Can the Use of Siri, Alexa, and Google Assistant for Medical Information Result in Patient Harm?

Bickmore TW, Trinh H, Olafsson S, et al. Patient and consumer safety risks when using conversational assistants for medical information: an observational study of Siri, Alexa, and Google Assistant. J Med Internet Res. 2018;20:e11510.

Study Overview

Objective. To determine the prevalence and nature of the harm that could result from patients or consumers using conversational assistants for medical information.

Design. Observational study.

Settings and participants. Participants were recruited from an online job posting site and were eligible if they were aged ≥ 21 years and were native speakers of English. There were no other eligibility requirements. Participants contacted a research assistant by phone or email, and eligibility was confirmed before scheduling the study visit and again after arrival. However, data from 4 participants was excluded after the participants disclosed that they were not native English speakers at the end of their study sessions. Participants were compensated for their time.

Each participant took part in a single 60-minute usability session. Following informed consent and administration of baseline questionnaires, each was assigned a random selection of 2 medication tasks and 1 emergency task (provided as written scenarios) to perform with each conversational assistant—Siri, Alexa, and Google Assistant—with the order of assistants and tasks counterbalanced. Before the participants completed their first task with each conversational assistant, the research assistant demonstrated how to activate the conversational assistant using a standard weather-related question, after which the participant was asked to think of a health-related question and given 5 minutes to practice interacting with the conversational assistant with their question. Participants were then asked to complete the 3 tasks in sequence, querying the conversational assistant in their own words. Tasks were considered completed either when participants stated that they had found an answer to the question or when 5 minutes had elapsed. At task completion, the research assistant asked the participant what they would do next given the information obtained during the interaction with the conversational assistant. After the participant completed the third task with a given conversational assistant, the research assistant administered the satisfaction questionnaire. After a participant finished interacting with all 3 conversational assistants, they were interviewed about their experience.

Measures and analysis. Interactions with conversational assistants were video recorded, with the audio transcribed for analysis. Since each task typically took multiple at-

tempts before resolution or the participant gave up, usability metrics were coded at both the task and attempt level, including time, outcomes, and error analysis. Participant-reported actions for each medical task were rated for patient harm by 2 judges (an internist and a pharmacist) using a scale adapted from those used by the Agency for Healthcare Research and Quality and the US Food and Drug Administration. Scoring was based on the following values: 0 for no harm; 1 for mild harm, resulting in bodily or psychological injury; 2 for moderate harm, resulting in bodily or psychological injury adversely affecting the functional ability or quality of life; 3 for severe harm, resulting in bodily or psychological injury, including pain or disfigurement, that interferes substantially with functional ability or quality of life; and 4 was awarded in the event of death. The 2 judges first assigned ratings independently, then met to reach consensus on cases where they disagreed. Every harmful outcome was then analyzed to determine the type of error and cause of the outcome (user error, system error, or both). The satisfaction questionnaire included 6 self-report items with response values on a 7-point scale ranging from "Not at all" to "Very satisfied."

Main results. 54 participants completed the study, with a mean age of 42 years (SD 18) and a higher representation of individuals in the 21- to 24-year-old category than the general US adult population (30% compared to 14%). Twenty-nine (54%) were female, 31 (57%) Caucasian, and 26 (50%) college educated. Most (52 [96%]) had high levels of health literacy. Only 8 (15%) reported using a conversational assistant regularly, while 22 (41%) had never used one, and 24 (44%) had tried one "a few times." Forty-four (82%) used computers regularly.

Of the 168 tasks completed with reported actions, 49 (29.2%) could have resulted in some degree of harm, including 27 (16.1%) that could have resulted in death. An analysis of 44 cases that potentially resulted in harm yielded several recurring error scenarios, with blame attributed solely to the conversational assistant in 13 (30%) cases, to the user in 20 (46%) cases, and to both the user and the conversational assistant in the remaining 11 (25%) cases. The most common harm scenario (9 cases, (21%) is one where the participant fails to provide all the information in the task description, and the conversational assistant responds correctly to the partial query, which the user then accepts as the recommended action to take. The next most common type of harm scenario occurs when the participant provides a complete and correct utterance describing the problem and the conversational assistant responds with partial information (7 cases, 16%). Overall self-reported satisfaction with conversational assistants was neutral, with a median rating of 4 (IQR 1-6).

Outcomes by conversational assistant were significantly different ($X^2_4 = 132.2$, $P < 0.001$). Alexa failed for most tasks (125/394 [91.9%]), resulting in significantly more attempts made but significantly fewer instances in which responses could lead to harm. Siri had the highest task completion rate (365 [77.6%]), in part because it typically displayed a list of web pages in its response that provided at least some information to the participant. However, because of this, it had the highest likelihood of causing harm for the tasks tested (27 [20.9%]). Median user satisfaction with the 3 conversational assistants was neutral, but with significant differences among them. Participants were least satisfied with Alexa and most satisfied with Siri, and stated they were most likely to follow the recommendations provided by Siri.

Qualitatively, most participants said they would use conversational assistants for medical information, but many felt they were not quite up to the task yet. When asked about their trust in the results provided by the conversational assistants, participants said they trusted Siri the most because it provided links to multiple websites in response to their queries, allowing them to choose the response that most closely matched their assumptions. They also appreciated that Siri provided a display of its speech recognition results, giving them more confidence in its responses, and allowing them to modify their query if needed. Many participants expressed frustration with the systems, but particularly Alexa.

Conclusion. Reliance on conversational assistants for actionable medical information represents a safety risk for patients and consumers. Patients should be cautioned to not use these technologies for answers to medical questions they intend to act on without further consultation from a health care provider.

Commentary

Roughly 9 in 10 American adults use the Internet,[1] with the ability to easily access information through a variety of devices including smartphones, tablets, and laptop computers. This ease of access to information has played an important role in shifting how individuals access health information and interact with their health care provider.[2,3] Online health information can increase patients' knowledge of, competence with, and engagement in health care decision-making strategies. Online health information seeking can also complement and be used in synergy with provider-patient interactions. However, online health information is difficult to regulate, complicated further by the wide range of health information literacy among patients. Inaccurate or misleading health information can lead patients to make detrimental or even dangerous health decisions. These benefits and concerns similarly apply to conversational assistants like Siri (Apple), Alexa (Amazon), and Google Assistant, which are increasingly being used by patients and consumers to access medical- and health-related information. As these technologies are voice-activated, they appear to address some health literacy limitations. However, they still pose important limitations and safety risks,[4] especially as conversational assistants are being perceived as a trustworthy parallel to clinical assessment and counseling systems.[5]

There has been little systematic research to explore potential risks of these platforms, as well as systematically characterize error types and error rates. This study aimed to determine the capabilities of widely used, general-purpose conversational assistants in responding to a broad range of medical questions when asked by laypersons in their own words and sought to conduct a systematic evaluation of the potential harm that could result from patients or consumers acting on the resulting recommendations. The study authors found that when asked questions about situations that require medical expertise, conversational assistants failed more than half of the time and led study participants to report that they would take actions that could have resulted in harm or death. Further, the authors characterized several failure modes, including errors due to misrecognition of study participant queries, study participant misunderstanding of tasks and responses by the conversation assistant, and limited understanding of the capabilities of the assistants to understand user queries. This misalignment of expectations by users that assistants can follow conversations/discourse led to frustrating experiences by some study participants.

Not only do these findings make important contributions to the literature of health information–seeking behaviors and limitations via conversational assistants, the study design highlights relevant approaches to evaluating interactions between users and conversational assistants and other voice-activated platforms. The authors designed a range of everyday task scenarios that real-life users may be experiencing and that can lead to querying home or smartphone devices to seek health- or medical-related information. These scenarios were also written with a level of real-life complexity that incorporated multiple facts to be considered for a successful resolution and the potential of harmful consequences should the correct course of action not be taken. In addition, they allowed study participants to interpret these task scenarios and query the conversational assistants in their own words, which further aligned with how users would typically interact with their devices.

However, this study also had some limitations, which the authors highlighted. Eligibility was limited to only English-speakers and the study sample was skewed towards younger, more educated individuals with high health literacy. Combined with the small convenience sample used, findings may not be generalizable to other/broader populations and further studies are needed, especially to highlight potential differences in population subgroups (eg, race/ethnicity, age, health literacy).

Applications for Clinical Practice

Because of the increased prevalence of online health-information–seeking behaviors by patients, clinicians must be prepared to adequately address, and in some cases, educate patients on the accuracy or relevance of medical/health information they find. Conversational assistants pose an important risk in health care as they incorporate natural language interfaces that can simulate and be misinterpreted as counseling systems by patients. As the authors highlight, laypersons cannot know what

the full, detailed capabilities of conversational assistants are, either concerning their medical expertise or the aspects of natural language dialogue the conversational assistants can handle. Therefore, it is critical that clinicians and other providers emphasize the limitations of these technologies to patients and that any medical recommendations should be confirmed with health care professionals before they are acted on.

—Katrina F. Mateo, MPH

References

1. Pew Research Center. Demographics of Internet and Home Broadband Usage in the United States [online]. Accessed at: http://www.pewinternet.org/fact-sheet/internet-broadband/.
2. Tonsaker T, Bartlett G, Trpkov C. Health information on the Internet: gold mine or minefield? *Can Fam Physician.* 2014;60:407-408.
3. Tan SS-L, Goonawardene N. Internet health information seeking and the patient-physician relationship: a systematic review. *J Med Internet Res.* 2017;19:e9.
4. Chung H, Iorga M, Voas J, Lee S. Alexa, can I trust you? *Computer (Long Beach Calif).* 2017;50:100-104.
5. Miner AS, Milstein A, Hancock JT. Talking to machines about personal mental health problems. *JAMA.* 2017;318:1217.

Hip Fracture in Nursing Home Residents with Advanced Dementia: An Opportunity for Palliative Care

Berry SD, Rothbaum RR, Kiel DP, et al. Association of clinical outcomes with surgical repair of hip fractures vs nonsurgical management in nursing home residents with advanced dementia. JAMA Intern Med. 2018;178:774-780.

Study Overview

Objective. To compare clinical outcomes (mortality, pain, physical restraint use, pressure ulcer, antipsychotic drug use) in long-term care nursing home (NH) residents with advanced dementia and hip fracture who underwent surgical repair or nonsurgical management.

Design. A retrospective cohort study utilizing nationwide Medicare (Parts A, B, D and hospice) claims data linked with Centers for Medicare & Medicaid Services Minimum Data Set (MDS version 2.0) assessments.

Setting and participants. Long-stay NH residents older than 65 years with advanced dementia (defined as being assigned to Cognitive Performance Scale category 5 or 6 and a diagnosis of dementia or Alzheimer disease) and without a do not hospitalize (DNH) directive before hip fracture were identified by using MDS assessments completed from January 1, 2008 to December 31, 2013. Medicare (Part A – inpatient, or Part B – outpatient) claims data was then used to identify those residents who experienced a hip fracture within 2 years of the full MDS assessment using the International Classification of Diseases, Ninth Revision diagnostic codes. Procedure codes were used to determine whether a resident who experienced hip fracture underwent surgical repair.

Main outcome measures. The main outcome measure was all-cause mortality after hip fracture ascertained by the Medicare Enrollment File through 2013. The secondary outcome measures were documented pain, physical restraint use, pressure ulcers, antipsychotic drug use, and ambulatory status in NH residents who survived 6 months after hip fracture. These outcome measures were captured from the first MDS assessment completed between 120 and 240 days following the fracture or Medicare Part D claims. Documented pain was determined using a validated MDS 2.0 nursing assessment pain instrument within 7 days preceding MDS assessment. Physical restraint use was defined by the use of trunk, limb, or chair restraint within 7 days prior to MDS assessment. Pressure ulcer was defined as any stage 2 to 4 pressure ulcer. Antipsy-

chotic drug use of any medication subclass was determined using Medicare Part D claims data and affirmative if drug was administered 180 days after hip fracture. Ambulatory status between 120 and 240 days following the fracture was determined in a subset of NH residents who were ambulatory before the hip fracture. The utilization of comfort-focused care after hip fracture was determined in NH residents who had a Medicare hospice claim or a new DNH directive in the 180 days after hip fracture.

The differences in survival among NH residents with advanced dementia and hip fracture were described by Kaplan-Meier curves. The association between surgical repair and survival was determined using multivariable Cox proportional hazards for all NH residents and stratified by pre-fracture ambulatory status. In those who survived 6 months after hip fracture, the associations between surgical repair and outcomes including documented pain, physical restraint use, pressure ulcers, antipsychotic drug use, and ambulatory status were determined using multivariable logistic regression models. Adjustment for differences in characteristics before hip fracture was performed using inverse probability of treatment weighting (IPTW) models.

Main results. 3083 long-stay NH residents with advanced dementia and hip fracture were included in the study. The cohort's mean age was 84.2 ± 7.1 years, 79.2% were female (n = 2441), and 28.5% were ambulatory before hip fracture (n = 879). Of these NH residents, 84.8% (n = 2615) underwent surgical repair and 15.2% (n = 468) received nonsurgical management. At 6 months after hip fracture, mortality was 31.5% in the surgical group compared to 53.8% in the nonsurgical group. The greatest mortality difference between groups occurred in the first 30 days after hip fracture (11.5% in surgical group versus 30.6% in nonsurgical group). Surgical repair was associated with a decreased risk of death (Cox proportional hazard ratio) in the unadjusted (hazard ratio [HR], 0.55 [95% confidence interval {CI}, 0.49-0.61), multivariable adjusted (adjusted HR, 0.56 [95% CI, 0.49-0.63]), and IPTW (adjusted HR, 0.88 [95% CI, 0.79-0.98]) models. Similarly, surgically treated NH residents were less likely to die than those managed non-surgically when mortality was stratified by pre-fracture ambulatory status.

Among NH residents who survived 6 months after hip fracture, those who underwent surgical repair compared with those who received nonsurgical management had less documented pain (29.0% versus 30.9%), fewer pressure ulcers (11.2% versus 19.0%), greater physical restraint use (13.0% versus 11.1%), and greater antipsychotic drug use (29.5% versus 20.4%). In the adjusted IPTW models, surgical repair was associated with less pain (adjusted HR, 0.78 [95% CI, 0.61-0.99]) and fewer pressure ulcers (adjusted HR, 0.64 [95% CI, 0.47-0.86]).

Overall, 21.5% of NH residents utilized comfort-focused care within 6 months after hip fracture, with a mean time to utilization of hospice care of 56 ± 49 days. In those who were managed surgically, 19.3% utilized hospice care, as compared with 33.8% in those who did not receive surgical intervention. In NH residents who survived 6 months after hip fracture, only 1.1% in both groups acquired a DNH directive.

Conclusion. In older long-stay NH residents with advanced dementia and hip fracture, surgical repair was associated with lower all-cause mortality, less documented pain, and fewer pressure ulcers compared to nonsurgical management. However, adverse clinical outcomes such as pain, physical restraint use, pressure ulcers, and antipsychotic drug use were common regardless of treatment modality. The high incidence of these adverse outcomes and hazardous interventions, coupled with low utilization of comfort-focused care and DNH directive, highlight an opportunity to improve the quality of care in this vulnerable population.

Commentary

Hip fracture is very common in NH residents, with an overall incident rate of 2.3 per 100 person years and is associated with a high mortality rate of 36.2% by 6 months after fracture.[1,2] Moreover, Neuman and colleagues have recently reported that among NH residents who have some degree of functional independence in locomotion prior to hip fracture, 54% either die or develop new total dependence in locomotion within 6 months of fracture and that severe cognitive impairment is a risk factor highly associated with these adverse outcomes.[3] Despite this emerging knowledge, surgical repair of hip fracture re-

mains the mainstay treatment in many NH residents in the hope of alleviating pain and improving mobility, and palliative care is considered only when patients are imminently dying or have deteriorated past the point of meaningful recovery. In cases of NH residents with advanced dementia whose life expectancy is limited and whose care goals may favor maintaining comfort, the health care proxies are frequently challenged with a difficult choice of either pursuing or foregoing surgical management—a complex medical decision to be made in the absence of sufficient evidence in this uniquely frail patient population.

The study reported by Berry and colleagues provides an important and timely investigation in examining associations of adverse clinical outcomes (mortality, pain, pressure ulcer) and hazardous interventions (physical restraint and antipsychotic drug use) in long-stay NH residents with advanced dementia and hip fracture who underwent surgical repair or nonsurgical management. The authors reported a 6-month mortality rate of 31.5% in NH residents who underwent surgical repair, an event rate similar to that reported by Neuman and colleagues. While surgical repair after hip fracture was associated with a decreased risk of death compared to nonsurgical management, high incidences of pain (29.0% to 30.9%) and pressure ulcers (11.2% to 19.0%), and frequent physical restraint use (11.1% to 13.0%) and antipsychotic drug use (20.4% to 29.5%) were noted in NH residents who survived 6 months after fracture regardless of treatment modality. These findings are consistent with the high rate of post-hip fracture functional disability previously reported by Neuman and colleagues, and highlight the trajectory of decline, frequent distressing symptoms, and prevalent use of pharmacologic and nonpharmacologic restraints in long-stay NH residents after hip fracture. Taken together, the low utilization of comfort-focused care (21.5%) and DNH directive (1.1%) in NH residents who survived 6 months suggest a missed opportunity to integrate palliative care in a patient population that stands to benefit from this intervention.

This study is the first to report the associations between hip fracture surgery and a reduction in adverse outcomes such as pain and pressure ulcer that commonly affect vulnerable NH residents with advanced dementia. This study was well designed and leveraged strengths of Medicare claims data linked with MDS assessments to capture outcome measures including pain, pressure ulcer, and restraint use that would otherwise be difficult to ascertain. However, as in all retrospective cohort design, there were limitations in this study. For instance, secondary outcomes were determined from a single time point (ie, first MDS assessment completed between 120 to 240 days following hip fracture) and thus data capture may be incomplete. Additionally, other conditions important to complex decision making in the care of frail older adults including postoperative complications (eg, delirium, infections, cardiac complications) and in-hospital mortality were not examined. Despite these limitations, this study has enhanced our understanding of the clinical course of long-term care NH residents with advanced dementia who endured hip fracture.

Applications for Clinical Practice

Patients' goals of care should guide medical decision making in the management of hip fracture in NH residents with advanced dementia. The increased survival benefit of surgical repair of hip fracture in this patient population should be considered in the medical decision-making process if life-prolongation is preferred. However, palliative and hospice care need to be an important facet of discussion given the high rates of mortality, pain, pressure ulcer, and restraint use in this vulnerable subset of older adults.

—Fred Ko, MD, MS

References

1. Berry SD, Lee Y, Zullo AR, et al. Incidence of hip fracture in U.S. nursing homes. J Gerontol A Biol Sci Med Sci. 2016;71:1230-1234.
2. Neuman MD, Silber JH, Magaziner JS, et al. Survival and functional outcomes after hip fracture among nursing home residents. JAMA Intern Med. 2014;174:1273-1280.
3. Berry SD, Rothbaum RR, Kiel DP, et al. Association of clinical outcomes with surgical repair of hip fractures vs nonsurgical management in nursing home residents with advanced dementia. JAMA Intern Med. 2018;178:774-780.

Procalcitonin, Will It Guide Us?

Huang DT, Yealy DM, Filbin MR, et al. Procalcitonin-guided use of antibiotics for lower respiratory tract infection. New Engl J Med. 2018;379:236-249.

Study Overview

Objective. To assess whether procalcitonin-guided antibiotic usage results in less exposure to antibiotics than usual care, without a significantly higher rate of adverse events.

Design. Multi-center 1:1 randomized trial.

Setting and participants. This study was conducted at 14 academic hospitals in the United States between 2014 and 2017 in which procalcitonin assay was not routinely used. All adult patients in the emergency department with an initial diagnosis of acute lower respiratory tract infection without a decision to give or withhold antibiotics because of uncertainty regarding the need for antibiotics were included in the study. Patients were excluded if antibiotics were unlikely to be held in their case, such as if there was a need for mechanical ventilation or known severe immunosuppression, and if procalcitonin could be falsely elevated (chronic dialysis, metastatic cancer, surgery in the past 7 days).

Intervention. Patients were randomly assigned to receive guideline-based care using procalcitonin (procalcitonin group) or usual care (usual-care group). In the procalcitonin group, the procalcitonin assay results, and the procalcitonin treatment guidelines were provided to the treating physician. The guideline used previously established cutoffs (procalcitonin level of < 0.1 μg/L, antibiotics were strongly discouraged; 0.1 to 0.25 μg/L, antibiotics were discouraged; 0.25 to 0.5 μg/L, antibiotics were recommended; and > 0.5 μg/L, antibiotics were strongly recommended). Procalcitonin was measured initially in the emergency department. If the patient was hospitalized, procalcitonin was again measured 6 to 24 hours later, and on hospital days 3, 5, and 7. To implement this intervention, a multifaceted approach was used, which included sending letters to local primary care providers describing the trial, ensuring rapid delivery of procalcitonin results by tracking and coordinating blood samples with routine morning draws, and embedding the procalcitonin results and guidelines into the sites' electronic health records. In the usual-care group, procalcitonin levels at enrollment were measured but not disclosed to clinicians. In both treatment groups, clinicians retained autonomy regarding care decisions.

Main outcome measures. The primary outcome was total antibiotic exposure, defined as the total number of antibiotic-days within 30 days after enrollment. The primary safety outcome was any adverse effects that could be attributable to withholding antibiotics in lower respiratory tract infections, within 30 days after enrollment. Secondary outcomes included admission to the intensive care unit (ICU), subsequent emergency department visits by day 30, and quality of life as assessed with the Airway Questionnaire 20.

Main results. 8360 patients with acute lower respiratory tract infection who presented to the emergency department were screened for eligibility; of these, 1664 patients underwent randomization. Ultimately, 1656 patients were included in the final analysis cohort (826 in the procalcitonin group and 830 in the usual-care group), because 8 patients withdrew. Of the cohort, 1345 (81.2%) patients completed the full 30-day follow up. Baseline characteristics were similar between the treatment groups. In the procalcitonin group, clinicians received the procalcitonin results for 95.9% of the patients. As a result of clinical care, 2.2% of the patients in the usual-care group also had procalcitonin testing. Clinicians adhered to the procalcitonin guideline recommendations for 64.8% of the procalcitonin group.

There was no significant difference in the intention-treat-treat analysis between the procalcitonin group and the usual-care group in antibiotic days during the first 30 days (mean antibiotic days, 4.2 and 4.3 days, re-

spectively [95% confidence interval {CI}, −0.6 to 0.5; P = 0.87]). Within 30 days there was no significant difference in the proportion of patients with adverse outcomes in the procalcitonin group and usual-care group (11.7% and 13.1%, respectively [95% CI, −4.6 to 1.7]; $P < 0.01$ for noninferiority). There was no significant difference between the procalcitonin and usual-care groups for any of the secondary outcomes.

Conclusion. A procalcitonin-directed antibiotic administration guideline did not result in fewer antibiotic days than did usual-care among patients with suspected lower respiratory tract infection.

Commentary

Procalcitonin is a serum biomarker synthesized in thyroid neuroendocrine cells and is the precursor to calcitonin.[1] It is undetectable in healthy human serum, but in the setting of systemic inflammation caused by bacterial infection, procalcitonin synthesis is induced in many tissues. Since its discovery in 1970, procalcitonin's potential utility has been sought in various settings, such as guiding the initiation and/or discontinuation of antibiotics.[2]

In a prospective randomized trial in patients with an acute chronic obstructive pulmonary disease (COPD) exacerbation, treatment success was not better with antibiotics than placebo in patients with a procalcitonin level < 0.1 µg/L.[3] Others replicated these results in COPD patients with acute exacerbation of COPD.[4] Another small randomized trial showed that using procalcitonin in intensive care patients reduced antibiotic duration.[5] Another small study found similar results in their critical care setting.[6] Procalcitonin-guided antibiotic treatment produced similar results in patients with aspiration pneumonia.[7] In summary, previously published studies nearly uniformly report reduced antibiotic duration or initiation using procalcitonin cutoffs without increasing adverse events.

In the current study, Huang and colleagues conducted a multi-center randomized trial in 14 academic US hospitals, while simultaneously attempting quality improvement methods for implementing and maximizing compliance with procalcitonin guidelines for local physicians. This study was able to achieve approximately 65% compliance with the guideline, which is relatively lower than in previously reported studies using procalcitonin guidelines. This study was larger and involved more hospitals than the other studies. Interestingly, this study did not find statistically significant differences in antibiotic usage or duration between the procalcitonin group compared to the usual-care group. While this result can be partially explained by the low rate of compliance with the guideline, the result may actually reflect the real-life pattern of procalcitonin guideline usage in clinicians. These results suggest that procalcitonin-based guidelines attempting to reduce antibiotic usage and exposure may be of low value, contrasting with findings from previous studies.

The Huang et al study is well-designed, had a low rate of follow-up loss and withdrawal, was conducted mostly at urban academic hospitals that had a high level of adherence to Joint Commission pneumonia core measures, and had appropriate statistical analyses; however, several factors should be considered when applying the results of this study to clinical practice. First, the large majority (80.1%) of the study cohort had final diagnoses of a COPD exacerbation, asthma exacerbation, or acute bronchitis. These patients had a moderate degree of disease (required hospitalization in 59% of patients with a mean hospital length of stay of 5 days), but their symptoms were severe enough for the patients to present to the emergency department. Patients with a suspected nonrespiratory infection or a milder degree of infection, especially in the ambulatory care setting, may have different antibiotic prescribing patterns. Also, patients in the ambulatory care setting likely have different causal organisms of their diagnosis. Second, this study excluded patients with severe disease who required ICU admission with either septic shock or respiratory failure, patients with pre-existing diseases that placed them at high risk (eg, immunosuppressed patients), and/or patients who had complications of their infection with either a lung abscess or empyema. This pattern of exclusion was widely similar to the other previous procalcitonin studies, which shows that procalcitonin guidelines should not be applied blindly in critically ill patients, even those not requiring ICU admission. Third, patients were excluded from the study if they were on chronic dialysis, had metastatic cancer, or had a recent surgery because of possible elevation of procalcitonin levels without a bacterial infection.

In conclusion, the current study did not find any difference in antibiotic exposure throughout the course of care (including discharge or hospitalization) of patients with a lower respiratory tract infection who presented to the emergency department when a procalcitonin guideline was implemented. The results of the current study raise questions regarding the new trend of widely accepting procalcitonin-based antibiotic usage.

Applications for Clinical Practice

Procalcitonin is a relatively new marker that is released during a systemic bacterial infection. While prior studies have supported systematic use of procalcitonin-based guidelines to initiate and discontinue antibiotics in order to limit antibiotic exposure, clinicians should be mindful that a procalcitonin antibiotic guideline may be useful in specific patients and should only be used in combination with usual clinical judgment. Clinicians must also recognize the medical conditions that may falsely elevate the procalcitonin level. Most important, the procalcitonin level should not be used as the sole indication to withhold antibiotics in an otherwise appropriately indicated clinical scenario.

—Minkyung Kwon, MD, Scott A. Helgeson, MD, and Vichaya Arunthari, MD
Pulmonary and Critical Care Medicine, Mayo Clinic Florida, Jacksonville, FL

References

1. Maruna P, Nedelnikova K, Gurlich R. Physiology and genetics of procalcitonin. *Physiol Res.* 2000;49:S57-S61.
2. Deftos LJ, Roos BA, Bronzert D, Parthemore JG. Immunochemical heterogeneity of calcitonin in plasma. *J Clin Endocr Metab.* 1975;40:409-412.
3. Wang JX, Zhang SM, Li XH, et al. Acute exacerbations of chronic obstructive pulmonary disease with low serum procalcitonin values do not benefit from antibiotic treatment: a prospective randomized controlled trial. *Int J Infect Dis.* 2016;48:40-45.
4. Corti C, Fally M, Fabricius-Bjerre A, et al. Point-of-care procalcitonin test to reduce antibiotic exposure in patients hospitalized with acute exacerbation of COPD. *Int J Chron Obstruct Pulmon Dis.* 2016;11:1381-1389.
5. Deliberato RO, Marra AR, Sanches PR, et al. Clinical and economic impact of procalcitonin to shorten antimicrobial therapy in septic patients with proven bacterial infection in an intensive care setting. *Diagn Microbiol Infect Dis.* 2013;76:266-271.
6. Najafi A, Khodadadian A, Sanatkar M, et al. The comparison of procalcitonin guidance administer antibiotics with empiric antibiotic therapy in critically ill patients admitted in intensive care unit. *Acta Med Iran.* 2015;53:562-567.
7. Tanaka K, Ogasawara T, Aoshima Y, et al. Procalcitonin-guided algorithm in nursing and healthcare-associated pneumonia. *Respirology.* 2014;19:220-220.