OPEN ACCESS

Medical judgement analogue studies with applications to spaceflight crew medical officer

Michele L McCarroll,[1] Rami A Ahmed,[2,3] Alan Schwartz,[4] Michael David Gothard,[5] Steven Scott Atkinson,[3] Patrick Hughes,[2] Jose Cepeda Brito,[2] Lori Assad,[2] Jerry Myers,[6] Richard L George[2,3]

[1]Clinical Medicine, Pacific Northwest University of Health Sciences, Yakima, Washington, USA
[2]Medical Education, Summa Health System, Akron, Ohio, USA
[3]Northeastern Ohio Medical University, Rootstown, Ohio, USA
[4]University of Illinois at Chicago, Chicago, Illinois, USA
[5]Biostats, Inc, East Canton, Ohio, USA
[6]NASA Glenn Research Center, Cleveland, Ohio, USA

Correspondence to
Dr Michele L McCarroll, College of Osteopathic Medicine, Pacific Northwest University of Health Sciences, 111 University Parkway, Suite 202, Yakima, WA 98901, USA; mmccarroll@pnwu.edu, mccarroll314@gmail.com

ABSTRACT

Background The National Aeronautics and Space Administration (NASA) developed plans for potential emergency conditions from the Exploration Medical Conditions List. In an effort to mitigate conditions on the Exploration Medical Conditions List, NASA implemented a crew medical officer (CMO) designation for eligible astronauts. This pilot study aims to add knowledge that could be used in the Integrated Medical Model.

Methods An analogue population was recruited for two categories: administrative physicians (AP) representing the physician CMOs and technical professionals (TP) representing the non-physician CMOs. Participants completed four medical simulations focused on abdominal pain: cholecystitis (CH) and renal colic (RC) and chest pain: cardiac ischaemia (STEMI; ST-segment elevation myocardial infarction) and pneumothorax (PX). The Medical Judgment Metric (MJM) was used to evaluate medical decision making.

Results There were no significant differences between the AP and TP groups in age, gender, race, ethnicity, education and baseline heart rate. Significant differences were noted in MJM average rater scores in AP versus TP in CH: 13.0 ($\pm$2.25), 4.5 ($\pm$0.48), p=<0.001; RC: 12.3 ($\pm$2.66), 4.8 ($\pm$0.94); STEMI: 12.1 ($\pm$3.33), 4.9 ($\pm$0.56); and PX: 13.5 ($\pm$2.53), 5.3 ($\pm$1.01), respectively.

Discussion There could be a positive effect on crew health risk by having a physician CMO. The MJM demonstrated the ability to quantify medical judgement between the two analogue groups of spaceflight CMOs. Future studies should incorporate the MJM in a larger analogue population study to assess the medical risk for spaceflight crewmembers.

INTRODUCTION

At the dawn of long-duration missions to Mars, the National Aeronautics and Space Administration (NASA) developed plans for more extensive use of high-fidelity medical simulation to prepare astronauts for potential emergency conditions from the Exploration Medical Conditions List.[1] NASA's participation in nearly 10 years of long-duration missions confirms that it is difficult to assess whether a loss of life, loss of mission or loss of function condition will occur on long-duration missions.

Applicant backgrounds to become an astronaut at NASA range from a breadth of scientific disciplines of engineering, biological science, physical science or mathematics.[2] In fact, the majority of International Space Station astronauts are from non-medical professions.[3][4] In an effort to mitigate conditions on the Exploration Medical Conditions List and maximise the astronaut's health throughout all phases of long-duration missions, NASA implemented a crew medical officer (CMO) designation for eligible astronauts.[5] The CMO astronaut training occurs during the 2-year period leading up to the actual space mission.[6] In order to serve as a CMO, future crew members receive 40–70 hours of medical training within 18 months before missions.[5]

Currently, NASA-STD-3001 mandates that for missions longer than 210 days, the CMO must be a physician astronaut—a requirement based on subject matters experts' opinion with the assumption that a quick return to earth is unlikely for missions of this duration.[7] Additionally, the subject matter experts determined that the medical decision-making ability or medical judgement associated with physician CMOs would be superior over non-physician CMOs resulting in mission-related medical risk reductions. Kienle and Kiene studied several approaches to assess causality associated with medical judgement and, while they concluded that it was important, did not provide any outcome or risk-based analysis.[8] Thus, the rationale for a physician CMO over a non-physician CMO for long-duration missions lasting 210 days or longer has not been quantified as an acceptable mitigation approach to reducing or recovering from diagnoses on the Exploration Medical Conditions List. In a similar fashion, the effect on mission risk of not requiring a physician CMO for shorter missions has also not been quantified.[9] In addition, medical decision making has not been determined to be better from a physician CMO over a non-physician CMO in the confines of space, especially since most physician CMOs are no longer involved in day-to-day medical practice.[10] It is possible that medical decision-making ability decay in a physician CMO could threaten long-duration missions.

Medical decision making (judgement) is not easily assessed.[11][12] Moreover, there remains significant variability in medical decision making by specialty and the amount of training a clinician has received.[13][14] With this in mind, NASA investigated the impact of incorporating medical simulation to analyse the impact on crew autonomy of a physician CMO as part of the team.[15] Unfortunately, the study did not quantitatively measure the effect on crew health risk by having a physician CMO or the physician CMO clinical judgement. Hence, this pilot study aims to add knowledge to the Exploration

Medical Systems element of NASA's Human Research Program to develop data that could be used in the Integrated Medical Model to quantify the change in risk posture associated with a physician CMO versus a non-physician CMO.

METHODS

Subjects

The Summa Health Institutional Review Board and the NASA Johnson Space Center Committee for the Protection of Human Subjects approved the study protocol. This study was a non-equivalent group's trial of a NASA astronaut physician and non-physician analogue population. The analogue population sample was recruited based on the current NASA astronaut pool demographic data of gender: 60% male, 40% female; age (SD): 48.04 (±5.21) years; and background: engineering, biological science, physical science or mathematics (4). The analogue cohorts were stratified into administrative physicians (AP) representing the physician CMOs (n=10, 6 men and 4 women) and technical professionals (TP) representing the non-physician CMO group (n=10, 6 men and 4 women). For the AP analogue group, the study team recruited board certified physicians; age 40–70 years (NASA astronaut age average ±3 STD); practising as administrators for more than 2 years; and performing clinical duties for less than 8 hours/week. For the TP analogue group, the study team recruited masters or doctoral prepared scientists in engineering, biological science, physical science or mathematics; aged (NASA astronaut age average ±1 STD); and with no medical training or experience.

Materials

Upon completion of the informed consent, each participant completed a demographic questionnaire to collect background information on education (for inclusion/exclusion criteria), physical activity,[16] musical instrument (musicianship),[17] video game,[18] aviation, managerial, simulation and boy/girl scout experience.[19] Each subject also completed standardised surveys consisting of General Self–Efficacy,[20] Big Five Inventory,[21] PROMIS Global Health Scale (QOL),[22] the Teamwork and Safety Climate Survey (SAQ)[23] and the NASA Task Load Index (NASA TLX).[24] The NASA TLX was completed after each medical simulation to assess whether workload was different between the four medical simulations.

A high-fidelity adult simulator was used for all simulations (METIman ECS). Each bay was stocked with a crash cart, LIFEPAK 12 defibrillator, standard hospital airway equipment, as well as a trauma cart with standard equipment for a US hospital including thoracostomy supplies, vascular access equipment including an EZ-IO intraosseous power driver and needles. The patient bay also had a simulated patient monitor with the ability to display the patients' vital signs, diagnostic imaging and ECGs.

Procedures

Participants completed one practice and four scored medical simulations focused on conditions from the Exploration Medical Conditions List.[1] The practice simulation was a patient with a diagnosis of deep vein thrombosis (a common and non-esoteric clinical situation) to reduce participant anxiety regarding study procedures, expose all participants to medical simulation regardless of background and incorporate all of the aspects of clinical assessment and care. The four scored simulations were abdominal pain: cholecystitis (CH) and renal colic (RC) and chest pain: cardiac ischaemia (STEMI; ST-segment elevation myocardial infarction) and pneumothorax (PX). Each subject was tested on CH, RC, STEMI and PX in a randomised order by using the random numbers generator in Microsoft Excel V.2007 (Redmond, Washington, USA). All medical simulations were conducted out of the Virtual Care Simulation Lab located at an American College of Surgeons verified level I trauma institution in the USA. The Virtual Care Simulation Lab used the simulator METI ECS with HPS6 software (Medical Education Technologies, Saint-Laurent, Quebec, Canada) for all simulations.

In each tested medical simulation, participants were informed that they were in a moderate-sized community hospital emergency department in the USA and were asked to care for the patient as best as possible by identifying the necessary screening, testing, treatment and diagnosis. Each participant had access to standard medical equipment and a nurse to use the equipment at the command of the participant. 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.

The participants had a maximum time limit of 15 min for each scenario to perform a history, physical examination, order diagnostic testing, medications, and make management decisions including performing life-saving procedures. The 15 min time limit for each scenario was felt to be an adequate amount of time and any additional time was unlikely to lead any change in results/management. The myocardial infarction case was designed to proceed to cardiac arrest after 8 min if no critical interventions were performed. The CH case and RC/pyelonephritis cases were designed to similarly have a change in mental status at about 8–10-min if key interventions were not performed reflecting a septic picture. For example, in the PX case, the participant must quickly evaluate a patient with a history of chest trauma and sudden shortness of breath. They must either confirm clinically (absent breath sounds, increasing pulse, decreasing blood pressure, decreasing pulse oximetry) that the patient has a tension PX, or quickly order a chest X-ray demonstrating an obvious tension PX with midline shift and immediately perform a needle decompression of the affected lung or place a chest tube (tube thoracostomy) in the affected lung. If this manoeuvre was not executed within 6 min of the start of the encounter the patient would progress to pulselessness (pulseless electrical activity) and require cardiopulmonary resuscitation.

After each simulation, the participant had a 5 min rest. Each participant wore a heart rate monitor (Polar Heart Rate Monitor; Polar Electro, Lake Success, New York) where baseline and maximum heart rates during each medical simulation were recorded. During the rest period, the participant filled out a NASA TLX survey and another resting heart rate was obtained prior to commencing the next simulation. Participants were audio and video recorded during the practice and scored medical simulations.

Statistical analysis

The Medical Judgment Metric (MJM), as well as a simulation scenario-specific critical action checklist, and a categorical determination of the patient's final outcome (stabilised, loss of function or loss of life) were used to evaluate medical decision making (judgement). The MJM is a tool that measures medical judgement in four clinical domains: history and physical, diagnostic, interpretation, and management, with a maximum score

of 4 in each domain on a 0.5 interval scale up to an overall maximum score of 16. [25]

All four medical raters had backgrounds in emergency, trauma and medical simulation and were trained in using the MJM. The medical raters independently scored participants using the MJM either live or using the video recording of the medical simulation. A minimum of three raters were required for each medical simulation. The simulation medical director also performed three pilot simulations prior to execution of the first simulation using the MJM to provide feedback to raters' scores on excellent, average and below average performance in an attempt to have greater inter-rate reliability in scoring. The raters were blinded to the subject's name and participant cohort. Demographics and baseline survey measures of the two cohorts, as well as each simulation's MJM, outcome determinations, maximum heart rate, and NASA TLX score were compared using non-parametric tests (Fisher's exact test for categorical variables or the Mann-Whitney U test for continuous variables). All statistical analyses were completed using STATA V.14, SPSS V.23.0 and Microsoft Excel V.2007.

RESULTS

There were no significant differences between the AP (n=10, degrees of freedom (*df*), n-1) and TP (n=10, *df*, n-1) groups in demographics of age: 59.6 years (±7.78) vs 52.3 years (±3.71) (p=0.143); gender: female n=4 (40%) and male n=6 (60%) (p=1.00); race: White n=9 (90%), Asian n=1 (10%) vs White n=10 (100%) (p=1.00); ethnicity: not Hispanic or Latino n=10 (100%) vs n=10 (100%) (p=1.00); education: doctoral degree n=10 (100%), master's degree n=0 (0%) vs master's degree n=2 (20%), doctoral degree n=8 (80%) (p=0.474); and heart rate: 81.8 bpm (±12.11) vs 81.1 bpm (±8.41) (p=0.912), respectively (table 1). The AP analogue group age 59.6 (±7.78)

Table 1 Demographics

Variable/Statistic	AP group (n=10)	TP group (n=10)	p Value
Age (years)			0.143
Mean (SD)	59.6 (7.78)	52.3 (3.71)	
Median	61.5	53.5	
Min-Max	49–70	45–56	
Gender, n (%)			1.000
Female	4 (40%)	4 (40%)	
Male	6 (60%)	6 (60%)	
Race, n (%)			1.000
Asian	1 (10%)	0 (0%)	
Asian and White	0 (0%)	0 (0%)	
Black	0 (0%)	0 (0%)	
White	9 (90%)	10 (100%)	
Ethnicity, n (%)			1.000
Not Hispanic or Latino	10 (100%)	10 (100%)	
Education, n (%)			0.474
Master's degree	0 (0%)	2 (20%)	
Doctoral degree	10 (100%)	8 (80%)	
Baseline heart rate (bpm)			0.912
Mean (SD)	81.8 (12.11)	81.1 (8.41)	
Median	79.5	80	
Min-Max	68–104	69–100	

p Value from Fisher's exact test for categorical data or Mann-Whitney U test for numerical data.
AP, administrative physician; Max, maximum; Min, minimum; TP, technical professional.

Table 2 Biliary colic simulation results

Variable/Statistic	AP group (n=10)	TP group (n=10)	p Value
Resting heart rate (bpm)			0.739
Mean (SD)	75.3 (13.73)	73.1 (11.34)	
Median	76.5	74	
Min-Max	53–100	48–88	
Max heart rate (bpm)			1.000
Mean (SD)	93.2 (15.50)	93.2 (14.75)	
Median	97.5	90.5	
Min-Max	65–114	66–115	
NASA TLX score			0.002*
Mean (SD)	54.4 (21.59)	83.8 (16.81)	
Median	52	81	
Min-Max	22–102	58–111	
Total simulation time (min)			0.315
Mean (SD)	10.3 (2.78)	8.9 (2.41)	
Median	10.7	8.8	
Min-Max	6.6–14.1	5.3–12.6	

p Value from Fisher's exact test for categorical data or Mann-Whitney U test for numerical data.
*Denotes significance at p<0.05.
AP, administrative physician; Max, maximum; Min, minimum; NASA TLX, NASA Task Load Index; TP, technical professional.

years was higher, though not significant, than the TP analogue group age 52.3 (±3.71) years due to recruitment criteria for the AP analogue cohort. There were no other significant differences in other baseline characteristics of interest. The only baseline characteristic with a significant difference (p=0.043) was the experience in medical simulations (4/10 experienced in the AP group vs 0/10 in the TP group). For baseline surveys and questionnaires, no significant differences were found. However, the AP group had a significantly higher SAQ scores (773.7±47.6) than the TP group (7.4±12.4, p=0.001).

In each of the four medical simulations, the AP analogue group reported significantly lower (better) NASA TLX ratings (CH: AP=54.4 (±21.59) vs TP 83.8 (±16.81), p=0.002 (table 2); RC: AP=52.2 (±31.11) vs TP 89.7 (±9.52), p=0.015 (table 3); STEMI: AP=54.3 (±26.04) vs TP 91.8 (±16.32), p=0.01 (table 4); PX: AP=50.5 (±30.04) vs TP 86.5 (±19.67), p=0.011 (table 5)) than the TP analogue group in each simulation. No other significant differences were found.

The AP analogue group outperformed the TP analogue group in the correct categorical determination (stabilised, loss of function or loss of life) of the simulated patient's final outcome (table 6). The AP analogue group had significantly higher (better) MJM scores compared with the TP analogue group for each simulation scenario. In every simulation in the TP group, except for one, each categorical determination by the raters was a loss of life of the simulated patient regardless of any intervention attempted by the TP participant. The PX simulation was the one exception where a TP participant prevented a loss of life in the patient and left the patient with a loss of function. In contrast, AP participants prevented loss of life in 21 of the 40 simulated cases, loss of function in 10 of 40 simulated cases, and loss of life in 9 of the 40 simulated cases.

DISCUSSION

The main result of the study indicates that there exist significant differences in medical judgement and simulation performance

Original research

Table 3 Renal colic simulation results

Variable/Statistic	AP group (n=10)	TP group (n=10)	p Value
Resting heart rate (bpm)			0.393
Mean (SD)	77.1 (12.09)	73.1 (10.47)	
Median	79.5	72	
Min-Max	56–96	54–89	
Max heart rate (bpm)			0.481
Mean (SD)	90.1 (15.39)	88.2 (10.36)	
Median	94.5	89	
Min-Max	59–106	67–101	
NASA TLX score			0.015*
Mean (SD)	52.2 (31.11)	89.7 (9.52)	
Median	46	89.5	
Min-Max	14–108	76–106	
Total simulation time (min)			0.481
Mean (SD)	9.2 (2.65)	8.8 (1.14)	
Median	9.4	8.9	
Min-Max	5.3–14.3	7.2–11.4	

p Value from Fisher's exact test for categorical data or Mann-Whitney U test for numerical data.
*Denotes significance at p<0.05.
AP, administrative physician; Max, maximum; Min, minimum; NASA TLX, NASA Task Load Index; TP, technical professional.

Table 5 Pneumothorax simulation results

Variable/Statistic	AP group (n=10)	TP group (n=10)	p Value
Resting heart rate (bpm)			0.579
Mean (SD)	75.3 (13.01)	73.4 (9.79)	
Median	80	70	
Min-Max	53–90	57–87	
Max heart rate (bpm)			0.579
Mean (SD)	87.9 (13.54)	85.6 (9.29)	
Median	90	83.5	
Min-Max	62–105	71–101	
NASA TLX score			0.011*
Mean (SD)	50.5 (30.04)	86.5 (19.67)	
Median	40.5	87.5	
Min-Max	15–91	60–118	
Total simulation time (min)			0.796
Mean (SD)	8.0 (2.36)	8.0 (1.57)	
Median	7.8	8.3	
Min-Max	4.5–12.2	4.3–9.8	

p Value from Fisher's exact test for categorical data or Mann-Whitney U test for numerical data.
*Denotes significance at p<0.05. AP, administrative physician; Max, maximum; Min, minimum; NASA TLX, NASA Task Load Index; TP, technical professional.

outcomes in spaceflight crew analogue groups of non-physician CMOs versus physician CMOs. This study compared two analogue groups, demonstrating an effect on crew health risk by having the medical background of the AP group as a physician CMO over the TP group. The ability to quantify medical judgement using the MJM and its impact (stabilised, loss of function or loss of life) in the analogue samples using medical simulation added missing knowledge to the existing Exploration Medical Systems element of NASA's Human Research Program.

NASA's Human Research Program and the Integrated Medical Model are focused on various components of crew health and how the qualifications of the CMO may impact crew health. While medical judgement is one aspect of assessing a CMO's impact, it may not be the only factor. Wang and Wu demonstrated that leadership qualities and crew

Table 4 STEMI simulation results

Variable/Statistic	AP group (n=10)	TP group (n=10)	p Value
Resting heart rate (bpm)			0.631
Mean (SD)	76.2 (12.40)	74.3 (11.11)	
Median	77	72.5	
Min-Max	58–94	56–88	
Max heart rate (bpm)			0.190
Mean (SD)	90.4 (15.41)	84.3 (10.68)	
Median	96.5	84	
Min-Max	61–109	61–98	
NASA TLX score			0.001*
Mean (SD)	54.3 (26.04)	91.8 (16.32)	
Median	53.5	99	
Min-Max	15–91	61–108	
Total simulation time (min)			0.631
Mean (SD)	8.0 (2.97)	8.4 (1.79)	
Median	7.1	8.4	
Min-Max	4.9–12.4	5.2–11.9	

p Value from Fisher's exact test for categorical data or Mann-Whitney U test for numerical data.
*Denotes significance at p<0.05.
AP, administrative physician; Max, maximum; Min, minimum; NASA TLX, NASA Task Load Index; STEMI, ST-segment elevation myocardial infarction; TP, technical professional.

Table 6 Medical Judgment Pathway Metric and outcomes

Variable/Statistic	AP group (n=10)	TP group (n=10)	p Value
Biliary colic outcome			0.001*
Stabilised	5 (50%)	0 (0%)	
Loss of function	3 (30%)	0 (0%)	
Loss of life	2 (20%)	10 (100%)	
MJM rater mean (SD)	13.0 (2.25)	4.5 (0.48)	<0.001*
Renal colic outcome			0.001*
Stabilised	5 (50%)	0 (0%)	
Loss of function	3 (30%)	0 (0%)	
Loss of life	2 (20%)	10 (100%)	
MJM rater mean (SD)	12.3 (2.66)	4.8 (0.94)	<0.001*
STEMI outcome			0.011*
Stabilised	4 (40%)	0 (0%)	
Loss of function	2 (20%)	0 (0%)	
Loss of life	4 (40%)	10 (100%)	
MJM rater mean (SD)	12.1 (3.33)	4.9 (0.56)	<0.001*
Pneumothorax outcome			<0.001*
Stabilised	7 (70%)	0 (0%)	
Loss of function	2 (20%)	1 (10%)	
Loss of life	1 (10%)	9 (90%)	
MJM rater mean (SD)	13.5 (2.53)	5.3 (1.01)	<0.001*

p Value from Fisher's exact test for categorical data or Mann-Whitney U test for numerical data.
*Denotes significance at p<0.05.
AP, administrative physician; MJM, Medical Judgment Metric; STEMI, ST-segment elevation myocardial infarction; TP, technical professional.

support were positively correlated with a crew's cohesion, expressiveness and involvement which impact crew psychosocial precepts.[26] The CMO leadership trait is also supported by Musson *et al* where simulated medical support operations in an earth-based analogue environment demonstrated that crew autonomy, crewmember status and crew communication were parallel to crew health.[27] Moreover, Kanas reported that decreasing leader support, tension and negative emotions impacted crew health, especially if the crew were from diverse cultures.[28] Thus, medical judgement scores using the MJM, in addition to interpersonal leadership styles and crew responses to stressors, will need careful interpretation, application and consideration when applying them to the IMM for long-duration missions. In addition to the above qualities, the ability to complete a complex task under pressure might impact the MJM scores even though we found no differences in physical measures (baseline heart rate, maximum heart rate). The physical measures may not be different due to levels of stress being similar in both groups where the AP group felt stress performing procedures that they have not completed in an extended period of time and the TP group completing procedures they have never completed.

As a pilot analogue study, there are several limitations in applying the results to spaceflight personnel during long-duration missions and in medical personnel in medical training. Perhaps the most glaring limitation is that the participants did not have the current 40–70 hours of NASA CMO training. Even with a practice simulation, the non-physician CMO analogue group exhibited nervous behaviour prior to medical simulation testing in anticipation of not knowing what to expect. One quote from a participant in the TP group after the medical simulation was, 'you sure know how to make a smart person feel dumb.' During the breaks between each medical simulation, the TP participant's disposition, anxiety and emotional vulnerability were observed while ruminating about whether they had the correct medical diagnosis, medications and courses of action to help the simulated patient. The results of the NASA TLX scores appear to reflect the distress in the TP group as they were significantly higher than the AP group. Additionally, the physician CMO analogue group demonstrated similar distress when faced with a medical scenario outside their practice specialty. Another point to consider is that there was no attempt to communicate with an analogue mission control flight surgeon or medical consultant as the focus of the study was to implement a new instrument measuring an individual's medical judgement without an additional support. One of the most glaring limitations of this first pilot study is the sample size of the two analogue groups. Since there were only 10 in each group, it is difficult to draw strong conclusions and further testing is necessary to obtain a level of inference. Lastly, the medical simulations did not assess the ability of the two analogue CMOs to effect treatment (such as a minor invasive procedure) and this may further describe the differences between the two analogue groups.

The study demonstrated that when placed in an unfamiliar situation outside of one's expertise, or profession entirely, these participants were able to, or unable to, make the correct medical decisions in a potentially life-threatening situation. While this pilot study was not powered to undertake a subset analysis of performance by specialty, attending physicians in the field of emergency medicine and general surgery displayed the most composure and the highest MJM scores during these Exploration Medical Conditions List specific scenarios.

Despite the rich history of medical education and training, measuring and quantifying competency in medical decision making presents significant challenges. However, the study demonstrated the ability of the Medical Judgment Pathway Metric to quantify the difference between untrained AP group and the untrained non-physician analogue groups. To truly assess whether the MJM is a robust enough tool for inserting the scores into probability analyses determining a change in risk posture, we first need to use the MJM in a larger trained analogue population of CMOs. Also, current CMO training schedules and medical simulations should incorporate the MJM to assess whether there are parallels to clinical skills and competencies. These future studies should also include a variety of participant backgrounds to improve predictability and applicability of reducing risk on long-duration missions.

REFERENCES

1 National Aeronautics and Space Administration (NASA). Space medicine Exploration Medical Condition list. https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20110008645.pdf.

2 National Aeronautics and Space Administration (NASA)http://nodis3.gsfc.nasa.gov/displayCA.cfm?Internal_ID=N_PR_3300_001B_&page_name=Chapter2

3 Hurst VW, Whittam SW, Austin PN, *et al*. Cardiopulmonary resuscitation during spaceflight: examining the role of timing devices. *Aviat Space Environ Med* 2011;82:810–3.

4 Parazynski SE. From model rockets to spacewalks: an astronaut physician's journey and the science of the United States' space program. *Trans Am Clin Climatol Assoc* 2006;117:273–81.

5 Flynn CF. An operational approach to long-duration mission behavioral health and performance factors. *Aviat Space Environ Med* 2005;76:B42–B51.

6 National Aeronautics and Space Administration (NASA) Human Research Program. *Human Research Roadmap. [Internet]* 2016 https://humanresearchroadmap.nasa.gov/gaps/gap.aspx?i=387.

7 Williams RS. Nasa space flight human-system standard volume 1, revision a: crew health. 2014 http://standards.globalspec.com/std/9858810/nasa-std-3001-vol-1.

8 Kienle GS, Kiene H. Clinical judgement and the medical profession. *J Eval Clin Pract* 2011;17:621–7.

9   Blue RS, Bridge LM, Chough NG, et al. Identification of Medical Training methods for exploration missions. 2014 https://humanresearchroadmap.nasa.gov/gaps/closureDocumentation/Blue-R_TM-2014-217384.pdf?rnd=0.242290917896801.

10 Stewart GE, Drudi L. Medical education for exploration class missions NASA Aerospace Medicine Elective at the Kennedy Space Centre. Mcgill J Med 2011;13:55.

11 Bleich MR. Teasing leadership development out of clinical scenarios. J Contin Educ Nurs 2015;46:105–6.

12 Murray DJ, Freeman BD, Boulet JR, et al. Decision making in trauma settings: simulation to improve diagnostic skills. Simul Healthc 2015;10:139–45.

13 Croskerry P, Petrie DA, Reilly JB, et al. Deciding about fast and slow decisions. Acad Med 2014;89:197–200.

14 Zendehrouh S. A new computational account of cognitive control over reinforcement-based decision-making: modeling of a probabilistic learning task. Neural Netw 2015;71:112–23.

15 Bridge LM, Watkins S. Impact of medical training level on medical autonomy for Long-Duration Space Flight. Hanover NASA STI Program Office.

16 Allan JF, McKenna J, Dominey S, Dinnan S. Degrees of resilience: profiling psychological resilience and prospective academic achievement in university inductees. Br J Guid Counc 2014;42:9–25.

17 Schroeder SR, Marian V, Shook A, et al. Bilingualism and musicianship enhance cognitive control. Neural Plast 2016;2016:1–11.

18 Cho BK, Aghazadeh F, Al-Qaisi S. Effect of video-game experience and position of flight stick controller on simulated-flight performance. Int J Occup Saf Ergon 2012;18:429–41.

19 Wang J, Ferris KA, Hershberg RM, et al. Developmental trajectories of Youth Character: a Five-Wave Longitudinal Study of Cub scouts and Non-Scout Boys. J Youth Adolesc 2015;44:2359–73.

20 Schwarzer R, Jerusalem M. Generalized Self-Efficacy scale. In: Weinman J, Wright S, Johnston M, Measures in health psychology: a user's portfolio; vol. Causal and control beliefs. Windsor, England: Nfer-Nelson, 1995.

21 Rammstedt B, John OP. Measuring personality in one minute or less: a 10-item short version of the big five inventory in English and german. J Res Pers 2007;41:203–12.

22 Hays RD, Bjorner JB, Revicki DA, et al. Development of physical and mental health summary scores from the patient-reported outcomes measurement information system (PROMIS) global items. Qual Life Res 2009;18:873–80.

23 Agency for Healthcare Research and Quality (AHRQ). Hospital Survey on Patient Safety Culture. 2010 http://www.ahrq.gov/professionals/quality-patient-safety/patientsafetyculture/hospital/index.html.

24 Hart SG, Staveland LE. Development of NASA-TLX (Task load index): Results of empirical and theoretical research. Advances in Psychology 1988;52:139–83.

25 Ahmed RA, McCarroll ML, Schwartz A, et al. Development, Validation, and implementation of A Medical Judgment Metric. Medical Decision Making Policy & Practice 2017.

26 Wang Y, Wu R, . Time effects, displacement, and leadership roles on a Lunar Space Station Analogue. Aerosp Med Hum Perform 2015;86:819–23.

27 Musson DM, Doyle TE, Saary J. Systems modeling of space medical support architecture: topological mapping of high level characteristics and constraints. Conf Proc IEEE Eng Med Biol Soc 2012;2012:6090–4.

28 Kanas N. Interpersonal issues in space: shuttle/Mir and beyond. Aviat Space Environ Med 2005;76(6 Suppl):B126–34.

Medical judgement analogue studies with applications to spaceflight crew medical officer

Michele L McCarroll, Rami A Ahmed, Alan Schwartz, Michael David Gothard, Steven Scott Atkinson, Patrick Hughes, Jose Cepeda Brito, Lori Assad, Jerry Myers and Richard L George

Updated information and services can be found at:

http://stel.bmj.com/content/early/2017/07/18/bmjstel-2017-000210

These include:

References	This article cites 19 articles, 0 of which you can access for free at: **http://stel.bmj.com/content/early/2017/07/18/bmjstel-2017-000210** **#BIBL**
Open Access	This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/
Email alerting service	Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.
Topic Collections	Articles on similar topics can be found in the following collections Open access (6)

Notes