Preliminary Study of the Effects of Sequential Hypoxic Exposures in a Simulated Flight Task

F. Eric Robinson; Dain Horning; Jeffrey B. Phillips

BACKGROUND: Previous studies of acute hypoxia have largely examined different altitudes in isolation. Pilots, however, receive two exposures during in-flight hypoxic emergencies (IFHEs): the initial exposure at altitude, followed by a second mild exposure after descending and removing the breathing mask. Conventional wisdom holds that performance recovers with blood oxygen saturation and that exposure to mild hypoxia is safe. This study examined the possibility that the effects of moderate hypoxia may linger to overlap with the effects of mild hypoxia during sequential exposures such as those experienced by pilots during an IFHE.

METHODS: Subjects performed a simulated flight task and secondary task while being exposed to normobaric hypoxia via the ROBD-2.

RESULTS: Average error on the flight task during exposure to 3048 m (10,000 ft) was marginally worse when preceded by exposure to 7620 m (25,000 ft; 7.40 $\pm$ 3.32) than when experienced in isolation (6.42 $\pm$ 3.82). Performance on the secondary task was likewise worse when the mild exposure followed the moderate exposure (0.27 $\pm$ 0.30 lapses per minute) than when the mild exposure occurred by itself (0.19 $\pm$ 0.20 lapses per minute). Minimum S_pO_2 showed a similar pattern of results (84.87 $\pm$ 4.37 vs. 86.61 $\pm$ 2.47).

DISCUSSION: We believe our results are most likely due to a failure to recover from the original moderate exposure rather than an additive effect between the exposures. Even so, our findings suggest that pilot impairment following an IFHE may be worse than previously believed.

KEYWORDS: normobaric hypoxia, recovery from hypoxia, moderate hypoxia, mild hypoxia, gas blending device.

Robinson FE, Horning D, Phillips JB. *Preliminary study of the effects of sequential hypoxic exposures in a simulated flight task*. Aerosp Med Hum Perform. 2018; 89(12):1050–1059.

Most research examining the effects of acute hypoxia on performance has studied single exposures in isolation. However, this may not accurately reflect the exposures received by pilots during an in-flight hypoxic emergency (IFHE). The present study therefore examined whether exposure to moderate hypoxia worsens the effects of subsequent exposure to mild hypoxia. For the purposes of this paper, "moderate" hypoxia will refer to a hypoxic exposure that is sufficient to cause notable symptoms and performance deficits but would not generally cause incapacitation within the first few minutes (as opposed to extremely rapid incapacitation typically expected at altitudes above 10,668 m or 35,000 ft).[8] "Mild" hypoxia will refer to a hypoxic exposure that does not typically lead to notable discomfort or incapacitation within the first half hour of exposure.

Conventional wisdom assumes that hypoxia-related performance deficits disappear following a return to normal blood oxygen saturation (S_pO_2), and that mild hypoxia does not impair performance. Based on these assumptions, pilots experiencing an IFHE breathe 100% oxygen to return to normal S_pO_2 while descending to a cabin altitude of 3048 m (10,000 ft), at which point they remove the flight mask and breathe cabin air. As a result of this procedure, pilots actually receive sequential exposures during an emergency: moderate hypoxia during the emergency itself, followed by mild hypoxia upon removing the flight mask.

From the Naval Medical Research Unit Dayton, Wright-Patterson AFB, OH, and Florida Institute for Human & Machine Cognition, Pensacola, FL.

This manuscript was received for review in December 2017. It was accepted for publication in August 2018.

Address correspondence to: F. Eric Robinson, 2624 Q St., Bldg 851, Area B, WPAFB, OH 45433; f.eric.robinson@gmail.com.

DOI: https://doi.org/10.3357/AMHP.5052.2018

Recent experimental data have challenged the notions that performance recovers quickly upon return to normal oxygen levels, and that mild hypoxia does not cause performance deficits.

Exposure to moderate hypoxia leads to deficits in several flight-related cognitive functions, including visual processing,[10] reaction time,[11,12] and motor control.[11] Accordingly, moderate hypoxia has been shown to affect pilots' ability to perform even simple tasks such as maintain a prescribed airspeed and altitude during simulated flight.[25] Further, although S_pO_2 typically returns to baseline within one minute of breathing normal oxygen levels, performance on certain tasks may not return to baseline levels for several hours. Auditory monitoring may remain impaired beyond the hypoxic exposure.[2] Similarly, the Flanker Arrow Task (designed to measure reaction time and attention in the presence of distracting stimuli) demonstrated impairment throughout a 10-min recovery period after exposure to 6096 m (20,000 ft) simulated altitude,[22] and simple and choice reaction time were impaired for at least 2.5 h after simulated exposure to 5486 m (18,000 ft).[23]

Altitudes below 3048 m are generally not considered to cause performance deficits, but research findings indicate that this assumption is not true in all circumstances. Exposure to 3048 m for as little as 15 min can cause color vision deficits under the lighting conditions encountered during night flying.[6] Exposure to 3048 m also increases reaction time[7] as well as procedural errors during simulated flight, particularly during descent and landing.[18] Finally, Legg and colleagues indicate that performance on difficult cognitive tasks such as complex logical reasoning or demanding memory tasks may begin to show marginal impairment after exposure to altitudes as low as 2438 m (8000 ft).[15,16] Although the deficits associated with mild hypoxia are relatively minor, even subtle impairments in vision, procedural execution, reasoning, or memory can increase the risk of a mishap following an in-flight hypoxic event.

Mild and moderate hypoxia may each compromise flight safety by themselves. However, performance deficits may not disappear upon return to normal S_pO_2, potentially leading to a situation in which the effects of moderate hypoxia linger to exacerbate the effects of mild hypoxia during a normal response to a hypoxic emergency. This possibility is especially concerning when coupled with the fact that the most pronounced effects of mild hypoxia are associated with the descent and landing phase of flight. Given that descent and landing are a pilot's primary objectives after initiating emergency procedures, pilots are exposed to mild hypoxia precisely when its effects are most serious. The risk of pilot error following a hypoxic event may therefore be greater than is currently believed. However, failure to recover has to date only been demonstrated using simple cognitive tasks, rather than more realistic flight tasks. The current study extends prior research by using a simulated flight task rather than basic cognitive tasks.

The present study evaluated whether sequential exposures to moderate and mild hypoxia such as those experienced by pilots in a hypoxic emergency lead to greater performance deficits than compared to a single altitude exposure. Due to the possibility of a "hangover" effect from a moderate exposure interacting with the effects of a mild exposure, we hypothesized that performance at the traditional "safe" altitude of 3048 m would be worse following a moderate exposure than when 3048 m was experienced in isolation.

METHODS

Subjects

This study was reviewed and approved by the Naval Medical Research Unit Dayton (NAMRU-D) IRB. All subjects gave informed consent after having the opportunity to read the consent form and ask any questions. A total of 21 active duty United States Air Force personnel assigned to Wright-Patterson Air Force Base, OH (WPAFB) successfully completed the entire study. This sample size was determined a priori to be sufficient to deliver greater than 80% power using G*Power software.[9] Subjects included 20 men and 1 woman, ranging in age from 22 to 37. Seven subjects were recruited via email solicitation of a high altitude research subject panel maintained at the United States Air Force School of Aerospace Medicine. The remaining subjects were recruited via flyers posted in break rooms and public spaces around the base, and through mass email solicitation to several units stationed on WPAFB.

Subjects were screened prior to participation to rule out any medical conditions or lifestyle issues that may have compromised safety or confounded the results (e.g., asthma, anemia, tobacco use, etc.). Our subjects had an average baseline heart rate of 70.36 ± 8.40 and S_pO_2 of 98.22 ± 1.05. Active duty Air Force personnel are expected to meet minimum physical fitness standards as measured by periodic fitness tests. These standards are adjusted for age and sex and include a 2.4 km (1.5 mi) timed run, number of push-ups and sit-ups completed within 1 min, and a body composition measurement.

None of the subjects were licensed pilots, but some did report an interest in flying and prior experience using flight simulators. There were 14 subjects who reported prior experience with hypoxia. An additional six subjects volunteered for the study but withdrew because they could not find sufficient time from their work schedules to complete all four sessions. No subjects withdrew due to adverse reactions to the hypoxic exposure. Subjects who withdrew were not included in the analysis, but their dropout did alter the counterbalancing scheme for the study.

Condition presentation order was predetermined using a Latin Square counterbalancing scheme, with a given presentation order assigned to a subject based on the order in which they volunteered (e.g., 1: order A, 2: order B, 3: order C, 4: order D, 5: order A, 6: order B…). This system facilitated enrolling and scheduling multiple subjects concurrently. However, multiple required visits meant that many subjects were in various stages of completion at any given time. A subject's withdrawal may not have been known until weeks after originally volunteering and several new volunteers had been enrolled. The number of participants dropping out in the process of

completing the study was unanticipated, and by the time the impact of the dropouts on the counterbalancing scheme was apparent we did not have enough volunteer slots left to fully correct the problem. Uneven dropout among subjects across condition presentation order caused the Control condition to be presented first nearly twice as often as the other conditions (eight times for Control vs. four or five times for the other three conditions). Thus, learning effects likely altered performance in the Control condition relative to the other conditions and made the Control condition an unreliable indication of baseline performance. This issue is addressed further in the Discussion section.

Equipment and Materials

Subjects were exposed to normobaric hypoxia via the Reduced Oxygen Breathing Device (ROBD-2; Environics, Inc., Tolland, CT). The ROBD-2 is a gas blending device that uses thermal mass flow controllers to deliver mixtures of compressed breathing air, nitrogen, and oxygen to simulate altitudes between ground level and 10,363 m (34,000 ft) without altering the barometric pressure experienced by subjects. Gas mixtures were delivered through a standard aviation mask attached to a flight helmet via bayonet clips.

Subjects performed tasks in a fixed-based flight simulator operated via X-Plane software emulating a T-6 Texan. The flight instruments were displayed on a 66-cm (26-in) diagonal ELO monitor, while the outside-the-cockpit view was displayed on a 152-cm (60-in) diagonal Samsung LED High Definition TV, providing an 87° wide by 49° high field of view. A FitPC3Pro drove the outside the window scene graphics. Subjects sat in an open cockpit on a SPARCO seat adjustable for height and seat back angle. Control inputs were made using a Thrustmaster Cougar joystick and Thrustmaster Warthog throttle. S_pO_2 was monitored using a standard pulse oximeter (Model 3900P, Datex Ohmeda Corp., Madison, WI) placed on the index finger of the left hand.

Subjects performed two tasks during the exposure profile: a flight task and a time estimation task. Subjects were not instructed to prioritize one task over the other. The primary flight task consisted of maintaining straight and level flight on a heading of 90° at an altitude of 3657 m (12,000 ft) and an airspeed of 150 kn. Subjects were instructed that all three parameters would count equally toward their performance score. Subjects used only the control stick and throttle to fly the aircraft – all other controls and cockpit switches were disabled. The aircraft was untrimmed and a steady quartering wind was blowing from 45° at 5 kn. Subjects flew over a simulation of the terrain around Fallon Naval Air Station, NV, with clear weather. Similar straight-and-level flight tasks in simulators have proven to be sensitive to the effects of hypoxia.[25]

Subjects also performed a secondary task consisting of estimating 10-s intervals. While flying, subjects received a prompt to "Begin counting 10 seconds now" displayed on the outside-the-cockpit monitor as well as broadcast through speakers mounted to the simulator. Prompts were randomly timed to occur between 20 and 30 s apart. After each prompt, subjects started the timer by pressing a button on the control stick. When the subject estimated that 10 s had elapsed, the subject pressed the same button again to stop the timer. Upon activation/deactivation of the timer, the perimeter of the outside-the-cockpit monitor flashed red to acknowledge the button press. Other than this indication that the timer had been successfully activated/deactivated, subjects did not receive feedback regarding the time estimation task. Similar time estimation tasks have been shown to be sensitive to workload and difficulty manipulations in flight simulators.[3–5]

Procedure

Subjects experienced four total exposure profiles over the course of four separate visits to the lab. These profiles will hereafter be referred to as the Control, Mild, Moderate, and Combined conditions (**Fig. 1**). Each exposure profile consisted of two altitudes. Altitude Equivalent 1 was either ground level (approximately 250 m/820 ft above sea level in the testing laboratory) or 7620 m (25,000 ft) normobaric equivalent. Altitude Equivalent 2 was either ground level or 3048 m normobaric equivalent. These altitudes meet the symptom and time to incapacitation criteria outlined in the introduction.[18,19] For all flight

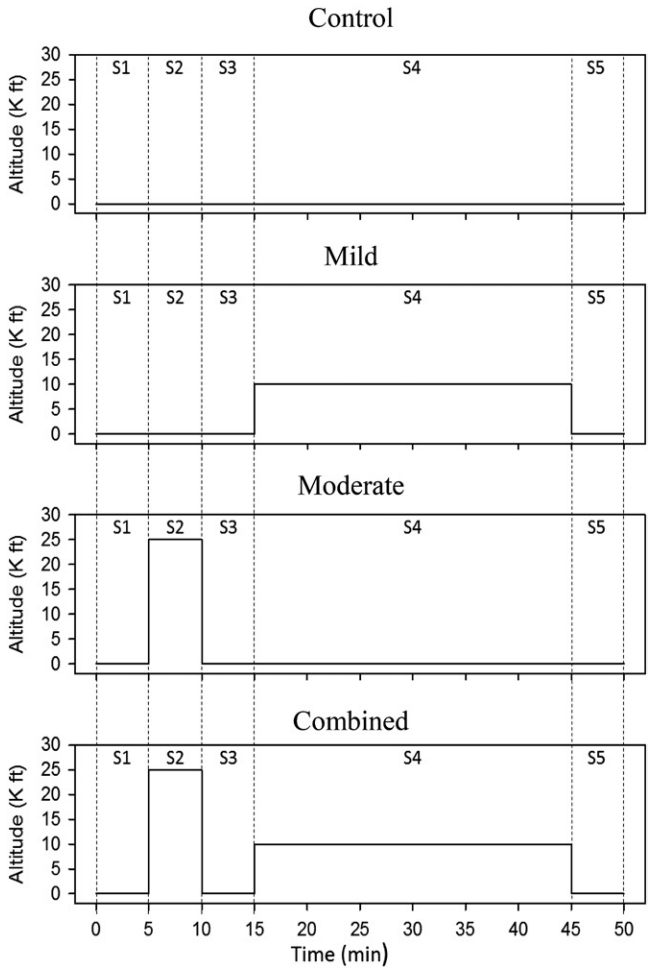

Fig. 1. Exposure profiles for the Control, Mild, Moderate, and Combined conditions, as well as the segments used for analysis.

profiles, subjects breathed ground level air for 5 min (Segment 1; S1), followed by Altitude Equivalent 1 for 5 min (Segment 2; S2), another 5 min of ground level air (Segment 3; S3), Altitude Equivalent 2 for 30 min (Segment 4; S4), and a final 5 min of ground level air (Segment 5; S5). Subjects were blinded regarding which flight profile they experienced on any given visit. The order of the flight profiles was counterbalanced across subjects using a Latin Square design. The exposure times for Altitudes 1 and 2 were within the limits listed in the Time of Useful Consciousness table,[8] and are reasonable estimates of how long exposure to each altitude may last in an aircraft as the pilot must first recognize hypoxia (Altitude Equivalent 1), descend, and then fly to an airfield to land after removing the flight mask (Altitude Equivalent 2).

Subjects reported to NAMRU-D on four separate occasions, experiencing a different exposure profile each visit. Visits were scheduled a minimum of 48 h apart in order to ensure complete recovery between visits. Upon arrival for their first visit to the laboratory, the subject had an opportunity to read the informed consent document and ask questions. After giving informed consent, subjects completed a brief questionnaire to confirm compliance with study requirements, followed by a blood pressure check and a blood draw to ensure normal levels of hematocrit and hemoglobin (subjects with a current flight physical were not required to undergo the blood pressure check or blood draw). Female subjects were given a urine pregnancy test to rule out pregnancy prior to any exposure. Subjects were then fitted for a flight helmet and flight mask.

After equipment fitting, subjects were trained on the performance tasks, including an explanation of the task, instructions on the flight controls, and a brief description of the general relationship between airspeed and altitude in a fixed wing aircraft. Subjects were then allowed to practice flying the simulator on a straight and level course until they felt comfortable with the task (approximately 5 min for most subjects). Subsequent visits followed the same procedure minus the informed consent and equipment fitting. Subjects were allowed to practice flying the simulator prior to exposure for every session.

Once subjects indicated that they were finished practicing the flight task, the simulator was reset and the subject donned the helmet, flight mask, and physiological monitoring equipment. At this point the experimenters confirmed that the subject was ready and began the exposure. Subjects performed both the flight task and the time estimation task for the entire duration of the flight. In addition to the physiological sensors, subjects were monitored via closed-circuit video as well as audio communication with the experimenters. Exposure to each altitude was terminated and the subject was advanced to the next ground level portion of the profile after the time limit was reached, if the subject's S_pO_2 dropped to 55%, if the subject became nonresponsive to verbal prompts, or if the subject requested to be brought back to ground level.

Statistical Analysis

ROBD-2 oxygen concentration and pulse-oximeter (PO) data were collected in LabView (v8.2, National Instruments, Austin, TX). X-Plane output and cognitive performance data were collected via a custom plugin and instructor operating station (IOS) written in the C# and C++ languages. All data processing, including time line-up and calculation of physiological and outcome measure statistics, was performed in MATLAB (MathWorks, Inc., Natick, MA). Repeated measures ANOVAs were performed in SPSS (IBM).

The outcome measure for the flight task was normalized root mean square error (NRMSE). For each flight parameter (i.e., heading, airspeed, and altitude), NRMSE was computed as:

$$NRMSE = \sqrt{\frac{\Sigma(actual\,value - target\,value)^2}{n}} \times \frac{1}{target\,value}$$

where n is the number of data points. NRMSE was then summed across each parameter within a given segment of the flight profile to derive a single value accounting for total error in airspeed, altitude, and heading during each segment of the flight profile (Total NRMSE). Total NRMSE will be hereafter referred to as flight-sim error (FSE).

Outcome measures for the time estimation task included lapses per minute (LPM) and standard deviation in the time estimates (TSD). Lapses were defined as any response pattern that did not match the prescribed order of "prompt – start timer – stop timer." In addition, because extremely low time estimates were observed to occur in conjunction with multiple failed starts/stops (indicating confusion), a trial was considered a lapse if the estimate was shorter than 3 SD below the subject's mean estimate for the profile. Lapses were standardized by total timespan of segments in minutes because the duration of S2 was often shorter in the Moderate and Combined conditions compared to the Control and Mild conditions due to subjects reaching physiological cutoffs prior to the 5-min cutoff.

The primary physiological measures of interest were S_pO_2 and heart rate (HR) as measured by the Datex-Ohmeda at the finger. A low pass Butterworth filter with a cutoff frequency of 0.08 Hz was applied to the raw HR data to reduce variance attributable to the 3-s averaging mode of the oximeter, thus stabilizing HR extremes. HR measures were calculated from filtered HR. Measures compared included minimum S_pO_2 (S_pO_2 Min), average S_pO_2 (S_pO_2 Avg), maximum HR (HR Max), and average HR (HR Avg), defined as follows:

3048-m equivalent exposure:
S_pO_2 Min: minimum S_pO_2 reached across all of S4.
S_pO_2 Avg: mean S_pO_2 for the last 15 min of S4.
HR Max: maximum HR across all of S4.
HR Avg: mean HR for the last 15 min of S4.

Subjects' S_pO_2 and HR typically required around 15 min to stabilize during exposure to the 3048 m equivalent altitude. We therefore averaged over the last 15 min to capture a more stable estimate.

7620-m equivalent exposure:

S_pO_2 Min: minimum S_pO_2 reached across all of S2 and S3.

S_pO_2 Avg: mean S_pO_2 in the 30-s interval [tmin-29, tmin], where tmin = time S_pO_2 Min occurred.

HR Max: maximum HR reached across all of S2 and S3.

HR Avg: mean HR in the 30-s interval [tmax-29, tmax], where tmax = time at which HR Max occurred.

S3 was included because there is a lag between the end of the hypoxic exposure and cessation of the downward trend in vital signs. Extreme values of S_pO_2 and HR commonly occurred during the 7620-m equivalent recovery segment after return to breathing normal oxygen levels.

We conducted a series of repeated measures ANOVAs followed by planned comparisons to test for differences in FSE, LPM, and TSD in segments of primary interest (S2 and S4) across the different exposure profiles. One-tailed tests ($\alpha = 0.05$) were employed for cases where directional a priori hypotheses existed; otherwise two-tailed tests were used. However, these analyses are necessary but not sufficient to establish the nature of effects during S4. Two possibilities would explain a difference in performance during S4 of the Mild and Combined conditions. One possibility is a simple failure to recover such that subject performance remained impaired following the 7620-m exposure, and the 3048-m exposure had no discernable effect. The other possibility is an interactive effect such that performance remained impaired after exposure to 7620 m, and these lingering performance deficits were further exacerbated by a second exposure to 3048 m. The former hypothesis implies that 3048 m does not cause additional harm in the context of an IFHE; the latter implies a physiological interaction between two distinct exposures and calls into question the suitability of current emergency procedures.

In order to identify an interactive hypoxic effect, we must examine whether performance is worse in the Combined condition compared to the Moderate condition and in the Moderate condition compared to the Mild condition. This could not be done simultaneously with planned comparisons. We used post hoc tests (with Sidak correction) to compare the Combined condition to the Moderate condition and the Moderate condition to the Mild condition because a condition cannot be entered into multiple planned comparisons.

Set notation will be used when two or more profiles are considered together in planned comparisons. Outliers were identified according to the Tukey hinges method. Individual outliers were investigated to determine causality. If no clear cause was determined, ANOVAs were run with and without outliers. Mean replacement was used to impute data points deemed spurious due to experimental issues (e.g., extreme practice effects, lack of task understanding). The replacement mean was calculated based on the mean of the data set within each time period/condition excluding the outlier value (e.g., calculated within S4 of the Moderate condition). Mean replacement and/or subject exclusion are noted below, identified in the format subject(profile) (i.e., 1(Control) means subject 1, Control condition).

A series of paired *t*-tests was performed to test for differences in physiological measures (S_pO_2 Avg., S_pO_2 Min, HR Avg., HR Min) between the Mild and Combined conditions during S4 and between the Moderate and Combined conditions during S2 and S3.

RESULTS

We first confirmed that our measures were sensitive to the effects of hypoxia. Mean replacement was used for two subjects in the FSE analysis [13(Control, Mild, Moderate) and 21(Control, Mild)], and two subjects in the LPM analysis [4(Combined) and 38(Moderate, Combined)]. We compared performance measures from S2 in the Control and Mild conditions (ground level) to performance measures from S2 in the Moderate and Combined conditions (7620-m equivalent). **Fig. 2** illustrates these comparisons as well as the means and SE of each condition. Planned comparisons for the 7620-m repeated measures ANOVA were as follows: {Control, Mild} vs. {Moderate, Combined}; Control vs. Mild; Moderate vs. Combined. We expected worse performance in {Moderate, Combined} compared to {Control, Mild} (Fig. 2). The comparisons Control vs. Mild and Moderate vs. Combined were performed to check for test-retest reliability in the outcome measures; we did not expect to see differences in these secondary comparisons. These expectations were confirmed, indicating that performance during the 7620-m simulated exposure was significantly worse than performance at ground level and supporting the validity of the performance outcome measures **(Table I**).

We now turn to the results of our analysis of the 3048-m exposure. Mean replacement was used for two subjects in the FSE analysis [13(Control) and 18(Mild)]. For LPM, mean replacement was used for 38(Moderate, Combined). Two subjects (15 and 16) were outliers across all profiles for LPM for reasons we could not identify; we therefore treated the data as valid but the LPM ANOVA was run both with and without subjects 15 and 16 included. There were no outliers for TSD.

For the 3048-m equivalent exposure repeated measures ANOVA, we compared performance measures within S4 across all four exposure conditions. Planned comparisons for the 3048-m equivalent exposure repeated measures ANOVA were as follows: Control vs. {Mild, Moderate, Combined}; Moderate vs. {Mild, Combined}; Mild vs. Combined. The comparison of primary interest to this study was Mild vs. Combined. **Fig. 3** illustrates the Mild vs. Combined comparison, as well as the means and SE for each condition. If performance during the 3048-m equivalent exposure was worse in the Combined condition compared to the Mild condition, an effect due to combined exposures is indicated. However, the other comparisons were also of interest for their a priori potential to aid in interpretation and validation of our results. Control vs. {Mild, Moderate, Combined} allowed us to establish hypoxic effects of our manipulations (though this comparison is largely rendered uninformative due to issues with our Control condition described in the Discussion section). Moderate vs. {Mild,

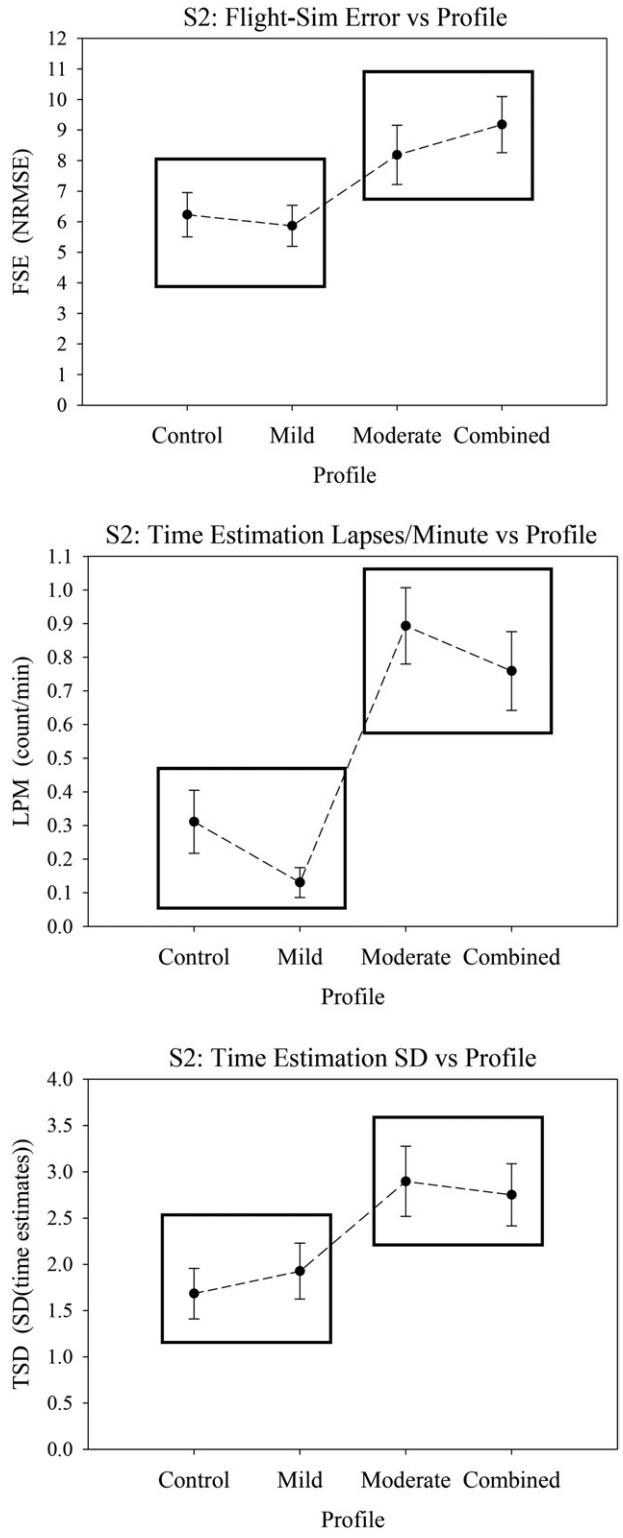

Fig. 2. {Control, Mild} to {Moderate, Combined} comparisons for S2 across the three performance measures. To assist in visualizing the analysis, boxes indicate the groups of conditions being compared to one another within each performance measure. Error bars represent standard error.

Combined} allowed us to compare a condition in which the subjects were at sea level during S4 to conditions in which the subjects were at 3048 m. This comparison offers some insight into performance effects of 3048 m. The results of these comparisons as well as post hoc tests comparing Moderate vs. Combined and Moderate vs. Mild are shown in **Table II**.

Table III displays means and SD during S4 along with *P*-values for the physiological comparisons using paired *t*-tests. Unsurprisingly, we found no significant differences within S4 between the Control and Moderate conditions, and we found significant differences within S4 among all four physiological measures between the Moderate and Combined conditions. One-tailed tests revealed a small (MD = 1.74) but significant difference in S_pO_2 Min between the Mild and Combined conditions in segment S4.

DISCUSSION

In contrast to most research examining the performance effects of acute hypoxia at different altitudes in isolation, this study examined performance under conditions of sequential exposures such as those that may occur during an IFHE. Performance on a time estimation task indicated that performance may not recover immediately after a hypoxic exposure. S_pO_2 data further indicated that the physiological effects of exposure to 3048 m simulated altitude may be exacerbated by prior exposure to moderate hypoxia. FSE data showed a similar nonsignificant trend indicating possible delayed recovery. The aggregate conclusion of these various outcome measures supports prior work challenging the conventional wisdom that performance recovers with S_pO_2 following a hypoxic exposure, and extends that work by demonstrating effects in a setting that better replicates the demands of flight.

When comparing S4 across the Mild and Combined conditions, we found evidence that LPM increased in the Combined condition compared to the Mild condition when all subjects were included in the analysis. FSE showed a similar trend, although this result only approached significance. These results imply that performance during a 3048-m simulated exposure was worse when preceded by exposure to a 7620-m simulated altitude than when the lower exposure occurred in isolation. Two hypotheses may explain this pattern of results. The first is that the moderate and mild exposures interact such that the effects of the first exposure linger and exacerbate/are exacerbated by the effects of the second exposure. Such a finding would imply a physiological interaction between two exposures separated in time, perhaps due to incomplete recovery at a cellular level despite global measures of S_pO_2 returning to normal. The other is that the effects of the first exposure never fully dissipated upon return to normal S_pO_2, but the second exposure had no effect on its own. This would imply that there is no physiological interaction between the two exposures and support the belief that 3048 m by itself is not necessarily detrimental in the context of recovery during an IFHE. We examined these hypotheses with post hoc tests comparing S4 in the Mild, Moderate, and Combined conditions. Performance during S4 of the Moderate condition was not statistically different than performance in either the Mild or Combined conditions, leaving us unable to distinguish between the

Table I. Comparisons During S2 for FSE, LPM, and TSD.

MEASURE	COMPARISON	*F*(1,19)	*P*	η_p^2
FSE	{Moderate, Combined};{Control, Mild}	11.25	< 0.01	0.37
	Control; Mild	0.30	0.59	0.02
	Moderate; Combined	0.73	0.40	0.04
LPM	{Moderate, Combined};{Control, Mild}	70.42	< 0.01	0.79
	Control; Mild	3.35	0.08	0.15
	Moderate; Combined	0.58	0.46	0.03
TSD	{Moderate, Combined};{Control, Mild}	14.03	< 0.01	0.43
	Control; Mild	0.64	0.43	0.03
	Moderate; Combined	0.10	0.76	0.01

two explanations statistically. We therefore speculate on the most likely cause based on patterns in the data.

The means for S4 in the Moderate condition across both of LPM indicated that performance in the Moderate condition was nearly identical to the Combined condition (Fig. 3). In addition, the mean for LPM indicated worse performance in the Moderate condition than the Combined condition, which should not happen in the presence of an interactive effect. Further, post hoc tests indicated a trend for FSE in the Moderate condition to be worse than the Mild condition, but no different from the Combined condition. The very conservative nature of the post hoc corrections likely prevented finding significant differences, but we believe that performance on our tasks in the Moderate condition is much more similar to the Combined condition than the Mild condition. We are therefore inclined to believe that the observed performance effects are due to a failure to recover completely

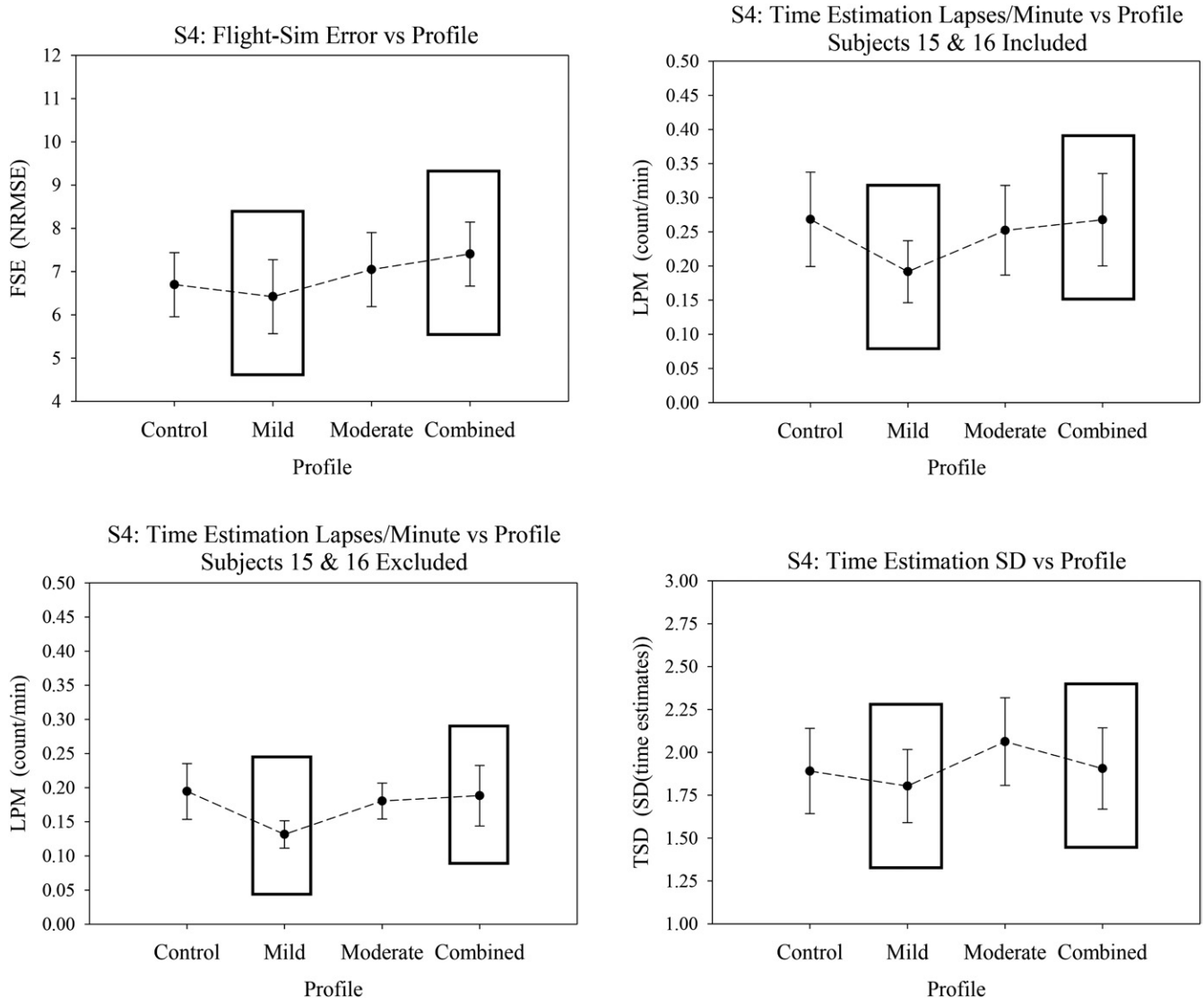

Fig. 3. Mild vs. Combined comparisons for S4 across each performance measure. To assist in visualizing the analysis, boxes indicate conditions being compared to one another within each performance measure. Error bars represent standard error. Note that these plots have been scaled for easy comparison with the S2 comparisons of Fig. 2.

Table II. Comparisons and Post Hoc Tests During S4 for FSE, LPM, and TSD.

MEASURE	COMPARISON	*F*(df) or *t*(df)	*P*	η_p^2; d†
FSE	Control vs. {Mild, Mod., Combined)	$F(1,19) = 0.17$	0.68	0.01
	Moderate vs. {Mild, Combined)	$F(1,19) = 0.10$	0.76	0.01
	Mild vs. Combined*	$F(1,19) = 2.81$	0.06	0.13
	Mild vs. Moderate (post hoc)	$t(19) = 1.13$	1.00	0.25
	Moderate vs. Combined (post hoc)	$t(19) = 0.75$	1.00	0.17
LPM	Control vs. {Mild, Mod., Combined)	$F(1,19) = 0.36$	0.56	0.02
(15 and 16 included)	Moderate vs. {Mild, Combined)	$F(1,19) = 0.29$	0.60	0.02
	Mild vs. Combined*	$F(1,19) = 3.31$	0.04	0.15
	Mild vs. Moderate (post hoc)	$t(19) = 1.58$	0.79	0.35
	Moderate vs. Combined (post hoc)	$t(19) = 0.28$	1.00	0.06
LPM	Control vs. {Mild, Mod., Combined)	$F(1,17) = 0.70$	0.41	0.04
(15 and 16 excluded)	Moderate vs. {Mild, Combined)	$F(1,17) = 0.59$	0.46	0.03
	Mild vs. Combined*	$F(1,17) = 1.69$	0.11	0.09
	Mild vs. Moderate (post hoc)	$t(17) = 2.11$	0.30	0.50
	Moderate vs. Combined (post hoc)	$t(17) = 0.18$	1.00	0.04
TSD	Control vs. {Mild, Mod., Combined)	$F(1,19) = 0.05$	0.82	<0.01
	Moderate vs. {Mild, Combined)	$F(1,19) = 2.30$	0.15	0.11
	Mild vs. Combined*	$F(1,19) = 0.59$	0.23	0.03
	Mild vs. Moderate (post hoc)	$t(19) = 1.71$	0.62	0.38
	Moderate vs. Combined (post hoc)	$t(19) = 1.02$	1.00	0.23

* This comparison used a one-tailed test.
† Cohen's d values were used for post hoc tests.

following exposure to moderate hypoxia rather than an interaction between the two exposures.

In addition to the performance measures, physiological measures also indicated that hypoxia may have effects beyond the duration of the exposure. Pairwise tests comparing S4 across the different exposure profiles indicated that vital signs returned to normal within minutes after exposure to moderate hypoxia when recovering at ground level, but not when recovering from moderate hypoxia at 3048 m simulated altitude. This finding is to be expected given the difference between ground level and 3048 m, and the S_pO_2 and HR observed at 3048 m simulated altitude during the Combined condition were not in the range that would cause concern. However, we did observe a small but statistically significant difference in S_pO_2 Min between the Mild and Combined conditions during S4. Subjects reached a lower minimum S_pO_2 when breathing at 3048 m simulated altitude if they had previously been exposed to 7620 m simulated altitude.

This result indicates that some physiological interaction between the two hypoxic exposures may have occurred such that the first exposure made our subjects more vulnerable to desaturation during the second exposure and lends some level of plausibility to the notion that hypoxic exposures may interact to exacerbate one another's effects if they occur in close enough succession. However, S_pO_2 data do not necessarily correspond to performance. Prior research has demonstrated continued impairment after hypoxic exposure despite returning to normal S_pO_2 levels.[2,22,23] The similarity in performance despite vastly different S_pO_2 levels during S4 seen between the Moderate and Combined conditions in this study further supports such a claim. We therefore believe that something other than change in S_pO_2 is driving the performance effects seen in this study, though it is not clear what underlying physiological factors may explain posthypoxia performance effects. Possibilities include posthypoxic inflammation, lingering changes in regional cerebral blood perfusion, or disruptions in axonal potentiation that linger beyond the hypoxic event itself. Future studies should investigate these possibilities further.

One of the main limitations that complicated interpretation of our results was the fact that our counterbalancing scheme was altered due to subject dropout. Condition presentation order was therefore confounded with experience in the simulator, possibly leading to artificially poor performance during the Control condition despite the training time given to subjects. This issue likely caused performance in the Control condition to appear slightly worse than the Mild condition in Fig. 2 and Fig. 3. The fact that our Control condition did not capture a valid estimate of unimpaired performance made it difficult to use the Control condition as a baseline to determine the performance effects of the Mild and Moderate conditions alone. However, we do not believe this issue affected our primary results or interpretation in a substantive way.

In addition, we did not use licensed pilots in this study due to practical concerns about our ability to recruit sufficient numbers of subjects and a previous history of using nonpilots in the simulator successfully. The novelty of the flight task may have been an issue for a nonpilot sample, leading to practice effects or otherwise affecting performance on the task. We attempted to provide sufficient practice time to guard against such effects, in addition to our efforts to counterbalance

Table III. Means, SDs, and *P*-Values for Physiological Comparisons During S4 Across Conditions.

					CONTROL VS. MOD.		MILD VS. COMBINED		MOD. VS. COMBINED	
MEASURE	CONTROL	MILD	MODERATE	COMBINED	*t*(19)	*P*	*t*(19)	*P*	*t*(19)	*P*
S_pO_2 Avg	99.31 (1.05)	90.91 (2.45)	99.59 (1.06)	90.06 (2.77)	0.87	0.39	1.14	0.13	16.8	<0.01
S_pO_2 Min	97.57 (1.34)	86.61 (2.47)	97.78 (1.04)	84.87 (4.37)	0.74	0.49	1.80	0.04	12.4	<0.01
HR Avg	74.57 (9.47)	80.57 (11.91)	73.71 (9.91)	78.92 (10.14)	0.49	0.63	0.95	0.18	2.83	0.01
HR Max	88.77 (11.87)	93.26 (13.94)	87.23 (12.09)	92.27 (10.92)	0.73	0.47	0.50	0.31	2.73	0.01

condition presentation order. These efforts appear to have been only modestly successful, however, due to the issues described above. We observed no obvious indicators that our nonpilot sample struggled unexpectedly with the flight task such as crashing or large oscillations in flight path. A straight and level flight task in the absence of other flight-related duties is arguably similar to the divided attention and tracking tasks common in the human performance literature and our subjects appeared to learn the task relatively quickly. Though some practice effects likely occurred, they appear to have been largely absorbed in the Control condition and we therefore do not believe them to be the primary influence on our results.

The third limitation of this study is a lack of fidelity in the recovery procedure. Subjects in our study recovered by breathing ground level oxygen concentrations rather than the 100% oxygen used in aircraft. We did not use 100% oxygen in order to avoid a possible confound in our results. Though 100% oxygen may reduce overall recovery time for some tasks,[20,21] it can also temporarily exacerbate symptoms in a phenomenon termed the "oxygen paradox."[14,17] Few, if any, studies have examined possible interactions between moderate and mild hypoxic exposures and we wanted to ensure that any observed performance deficits were due to the hypoxic exposure rather than the recovery procedure. Future work should investigate whether 100% oxygen leads to faster recovery and reduces the deficits observed in this study.

We also note the clear imbalance of men to women among our subjects. A higher ratio of male to female volunteers would be expected from the active duty Air Force population (which is roughly 80% male),[1] but not necessarily to the degree observed in our study. We are uncertain why this occurred, but note that our sample is somewhat representative of the ratio of men to women among Air Force pilots, as only 2% of Air Force pilots were women in 2013.[24] Although we would certainly be cautious generalizing our results to the broader civilian population, the sex distribution of our sample is largely representative of the pilot population of primary interest.

It is also important to point out that normobaric hypoxia such as that induced by the ROBD is not necessarily the same as hypobaric hypoxia such as that experienced in an aircraft or altitude chamber. A recent study comparing ROBD-induced hypoxia to that induced by a hypobaric camber and an additional normobaric device that filters oxygen from ambient air (the Reduced Oxygen Breathing Environment; ROBE) found that the ROBD was associated with the deepest oxygen desaturation and the most severe symptoms.[13] This difference may be due to the increased breathing resistance of the ROBD compared to the other methods, potentially leading to disruption in CO_2 levels and acid-base stasis.

On the aggregate, our data strongly imply that performance deficits may persist beyond the period of hypoxic exposure, even after heart rate and S_pO_2 return to normal. However, this effect was inconsistent across our performance measures. The results presented here are in line with previous work, but we note that the literature is similarly inconsistent in this area. Some cognitive tasks do not show any delay in recovery,[2,23] and some tasks such as reaction time have demonstrated conflicting findings across studies.[7,23] The reasons behind these differential findings are unclear, but may be due to differences in the cognitive requirements of the tasks as well as the ways in which the exposures and task presentations are executed. Further work is needed to determine what aspects of performance are likely to show continued impairment after a hypoxic event and under what conditions. We must then replicate this effect using more sensitive measures of these aspects of performance in order to better understand the ways in which pilots are likely to remain impaired following an in-flight emergency.

Although the performance data indicated that our findings are most likely the result of a failure to recover from the original moderate exposure, the physiological data indicated that a small additive effect of the moderate and mild exposures on S_pO_2 was possible. Future work should examine these two explanations to try to tease apart the mechanisms of prolonged impairment following a hypoxic exposure in the context of current emergency procedures. If multiple exposures do in fact have an additive effect, emergency procedures may need to be modified to minimize this impact.

The data in this study indicate that at least one of the implicit assumptions underlying current emergency procedures may not hold true: physiological recovery as assessed by S_pO_2 and heart rate does not necessarily indicate cognitive recovery. We found that performance may be impaired during a mild exposure to hypoxia when preceded by a moderate hypoxic exposure. Based on prior research findings and the patterns seen in our own data, we believe this effect is due to a failure to recover from the moderate exposure. However, we are currently unable to rule out the possibility of an additive effect between the two exposures with 100% certainty. Regardless of the cause, the findings of this study challenge the notion that pilots are able to recover to baseline after descending following a hypoxic emergency. More work is needed to ascertain the precise nature of hypoxia's effect on performance in order to better predict what aspects of performance are likely to remain impaired following an IFHE. If performance remains significantly impaired following exposure to hypoxia, more focus should be turned to preventing exposure via in-cockpit sensors to proactively alert the pilot before physiology is affected.

ACKNOWLEDGMENTS

We wish to acknowledge and thank the members of the research team who contributed to this effort. Ms. Jacqueline Gomez and Ms. Bernadette McCann were integral to data collection efforts. Dr. Leslie Drummond provided valuable support with data collection and lab operation. Mr. Robbie Powell created our flight simulation tasks and interfaces. Mr. Matt Lee provided software and hardware technical support. Mr. Dan Geyer provided phlebotomy services.

The study protocol was approved by the Naval Medical Research Unit Dayton Institutional Review Board in compliance with all applicable Federal regulations governing the protection of human subjects.

This work was supported/funded by work unit number 5ZPIV3.

The views expressed in this article are those of the authors and do not necessarily reflect the official policy or position of the Department of the Navy, Department of Defense, nor the U.S. Government.

Authors and affiliations: Frank Eric Robinson, M.S., Ph.D., and Dain Stephen Horning, B.S., Naval Medical Research Unit Dayton, Wright-Patterson AFB, OH; and Jeffrey Brooks Phillips, M.A., Ph.D., Naval Medical Research Unit Dayton/Institute for Human & Machine Cognition, Pensacola, FL.

REFERENCES

1. Air Force's Personnel Center. Air Force Demographics [Internet]. 31 Mar 2018 [Accessed 1 May 2018.] Available from: http://www.afpc.af.mil/About/Air-Force-Demographics/.
2. Beer JM, Shender B, Chauvin D, Dart T, Fischer J. Cognitive deterioration in moderate and severe hypobaric hypoxia conditions. Aerosp Med Hum Perform. 2017; 88(7):617–626.
3. Bortolussi MR, Hart S, Shively R. Measuring moment-to-moment pilot workload using synchronous presentations of secondary tasks in a motion-based trainer. Aviat Space Environ Med. 1989; 60:124–129.
4. Bortolussi MR, Kantowitz B, Hart S. Measuring pilot workload in a motion base trainer: a comparison of four techniques. Appl Ergon. 1986; 17(4):278–283.
5. Casali JG, Wierwille W. A comparison of rating scale, secondary-task, physiological, and primary-task workload estimation techniques in a simulated flight task emphasizing communications load. Hum Factors. 1983; 25(6):623–641.
6. Connolly DM, Barbur J, Hosking S, Moorhead I. Mild hypoxia impairs chromatic sensitivity in the mesopic range. Invest Ophthalmol Vis Sci. 2008; 49(2):820–827.
7. Dart T, Gallo M, Beer J, Fischer J, Morgan T, Pilmanis A. Hyperoxia and hypoxic hypoxia effects on simple and choice reaction times. Aerosp Med Hum Perform. 2017; 88(12):1073–1080.
8. DeHart R, editor. Fundamentals of aerospace medicine. Philadelphia (PA): Lea & Febiger; 1985.
9. Faul F, Erdfelder E, Lang A-G, Buchner A. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods. 2007; 39(2):175–191.
10. Fowler B, Banner J, Pogue J. The slowing of visual processing by hypoxia. Ergonomics. 1993; 36(6):727–735.
11. Fowler B, Taylor M, Porlier G. The effects of hypoxia on reaction time and movement time components of a perceptual-motor task. Ergonomics. 1987; 30(10):1475–1485.
12. Fowler B, White P, Wright G, Ackles K. The effects of hypoxia on serial response time. Ergonomics. 1982; 25(3):189–201.
13. Gallo M, Funke M, Phillips J, Drummond L, Golich F, et al. Comparison of symptomology, physiologic response, and reaction time in three hypoxic environments [Abstract #119]. 88th Annual Scientific Meeting of the Aerospace Medical Association; April 29-May 4, 2017; Denver, CO. Aerosp Med Hum Perform. 2017; 88(3):197.
14. Gradwell D, Rainford D, editors. Ernsting's aviation and space medicine. Boca Raton (FL): CRC Press; 2016.
15. Legg S, Hill S, Gilbey A, Raman A, Schlader Z, Mundel T. Effect of mild hypoxia on working memory, complex logical reasoning, and risk judgment. Int J Aviat Psychol. 2014; 24(2, Pt. 2):126–140.
16. Legg S, Hill S, Mundel T, Gilbey A, Schlader Z, Raman A. Could mild hypoxia impair pilot decision making in emergencies? Work. 2012; 41:198–203.
17. Malle C, Bourrilhon C, Quinette P, Laisney M, Eustache F, Pierard C. Physiological and cognitive effects of acute normobaric hypoxia and modulations from oxygen breathing. Aerosp Med Hum Perform. 2016; 87(1):3–12.
18. Nesthus T, Rush L, Wreggit S. Effects of mild hypoxia on pilot performance at general aviation altitudes. Washington (DC): Federal Aviation Administration Office of Aviation Medicine; 1997. Report No: DOT/FAA/AM-97-9.
19. Phillips J, Drummond L, Robinson F, Funke M. Hypoxia: exposure time until significant performance effects. Dayton (OH): Naval Medical Research Unit Dayton; 2016. Report No: NAMRU-D 16-19.
20. Phillips J, Drummond L, Robinson F, Warner S, VerStraten M, Funke M. Differing oxygen concentrations and the effect on post-hypoxia recovery. [Abstract #388] 87th Annual Scientific Meeting of the Aerospace Medical Association; April 24–28, 2016; Atlantic City, NJ. Aerosp Med Hum Perform. 2016; 87(3):289.
21. Phillips J, Horning D, Robinson F, Warner S, Gomez J, et al. Recovery from hypoxic exposure. [Abstract #176] 86th Annual Scientific Meeting of the Aerospace Medical Association. May 10–14, 2015; Orlando, FL. Aerosp Med Hum Perform. 2015; 86(3):213.
22. Phillips J, Simmons R, Florian J, Horning D, Lojewski R, Chandler J. 2009. Moderate intermittent hypoxia: effect on two-choice reaction time followed by a significant delay in recovery. Proceedings of the Human Factors and Ergonomics Society 53rd Annual Meeting; Oct 19–23, 2001; San Antonio, TX.
23. Phillips J, Simmons R, Horning D. 2012. Post-hypoxic recovery of cognitive and perceptual function. [Abstract #229]. 83rd Annual Scientific Meeting of the Aerospace Medical Association; May 13–17, 2012; Atlanta, GA. Aviat Space Environ Med. 2012; 83(3):274.
24. Scarborough R. She has the right stuff: female combat pilots have been flashing their skills for 20 years. Washington Times. 2013 Mar 25. [Accessed December 2017.] Available from: https://www.washingtontimes.com/news/2013/mar/25/more-than-a-sound-barrier-female-pilots-have-displ/.
25. Temme LA, Still D, Acromite M. Hypoxia and flight performance of military instructor pilots in a flight simulator. Aviat Space Environ Med. 2010; 81(7):654–649.