Medical Decision Making in the Physician Hierarchy: A Pilot Pedagogical Evaluation

John Rosasco, OMS II,[1] Michele L. McCarroll, Ph.D.,[1] M. David Gothard, M.S.,[2] Jerry Myers, Ph.D.,[3]
Patrick Hughes, D.O., M.E.H.P.,[4] Alan Schwartz, Ph.D.,[5] Richard L. George, M.D., M.S.P.H.,[4] and Rami A. Ahmed, D.O., M.H.P.E.[4]

[1]Pacific Northwest University, [2]Biostats, Inc., [3]NASA Glenn Research Center
[4]Summa Health, [5]University of Illinois at Chicago

Disclosure

Financial support for this study was provided entirely by a sub-contract grant with Zin Technologies, Inc. The funding agreement ensured the authors' independence in designing the study, interpreting the data, writing, and publishing the report.

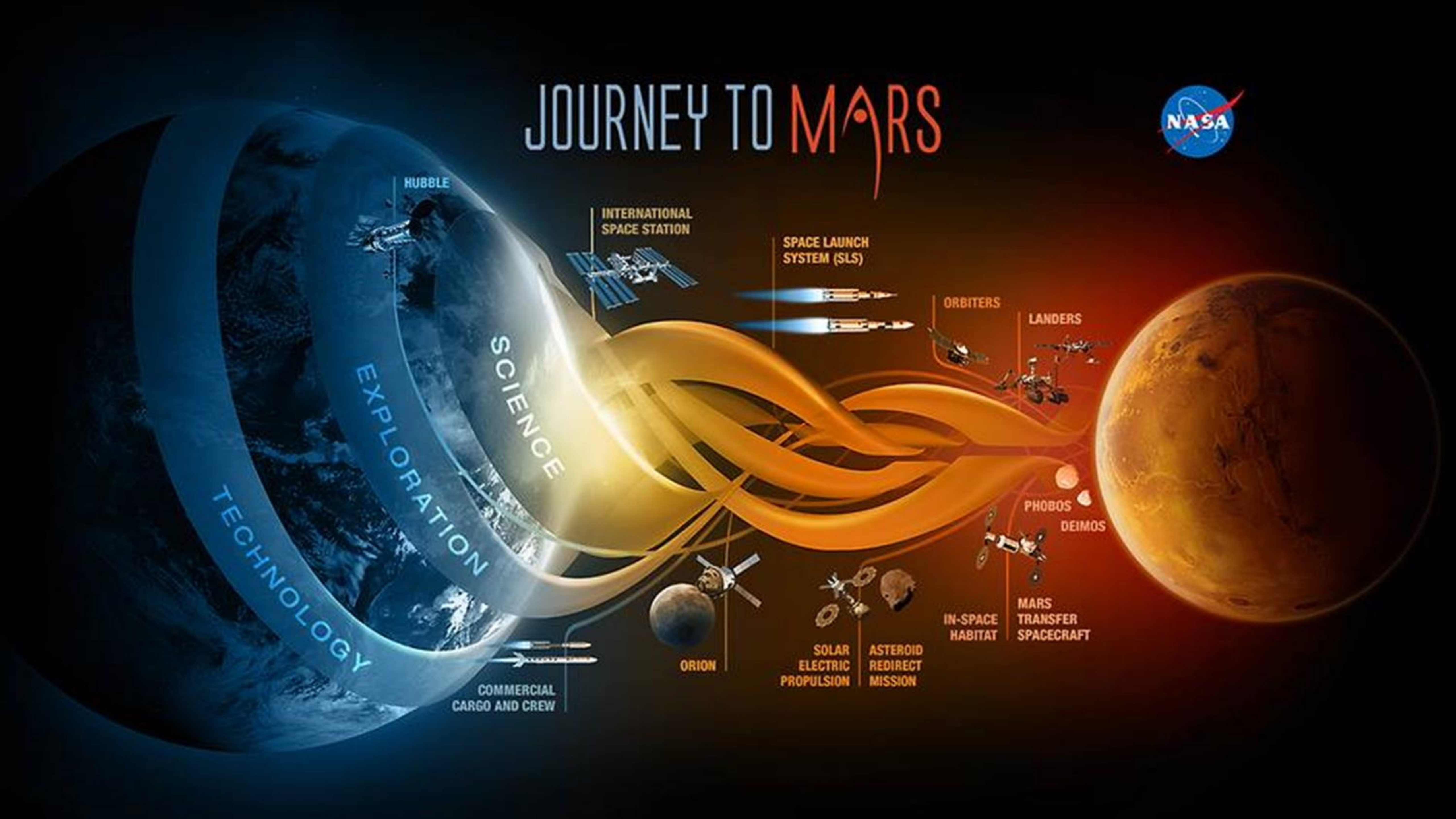
Contributors

Pacific Northwest University of Health Sciences, Yakima, WA

NASA Glenn Research Center, Cleveland, OH

Summa Health System, Akron, OH

Northeast Ohio Medical University, Rootstown, OH


JOURNEY TO MARS
NASA
HUBBLE
INTERNATIONAL SPACE STATION
SPACE LAUNCH SYSTEM (SLS)
ORBITERS
LANDERS
SCIENCE
EXPLORATION
TECHNOLOGY
PHOBOS
DEIMOS
ORION
SOLAR ELECTRIC PROPULSION
ASTEROID REDIRECT MISSION
IN-SPACE HABITAT
MARS TRANSFER SPACECRAFT
COMMERCIAL CARGO AND CREW

Background

Missions longer than 210 days, the Crew Medical Officer (CMO) must be a physician

While, perhaps intuitive, the above statement is not evidence based:

The quantity of days has not been verified

The requisite qualifications of the CMO have not been verified

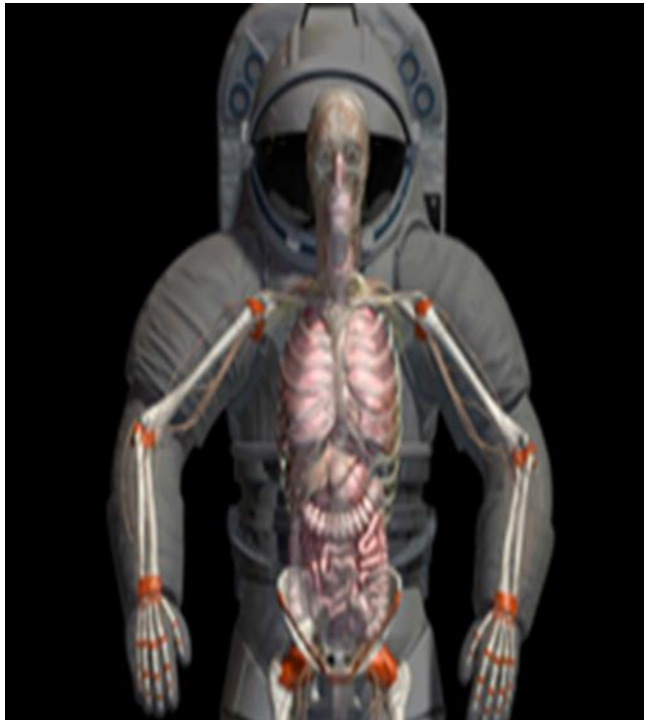


Photo Credit: NASA Timeline Photo 12/4/2014

Figure 1. Levels of evidence

Figure 1. Levels of evidence

Figure 1. Levels of evidence

Figure 1. Levels of evidence

Figure 1. Levels of evidence

Pacific Northwest University of Health Sciences

Figure 1. Levels of evidence

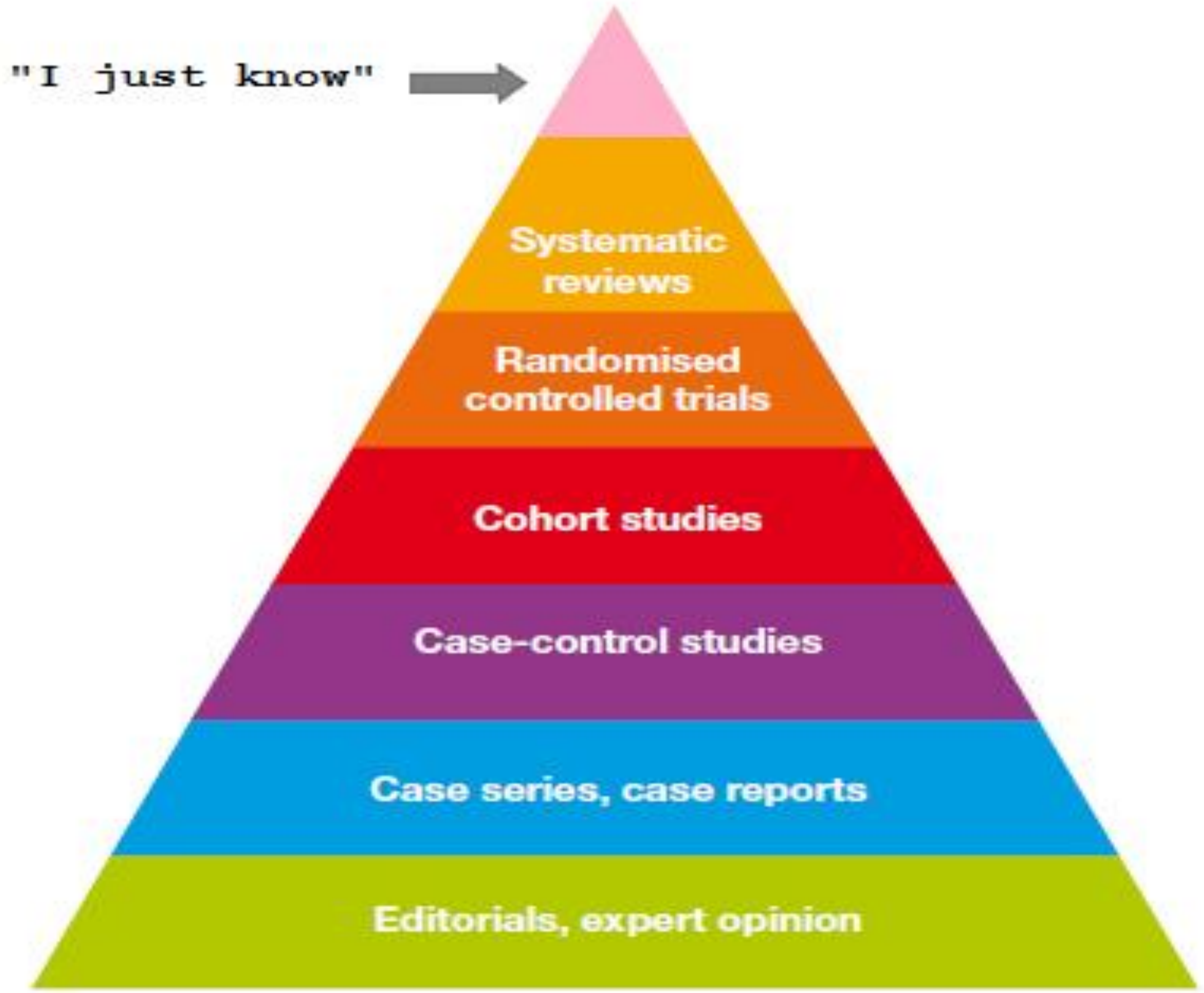


Figure 1. Levels of evidence

Background

- What makes the best doctor?
  - All of us know technically gifted proceduralists or surgeons, but are these always the best persons to make decisions for our loved ones in critical situations?
  - In multi-dimensional situations, particularly in acute care settings, the complexities of medical decision making may exceed linear decision making
  - So in these situations how can differentiate between a doctor who can make difficult decisions and those who can only follow a recipe?

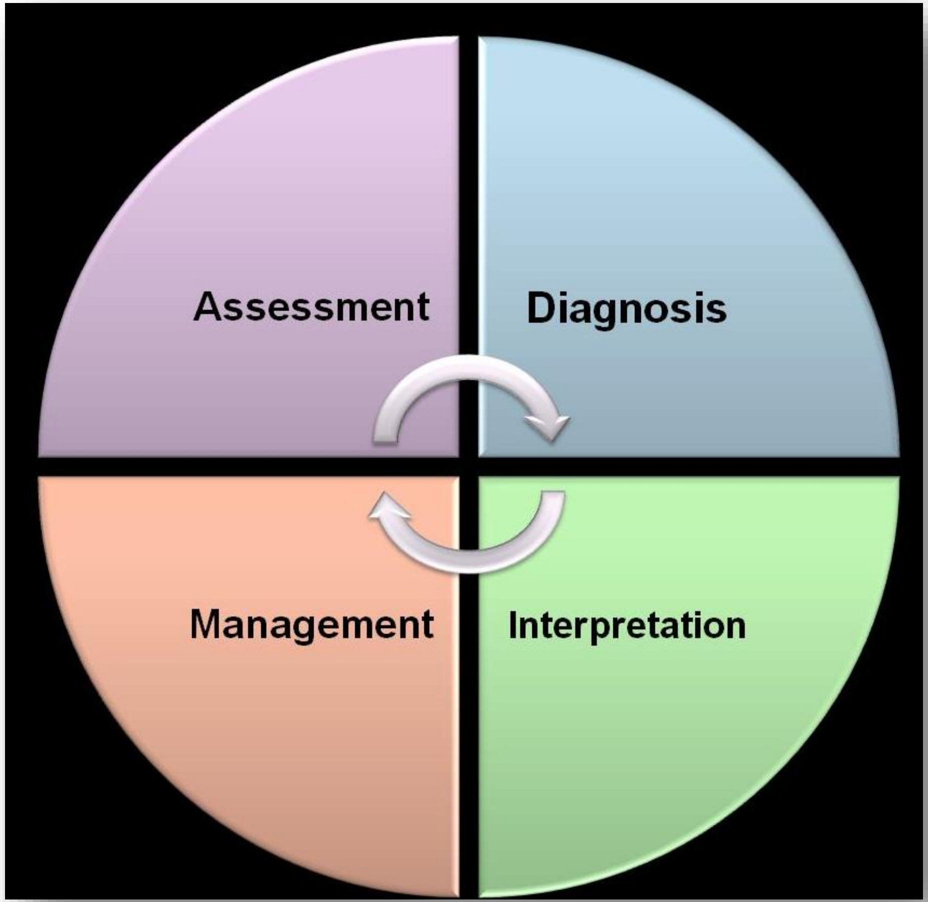
Background

- Recently, the American College of Graduate Medical Education (ACGME) has included medical decision making as a core competency in several specialties. [1,2,3]

Background

- What is Medical Decision Making?

- Medical decision making describes the ability to build ties between data obtained from history, physical exam, imaging results, and laboratory studies in order to formulate assessments and plans for patients. [4]



Background

- Good medical decision making leads to improved outcomes for patients whereas poor medical decision making is associated with morbidity and mortality in patients [5]
- Promoting strong medical judgment skills is a unique challenge in the evaluation of physician training, particularly as medical education continues to develop [6]
- Thus, standardized criteria are needed to evaluate medical decision making not only among medical students but also amongst post-graduate physicians [7]

Background

- How have we been teaching medical decision making during physician training?

- In medical education, there are various levels of pedagogical transitions including undergraduate medical education (UGME), graduate medical education (GME), and continuing medical education (CME) where evidenced-based and scientifically-based knowledge are applied to ensure quality performance and there is a vertical transmission of knowledge. [8]

Background

- How do we assess this transmission of knowledge?

- Physician competency has been assessed via standard training checkpoints such as state licensing requirements, national board exams, CME requirements for specialties, and board re-certification standards; however, accurately measuring and determining competency in decision making from these assessments remains a challenge. [9]

Background

- What are we doing to address that challenge?
  - Medical education moved to integrating key competencies in medical simulation, problem based learning (PBL), and case-based discussions in an attempt to measure competency and judgment.[10, 11, 12. 13]

Attending’s opinion

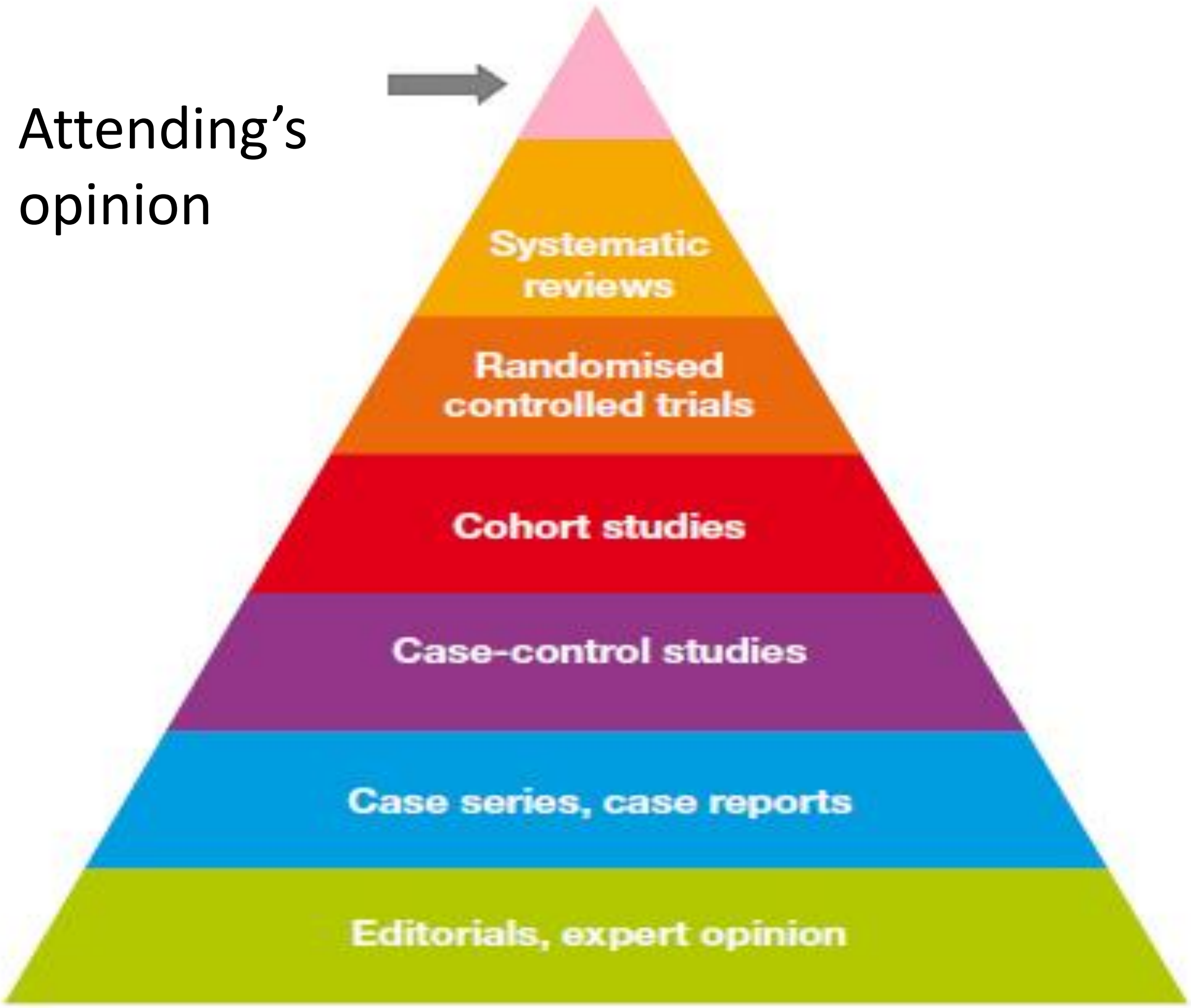


Figure 1. Levels of evidence

Background

- Medicine lacks a reproducible environment married with a consistent assessment tool that allows measurement of safe medical decision making, particularly we lack a quantitative assessment of this challenging topic. [14]

- To address the needs of the environment the simulation lab has emerged as a reproducible environment that provides experience and training in unpredictable or rare situations. [15]

Objective/Aims

Aim 1: To develop a Medical Judgment Pathway Metric (MJPM) assessing clinical judgment based on evidence-based practice for two categories from the Exploration Medical Conditions List[16]:

Aim 2: Further define a role for the MJM as a tool for the analysis of competency in medical decision making.

Objective/Aims

Aim 3: To implement the MJPM in medical simulations across an analog study population:

- An advanced group of medical professionals
- An engineering group with very basic medical experience

Objective/Aims

Aim 4: Evaluate the medical decision making of physicians in different stages of practice in acute care simulations utilizing the MJM

?

Background - MJM

- The Medical Judgment Metric (MJM)
    - "A numerical rubric to quantify good decisions in practice in simulated environments" [17]
    - The individual judgment items, clinical domains, and competency sections of the initial MJM were created based off:
        - Existing framework by Weber *et al.* as well as the *NOTECHS* scale[18, 19, 20]
        - ACGME Clinical Competency Committees [21]
        - The Association of American Medical Colleges conceptual frameworks for clinical judgment. [22]

Background - MJM

- 4 domains consisting of findings from:
  - Health and Physical
  - Interpretation
  - Diagnostic
  - Management

- Items within each domain were stratified into competency levels: Novice, Intermediate, Proficient, and Advanced corresponding to a grade of 1-4 for each

- Then each domain is scored on a 0.5 interval scale from 1-4 for a total score of up to 16

Health & Physical

Health & Physical	Level 1 (Novice)	Level 2 (intermediate)	Level 3 (Proficient)	Level 4 (Advanced)
	Does not collect accurate historical data	Limited ability to acquire accurate historical information in an organized fashion	Demonstrates ability to acquire accurate and relevant histories from patient	Proficient at acquiring accurate histories from the patients in an efficient prioritized and hypothesis driven fashion
	Does not use Physical exam to confirm history	Limited thorough physical exam or misses key physical exam findings	Demonstrates ability to perform accurate and appropriately thorough physical exams	Proficient at performing accurate physical exams that are targeted to the patients complaints
	Does not recognize potentially life threatening presentations	Limited ability to seek secondary data (e.g. medical records)	Demonstrates ability to seek and obtain data from secondary sources (e.g. medical records) when needed	Proficient at synthesizing history and physical exam findings to generate a prioritized differential diagnosis and problem list
	Does not recognize patients central clinical problem	Limited ability to recognize patient's central clinical problem or develop limited differential diagnosis	Demonstrates ability to use history and exam findings to define patient's central clinical problems	Proficient at using history and physical exam effectively to minimize the need for further diagnostic

Overall Score	1	1.5	2	2.5	3	3.5	4

Figure 1: Health and Physical Domain of MJM. Grades from each category in the domain are scored from 1-4. Then each domain is given an overall score of 1-4 on a 0.5 interval scale. [16]

Aims 2 and 3

- The initial MJM pilot examined the use of the MJM to grade 40 subjects of four levels of medical experience/skill
  - Goal was to establish validity and reliability (amongst metric operators)

- Next the metric was explored as a tool to measure the decision making of NASA spaceflight chief medical officers (CMOs) against their non-medically trained technical professional peers (TPs). [23]

  - This study examined the decision making of administrative physicians (to represent CMOs) verses technical professionals (to represent non CMO spaceflight officers) across 4 simulated acute care scenarios
  - Saw a significant difference between CMO and TP groups in MJM scores

Design/Methods - Subjects

- Summa Health Institutional Review Board and NASA Johnson Space Center Committee for the Protection of Human Subjects were responsible for approving the study protocol

- Subjects were recruited across three physician groups to represent various levels of pedagogical transitions: administrative, mastery level, and resident physicians. (n=10 for each)

Design/Methods - Subjects

- Administrative physicians (AP) are considered the most experienced group but those whose current duties are largely administrative, performing a maximum of eight hours of clinical care per week
- mastery level physicians (MP) are attending physicians who have completed postgraduate education and whose responsibilities are entirely clinical. These physicians were board-certified or board-eligible in their specialty.
- Resident physicians (RP) were selected from those who are in the third or fourth year of their post-graduate education (PGY) and, thusly, are in their final stages of training.

Design/Methods - Simulations

- Each group of participants (AP, MP, and RP) performed one practice simulated scenario (acute Deep Vein Thrombosis) to reduce anxiety and introduce participants to the process.

- Then in random order they were assigned to perform 4 graded scenarios across 2 main chief complaints: Chest pain and abdominal pain.

Pacific Northwest University of Health Sciences

Design/Methods - Simulations

- Abdominal Pain scenarios were:
  - Biliary Colic (SIM_BC)
  - Renal Colic (SIM_RC)

- Chest Pain scenarios were:
  - cardiac ischemia with ST-segment elevation myocardial infarction (SIM_STEMI)
  - tension pneumothorax (SIM_PTX).

Design/Methods – Simulation Materials

- The simulations were carried out in a mature medical simulation lab in an American College of Surgeons verified Level I trauma center in the United States.

- The simulation labs utilized the METI ECS mannequin simulation technology with HPS6 software (Medical Education Technologies, Saint-Laurent, Quebec, Canada). All simulations were audio and video recorded in their entirety

Design/Methods – Simulation Scenario

- Participants were informed the scenarios would take place in a moderate-sized community hospital emergency department in the USA

- Participants were informed they had 15 minutes to obtain relevant history, perform physical examination, order diagnostic testing, medications, and make management decisions including performing life-saving procedure

- Each participant had access to standard medical equipment and a nurse to use the equipment at the command of the participant.

Design/Methods – Simulation Scenario

- The full body METI mannequin was capable of receiving any of the tests and manoeuvres as directed by the participant whereas verbal and/or visual feedback was provided by Virtual Care Simulation Lab staff.

- All laboratory values, radiographs, electrocardiographs and ultrasound images were provided without interpretation beyond reference laboratory values and in a scaled time-delay fashion.

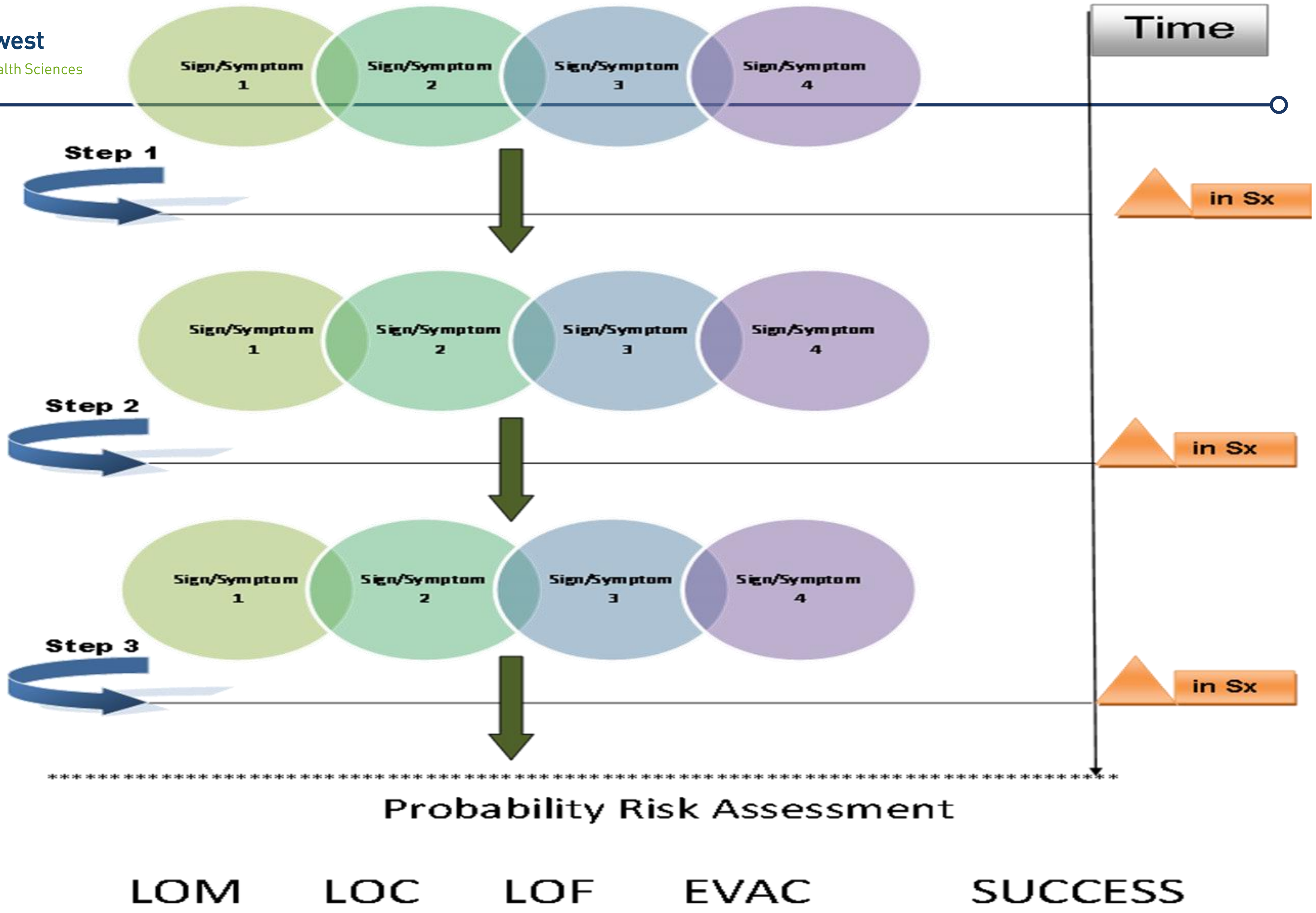
Design/Methods - Simulations

- Participant heart rates, NASA Task Load Score (NASA_TLX) and total sim time were recorded.

- Medical decision making in each of the scenarios was evaluated using 3 assessments:
- Scenario-specific critical action checklist (similar to that used in previous MJM experiments)
- Categorical determination of patient outcome (loss of function, loss of life, or stabilized)
- Medical Judgement Metric (MJM).

Pacific Northwest University of Health Sciences

Differential Diagnosis – Simulation



Design/Methods - Simulations

- Four reviewers were utilized to score participants using the MJM
- Each of the reviewers possessed post-graduate training in either emergency medicine, general surgery with surgical critical care fellowship training, and/or fellowship training in medical simulation in addition each was trained in how to use the MJM.
- The raters were blinded to both the subject's name and cohort.

Pacific Northwest University of Health Sciences

Results – Demographics!

N=30 participants (10 Residents, 10 Attendings,10 Admin)

Age

Table 1: Demographics

	Study Group		
	PGY level 3 or 4	Board Certified Physician	Administrative Physician
Variable/Statistic	(n=10)	(n=10)	(n=10)
Age (Years)			
Mean (SD)	28.6 (1.43)	42.3 (10.83)	59.6 (7.78)
Median (IQR)	28.5 (27.8 - 29)	38.5 (33.8 - 55)	61.5 (51.8 - 66)
Min - Max	27 - 32	28 - 57	49 - 70
Gender - n(%)			
Female	4 (40%)	4 (40%)	4 (40%)
Male	6 (60%)	6 (60%)	6 (60%)
Race - n (%)			
Asian	1 (10%)	0	1 (10%)
Asian/White	2 (20%)	0	0
Black	0	1 (10%)	0
White	7 (70%)	9 (90%)	9 (90%)
Education - n (%)			
Doctoral Degree	10 (100%)	10 (100%)	10 (100%)
Specialty - n (%)			
Emergency Medicine	4 (40%)	7 (70%)	3 (30%)
General Surgery	4 (40%)	1 (10%)	0
OB/GYN	2 (20%)	1 (10%)	1 (10%)
Radiology	0	1 (10%)	0
Family Medicine	0	0	1 (10%)
Internal and Pediatrics	0	0	1 (10%)
Internal Medicine	0	0	1 (10%)
Internal Medicine and Infectious Diseases	0	0	1 (10%)
OB/Gyn and Preventative	0	0	1 (10%)
Pathology	0	0	1 (10%)

Results

NASA TLX total score

Title	Endpoints	Descriptions
MENTAL DEMAND	Low/High	How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?
PHYSICAL DEMAND	Low/High	How much physical activity was required (e.g., pushing, pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?
TEMPORAL DEMAND	Low/High	How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?
PERFORMANCE	good/poor	How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?
EFFORT	Low/High	How hard did you have to work (mentally and physically) to accomplish your level of performance?
FRUSTRATION LEVEL	Low/High	How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?

Results – Physiologic data

Column1	Column2	Column3	Column4	Column5
Table 2: Physiological Stress Metrics				
	Study Group			
	PGY level 3 or 4	Board Certified Physician	Administrative Physician	Between Groups
Simulation Number/Metric/Statistic	(n=10)	(n=10)	(n=10)	P-value
Between Group Effect				
NASA TXL Score				0.161
Resting Heart Rate				0.907
Maximum Heart Rate				0.283
Total Sim Time				0.204
1				
NASA TXL Score Mean (SD)	43.7 (16.93)	37.6 (13.91)	54.4 (21.59)	
Resting Heart Rate Mean (SD)	77.5 (7.86)	76.0 (14.78)	75.3 (13.73)	
Maximum Heart Rate Mean (SD)	99.4 (12.79)	88.3 (14.48)	93.2 (15.50)	
Total Sim Time Mean (SD)	6.2 (1.58)	8.7 (2.46)	10.3 (2.78)	0.002
2				
NASA TXL Score Mean (SD)	53.6 (25.76)	38.3 (20.23)	52.2 (31.11)	
Resting Heart Rate Mean (SD)	77.9 (8.05)	76.5 (9.85)	77.1 (12.09)	
Maximum Heart Rate Mean (SD)	95.7 (9.26)	87.2 (14.49)	90.1 (15.39)	
Total Sim Time Mean (SD)	8.6 (2.78)	7.6 (1.94)	9.2 (2.65)	0.339
3				
NASA TXL Score Mean (SD)	47.2 (25.71)	43.2 (25.57)	54.3 (26.04)	
Resting Heart Rate Mean (SD)	77.9 (9.97)	74.1 (11.13)	76.2 (12.40)	
Maximum Heart Rate Mean (SD)	94.6 (12.42)	87.2 (11.78)	90.4 (15.41)	
Total Sim Time Mean (SD)	7.9 (2.93)	7.9 (1.40)	8.0 (2.97)	0.994
4				
NASA TXL Score Mean (SD)	42.1 (19.43)	28.6 (11.44)	50.5 (30.04)	
Resting Heart Rate Mean (SD)	75.3 (9.29)	73.6 (13.10)	75.3 (13.01)	
Maximum Heart Rate Mean (SD)	93.8 (8.09)	84.3 (13.98)	87.9 (13.54)	
Total Sim Time Mean (SD)	7.3 (2.13)	7.2 (1.74)	8.0 (2.36)	0.626
P-values from between group factor of repeated measures ANOVA.				

Note: For total sim time a significant (p=0.033) time*group interaction term was present; hence the between group comparisons at each time period.

P-value for total sim time from single factor ANOVA with post-hoc Tukey HSD test. PGY 3 or 4 differed significantly at sim 1 from administrative group.

Pacific Northwest University of Health Sciences

Results – MJM scores

Column1	Column2	Column3	Column4	Column5
Table 3: Analysis of MJPM Scores				
	Study Group			
	PGY level 3 or 4	Board Certified Physician	Administrative Physician	
Simulation/Statistic	(n=10)	(n=10)	(n=10)	P-value
Renal Colic				0.010
Mean (SD)	14.7 (1.15)	14.7 (1.48)	12.3 (2.66)	
Median (IQR)	14.8 (14.2 - 15.7)	15.4 (13.5 - 15.7)	12.4 (10.6 - 14.5)	
Min - Max	12.2 - 15.8	12 - 16	8.2 - 16	
Biliary Colic				0.269
Mean (SD)	14.5 (1.16)	14.2 (2.68)	13.0 (2.25)	
Median (IQR)	14.8 (13.5 - 15.2)	15.5 (12.6 - 15.9)	13.1 (11.5 - 14.9)	
Min - Max	12.5 - 16	7.7 - 16	8.8 - 16	
STEMI				0.175
Mean (SD)	14.2 (1.93)	14.1 (2.71)	12.1 (3.33)	
Median (IQR)	14.6 (12.7 - 16)	15.7 (10.5 - 16)	11.8 (9.5 - 15.4)	
Min - Max	11 - 16	9.5 - 16	7 - 16	
TP				0.336
Mean (SD)	14.4 (1.75)	14.8 (1.12)	13.5 (2.53)	
Median (IQR)	15.3 (13.2 - 15.9)	14.8 (14 - 15.8)	14.3 (11.5 - 15.7)	
Min - Max	11 - 16	12.7 - 16	9 - 15.8	

P-values from single factor ANOVA F test with post hoc Tukey HSD test.

For the renal colic simulation the administrative study group was statistically distinct ($p<0.001$) in their MJPM scores relative to the other two study groups which were statistically indistinguishable.

Pacific Northwest University of Health Sciences

Results - outcomes

Column1	Column2	Column3	Column4	Column5
Table 4: Analysis of Outcomes				
	Study Group			
	PGY level 3 or 4	Board Certified Physician	Administrative Physician	
Simulation/Outcome	(n=10)	(n=10)	(n=10)	P-value
Renal Colic				0.040
Loss of Function	0	1 (10%)	3 (30%)	
Loss of Life	0	0	2 (20%)	
Stabilized	10 (100%)	9 (90%)	5 (50%)	
Biliary Colic				0.457
Loss of Function	2 (20%)	1 (10%)	3 (30%)	
Loss of Life	0	1 (10%)	2 (20%)	
Stabilized	8 (80%)	8 (80%)	5 (50%)	
STEMI				0.610
Loss of Function	1 (10%)	0	2 (20%)	
Loss of Life	4 (40%)	3 (30%)	4 (40%)	
Stabilized	5 (50%)	7 (70%)	4 (40%)	
TP				0.877
Loss of Function	0	1 (10%)	2 (20%)	
Loss of Life	2 (20%)	1 (10%)	1 (10%)	
Stabilized	8 (80%)	8 (80%)	7 (70%)	
P-values from Fisher's exact tests with post-hoc Bonferroni adjusted z tests.				
Bolded cells indicate cells with significant Bonferroni adjusted z tests.				

Conclusions

Analysis of the results indicate that significant differences in clinical judgment via simulated patient outcomes exist between physician groups whose responsibilities remain largely clinical as attending physicians versus those physicians that take on largely administrative duties, but not for all scenarios tested.

Drawbacks

Pilot study

Future Aims

- Is the MJM generalizeable?

- Future work will need to focus on the use of the MJM to assess competencies amongst a larger cohort of peers, particularly within the same level of training and specialty, before its utility can begin to be examined outside of the simulation lab.

Pacific Northwest University of Health Sciences

Questions?

John Rosasco (Jrosasco@pnwu.edu)

Michele McCarroll (Mmccarroll@pnwu.edu)

Pacific Northwest University of Health Sciences

References

1. Lyss-Lerman P, Teherani A, Aagaard E, Loeser H, Cooke M, Harper GM. What training is needed in the fourth year of medical school? Views of residency program directors. Acad Med. 2009; 84(7):823–9.
2. Frischknecht AC, Boehler ML, Schwind CJ, et al. How prepared are your interns to take calls? Results of a multi-institutional study of simulated pages to prepare medical students for surgery internship. Am J Surg. 2014;208(2):307–15.
3. Chamberland M, Mamede S, St-Onge C, Setrakian J, Schmidt HG. Does medical students' diagnostic performance improve by observing examples of self-explanation provided by peers or experts? Adv Health Sci Educ Theory Pract. 2015;20(4): 981–93.
4. Sox, Harold C. (2013) Medical decision making. Chichester, West Sussex, UK John Wiley & Sons.
5. Roter D. The patient-physician relationship and its implications for malpractice litigation. J Health Care L & Pol'y. 2006;9:304–14.
6. Cohen JJ. Medical education in an era of health-care reform. Perspect Biol Med. 2011;54(1):61-7.
7. Croskerry P, Petrie D, Reilly J, Tait G. Deciding about fast and slow decisions. Acad Med. 2014;89(2):197–200.
8. American College of Graduate Medical Education (ACGME) Accessed December 9, 2018. Available at: http://www.accme.org/accreditation-rules/policies/cme-content-definition-and-examples

References

9. Schwartz A. Medical decision making and medical education: challenges and opportunities. Perspect Biol Med. 2011;54(1):68-74.
10. Okuda Y, Bryson EO, Demaria S, et al. The utility of simulation in medical education: what is the evidence?. Mt Sinai J Med. 2009;76(4):330-43.
11. Onyon C. Problem-based learning: a review of the educational and psychological theory. Clin Teach. 2012;9(1):22-6.
12. Mclean SF. Case-Based Learning and its Application in Medical and Health-Care Fields: A Review of Worldwide Literature. J Med Educ Curric Dev. 2016;3.
13. Morrow CE, Reed VA, Eliassen MS, Imset I. Shared decision making: skill acquisition for year III medical students. Fam Med. 2011;43(10):721-5.
14. Murray DJ, Freeman BD, Boulet JR, Woodhouse J, Fehr JJ, Klingensmith ME. Decision making in trauma settings: simulation to improve diagnostic skills. Simul Healthc. 2015;10(3):139– 4.
15. Lateef F. (2010). Simulation-based learning: Just like the real thing. Journal of emergencies, trauma, and shock, 3(4), 348–352. doi:10.4103/0974-2700.70743

Pacific Northwest University of Health Sciences

References

16. “Exploration Medical Condition List”, (2013), JSC-65722.

17. Ahmed, R. A., McCarroll, M. L., Schwartz, A., Gothard, M. D., Atkinson, S. S., Hughes, P. G., George, R. L. (2017). Development, Validation, and Implementation of a Medical Judgment Metric. MDM Policy & Practice.

18. Weber JM, Kopelman S, Messick DM. A conceptual review of decision making in social dilemmas: applying a logic of appropriateness. Pers Soc Psychol Rev. 2004;8(3):281–307.

19. Mishra A, Catchpole K, McCulloch P. The Oxford NOTECHS System: reliability and validity of a tool for measuring teamwork behavior in the operating theatre. Qual Saf Health Care. 2009; 18(2):104–8.

20. Steinemann S, Berg B, DiTullio A, et al. Assessing teamwork in the trauma bay: introduction of a modified “NOTECHS” scale for trauma. Am J Surg. 2012;203(1):69–75.

21. Andolsek K, Padmore J, Hauer K, Holmboe E. Clinical Competency Committees. A Guidebook for Programs. Available at: https://www.acgme.org/acgmeweb/Portals/0/ACGMEClinical CompetencyCommitteeGuidebook.pdf.

22. Association of American Colleges of Medicine. Situational judgment. Available at: https://www.aamc.org/admissions/admi ssionslifecycle/409100/situationaljudgmenttest.html.

23. McCarroll ML, Ahmed RA, Schwartz A, et al. Medical judgement analogue studies with applications to spaceflight crew medical officer. BMJ Simul Technol Enhanc Learn. 2017;3(4):163-168.