusual during the shift that you are going to have questions about triaging a patient to where the most appropriate place for this patient to go is and so, it felt like you kind of had to take a physical break from the workspace to kind of find somebody or talk to somebody about what was going on, as opposed to just feeling like it was in the flow of the work room itself. (Participant 3)

I didn't see anybody use [the headphones]. I never used them because you can't wear earmuffs when you are on the phone. You can't wear earmuffs when you have to hear a pager. (Participant 1)

While not all participants found breathing exercises helpful, several noted that they utilized mindfulness-based practice as a reminder to slow down throughout a busy shift.

I used it as motivation – I am going to write four notes and then, I'm going to take a 15-minute break ... I will get up and walk around. That's how I will pace myself. (Participant 5)

I have an app on my phone ... and every so often, it will tell me to take a few seconds to breathe and I make a conscious effort ... it just tells me ... I need 30 seconds of just deep breathing. So, those things help me. (Participant 2).

Feedback on the use of an application that offered a diagnostic checklist was mixed. Participants noted little value of the checklist mechanism, though for a variety of reasons. For instance, some individuals felt that the checklist was utilized too late in the diagnostic process or was too conceptually rudimentary, whereas others noted that they have their own process (that at times already included checklists).

I mean, that's third year medical school stuff of what you do – what I do before I even see a patient ... you know, we have some people in our group who have been doing this for 20 years ... reviewing lab data, that's what you learn to do as a third year medical student ... and maybe that's important for a checklist but it just seems so ... trivial and like the low-functioning part of the job that if somebody is really needing a checklist to tell them to do that, they are not going to make it the first day. (Participant 1)

When I get to my documentation action and plan, that's really my checklist. I have a system that I follow, and I tie it back in with every lab or every complaint that maybe, okay, does this diagnosis make sense. So, yeah, I don't use checklists. It's kind of in here [points to brain] after having done it for so many years. (Participant 2)

Discussion

In this pilot study evaluating implementation of a multi-component intervention aimed at reducing diagnostic error, we found limited intervention uptake among hospitalists as they delivered clinical care. We found no significant differences in pre- and post-intervention subjective survey scores,

diagnostic confidence, differential diagnoses, or resource utilization aside from of a small increase in lab test ordering in the post-intervention period. Furthermore, outside of the increase in lab testing, no meaningful differences in outcomes with use of any one or more intervention components or any single individual intervention component were observed. These findings highlight the challenges of not only influencing diagnostic errors, but also conducting real world interventions in busy clinical arenas.

Although a negative study, this pilot informs the field of improving diagnostic errors in the “real world” in several ways. First, our study highlights many of the challenges associated with studying diagnostic errors within the clinical space. Our interventions were based on hundreds of hours of observations during which study team members embedded with medicine teams to understand physician workflow and both systems and cognitive challenges related to diagnosis, as well as dozens of hours of focus groups and interviews. Evident from those sessions was a need for balance between space for reflection and space for collaboration and a recognition of the timing in which diagnosis occurs (typically during documentation). Our intervention was built with these theoretical ideals in mind (e.g., privacy screens that were within the team environment and without complete obstruction, headphones that could be donned and doffed, a checklist to incorporate during time of documentation). That no component was utilized by more than half of the participants suggests that where theory meets practice is highly variable.

Insights on how and why our intervention was unsuccessful may be derived from the field of implementation science. The first comes in an assessment of organizational readiness to change. Readiness for change requires a shared resolve to implement change [12]. Errors in diagnosis receive little of the attention of many other patient safety challenges and an understanding of the gravity of the problem cannot be assumed. Individually, that diagnostic errors and consequences resulting from such error are disjointed in time and space often means that providers are unaware that an error occurred. This phenomenon contributes to the underappreciation of the impact of diagnostic errors on patient safety. As a result, it is incumbent on change agents (or those designing interventions to curb error) to clearly articulate and demonstrate the issues importance. Raising further awareness of the perils of diagnostic error within the group may have created that shared resolve and could have improved intervention uptake.

Next, organizational change requires favorable appraisal of task demands, resources available, and situational factors [12]. Participants in interviews cited that the intervention, particularly use of the checklist, was both outside of their

standard workflow and perhaps too rudimentary. For example, multiple individuals cited their professional experience as a reason why they may not need a reminder to “review labs independently.” In a large, diverse group, working within a complex healthcare system, assessment of task demands of an intervention and the resources available are unlikely to be uniform – while some found the components of the intervention beneficial, others found them prohibitively time consuming.

While our intervention was unsuccessful within the clinical setting, several interventions have shown benefit. For example, two studies in which physicians discussed cases with colleagues in a formalized way resulted in improvement – the first, a nearly 5 % absolute risk reduction for adverse events in the emergency department [13] and the second a change in patient plan in over 50 % of cases in which a hospitalist consultant was involved [14]. Additional tools and cognitive aids (e.g., checklists, reflection tools) have demonstrated modest improvement in outcomes [15], though most have yet to be rigorously studied via randomized controlled trials [8]. Technological developments now offer new hope for improving diagnosis in the clinical setting. While the application of artificial intelligence to diagnosis remains early, large language models have been used to facilitate ward-based education through “information querying, second-order content exploration, and engaged team discussion regarding auto-generated responses,” [16] and have shown promise in vignette-based studies [17].

Our study has several limitations. For example, the interventions were derived based on observations of clinical workflow within a single healthcare system and interventions were implemented within the same system. Neither the observations that served as intervention bases nor the outcomes of the intervention implementation may be generalizable beyond our facility. Next, we were unable to assess for the presence of diagnostic error, but rather for provider behaviors that may have been associated with error. How and whether these factors are truly associated with error warrants further evaluation. Nonetheless, our study has important strengths. The study was designed with both an extensive understanding of the underlying facility contextual factors and was based on local provider feedback and supportive literature. Next, the intervention targeted both the individual cognitive factors as well as systems factors (e.g., distractions) that may contribute to error. Finally, we supplemented our quantitative findings with qualitative data to better understand the why behind our findings.

With an understanding of the contextual factors, we have several ideas on how interventions to curb diagnostic

errors may be implemented and assessed. First, take time to build a sense of urgency – while the impacts of diagnostic errors are becoming more recognized, personalizing the message may drive support for the intervention. Second, interventions should be focused. Multi-component interventions may increase participation broadly by allowing individuals to participate in only part of the intervention; however, the distribution of use across components may limit interpretation and generalizability of any findings. Next, interventions should be tightly integrated into standard workflow to improve compliance. Finally, when available, interventions should be designed and enacted with the aid of those trained in implementation science and organizational change after carefully understanding the clinical context in which the intervention will occur.

Diagnostic error continues to be called the “next front in patient safety”, tangible interventions to improve healthcare outcomes remain elusive. Moving from the laboratory to the bedside is imperative owing to the multitude of complexities within healthcare settings. Studies, such as this pilot study, thus, are essential to understand barriers and opportunities ahead.

Research ethics: This study was reviewed and approved by the Institutional Review Board at the University of Michigan Health System (HUM00145793) and was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

Informed consent: Informed consent was obtained from all individuals included in this study.

Author contributions: The authors have accepted responsibility for the entire content of this manuscript and approved its submission. All authors listed have contributed sufficiently to the project to be included as authors, and all have reviewed and given approval for submission of this manuscript.

Use of Large Language Models, AI and Machine Learning Tools: None declared.

Conflict of interests: The authors state no conflict of interest.

Research funding: This project was supported by grant number P30HS024385 from the Agency for Healthcare Research and Quality. The funding source played no role in study design, data acquisition, analyses, or decision to report these data. Dr. Greene receives funding support from the Agency for Healthcare Research and Quality and the Department of Veterans Affairs. Dr. Chopra is supported by funding the Agency for Healthcare Research and Quality (1 R18 HS025891-01).

Data availability: The raw data can be obtained on request from the corresponding author.

References

1. Chopra V, Harrod M, Winter S, Forman J, Quinn M, Krein S, et al. Focused ethnography of diagnosis in academic medical centers. *J Hosp Med* 2018;13:668–72.
2. Graber ML, Franklin N, Gordon R. Diagnostic error in internal medicine. *Arch Intern Med* 2005;165. <https://doi.org/10.1001/archinte.165.13.1493>.
3. Gupta A, Harrod M, Quinn M, Manojlovich M, Fowler KE, Singh H, et al. Mind the overlap: how system problems contribute to cognitive failure and diagnostic errors. *Diagn (Berl)*. 2018;5:151–6.
4. Gupta A, Snyder A, Kachalia A, Flanders S, Saint S, Chopra V. Malpractice claims related to diagnostic errors in the hospital. *BMJ Qual Saf* 2017;27: 53–60.
5. Graber ML, Kissam S, Payne VL, Meyer AND, Sorensen A, Lenfestey N, et al. Cognitive interventions to reduce diagnostic error: a narrative review. *BMJ Qual Saf* 2012;21:535–57.
6. Singh H, Graber ML, Kissam SM, Sorensen AV, Lenfestey NF, Tant EM, et al. System-related interventions to reduce diagnostic errors: a narrative review. *BMJ Qual Saf* 2012;21:160–70.
7. Abimanyi-Ochom J, Bohingamu Mudiyansele S, Catchpool M, Firipis M, Wanni Arachchige Dona S, Watts JJ. Strategies to reduce diagnostic errors: a systematic review. *BMC Med Inform Decis Mak* 2019;19:174.
8. Dave N, Bui S, Morgan C, Hickey S, Paul CL. Interventions targeted at reducing diagnostic error: systematic review. *BMJ Qual Saf* 2021;31: 297–307.
9. Quinn M, Forman J, Harrod M, Winter S, Fowler KE, Krein SL, et al. Electronic health records, communication, and data sharing: challenges and opportunities for improving the diagnostic process. *Diagn (Berl)* 2019;6:241–8.
10. Croskerry P, Singhal G, Mamede S. Cognitive debiasing 2: impediments to and strategies for change. *BMJ Qual Saf* 2013;22:ii65–72.
11. Taylor B, Henshall C, Kenyon S, Litchfield I, Greenfield S. Can rapid approaches to qualitative analysis deliver timely, valid findings to clinical leaders? A mixed methods study comparing rapid and thematic analysis. *BMJ Open* 2018;8:e019993.
12. Weiner BJ. A theory of organizational readiness for change. *Implement Sci* 2009;4:67.
13. Freund Y, Goulet H, Leblanc J, Bokobza J, Ray P, Maignan M, et al. Effect of systematic physician cross-checking on reducing adverse events in the emergency department: the CHARMED cluster randomized trial. *JAMA Intern Med* 2018;178:812–9.
14. O'Neill LB, Bhansali P, Rush M, Stokes S, Todd S, Shah NH. Development and implementation of a peer curbside consult service for pediatric hospitalists. *Hosp Pediatr* 2022;12:e330–8.
15. Staal J, Hooftman J, Gunput STG, Mamede S, Frens MA, Van den Broek WW, et al. Effect on diagnostic accuracy of cognitive reasoning tools for the workplace setting: systematic review and meta-analysis. *BMJ Qual Saf* 2022;31:899–910.
16. Skryd A, Lawrence K. ChatGPT as a tool for medical education and clinical decision-making on the wards: case study. *JMIR Form Res* 2024; 8:e51346.
17. Jabbour S, Fouhey D, Shepard S, Valley TS, Kazerooni EA, Banovic N, et al. Measuring the impact of ai in the diagnosis of hospitalized patients: a randomized clinical vignette survey study. *JAMA* 2023;330:2275–84.

Supplementary Material: This article contains supplementary material (<https://doi.org/10.1515/dx-2024-0099>).