Research Report

SEAN ROBSON, DANIEL HICKS, ELIZABETH CHASE, AMY NABEL, CARRA S. SIMS, MAX STEINER

Setting Higher Close Combat Standards for the Army Combat Fitness Test

Evidence and Recommendations

For more information on this publication, visit **www.rand.org/t/RRA3092-1**.

Published by the RAND Corporation, Santa Monica, Calif.

Library of Congress Cataloging-in-Publication Data is available for this publication.

ISBN: 978-1-9774-1443-4

About This Report

This report documents research and analysis conducted as part of a project entitled *Tailoring Fitness Assessments and Standards to Combat MOS*, sponsored by the Deputy Chief of Staff, G-3/5/7, U.S. Army. The purpose of the project was to provide a research framework and analysis to support the identification and development of combat military occupational specialty (MOS)-specific fitness assessments and standards.

This research was conducted within RAND Arroyo Center's Personnel, Training, and Health Program. RAND Arroyo Center, part of RAND, is a federally funded research and development center (FFRDC) sponsored by the United States Army.

RAND operates under a "Federal-Wide Assurance" (FWA00003425) and complies with the *Code of Federal Regulations for the Protection of Human Subjects Under United States Law* (45 CFR 46), also known as "the Common Rule," as well as with the implementation guidance set forth in DoD Instruction 3216.02. As applicable, this compliance includes reviews and approvals by RAND's Institutional Review Board (the Human Subjects Protection Committee) and by the U.S. Army. The views of sources utilized in this report are solely their own and do not represent the official policy or position of DoD or the U.S. government.

Acknowledgments

We thank the many military and civilian employees of the U.S. Army who supported this project. Their feedback, insights, and assistance were instrumental to our ability to complete this analysis. Specifically, we thank the Deputy Chief of Staff, G-3/5/7, U.S. Army for sponsoring this project, and we highlight the support and insights of our action officer, COL Robert Underwood (Chief of Staff, Army Training Directorate). These efforts were key to our work: managing requests for necessary information, providing invaluable feedback, and coordinating our access to multiple Army organizations, units, and stakeholders. Furthermore, we thank the many individuals in the Army who collected and compiled the data on which our work depends, including Michael Higginbottom, Paul La Gant, and those working with the Digital Training Management System. We also extend our appreciation to MAJ Charles Harris for helping to coordinate and schedule our discussions and briefings. We also thank the stakeholders who participated in our workshop and provided their considered viewpoints on implementation of Army Combat Fitness Test standard alternatives.

Within the RAND Corporation, we recognize the guidance and support of Heather Krull, director of Arroyo's Personnel, Training, and Health program. We also recognize the programming support provided by Christina Panis, Judith Mele, and Pete Schirmer. We thank our reviewers, Michael Linick and Lou Mariano, for their feedback and guidance throughout this project.

Summary

The Army Combat Fitness Test (ACFT) was implemented as the official fitness test for the Regular Army on October 1, 2022, and for the reserve components in April 2023. The six-event test was designed to evaluate a soldier's physical fitness (Table S.1). The ACFT was initially designed with age- and gender-neutral standards,[1] which would apply common standards tied to tiers that reflected the physical demands of a job regardless of age or gender. But before official implementation, the scoring was revised and tailored to different age and gender groups.

Tailoring standards to age and gender raised concerns about whether soldiers will be physically ready for their duties, particularly in specialties such as close combat that are highly physically demanding. The National Defense Authorization Act (NDAA) for Fiscal Year (FY) 2024 aimed to address these concerns by calling for higher minimum fitness standards for soldiers in close combat military occupational specialties (MOSs).

The Army sought the assistance of RAND Arroyo Center, the U.S. Army's federally funded research and development center for studies and analysis, to develop courses of action to address the requirements outlined by Congress. To develop the options, we addressed the following questions:

- What evidence is available to establish gender- and age-neutral ACFT standards for close combat MOSs?
- What evidence is available to establish higher ACFT standards to maintain or strengthen general fitness among soldiers in close combat MOSs?
- What impact will higher ACFT standards have on different subgroups of soldiers, including gender, age group, MOS, and component (Regular Army, Army National Guard [ARNG], and U.S. Army Reserve [USAR])?

We took a systematic approach to answering these questions, which involved reviewing relevant research on ACFT events, task performance, and injuries; analyzing historical ACFT performance; developing options that are defensible, achieve Army objectives, and comply with the FY 2024 NDAA requirements; supporting implementation of a Practice Phase to test a set of higher ACFT standards and analyzing the results; and

TABLE S.1
Description of the Six Army Combat Fitness Test Events

ACFT Event	Description
3 Repetition Maximum Deadlift (MDL)	Deadlift the maximum weight possible three times.
Standing Power Throw	Throw a 10-pound medicine ball backward and overhead for distance
Hand Release Push-Up—Arm Extension	Complete as many hand-release push-ups as possible in two minutes, which requires extending arms to a T-position for each repetition when in down position
Sprint-Drag-Carry	Conduct five 50-meter shuttles for time: sprint, drag, lateral, carry, and sprint
Plank	Maintain a proper plank position for as long as possible
Two-Mile Run	Run two miles for time on a measured, generally flat outdoor course

SOURCE: U.S. Army, "Army Combat Fitness Test," webpage, undated.

[1] Although gender is often used to describe an individual's identity or presentation and sex to describe biological characteristics, we use *gender* instead of *sex* throughout the report when distinguishing between male and female soldiers.

conducting a workshop to document soldiers' concerns about higher ACFT standards and elicit strategies for mitigating concerns—particularly those about injuries and how to best close the gaps in pass rates for subgroups with scores that do not meet higher standards.

Our goal was not to recommend a single option for higher ACFT standards for close combat MOSs but instead to provide the Army with analysis of a set of defensible options and the trade-offs among them.

Initial Evidence for Higher ACFT Standards

There is sufficient evidence to inform the Army's decision on raising ACFT standards for close combat MOSs:

- The ACFT as a complete test with current administration protocols and standards has been evaluated against injury outcomes. Passing any of the six ACFT events is associated with reduced injury risks. Higher performance levels are associated with lower injury risk for all events except the Standing Power Throw. However, selecting minimum standards based on injury risk requires the Army to first specify an acceptable injury risk threshold.
- Higher ACFT standards for select combat MOSs, such as infantry and combat engineers, can be implemented as a strategy to ensure that overall fitness is maintained at the higher levels demonstrated by soldiers in close combat MOSs. Our analysis indicated that overall ACFT scores for soldiers in close combat MOSs are, on average, 6.5 percent higher than those for soldiers in the general-purpose force based on historical performance since implementation for Regular Army soldiers on October 1, 2022.
- For the MDL, there is strong evidence that can inform a specific, higher minimum standard for close combat MOSs based on combat task performance. Although evidence is available to support additional individual ACFT events, the quality and quantity of evidence varies by event.

Practice Phase

The Army undertook an ACFT Practice Phase to pilot higher potential standards among a sample of soldiers in close combat MOSs. This Practice Phase provided an opportunity to collect evidence in support of higher standards for the close combat force. Our analysis included Practice Phase data collected from May 1, 2024, to August 8, 2024. The Practice Phase continued through August 31, 2024, during which more than 44,000 soldiers participated, including 1,514 female soldiers and 9,155 soldiers from the reserve components. We provided two primary approaches for setting standards for the Practice Phase: (1) increase the overall ACFT point standard using the existing age- and gender-normed scoring system and (2) increase standards on the MDL using an age- and gender-neutral standard. Building on these concepts, we developed eight options, shown in Table S.2. The Army selected the final hybrid option, 3D, for the Practice Phase, which would increase the total ACFT point standard to 450 points overall (90 points above the minimum total passing score of 360 for the general-purpose force), along with a requirement of 150 pounds on the MDL (an increase of 10 pounds from the minimum standard of 140 pounds for males and increase of 30 pounds from current minimum of 120 pounds for females).

The results of the Practice Phase highlight trade-offs between alternative standards:

- Evidence suggests that a hybrid approach adopting an overall ACFT score minimum of 450 points and 150 pounds on the MDL may be too high to achieve a 95 percent overall pass rate and 90 percent pass rate for select subgroups in the short term. Under this standard, soldiers in the Regular Army exhibited a 91.4 percent pass rate, and several subgroups, including female soldiers and certain age and MOS

TABLE S.2
ACFT Standards to Consider for the Practice Phase

Option	Alternative Standard
1A	Increase overall ACFT score from 360 to 420 points
1B	Increase overall ACFT score from 360 to 450 points
2A	Increase the MDL standard for female soldiers from 120 to 140 pounds
2B	Increase the MDL standard from 120 (female) and 140 (male) to 150 pounds
3A	Increase overall score to 420 points and MDL standard to 140 pounds
3B	Increase overall score to 420 points and MDL to 150 pounds
3C	Increase overall score to 450 points and MDL to 140 pounds
3D (Practice Phase standard)	Increase overall score to 450 points and MDL to 150 pounds

subgroups, had pass rates below 90 percent. Pass rates for soldiers in the ARNG and USAR are significantly lower under this standard. Other options for setting higher standards achieve an overall pass rate greater than 95 percent and have less impact on Regular Army, ARNG, and USAR subgroups.

- Which higher standards are chosen affects ACFT pass rates differently across subgroups. Younger, female soldiers, particularly, have lower pass rates when the MDL standard is increased. Increasing the standard for total ACFT points to 450 results in pass rates below 90 percent for several MOSs. At this standard, pass rates for soldiers in the ARNG and USAR fall below 75 percent.

There are additional factors to consider in evaluating the results of the Practice Phase. Prior research shows that soldiers' performance improves with experience. Moreover, we found evidence of performance improvement in all six ACFT events in the Practice Phase—particularly among soldiers who did not meet the standard in their previous ACFT attempts and soldiers with less experience taking the ACFT. We modeled future performance under the alternative standards, and our results suggest that pass rates observed in the Practice Phase understate the true extent to which soldiers will continue to improve toward new standards. Taking these improvements into account, fewer subgroups, particularly those in the Regular Army, fall below a 90 percent pass rate.

Perspectives on Higher Standards

As part of the research on higher standards for close combat MOSs, we conducted a workshop with soldiers who reviewed the alternative standards and raised numerous concerns about the higher proposed ACFT standards for close combat MOSs. Few participants were concerned about which higher standard would be selected, though the discussion generally focused on the Practice Phase option.

Instead, much of the discussion focused on concerns around implementation of a higher standard. Principal among them were implementation timelines—specifically, the need for graduated performance goals that are designed to minimize injury while maximizing long-term performance and the need to resource supports to help soldiers prepare for the new standards. Among various concerns, pass rates and injuries were predominant concerns for all soldiers who would be affected by higher standards, but also for specific subgroups, including female soldiers, older and younger soldiers, and soldiers in the reserve components.

Related concerns involved messaging related to the ACFT: why MDL was identified as a "key" close combat individual event, the role of the ACFT for general fitness and as a metric for combat readiness, and similar

questions. Communication about raising ACFT standards for close combat MOSs will need to incorporate a clear rationale for higher standards, transparent goals and incentives to achieve the higher standards, and clarity regarding associated policies and expectations—all of which will be key to successful adoption of a higher standard.

Workshop participants also discussed strategies to mitigate these concerns, which informed our recommendations.

Recommendations to Facilitate Implementation of Higher Standards

As the results of the workshop indicated, how the Army implements the higher standards will be essential to success—regardless of which higher standard is adopted. Thus, our recommendations focus on three aspects of implementation that the Army should adopt:

1. **Supporting policy should be implemented to facilitate the movement to higher standards.** We identified five opportunities for mitigation measures:
    a. **Glide paths:** Provide glide paths to allow soldiers sufficient time to train and improve with minimal increases to injury risk.
    b. **Remedial assessments:** Use MOS-specific physically demanding tasks as a test for soldiers who do not meet the higher ACFT standard (these task-based tests should be standardized for each MOS).
    c. **Re-screening:** Administer the Occupational Physical Assessment Test again to confirm readiness levels (e.g., during basic combat training [BCT]).
    d. **Training resources:** Continue rolling out Holistic Health and Fitness resources to units and reserve components; invest in training for local levels (squad, company).
    e. **Clarify ACFT goals:** Promote acceptance of a higher ACFT standard by establishing a clear and consistent message across all organizational levels.
2. **Consider re-norming the scoring system at the same time as adopting higher standards.** The existing age- and gender-normed scoring system was developed using diagnostic test records that were not used for personnel decisions and should be updated to reflect current age and gender norms throughout the Army and to ensure that norms are appropriately set for soldiers in both noncombat and combat MOSs.
3. **Collect further evidence for minimum ACFT standards through criterion-related validation studies.** These studies should be designed to examine the relationship between each event in the ACFT as currently administered and important organizational outcomes, such as combat task performance, fitness, retention rates, and other relevant metrics. The studies should include (1) broad and diverse samples of soldiers (e.g., female and older age groups) to ensure that the findings are generalizable across target MOSs and (2) a longitudinal component to assess how changes in ACFT performance over time relate to changes in the key organizational outcomes.

Contents

About This Report iii
Summary v
Figures and Tables xi

CHAPTER 1
Introduction 1
Research Questions and Approach 2
Structure of This Report 5

CHAPTER 2
Initial Evidence for Higher ACFT Standards 7
ACFT Events and Combat Task Performance 8
ACFT Events and General Fitness 12
ACFT and Injury Prevention 14
ACFT Scoring System 14
Conclusions 15

CHAPTER 3
Practice Phase Design and Results 17
Setting Standards for the Practice Phase 17
Practice Phase Results 19
Comparison of Practice Phase and Prior Performance 20
Past Performance and Potential for Improvement 21
Subgroup Performance 22
Summary 23

CHAPTER 4
Perspectives on Higher ACFT Standards for Close Combat MOSs 27
Top Concerns Associated with Higher Standards 28
Facilitating Implementation of Higher Standards 30
Conclusion 31

CHAPTER 5
Conclusions and Courses of Action 33
Summary of Main Findings 33
Recommendations to Facilitate Implementation of Higher Standards 36
Final Thoughts 36

APPENDIXES
A. Additional Background on Sources of ACFT Evidence 39
B. Additional Details on Methodology 49
C. Additional Details on Workshop Methodology and Results 57
D. Pass Rates by Subgroups for Each Alternative Standard 63

Abbreviations 85
References 87

Figures and Tables

Figures

2.1. Overall ACFT Scores for Close Combat and Non–Close Combat MOSs 12
2.2. ACFT Event Performance, by Combat MOS Status 13
3.1. ACFT Performance During the Practice Phase (Regular Army) 19
3.2. Adjusted Practice Phase Pass Rates and Confidence Intervals for Regular Army 24
3.3. Adjusted Practice Phase Pass Rates and Confidence Intervals for ARNG and USAR 25
4.1. Ratings of Concerns Related to the Practice Phase Standard for Relevant Subgroups 29
A.1. Timeline of Major ACFT and Related Studies 41
A.2. Hypothetical Relationship Between Squat Lift and Casualty Drag Speed 46
B.1. Business Rules Used to Construct Historical and Practice Phase ACFT Samples 51
B.2. Effect Estimates from the Two ACFT Improvement Models 56
D.1. Pass Rates and Confidence Intervals for Regular Army, by Gender and Age Group 64
D.2. Pass Rates and Confidence Intervals for Regular Army, by MOS 65
D.3. Pass Rates and Confidence Intervals for ARNG and USAR, by Gender and Age Group 74
D.4. Pass Rates and Confidence Intervals for ARNG and USAR, by MOS 75

Tables

S.1. Description of the Six Army Combat Fitness Test Events v
S.2. ACFT Standards to Consider for the Practice Phase vii
1.1. Description of the Six Army Combat Fitness Test Events 1
1.2. MOSs Included or Excluded in Pass Rate Analysis 2
2.1. Key Findings and Limitations for Identifying Standards for Combat Task Performance 9
3.1. ACFT Standards to Consider for the Practice Phase 18
3.2. Mean Overall ACFT Score Increases: Practice Phase Versus Past ACFT 20
3.3. Regular Army Outcomes Under Alternative ACFT Standards 22
4.1. Summary of Proposed Facilitation Strategies 30
5.1. Decision Criteria for Potential Courses of Action 34
A.1. Conversions to Estimate a 3 Repetition Maximum Deadlift 43
A.2. Outcomes Using Hypothetical Squat Lift Standard 47
B.1. Recommended Practice Phase Sample Sizes, by Component and Gender 53
D.1. Pass Rates and Confidence Intervals for Regular Army—Standard 1A: 420 Total 66
D.2. Pass Rates and Confidence Intervals for Regular Army—Standard 1B: 450 Total 67
D.3. Pass Rates and Confidence Intervals for Regular Army—Standard 2A: 140 MDL 68
D.4. Pass Rates and Confidence Intervals for Regular Army—Standard 2B: 150 MDL 69
D.5. Pass Rates and Confidence Intervals for Regular Army—Standard 3A: 420 Total + 140 MDL 70
D.6. Pass Rates and Confidence Intervals for Regular Army—Standard 3B: 420 Total + 150 MDL 71
D.7. Pass Rates and Confidence Intervals for Regular Army—Standard 3C: 450 Total + 140 MDL 72
D.8. Pass Rates and Confidence Intervals for Regular Army—Standard 3D: 450 Total + 150 MDL 73
D.9. Pass Rates and Confidence Intervals for ARNG + USAR—Standard 1A: 420 Total 76
D.10. Pass Rates and Confidence Intervals for ARNG + USAR—Standard 1B: 450 Total 77
D.11. Pass Rates and Confidence Intervals for ARNG + USAR—Standard 2A: 140 MDL 78
D.12. Pass Rates and Confidence Intervals for ARNG + USAR—Standard 2B: 150 MDL 79
D.13. Pass Rates and Confidence Intervals for ARNG + USAR—Standard 3A: 420 Total + 140 MDL 80

D.14. Pass Rates and Confidence Intervals for ARNG + USAR—Standard 3B: 420 Total + 150 MDL … 81
D.15. Pass Rates and Confidence Intervals for ARNG + USAR – Standard 3C: 450 Total + 140 MDL… 82
D.16. Pass Rates and Confidence Intervals for ARNG + USAR – Standard 3D: 450 Total +150 MDL … 83

CHAPTER 1

Introduction

The Army Combat Fitness Test (ACFT), consisting of six events (Table 1.1), was implemented as the official fitness test in the Regular Army on October 1, 2022, and for the reserve components in April 2023. Initially designed with gender- and age-neutral standards, the first iteration of the ACFT was tailored to military occupational specialties (MOSs) in three tiers based on the physical demands of the job: heavy, significant, and moderate. However, following concerns from Congress and an independent evaluation by RAND, changes were made prior to full implementation, including a new scoring system.[1] This revised scoring mirrors the previous Army Physical Fitness Test (APFT), which tailored scoring and standards to different age and gender groups.

Adopting age- and gender-normed scoring led to renewed concerns about whether soldiers will be physically ready for their MOS duties. The Fiscal Year (FY) 2024 National Defense Authorization Act (NDAA)

TABLE 1.1
Description of the Six Army Combat Fitness Test Events

ACFT Event	Description
3 Repetition Maximum Deadlift (MDL)	Deadlift the maximum weight possible three times.
Standing Power Throw (SPT)	Throw a 10-pound medicine ball backward and overhead for distance
Hand Release Push-Up—Arm Extension (HRP)	Complete as many hand-release push-ups as possible in two minutes, which requires extending arms to a T-position for each repetition when in down position
Sprint-Drag-Carry (SDC)	Conduct five 50-meter shuttles for time: sprint, drag, lateral, carry, and sprint
Plank (PLK)	Maintain a proper plank position for as long as possible
Two-Mile Run (2MR)	Run two miles for time on a measured, generally flat outdoor course

SOURCE: U.S. Army, "Army Combat Fitness Test," webpage, undated.

[1] Chaitra M. Hardison, Paul W. Mayberry, Heather Krull, Claude Messan Setodji, Christina Panis, Rodger Madison, Mark Simpson, Mary Avriette, Mark E. Totten, and Jacqueline Wong, *Independent Review of the Army Combat Fitness Test: Summary of Key Findings and Recommendations*, RAND Corporation, RR-A1825-1, 2022.

As stated in Section 598 of the National Defense Authorization Act (NDAA) for Fiscal Year (FY) 2021:

> The Secretary of the Army may not implement the Army Combat Fitness Test until the Secretary receives results of a study, conducted for purposes of this section by an entity independent of the Department of Defense, on the following: (1) The extent, if any, to which the test would adversely impact members of the Army stationed or deployed to climates or areas with conditions that make prohibitive the conduct of outdoor physical training on a frequent or sustained basis. (2) The extent, if any, to which the test would affect recruitment and retention in critical support military occupational specialties of the Army, such as medical personnel. (Public Law 116-283, William M. [Mac] Thornberry National Defense Authorization Act for Fiscal Year 2021, Section 598, Limitations on Implementation of Army Combat Fitness Test, January 1, 2021)

aimed to address these concerns by calling for higher minimum fitness standards for soldiers in close combat MOSs. Specifically, Section 577 of the NDAA states the following:

> (a) Implementation.--Not later than 18 months after the date of the enactment of this Act, the Secretary of the Army shall implement increased minimum fitness standards as part of the Army Combat Fitness Test for all soldiers of the following military occupational specialties or areas of concentration:
>
> (1) 11A. (2) 11B. (3) 11C. (4) 11Z. (5) 12A. (6) 12B. (7) 13A. (8) 13F. (9) 18A. (10) 18B. (11) 18C. (12) 18D. (13) 18E. (14) 18F. (15) 18Z. (16) 19A. (17) 19C. (18) 19D. (19) 19K. (20) 19Z.
>
> (b) Briefing.--Not later than 365 days after the date of the enactment of this Act, the Secretary of the Army provide a briefing to the Committees on Armed Services of the Senate and House of Representatives describing the methodology used to establish standards under subsection (a). (Public Law 118-31, National Defense Authorization Act for Fiscal Year 2024, December 22, 2023)

Although Special Forces MOSs (18 series) were included in the mandate, we excluded these MOSs in the evaluation of pass rates because our analysis showed that soldiers within the Special Forces community have uniformly high ACFT scores. By restricting our analyses to the remaining close combat branches (Table 1.2), a more accurate assessment can be made about the impact of different ACFT standards on pass rates. Also, although not included in the FY 2024 NDAA, we include 13B Cannon Crewmember among the MOSs in our analysis because of their MOS-specific task requirements.

Research Questions and Approach

The Army asked us to provide courses of action (COAs) to help address the requirements outlined in FY 2024. To develop COAs, we addressed the following questions:

- What evidence is available to establish gender- and age-neutral ACFT standards for close combat MOSs?
- What evidence is available to establish higher ACFT standards to maintain or strengthen general fitness among soldiers in close combat MOSs?
- What impact will higher ACFT standards have on different subgroups of soldiers, including gender, age group, MOS, and component (Regular Army, Army National Guard [ARNG], and U.S. Army Reserve [USAR])?

TABLE 1.2
MOSs Included or Excluded in Pass Rate Analysis

Included MOSs	Excluded MOSs
11A: Infantry Officer	18A: Special Forces Officer
11B: Infantryman	18B: Special Forces Weapons Sergeant
11C: Indirect Fire Infantryman	18C: Special Forces Engineer Sergeant
11Z: Infantry Senior Sergeant	18D: Special Forces Medical Sergeant
12A: Engineer Officer	18E: Special Forces Communications Sergeant
12B: Combat Engineer	18F: Special Forces Intelligence Sergeant
13A: Field Artillery Officer	18Z: Special Forces Senior Sergeant
13B: Cannon Crewmember	
13F: Joint Fire Support Specialist	
19A: Armor Officer	
19C: M2 Bradley Crewmember	
19D: Cavalry Scout	
19K: M1 Armor Crewman	
19Z: Armor Senior Sergeant	

We took a systematic approach to answering these questions, which involved the following steps:

1. reviewing published and recent RAND research on ACFT events, combat task performance, and injuries
2. analyzing historical ACFT performance Army-wide and within the close combat MOSs
3. developing COAs that are defensible, achieve Army objectives, and comply with FY 2024 requirements
4. recommending a sampling approach for a Practice Phase to "test" a sample of soldiers in close combat MOSs against a set of higher ACFT standards
5. evaluating Practice Phase results and developing a model to predict improvement
6. conducting a workshop with soldiers to document concerns about the different COAs and to elicit strategies for closing any gaps in subgroup pass rates.

Two Types of Standards

To consider the full range of COAs for higher standards, we evaluated two broad types of physical fitness standards: criterion-referenced and norm-referenced standards.[2] The first, criterion-referenced standards, or Tier II standards, assess an individual's physical ability to perform specific tasks using established benchmarks for minimally acceptable performance. The goal is to determine whether a soldier has sufficient ability to complete tasks at a required level, such as dragging a casualty 15 meters in one minute. Developing these standards requires detailed data on task demands and corresponding minimally acceptable performance levels. These standards are applied uniformly regardless of age or gender (i.e., age- and gender-neutral) to ensure that soldiers can perform tasks to the same minimally acceptable levels.

The second, norm-referenced standards, or Tier I standards, evaluate an individual's fitness relative to a target population. This approach uses a scoring system based on percentiles to compare individual performance to performance norms of subgroups, assessing general health and fitness. For example, the ACFT scores are age- and gender-normed to reflect the Army's goal of promoting overall fitness and health.

Each approach presents distinct trade-offs. Criterion-referenced standards offer a transparent measure of capability, ensuring that all soldiers have the same probability of performing a task to an acceptable level (see Appendix A for more details and an illustrative example). However, they may not effectively indicate soldiers' general health and fitness because of variations in physiology across age groups or gender. While the ACFT is age- and gender-normed to reflect the Army's broader goal to maintain and promote general fitness and health, the Army uses a variety of criterion-referenced standards. For example, the Army uses the Occupational Physical Assessment Test (OPAT) to prescreen and qualify recruits into physically demanding MOSs and high physical demand tasks (HPDTs), which soldiers are required to complete in order to graduate from MOS training.[3]

---

[2] Society for Industrial and Organizational Psychology, *Principles for the Validation and Use of Personnel Selection Procedures*, 5th ed., August 2018.

[3] Consisting of the Standing Long Jump, Seated Power Throw, Strength Deadlift (SDL), and Interval Aerobic Run (IAR), the OPAT is a pre-enlistment test designed to ensure trainees can meet the physical demands of their MOS.

Specific MOS HPDTs are managed by the relevant Centers of Excellence based on the physical requirements identified in the DA PAM 611-21 Smartbook, specifically Chapters 3A (Officers) and 10B (Enlisted) (Department of the Army Pamphlet 611-21 [Smartbook], *Military Occupational Classification and Structure*, Headquarters Department of the Army, November 2023). As an example, the seven HPDTs required to be awarded the 19K MOS, Armor Crewman, include a 12-mile foot march, casualty drag, 35m hand grenade throw, 69-lb. duffle bag lift/carry, 35-lb. sandbag lift/carry, 120mm MPAT round lift/carry, and 120mm MPAT round loading.

Highlighting the differences between Tier I and Tier II standards explains why some standards can be age- and gender-normed while others are age- and gender-neutral, which would apply common standards that reflect the physical demands of a job regardless of age or gender. Tier I standards support the overarching goal of general fitness, while Tier II standards ensure physical readiness to complete specific tasks, which can be further tailored for specific purposes (e.g., return to duty assessments).

Decision Criteria

Following professional guidelines, we considered several approaches to set higher ACFT standards for close combat MOSs.[4] To develop and compare the trade-offs among each approach, we used the following decision criteria:

1. **Defensibility** requires that data are available to support the need for higher standards across all six ACFT events. Higher standards can be supported by demonstrating statistically significant relationships (i.e., correlation) between performance on the ACFT event(s) and relevant outcome(s), such as injuries, general fitness, and combat task performance. For example, lifting more weight on a deadlift may be associated with how quickly a soldier can evacuate a casualty to safety. Demonstrating that ACFT event(s) work equally well for different subgroups (e.g., female and older soldiers) is also an important factor when evaluating the overall strength of validity evidence.[5]
2. **Combat MOS fitness** focuses on upholding and potentially enhancing the fitness and physical readiness levels observed across soldiers in combat MOSs, including individual close combat MOS, gender, and age subgroups.
3. **Combat task performance** involves the amount of evidence available to support the development of gender- and age-neutral ACFT event standards that are aligned with minimally acceptable levels of combat task performance. To establish such criterion-referenced standards, specific values defining

[4] American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, *Standards for Educational and Psychological Testing*, American Educational Research Association, 2014.

The *Principles for the Validation and Use of Personnel Selection Procedures*, 5th ed. (American Psychological Association, August 2018), provides additional guidance and definitions on these issues to include the following concepts:

> Fairness – There are multiple perspectives on fairness. There is agreement that issues of equitable treatment, predictive bias, and scrutiny for possible bias when subgroup differences are observed are important concerns . . . (p. 47)
>
> Differential prediction – The case in which use of a common regression equation results in systematic nonzero errors of prediction for subgroups. (p. 47)
>
> Bias – In a statistical context, a systematic error in a score. In discussing fairness, bias refers to variance due to contamination or deficiency that differentially affects the scores of different groups of individuals. (p. 46)

[5] Department of Defense Instruction (DoDI) 1308.03, *DoD Physical Fitness/Body Composition Program*, Office of the Under Secretary of Defense for Personnel and Readiness, U.S. Department of Defense, March 10, 2022, may not support some approaches to setting higher ACFT standards. This policy states that "standards may be adjusted for age and gender unless they are occupationally-specific" (DoDI 1308.03 3.2 [f][1]). However, this requirement is ambiguous. While this section can be read to mean that *any* standard related to specific occupations cannot be tested using an age- or gender-normed scale, other sections of the policy seem to belie this interpretation. Section 3.1 (1) of the DoDI distinguishes between health and fitness standards (i.e., Tier I standards) and "occupationally-specific, operationally-relevant physical requirements for physically demanding career fields" (i.e., Tier II standards). Under this reading, even when applied to raising minimum fitness standards for close combat MOSs, only criterion-based assessments specifically tied to the MOS's physically demanding tasks must be tested using an age- and gender-neutral standard. In addition to meeting the FY 2024 NDAA, the Army may also need to determine whether its selected approach for higher standards adheres to DoD policy.

minimally acceptable task performance are needed.[6] Once established, this type of standard increases confidence that soldiers have the physical abilities required to perform specific combat tasks.

4. **Simplicity** ensures that processes for establishing higher standards are straightforward, allowing for easy explanation and implementation with minimal administrative burden.
5. A **compensatory** system provides soldiers with the flexibility to meet the overall performance standard across events, by raising the total ACFT score requirement without mandating higher scores in specific events. A compensatory system encourages each soldier to maximize their potential by focusing on physical abilities required for different combat roles (e.g., carrying heavy weapons versus maneuvering).
6. **Subgroup outcomes** assess the extent to which standards are elevated without disproportionately affecting the pass rates of these subgroups, unless justified by evidence. That is, absent any performance evaluation data that indicate performance deficiencies, it is assumed that soldiers currently serving in close combat MOSs are performing to a minimally acceptable standard. Therefore, any differences in subgroup pass rates on new standards would need to be justified by rigorous evidence collected from validation studies.

These criteria were designed to guide the evaluation and selection of appropriate approaches for identifying higher ACFT standards for close combat MOSs. Because of the limited time and availability of Army subject-matter experts, we received only general feedback about the criteria, their definitions, and relative importance. Using this feedback, we prioritized defensibility above all other criteria.

Structure of This Report

The results of these various analyses are presented in the following chapters. Chapter 2 provides the initial evidence for setting higher ACFT standards for close combat MOSs. Chapter 3 provides the results of a Practice Phase that administered the ACFT with a set of higher standards to a large sample of soldiers. Chapter 4 summarizes concerns and feedback on various COAs for higher standards from soldiers during a stakeholder workshop. The report concludes in Chapter 5 with a summary of findings, COAs, and recommendations.

The report contains four supplemental appendixes. Appendix A provides additional details about the research we reviewed to inform our initial recommendations for the Practice Phase standard. Appendix B provides technical details on sampling methodology for the Practice Phase, statistical analysis, and models we developed to support our evaluation of different COAs. Appendix C describes the stakeholder workshop design, participants, and results in more detail, and Appendix D provides a series of figures summarizing the pass rates for different COAs for the Regular Army and for the ARNG and the USAR.

---

[6] For example, see Marilyn A. Sharp, Stephen A. Foulis, Jan E. Redmond, Maria C. Canino, Bruce S. Cohen, Keith Hauret, Peter N. Frykman, Joseph R. Pierce Jr., Richard B. Westrick, Brooke M. Pacheco, Deborah L. Gebhardt, and Edward J. Zambraski, *Longitudinal Validation of the Occupational Physical Assessment Test (OPAT)*, U.S. Army Institute of Environmental Medicine, USARIEM Technical Report T18-05, 2018. In this study, a series of combat tasks were designed with minimally acceptable performance standards (MAPSs) to be used as outcome measures. One task involved dragging a simulated casualty; "[t]he casualty drag task required recruits to drag a simulated casualty (approximately 270 lb/122.7 kg) 15 meters as fast as possible with an upper limit of 60 seconds while wearing a fighting load with a weapon (approximately 83 lb/38 kg). . . . Scores were calculated as the velocity ($m \cdot s^{-1}$) at which the dummy was moved. The MAPS was 15 meters in 60 seconds" (p. 21).

CHAPTER 2

Initial Evidence for Higher ACFT Standards

In this chapter, we explore how ACFT events are primarily linked to combat task performance, with a secondary focus on general fitness and injury prevention.[1] We evaluated the evidence available to support an increase for an ACFT event standard (or the overall point standard) and the evidence needed to recommend a specific higher age- and gender-neutral standard. The distinction between these two concepts is important. Evidence may indicate that an ACFT event is positively associated with combat task performance, but we may lack sufficient evidence to determine precisely how much higher the ACFT event standard should be set. We expand on these points below (see Appendix A for additional discussion and evidence for each event).

1. **Evidence required to increase ACFT Event and overall point standards:** We examined whether higher scores on ACFT events (or the overall test) correlate with improved outcomes, such as enhanced combat performance, better fitness and health, or reduced injury risk. This evidence could support higher standards as a means to improve these outcomes. However, simply knowing that an ACFT event correlates with an important outcome is insufficient to establish a specific standard (see next step). Higher standards can also help achieve organizational objectives by maintaining higher fitness levels among soldiers in close combat MOSs. To support these objectives, data should demonstrate that these soldiers achieve higher ACFT scores compared with soldiers in other MOSs. It is important to note that the Army can set event or overall score standards as high as needed to meet organizational goals. However, excessively high standards may have negative consequences on other objectives, such as increased injury risk, increased recruiting and retention costs to accommodate the need for more–physically fit soldiers, and reduced time available for training on other critical, non-physical tasks.
2. **Evidence required to select a specific, higher standard:** A more effective approach is to set specific standards based on defined requirements. We considered the additional requirements necessary to establish specific higher age- and gender-neutral standards for individual ACFT events. To achieve this, the Army needs to define minimally acceptable levels for desired outcomes. For example, soldiers might be expected to drag a casualty 10 meters within one minute. While some minimally acceptable performance levels have been defined by the Army, this information is often lacking, making it difficult to recommend precise standards.[2]

In the following sections, we provide an overview of the research findings that can be used to support decisions to raise ACFT standards or to set a specific higher age- and gender-neutral standard. It is important

[1] For additional details on an independent evaluation of the initial ACFT designs, see Hardison et al., 2022.

[2] For example, the Baseline Soldier Physical Readiness Requirements Study (BSPRRS) conducted by the Army to develop and validate the ACFT administered the Warrior Task and Battle Drill Simulation as a proxy for combat task performance (Whitfield B. East, David DeGroot, and Stephanie Muraca-Grabowski, *Baseline Soldier Physical Readiness Requirements Study*, Research and Analysis Division, U.S. Army Center for Initial Military Training, Technical Report T19.041-13.1, November 2019).

to note that insufficient evidence to set an age- and gender-neutral standard does not mean the event is a poor test event. As we will show, the majority of research shows direct or indirect evidence linking ACFT events to combat task performance, fitness, and injuries. However, recommending a specific age- and gender-neutral event standard is complicated by three main factors.

First, following the Army's validation data collection, significant changes were made to the ACFT, including replacing the Leg Tuck with the Plank, in addition to changes in how events were administered.[3] For example, traditional push-ups were administered during the Baseline Soldier Physical Readiness Requirements Study (BSPRRS) data collection. The hand-release component was added later to improve test administrators' ability to accurately count successful push-up repetitions.[4] Similar administration protocol changes were also made to the MDL and SDC. Such changes increase the difficulty in estimating specific defensible standards.

Second, the Army did not set a minimally acceptable performance level (i.e., maximum acceptable time) on the Warrior Task and Battle Drill—Simulation Test (WTBD-ST), which was used as "a realistic measure of the individual physical fitness expected in a combat environment."[5] As described previously, this minimally acceptable performance level is necessary to ensure that the ACFT event standard is set at the right level.

Finally, while the Army established minimally acceptable performance levels for similar combat tasks when developing and validating the OPAT, the OPAT events differ enough from the ACFT events to create challenges in converting standards. Conversion formulas provided clear and consistent recommendations for the MDL, but were less successful for the 2MR.

ACFT Events and Combat Task Performance

For each ACFT event, we discuss the primary strengths and limitations of available data to establish higher standards based on linkages to combat task performance. We organize the events by the order in which they are administered. Other than the MDL, we generally lack information required to set a specific gender- and age-neutral standard. However, the lack of evidence for setting a specific cutoff score does not mean that the event is not a valid indicator of physical fitness for combat, health, or general fitness. Key findings and limitations for setting standards are summarized in Table 2.1 and discussed in more detail in the following subsections.

3-Repetition Maximum Deadlift

The MDL is a measure of muscular strength. Muscular strength is an important fitness component for combat task performance, critical to many soldier tasks, including casualty extraction and lifting tasks. Studies have shown a moderate to strong correlation between the ability to lift more weight and higher performance on tasks related to combat performance.[6]

[3] The BSPRRS Study has multiple components: systematic review and meta-analysis by the U.S. Army Public Health Center, identification of the physical requirements of critical Warrior Tasks and Battle Drills (WTBD) and Common Soldier Tasks (CST), evaluation of the APFT, and evaluation of alternative tests against WTBD/CST performance. Additional details on these sources of evidence are provided in Appendix A and in East et al., 2019.

[4] Hardison et al., 2022.

[5] Hardison et al., 2022, p. 8.

[6] Whitfield, DeGroot, and Muraca-Grabowski, 2019; Stephen A. Foulis, Jan E. Redmond, Bradley J. Warr, Edward J. Zambraski, Peter N. Frykman, and Marilyn A. Sharp, *Development of the Occupational Physical Assessment Test (OPAT) for*

TABLE 2.1
Key Findings and Limitations for Identifying Standards for Combat Task Performance

Event	Key Findings	Key Limitations for Setting Specific Minimum Standard
3 Repetition Maximum Deadlift (MDL)	Strong evidence suggests that specific, higher MDL minimum standards could be established for close combat MOSs based on combat task performance.	Minimum standards have been validated for the OPAT Strength Deadlift (SDL). However, determining the ACFT standard requires using a conversion formula to estimate an equivalent weight for the MDL.
Standing Power Throw (SPT)	Evidence for higher minimum standards based on combat task performance is not yet established.	Soldier height is associated with SPT performance, which warrants caution for setting a gender-neutral standard.
Hand Release Push-Up (HRP)	Indirect evidence from traditional push-ups provides linkages to combat task performance.	Minimum standards based on combat task performance requires additional data and assumptions because of changes in the administration protocol that added the hand-release component.
Sprint-Drag-Carry (SDC)	Indirect evidence shows that similar events, such as the 300m sprint, relate to combat task performance.	Because of changes in the administration protocol, data do not exist to evaluate correlations between the SDC event and combat task performance.
Plank (PLK)	Validity evidence is limited.	This event replaced the Leg Tuck and was not included in the original validation studies.
Two-Mile Run (2MR)	Higher 2MR minimum standards could be established for close combat MOSs based on aerobic requirements corresponding to the OPAT Interval Aerobic Run (IAR), also known as the "beep test."	Conversions from the OPAT IAR times to the 2MR are imprecise, resulting in uncertainty for a common standard based on combat task performance.

Despite the cumulative validity evidence, information to establish a specific cutoff score is more limited. Using correlational data from OPAT development studies, the U.S. Army Training and Doctrine Command Center for Initial Military Training set a minimum score of 160 pounds for the OPAT Strength Deadlift (SDL). An OPAT validation study found that the ability to lift 140 pounds on the SDL predicted with 76 percent accuracy whether a recruit would successfully complete combat-related tasks at the end of their initial entry training.[7]

In addition to the 160-pound SDL standard that the Army selected as a minimum acceptable cutoff score, we also considered a 180-pound SDL standard. The OPAT validation study suggested that the minimum standard could be increased to as high as 200 pounds.[8] Given the considerable increase this represents from the selected cutoff score, we modeled a more conservative 180-pound SDL standard.

Combat Arms Soldiers, U.S. Army Research Institute of Environmental Medicine, USARIEM Technical Report T16-2, 2015; and Sharp et al., 2018.

[7] Sharp et al., 2018. Because this study increased SDL weight by 40-pound increments during the test, the 160-pound OPAT standard was not tested. In addition, numbers cited are for MOSs 11B, 11C, 12B, 13F, and 19D. While 19Ks showed similar results, results from 13B recruits were less predictive. Increasing SDL weight to 220 pounds increased predictive value but led to more "false fails"—that is, recruits who did not meet the SDL standard but went on to pass the combat task testing.

[8] Sharp et al., 2018. The study found that a 180-pound score increased this predictive accuracy to 83 percent, with a 16 percent *false pass* rate (i.e., the rate if at which recruits met the SDL standard but went on to fail the combat task testing) and a 1 percent *false fail* rate (i.e., the rate at which recruits did not meet the SDL standard but went on to pass the combat task testing). A 220-pound score increased predictive accuracy to 88 percent and decreased the false pass rate to 5 percent; however, the false fail rate increased to 7 percent. The authors suggested that increasing the cut-score to 200 pounds would increase predictive value while reducing false passes.

The OPAT SDL standard is a one-repetition deadlift, whereas the ACFT MDL is a three-repetition deadlift. Using a mathematical model estimating correspondence in difficulty across these two events, we identified a potential MDL cutoff score between 145 and 165 pounds.[9] Although each conversion provides slightly different estimates, we recommend that the Army consider an MDL standard of 140 or 150 pounds, for two reasons: (1) Estimates based on OPAT standards for an untrained soldier are imprecise but suggest a standard ranging from 145 to 151 pounds, and (2) the overall pass rates for close combat soldiers at an MDL standard of 150 pounds is above 95 percent.

Because the FY 2024 NDAA requires raising a standard for all soldiers, if chosen as the only increase in standards, an MDL standard of 150 pounds would be necessary to increase the minimum for male soldiers.

Standing Power Throw

The SPT is a measure of power. The event measures total body coordination and power, which are abilities that support tasks that require jumping, throwing, or sprinting. The SPT was positively linked to combat task performance in the Army's studies used to develop and validate the ACFT. There are insufficient data to inform specific gender- and age-neutral standards. Furthermore, attempts to establish such standards would need to account for findings that show that SPT scores are influenced by soldier height and technique—factors unrelated to a soldier's physical fitness.

Hand Release Push-Up

The HRP is a measure of upper body muscular endurance relevant to combat tasks that require crawling, hand-to-hand contact, and pushing obstacles. Although traditional push-ups were less strongly correlated with combat tasks in two Army studies of fitness test events and combat task performance conducted at Fort Riley and Fort Benning, previous meta-analyses provide stronger support with linkages to a range of combat tasks.[10] Despite these positive linkages, the lack of information on minimally acceptable performance for the WTBD-ST limits the evidence available to recommend a specific higher gender- and age-neutral standard for close combat MOSs.[11]

Sprint-Drag-Carry

The SDC is a measure of muscular endurance and strength, as well as anaerobic endurance and power. The specific components (i.e., sprint, drag, and carry) are relevant measures for combat task performance. The SDC approximates tasks that require soldiers to react to direct and indirect fire, build a hasty fighting position, and extract a casualty. Although the SDC may have high content validity (approximates movement patterns required on the battlefield), the specific combination of sprint, drag, and carry as a composite event was not empirically evaluated in the Army's studies to develop the ACFT. However, more recent evidence has demonstrated strong correlations with foot march times in a small sample (n = 29) of cadets from Reserve

[9] Additional information on the SDL to MDL conversion can be found in Appendix A.

[10] Appendix A contains more information on the Fort Riley and Fort Benning studies.

[11] Push-up data were collected in the OPAT studies and could be used as a first step toward estimating gender- and age-neutral push-up standards that would then be converted to HRP standards, if the original data were available for re-analysis. However, a much stronger approach that would allow for more precise standards to be set requires collecting more recent data using the specific HRP protocol administered during the ACFT and relevant combat task performance data from the same sample of close combat soldiers.

Officers' Training Corps.[12] Despite the potential value of the SDC to close combat MOSs, additional data are needed to recommend a specific, higher standard. Considering that the Army may already have a standard for maximum acceptable time to complete a tactical foot march, an SDC event standard could be set if a sufficient sample of close combat soldiers' tactical foot march times were available.

Plank

The PLK is a measure of muscular endurance. Muscular endurance is an important fitness component for combat task performance and is supported by a limited number of studies linking PLK hold times to different fitness and health measures.[13] However, data are currently lacking to ensure that higher PLK hold times support combat and fitness objectives for close combat MOS soldiers.

Two-Mile Run

The 2MR is a measure of aerobic endurance. Aerobic endurance is an important fitness component for combat task performance, and studies have shown moderate to strong correlations between aerobic endurance and the ability to perform combat tasks.[14] Despite the cumulative validity evidence, information to establish a specific cutoff score is more limited but could be estimated by converting the OPAT IAR, or "beep test," to a 2MR time with equivalent aerobic requirements. However, this conversion is less straightforward than the SDL to MDL conversion. Data are available to explore two types of conversions. In the first case, the IAR requirements are converted first to VO2max requirements and then from VO2max to a 2MR equivalent. The second case involves a direct conversion using regression formulas that predict 2MR directly from IAR shuttle completions.

These conversions resulted in a wide range of 2MR time estimates, with no clear and consistent gender- and age-neutral standard.[15] Even if a more precise 2MR time could be estimated using IAR shuttles, additional data would also be needed to determine how much to adjust the 2MR standard to account for completing it as the last event in the ACFT. Although we are confident that faster 2MR times are associated with higher combat task performance and lower injury risk, we do not have sufficient data to recommend a specific higher gender- and age-neutral standard for close combat MOSs.

---

[12] Kevin L. Withrow, Daniela A. Rubin, J. Jay Dawes, Robin M. Orr, Scott K. Lynn, and Robert G. Lockie, "Army Combat Fitness Test Relationships to Tactical Foot March Performance in Reserve Officers' Training Corps Cadets," *Biology*, Vol. 12, 2023.

[13] For example, see Megan Sax van der Weyden, Michael Toczko, Marcie Fyock-Martin, and Joel Martin, "Relationship Between a Maximum Plank Assessment and Fitness, Health Behaviors, and Moods in Tactical Athletes: An Exploratory Study," *International Journal of Environmental Research and Public Health*, Vol. 19, 2022.

[14] Veronique Hauschild, David DeGroot, Shane Hall, Karen Deaver, Keith Hauret, Tyson Grier, and Bruce Jones, *Correlations Between Physical Fitness Tests and Performance of Military Tasks: A Systematic Review and Meta-Analyses*, U.S. Army Public Health Command, PHR No. 12-02-0614, June 2014; East, DeGroot, and Muraca-Grabowski, 2019; Foulis et al., 2015; and Sharp et al., 2018.

[15] Predicting 2MR times from IAR shuttles would result in different minimum standards for male and female soldiers. Furthermore, the predictions were relatively imprecise with an average error of approximately +/– 1 minute in predicted 2MR times. See Maria C. Canino, Bruce S. Cohen, Jan E. Redmond, Marilyn A. Sharp, Edward J. Zambraski, and Stephen A. Foulis, "The Relationship Between Soldier Performance on the Two-Mile Run and the 20-m Shuttle Run Test," *Military Medicine*, Vol. 183, No. 5/6, 2018.

Studies indicate that VO2 requirements for 43 shuttles on the IAR ranged from 33.6 to 42.88. See Alan M. Nevill, Roger Ramsbottom, Gavin Sandercock, Carlos Eduardo Bocachica-González, Robinson Ramírez-Vélez, and Grant Tomkinson, "Developing a New Curvilinear Allometric Model to Improve the Fit and Validity of the 20-m Shuttle Run Test as a Predictor of Cardiorespiratory Fitness in Adults and Youth," *Sports Medicine*, Vol. 51, 2021.

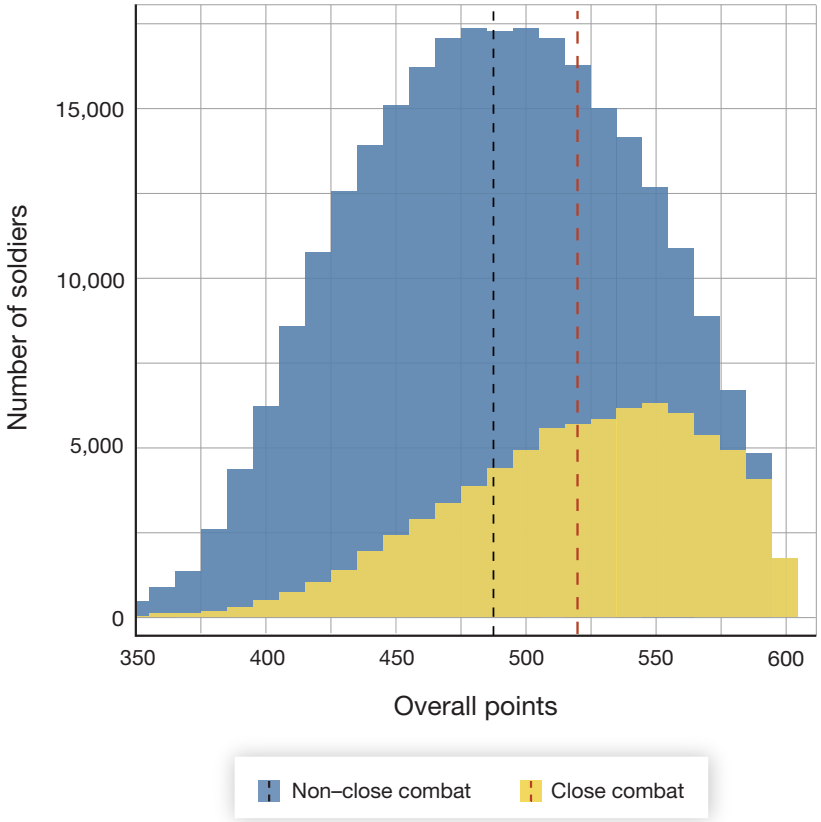
ACFT Events and General Fitness

For soldiers in close combat MOSs, overall for-record ACFT scores are, on average, 6.5 percent higher than those for soldiers in the general-purpose force. Higher ACFT standards for select combat MOSs can be implemented as a strategy to ensure that overall fitness is maintained at the higher levels demonstrated by soldiers in close combat MOSs. Therefore, setting higher ACFT standards could help maintain close combat MOS soldiers' higher fitness levels.

In Figure 2.1, we provide a histogram of overall ACFT scores. The height of each bar indicates the number of soldiers in close combat MOSs (in yellow) and soldiers in non–close combat MOSs (in blue) achieving each score. The figure shows that soldiers in close combat MOSs perform significantly better than soldiers in non–close combat MOSs. Dashed lines indicate mean performance levels, with soldiers in close combat MOSs averaging 520 points and soldiers in non–close combat MOSs averaging 488 points.[16]

Close combat soldiers, on average, outperform non–close combat soldiers on every ACFT event. Figure 2.2 presents a visual distribution of raw event performances, showing outcomes before applying the scoring point system. These raw performances are presented in pounds, repetitions, and completion time. Boxplots

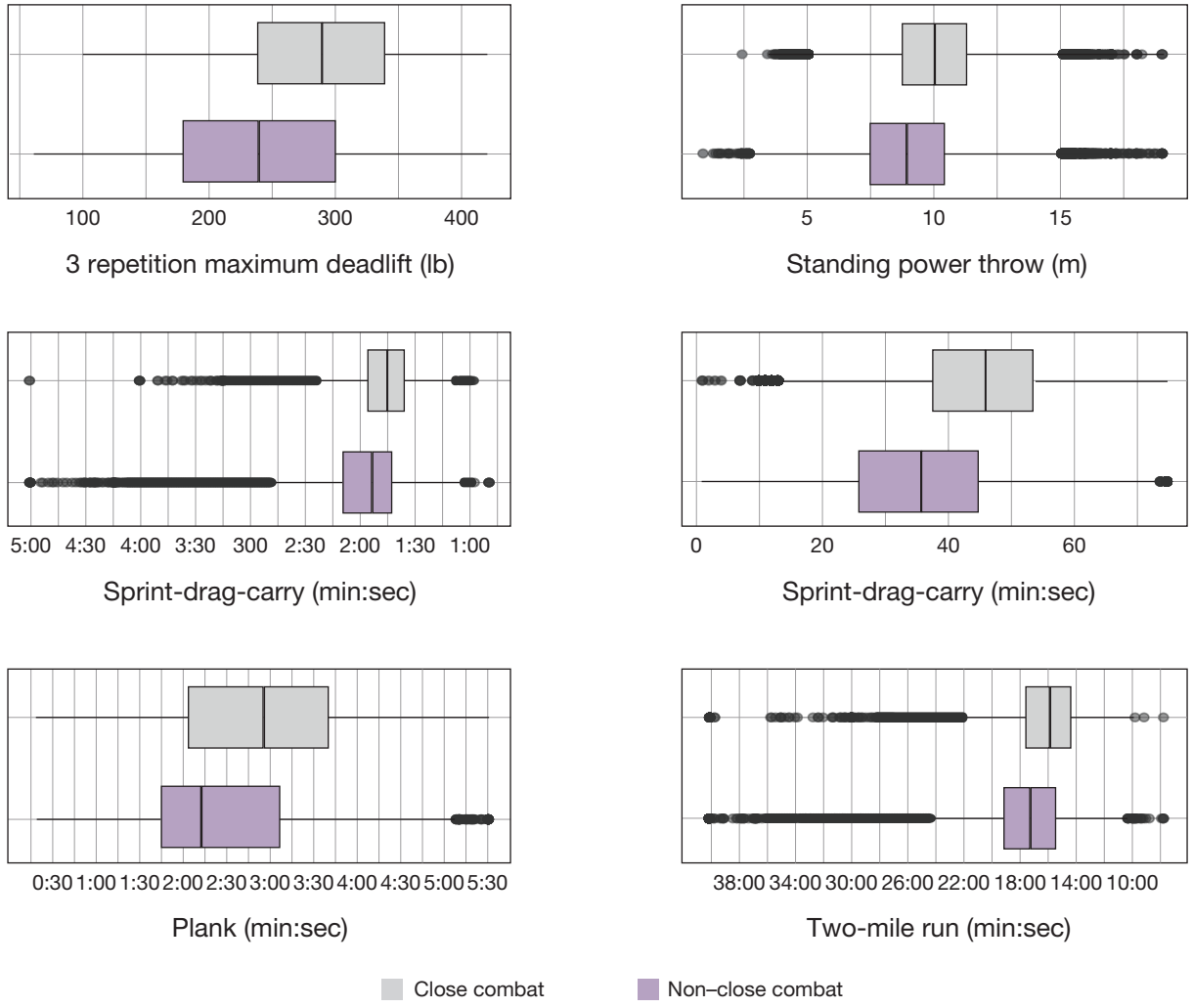
FIGURE 2.1
Overall ACFT Scores for Close Combat and Non–Close Combat MOSs



NOTE: Business rules and sample information are included in Appendix B.

[16] Soldiers in close combat MOSs had a median overall score of 526 points; soldiers in non–close combat MOSs had a median overall score of 489 points.

FIGURE 2.2
ACFT Event Performance, by Combat MOS Status



NOTE: Business rules and sample information are included in Appendix B.

display the data for soldiers in close combat MOSs (in gray) and in non–close combat MOSs (in purple). The solid line within each box represents the median raw score, while the entire box itself represents the performance of half of all soldiers on each event (the difference between the 25th and 75th percentile). The small dots outside the horizontal line are outliers—that is, scores outside the typical performance range.[17]

[17] The typical performance range is defined to be 1.5 times the interquartile range (IQR) above or below the 25th and 75th percentiles of performance. The IQR is the difference between the 75th and 25th percentiles of performance. Thus, an outlier would be any point that falls below $q25 - 1.5(q75 - q25)$ or above $q75 + 1.5(q75 - q25)$, where $q75$ and $q25$ represent the 75th and 25th percentiles, respectively.

ACFT and Injury Prevention

Increases in physical fitness standards may be justified as an organizational goal if those standards are associated with improvements in health, such as a reduction in injury rates. For instance, over the course of 2021, over half of all U.S. Army soldiers sustained an injury, three-quarters of which were overuse injuries.[18] Injuries have a notable impact on readiness, and prior research has estimated that at any given time, approximately 10 percent of soldiers cannot be deployed for associated medical and administrative reasons.[19]

The ACFT as a complete test with current administration protocols and standards has been evaluated against injury outcomes by RAND Arroyo Center researchers.[20] This research demonstrated that passing any of the six ACFT events is associated with reduced injury risks. Furthermore, among many ACFT events, soldiers with higher performance levels exhibit lower rates of injury risk, although these findings are nuanced. For example, the observed association of increased performance with reduced risk is strongest for the 2MR. It also exists for the MDL, HRP, PLK, and SDC.[21] One exception is the SPT, for which injury rates were higher among soldiers with higher performances.

Although the Army may raise standards on the basis of injury risk reduction, choosing a precise minimum standard requires the Army to first specify an acceptable injury risk threshold. For example, if the Army were to choose a precise target for reduced musculoskeletal injury rates—such as cutting such events by 5 percentage points from 52 percent annually to 47 percent annually—then future analysis could examine improved minimum standards for fitness as one approach to achieve this goal.

ACFT Scoring System

Approaches that raise ACFT standards for close combat soldiers within the existing age- and gender-normed system may face a number of challenges. Under an age- and gender-normed scoring scale, the performance of soldiers is compared with others of their respective age group and gender. This means that soldiers from different age and gender groups can earn the same number of event points but have different raw score performance. For example, a male soldier between 37 and 41 years old can earn 90 points in the MDL by lifting 270 pounds, but a younger male soldier between 27 and 31 years old needs to lift 310 pounds to earn the same 90-point score. If fully age- and gender-normed, we would expect ACFT event averages to be relatively similar across age and gender even though raw performance on each event differs.

The existing scoring system, "based on ACFT scores reported during FY 21," used diagnostic test records from all MOSs.[22] Several factors are worth consideration:

---

[18] U.S. Army, *2022 Health of the Force*, 2022.

[19] See Tanekkia M. Taylor-Clark, Lori A. Loan, Pauline A. Swiger, Larry R. Hearld, Peng Li, and Patricia A. Patrician, "Predictors of Temporary Profile Days Among U.S. Army Active Duty Soldiers," *Military Medicine*, Vol. 188, No. 5-6, May–June 2023.

[20] For details, see Daniel Hicks, Carra S. Sims, Mary Avriette, Max Steiner, and Sarah Baker, *The Army Combat Fitness Test (ACFT) and the Health of the Active Component: Understanding the Link Between the ACFT and Personnel Health and Injuries,* RAND Corporation, RR-A2005-1, forthcoming.

[21] Observed associations were shown to differ between male and female soldiers. In addition, acute trauma injuries, which are rare and low in baseline frequency, were slightly more common among the highest performers of MDL and SDC, potentially arising from behavioral factors such as overexertion. Additional event specific detail can be found in Appendix A.

[22] Whitfield B. East, *Fit to Serve: A History of US Army Physical Readiness*, Army University Press, U.S. Army Combined Arms Center, 2024, p. 263.

1. **ACFT points awarded for performance do not reflect current Army subgroup norms.** Under the current system, pass rates and scores are not consistently normed across age groups. For example, younger soldiers, particularly those ages 17–21, exhibit lower pass rates and mean scores than older soldiers. Furthermore, both pass rates and mean scores for female soldiers are consistently lower than for male soldiers across all age groups.
2. **Event standards do not always reflect the physical fitness levels of soldiers at different ages.** An age- and gender-normed system should account for variation across subgroups in underlying fitness. For instance, if younger soldiers need time to develop fitness, this accommodation should be reflected in the scoring system, and the norms assigned to these groups should have minimal to no gaps in pass rates and scores across subgroups.
3. **The scoring system does not equally incentivize improvement across the full range of fitness.** Some event-specific increases for select combat MOSs might require substantially different changes in fitness across age and gender groups, potentially affecting perceptions of equity or fairness. In addition, there is a lack of specific point values at key thresholds for some events (e.g., most groups cannot score exactly 70 points on the MDL). The following examples illustrate these points:
    a. Improving the 2MR from 60 to 61 points requires as much as about 1 minute of improvement, whereas increasing from 60 to 61 on the PLK requires only a 3-second longer hold.
    b. Increasing by a set number of event points does not result in similar rates of performance improvement among subgroups. For example, to move from 60 to 70 points, a 21-year-old male soldier must increase by 18 repetitions on the HRP, whereas a 27-year-old female soldier only needs 4 additional repetitions.

As these concerns illustrate, features of the existing scoring system make it significantly more complicated to select higher event-specific age- and gender-normed minimum standards for close combat soldiers. Because these issues may add complexity to the implementation of higher standards for the close combat force, updating the scoring system at the same time as raising standards would help ensure that norms are appropriately set for soldiers in both close combat and non–close combat MOSs.

Conclusions

After reviewing existing research and comparing ACFT performance between soldiers in close combat MOSs and non–close combat MOSs, we provided the following preliminary conclusions:

- Although combat task performance evidence is available to support individual ACFT events, the quality and quantity of evidence varies by event.
- The ACFT as a complete test with current administration protocols and standards has been evaluated against injury outcomes but not combat task performance. Passing any of the six ACFT events is associated with reduced injury risks. Higher performance levels are associated with lower injury risk, but only for some events and populations. However, setting standards based on injury risk requires the Army to first specify an acceptable injury risk threshold. In Table 2.1, we summarize key findings and limitations focused on ACFT event performance and combat task performance.
- Higher ACFT standards for select combat MOSs can be implemented as a strategy to ensure that overall fitness is maintained at the higher levels demonstrated by soldiers in close combat MOSs. Our analysis indicated that overall ACFT scores for soldiers in close combat MOSs are, on average, 6.5 percent higher

than those for soldiers in the general-purpose force based on historical performance since implementation for Regular Army soldiers on October 1, 2022.

- Furthermore, the current ACFT scoring system presents significant limitations for options that would increase event standards using points on specific events.

CHAPTER 3

Practice Phase Design and Results

To collect evidence in support of higher standards for the close combat force, the Army undertook an ACFT Practice Phase, during which higher potential standards were piloted among a sample of soldiers in close combat MOSs.[1] The Army initiated the Practice Phase with an April 30, 2024, executive order indicating that the phase would run from May 1, 2024, to June 30, 2024. Ultimately, the Army continued data collection through August 2024, although the results presented here go from May 1, 2024, to August 8, 2024. The Army targeted specific Regular Army and ARNG units for data collection. RAND Arroyo Center recommended that the Army sample units from each component to reach approximately 21,000 soldiers in total (see Appendix B for an overview of the sampling plan).

Setting Standards for the Practice Phase

Based on the existing evidence, there are two primary approaches that can be considered for evaluation in the Practice Phase designed to test a sample of close combat soldiers against a higher standard. Each has its own set of implications, and both were developed to address concerns regarding the existing ACFT scoring system and general guidelines for an overall close combat MOS pass rate of 95 percent and subgroup pass rate of 90 percent or higher. These guidelines were set based on initial discussions with Army leaders and can be adjusted to meet a variety of Army objectives to include force structure, injury prevention, general fitness, resource requirements, and organizational culture. There are often trade-offs in setting standards based on expected pass rates. For example, higher standards may support objectives to reinforce a fitness culture but may require more resources for training to achieve those standards.

1. **Increase the overall point standard using the existing age- and gender-normed scoring system.** Based on recent close combat soldier performance on the ACFT and linkages to general health and fitness, we recommend raising the total required ACFT score from 360 to 420 overall points for soldiers in select combat MOSs. This recommendation does not change event standards for any individual event but rather increases the overall point threshold that soldiers must achieve. This method retains the age- and gender-normed scoring system, which can be used to adjust for age and gender differences in physical performance, while also elevating general fitness and combat readiness standards across the board.
2. **Increase standards on MDL using a gender- and age-neutral standard.** Strong evidence supports setting a minimum gender- and age-neutral standard of 140 or 150 pounds for the MDL event, based on combat task requirements. This approach emphasizes the establishment of a specific, uniform

[1] On April 30, 2024, an execute order (Headquarters Department of the Army, EXORD 134-24, "Practice Physical Fitness Test," April 30, 2024) was sent out to commence the Practice Phase. The Army has opted to extend the duration of the Practice Phase to continue increasing the pool of available data to inform decisions.

standard for the MDL event that applies to all soldiers, regardless of gender or age, aligning with the objective of maintaining combat readiness.

The Army's Decision for the Practice Phase Standard

In response to our recommendations, the Army inquired about the potential benefits of applying a higher standard of 450 overall points on the ACFT during the Practice Phase. The primary advantage of a Practice Phase is to develop better estimates of soldier capabilities. We determined that raising the standard to 420 overall points would yield limited additional information, because most close combat subgroups were already achieving pass rates exceeding 90 percent at a 420 overall-point standard. However, there would be less certainty about pass rates if the standard was set at 450 overall points. Additionally, increasing the standard during the Practice Phase could heighten the risk of injury, especially with limited time for training or improvement. Ultimately, the Army decided to implement a Practice Phase standard of 450 overall points, along with a requirement of 150 pounds on the MDL during the Practice Phase.

As we will discuss in Chapter 4, in the Standards Workshop held in September 2024, although the April 30 execute order commencing the Practice Phase was distributed, numerous attendees indicated that they had been directed to take an ACFT between May and August 2024, consistent with the Practice Phase, but that they had not been informed of the need to reach a higher standard or of the specific standard to target. Participants in the Practice Phase may have performed better had they been aware of the higher goal standard. Therefore, results from the Practice Phase should be regarded as a lower bound estimate of future pass rates for standards up to 450 overall points and an MDL requirement of 150 pounds.

In the following sections, we present the results of the Practice Phase in addition to several alternative standards (Table 3.1). These alternatives explore raising only the overall point standard (1A, 1B), raising only the MDL (2A, 2B), and raising both the overall and MDL by different amounts (3A–3D).[2]

TABLE 3.1
ACFT Standards to Consider for the Practice Phase

Option	Standard-Setting Approach	Standard
1A	Norm referenced	420 total
1B	Norm referenced	450 total
2A	Criterion referenced	140 MDL
2B	Criterion referenced	150 MDL
3A	Norm + criterion referenced	420 total + 140 MDL
3B	Norm + criterion referenced	420 total + 150 MDL
3C	Norm + criterion referenced	450 total + 140 MDL
3D	Norm + criterion referenced (Practice Phase)	450 total + 150 MDL

NOTE: Norm-referenced standards are based on age and gender groups. Criterion-referenced standards are based on tasks performed on the job.

[2] We do not include options that raise the standards only for single events based on point values within the existing age- and gender-normed scoring system because of challenges with this approach discussed in Chapter 2.

Practice Phase Results

The Army provided RAND Arroyo Center with 47,744 Practice Phase test records spanning May 1, 2024–August 8, 2024. Limiting to one test per soldier, as of August 8, 2024, the Army was able to obtain 35,138 test records from the Regular Army, 8,653 soldiers in the ARNG, and 502 soldiers in the USAR.

The overall pass rate under the 450 overall points + MDL 150 pounds Practice Phase standard for Regular Army soldiers in close combat MOSs was 91.4 percent (91.1%–91.6% CI[3]). Of these, 33,967 records were from male soldiers with an estimated pass rate of 91.5 percent (91.2%–91.7% CI) and 1,171 were from female soldiers with an estimated pass rate of 88.5 percent (86.9%–90.0% CI). Performance data from the ACFT Practice Phase for the overall ACFT and the MDL event are visually depicted in Figure 3.1 for male and female soldiers separately (male soldiers, left panel; female soldiers, right panel). Each hexagon represents a combination of raw scores on the MDL (x-axis) and overall ACFT points (y-axis), and shading of the hexagon represents the number of soldiers scoring in that range. Practice Phase standards are indicated by dashed lines: The horizontal dashed line denotes an overall score of 450, and the vertical dashed line denotes an overall score of 150 on the MDL.

FIGURE 3.1
ACFT Performance During the Practice Phase (Regular Army)

NOTE: Business rules and sample information are included in Appendix B. Colors indicate the number of soldiers scoring in a specific range. Dashed lines reflect potential cutoffs of 450 points overall and 150 pounds on MDL performance. There were 171 male and 7 female soldiers who met the Practice Phase standard for the MDL but failed an event other than the MDL; these soldiers appear in the top right quadrant, but ultimately failed the Practice Phase standard.

[3] *CI* refers to a 95 percent confidence interval. The CI quantifies how much uncertainty there is around our estimate of the pass rate. If we imagined the Practice Phase simultaneously happening an infinite number of times with different samples of soldiers and estimated a CI on each of these simultaneous Practice Phases, we would expect 95 percent of those estimated CIs to contain the true pass rate.

Each figure is annotated to indicate the number and percentage of male and female soldiers in four quadrants:

- top left = failed MDL standard but passed overall ACFT score standard
- top right = passed both MDL standard and overall ACFT score standard
- bottom left = failed both MDL standard and overall ACFT score standard
- bottom right = passed MDL standard but failed overall ACFT score standard.

By summing the two right quadrants in each chart of Figure 3.1, over 99 percent of male soldiers and almost 93 percent of female soldiers pass the MDL standard. By summing the top two quadrants in each chart, just over 92 percent of male soldiers and 93 percent of female soldiers pass the overall 450 overall points standard. When considering only the Practice Phase standard of both 450 points overall and 150 pounds on the MDL (but not performance on the other five ACFT events), we observe pass rates of 92 percent for men and 89.1 percent for women, as depicted in the top right quadrant only.[4] Subgroup scores are detailed in Appendix D.

Comparison of Practice Phase and Prior Performance

Evidence of improvement on the ACFT was observed when comparing close combat soldiers' Practice Phase ACFT scores with their most recent for-record attempt taken 10–14 months prior to the Practice Phase.[5] These patterns are depicted in Table 3.2. Among all Practice Phase participants in the Regular Army, ACFT

TABLE 3.2
Mean Overall ACFT Score Increases: Practice Phase Versus Past ACFT

Gender	MOS	Age Group	Component	Participants Who Did Not Previously Meet Standard	All Participants
Male and female	All close combat	All	Regular Army	15.8	1.8
Male and female	All close combat	All	ARNG + USAR	3.2	–3.0
Male and female	19K	All	Regular Army	2.4	–2.3
Female	All close combat	All	Regular Army	43.5	8.4
Female	All close combat	All	ARNG + USAR	0.9	–4.8
Female	All close combat	17–21	Regular Army	51.8	15.1

NOTE: Business rules and sample information are included in Appendix B.

[4] These pass rates consider only passing the overall points and MDL requirements. A small number of soldiers met the overall points and MDL requirements but failed one of the other five ACFT events and thus did not meet the full Practice Phase standard (nor the existing ACFT standard). Pass rates that account for these soldiers are 91.5 percent for male soldiers and 88.5 percent for female soldiers, as presented above.

[5] For-record tests are eligible to be used for personnel actions, whereas diagnostic tests are not. For this reason, for-record tests may be more representative of soldier's fitness outcomes when fully incentivized to perform.

performance increased by only 1.8 points relative to their previous for-record performance.[6] ACFT scores in the Practice Phase were worse, on average, for 19K (–2.3 points) and ARNG and USAR (–3.0 points) close combat soldiers.

Participants Who Did Not Previously Meet Practice Phase Standard

We observed greater improvements among a subset of Practice Phase soldiers who previously scored below 450 points overall and lifted less than 150 pounds on the MDL in their most recent for-record test. These soldiers who previously did not meet the higher standard improved, on average, by 15.8 points relative to their most recent test 10–14 months pre-Practice Phase, with a 51.8-point increase for female soldiers ages 17–21 in the Regular Army. Score improvements were smaller among ARNG and USAR soldiers in this subset. The variation in the rate of performance may need to be further explored by the Army to determine whether certain factors limit ARNG and USAR soldiers' ability to improve at the same pace as Regular Army soldiers (e.g., training resources).

An important caveat is that we cannot attribute all observed performance improvements to soldier responses to the higher Practice Phase standard. Soldiers have been shown to improve with experience,[7] and soldiers with less experience were far more commonly in the group that did not previously meet the standard—so a portion of the improvement shown in these results may have been expected to occur even in the absence of the Practice Phase. The next section examines additional evidence on potential improvement more directly.

Past Performance and Potential for Improvement

Additional analysis suggests that future pass rates may be higher than performance outcomes during the Practice Phase. In Table 3.3, we summarize performance outcomes under the eight alternative standard options introduced in Table 3.1. These outcomes are based on potential pass rates, which adjust for prior performance and potential improvement. Each row depicts an alternative set of higher ACFT standards. To aid interpretation, we provide a brief explanation of the data in the last two columns:

- **+ Crediting Prior Performance:** Presents overall adjusted pass rates that assume soldiers who did not meet the standard during the Practice Phase but met the standard in a previous for-record test would pass again in the future.
- **+ Possible Improvements:** Presents overall adjusted pass rates that credit prior performance and improvements expected by soldiers who are predicted to be within range of meeting the standard within one year.[8]

---

[6] Among those soldiers who had a for-record test 10–14 months prior,

- 51 percent of soldiers improved their total score in the Practice Phase relative to their prior performance
- 35 percent improved their raw MDL performance in the Practice Phase relative to their prior performance.

[7] See Daniel Hicks and Sean Robson, *Implementation of the Army Combat Fitness Test: An Updated Analysis and Future Directions*, RAND Corporation, RR-A2232-1, forthcoming.

[8] To assess potential for continued improvement, we modeled ACFT performance for individual soldiers over time in order to predict ACFT performance in September 2025 (one-year post-Practice Phase). The assumption is that within one year participants will have had more time to train to a higher standard. This model incorporates date of first ACFT; age at first ACFT; performance (either total score or raw MDL) on first ACFT; passage of time since first ACFT; how many ACFTs had been taken and whether each ACFT was for-record or diagnostic; gender, MOS, and component; interactions between gender and age; passage of time since first ACFT and age; passage of time since first ACFT and gender; and passage of time since first

TABLE 3.3
Regular Army Outcomes Under Alternative ACFT Standards

Options			Practice Phase		Adjusted Overall Pass Rates	
Option Label	Standard-Setting Approach	Standard	Pass Rate	Subgroups Below 90% Pass Rate	+ Crediting Prior Performance	+ Possible Improvements
1A	Norm-referenced	420 Total	96.3%	19K	98.7%	99.2%
1B	Norm-referenced	450 Total	91.6%	Female soldiers ages 17–21; 12B, 13B, 13F, 19D, 19K	96.0%	97.6%
2A	Criterion-referenced	140 MDL	98.0%	—	99.4%	99.4%
2B	Criterion-referenced	150 MDL	97.2%	Female soldiers ages 17–21	99.1%	99.3%
3A	Norm + criterion-referenced	420 Total +140 MDL	96.2%	19K	98.7%	99.2%
3B	Norm + criterion-referenced	420 Total +150 MDL	95.9%	Female soldiers ages 17–21, male soldiers ages 52–56, 19K	98.5%	99.1%
3C	Norm + criterion-referenced	450 Total +140 MDL	91.6%	Female soldiers ages 17–21; 12B, 13B, 13F, 19D, 19K	96.0%	97.6%
3D	Norm + criterion-referenced	450 Total +150 MDL	91.4%	Female soldiers ages 17–21, male soldiers ages 47–51 and 52–56: 12B, 13B, 13F, 19D, 19K	95.9%	97.5%

NOTE: Norm-referenced standards are based on age and gender groups. Criterion-referenced standards are based on task performed on the job. Crediting prior performance and possible improvements adjust Practice Phase pass rates by making additional assumptions about how well soldiers will perform when taking the ACFT in the future.

The adjusted pass rates that credit past for-record performance and predicted improvement should be interpreted as upper bounds on what might be possible assuming no other changes in key policy, such as training time. For example, some soldiers might have hit a plateau such that significant fitness gains would be more difficult to achieve. Some degree of oscillation in fitness levels should also be expected. Finally, these pass rates reflect current training approaches. Changing the standard could result in soldiers changing how they train and developing additional test-taking strategies, which can have both positive and negative impacts that are not foreseeable until full implementation of the new standards. The Army should carefully consider the risk-benefit trade-offs associated with various options for higher standards.

Subgroup Performance

Because ACFT performance can vary for subgroups, we examined pass rates by MOS, gender, and age by gender in the Regular Army. We also examined pass rates for the same subgroups for soldiers in the ARNG and USAR.

ACFT and performance on first ACFT estimate. See Appendix B for a detailed discussion of the methodology employed to estimate predicted improvement over 12 months.

MOS, Gender, and Age by Gender Subgroups

Although the Army was able to obtain a large overall sample for the Practice Phase, some subgroups, particularly older soldiers, had fewer test records. As a result, the confidence intervals surrounding pass rate estimates for these groups are noticeably wider. To examine potential pass rates under new higher standards, we provide estimates in Figure 3.2 under three representative approaches to higher standards: 1A: 420 Overall, depicted in the left panel; 2B: 150 MDL, depicted in the center panel; and 3D: 450 Overall and 150 MDL, depicted in the right panel. Points represent estimated pass rates, while the bars reflect 95 percent confidence intervals. These figures credit soldiers for prior performance as discussed in the previous section.

Most subgroups exhibit pass rates above 90 percent, many with confidence intervals that do not extend below this level. The confidence interval can be interpreted as a test of statistical significance. If we hypothesized that the true pass rate was 90 percent, but the estimated confidence interval fell entirely above 90 percent, we could reject the hypothesis that the true pass rate is 90 percent. Subgroups with confidence intervals that fall entirely below 90 percent or below are highlighted in red, indicating that these subgroups are likely to have a true pass rate below 90 percent. Under the higher set of potential standards used in the Practice Phase, 3D, several subgroups, notably MOS 19K and female soldiers 17–21, exhibit pass rates that lag behind that of other subgroups.

ARNG and USAR Subgroups

While subgroup concerns are limited for most Regular Army subgroups once prior performance is incorporated into estimates of outcomes, the same is not true for soldiers in the ARNG and USAR. Figure 3.3 depicts subgroup pass rates among ARNG and USAR soldiers. Under all three potential standards, multiple age and MOS subgroups exhibit pass rates below 90 percent, with pronounced challenges occurring for female soldiers and multiple MOSs under option 3D. These findings suggest that ARNG and USAR may have pass rates that are considerably lower than the Regular Army.

Summary

The evidence collected during the Practice Phase suggests that among Regular Army soldiers, several subgroups exhibit pass rates below 90 percent. Adjusting pass rates based on predictions about how much soldiers are likely to improve suggests that Regular Army soldiers will perform quite well against all COAs. However, even after considering potential improvement, our analysis suggests that many subgroups within the ARNG and USAR will struggle to meet higher standards, particularly those requiring improvement above 420 overall points. Finally, while the Army was able to obtain a substantial number of Practice Phase tests, some degree of uncertainty in the precision of pass rate estimates remains for certain subgroups, including older soldiers and soldiers in the ARNG and USAR.

FIGURE 3.2

Adjusted Practice Phase Pass Rates and Confidence Intervals for Regular Army

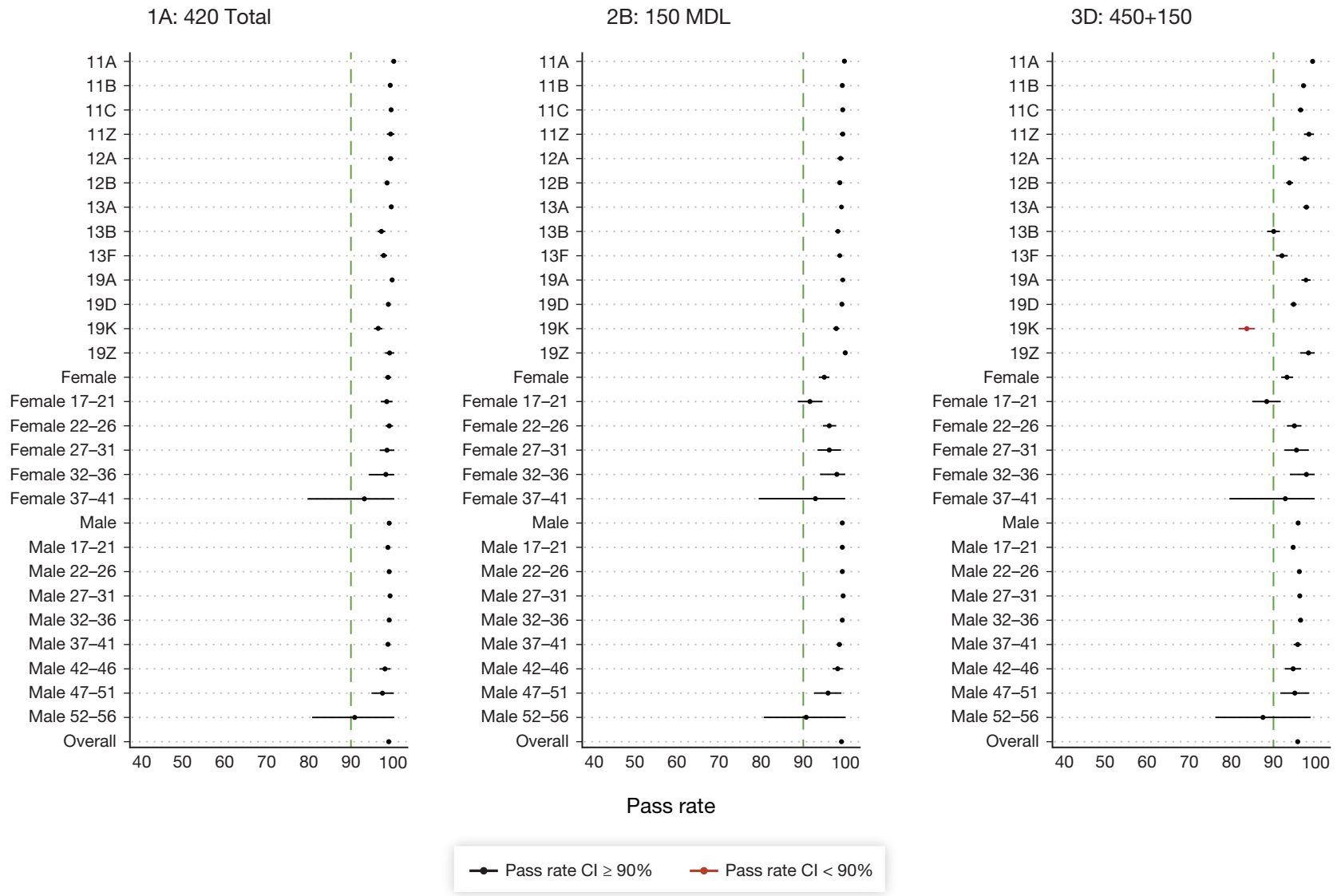
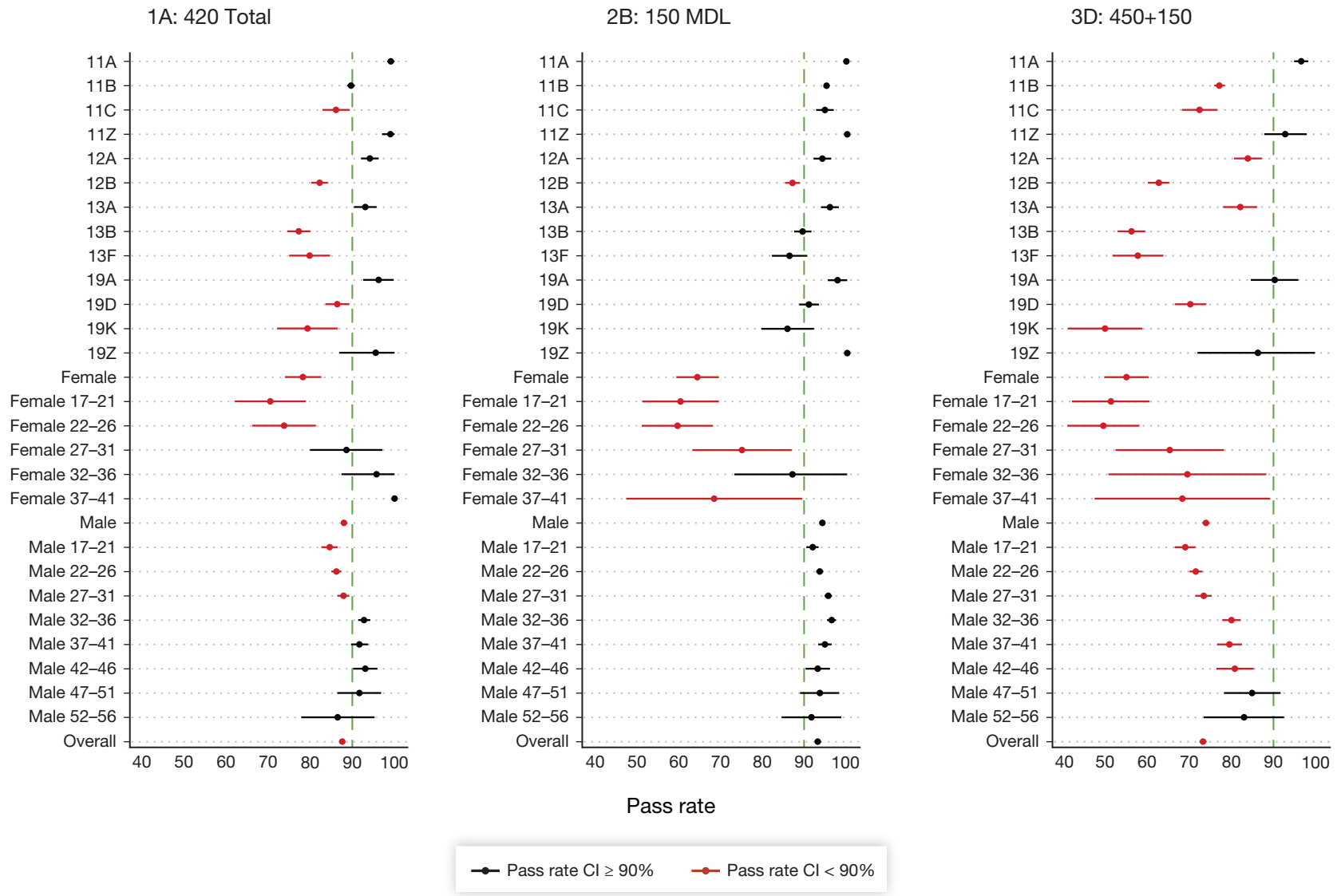


FIGURE 3.3

Adjusted Practice Phase Pass Rates and Confidence Intervals for ARNG and USAR



CHAPTER 4

Perspectives on Higher ACFT Standards for Close Combat MOSs

As part of this research on higher standards for close combat MOSs, we conducted a workshop with soldiers who reviewed the different alternative standards and offered strategies to close any gaps in subgroup pass rates. The workshop had two main objectives: (1) Identify concerns associated with implementation options for higher standards, and (2) identify policies to facilitate implementation.

To prepare for the workshop, we generated a short list of concerns that could be raised during ACFT alternative standard implementation, which was supplemented by workshop participants. The concerns examined during the workshop included the following:

- injuries (team generated)
- pass rates (team generated)
- problematic communication about ACFT scoring requirements (team generated)
- administrative simplicity (team generated)
- acceptability (team generated)
- insufficient time to train to new standards (workshop generated)
- appropriate supportive culture and incentives (workshop generated)
- challenges to recruitment and retention (workshop generated)
- impact on incentives around profiles (workshop generated)
- leadership challenges for implementing new standards (workshop generated).

Stakeholder perceptions of the variety of standards under consideration were solicited using a rating form that assessed views on the current 360-point (6 x 60) status quo, an MDL 140- or 150-pound requirement, a 420- or 450- overall point requirement, and combinations of the above, including the Practice Phase standard of 450 overall points + MDL 150 pounds (3D). Subsequently, stakeholders were presented with the most frequently chosen concerns about the various standards and discussed strategies to facilitate implementation and mitigate these concerns.

The workshop included a sample of 30 male and 8 female participants. The full range of ages were represented, including six 17–21-year-old soldiers and nine soldiers representing ages 42 and older. Relatively few participants were from the ARNG or USAR. Career management fields 11, 12, 13, and 19 were each represented by six or more soldiers. All soldiers had taken the ACFT previously, and some were or had been drill sergeants, but none were experts in exercise science per se and so represented perspectives that would be typical of the soldiers expected to implement any new standard in the field. Potential shortcomings were not insurmountable: For example, the data collection and research team included National Guard and Reserve members. We also incorporated notable expertise in exercise science on the study team to be able to speak to the feasibility of any specific training program ideas that would potentially be surfaced as part of the discussion of how best to facilitate higher standards. Thus, the assembled workshop participants were sufficient to

generate the type of information we were seeking—basic concerns about the implementation of higher ACFT standards and practical feedback on how to make that implementation go smoothly.

Top Concerns Associated with Higher Standards

Participants raised multiple concerns about the higher proposed ACFT standards for close combat MOSs. Few participants expressed concern about the impact of *which* higher standard would be selected, perhaps because so many met or exceeded the options under consideration. Thus, discussion was more generally focused on the 450 overall points + MDL 150 pounds Practice Phase standard option (3D). However, both discussion and ratings did reveal some concerns with, for example, higher MDL standards.

One repeated concern in the discussion focused on implementation timelines for the total force and entrants into the Army. Discussion centered around the need for graduated performance goals that are designed to minimize injury while maximizing long-term performance. Participants also discussed the need for appropriate resourcing to support graduated performance increases (a "glide path") to meet a new standard. For example, Holistic Health and Fitness (H2F) rollout for USAR and ARNG is still in the planning phase, and many soldiers commented favorably on the helpfulness of various H2F resources in improving ACFT performance. However, even some Regular Army participants noted that company commanders and other leaders need to be aware of resources and request them for their units, and that H2F resources are underutilized in some units.

Other concerns involved messaging relating to the ACFT. There was confusion, for example, about why the MDL was identified as a "key" close combat individual event and about the role of the ACFT as both a general fitness assessment and a metric for combat readiness. These concerns suggested that clear communication about the ACFT that incorporates a clear rationale for higher standards, transparent goals and incentives to achieve the higher standards, and clarity regarding associated policy and expectations will be essential to the success of a new standard for close combat MOSs. For example, if the ACFT is used as a retention standard to ensure that job incumbents maintain required physical fitness levels, clear and transparent communication would include describing consequences of failure. This was considered particularly important for older soldiers, who might have issues regaining fitness after injury and going on profile, for example. If a glide path to improvement is standardized for soldiers, the goals set along that glide path should be specified.

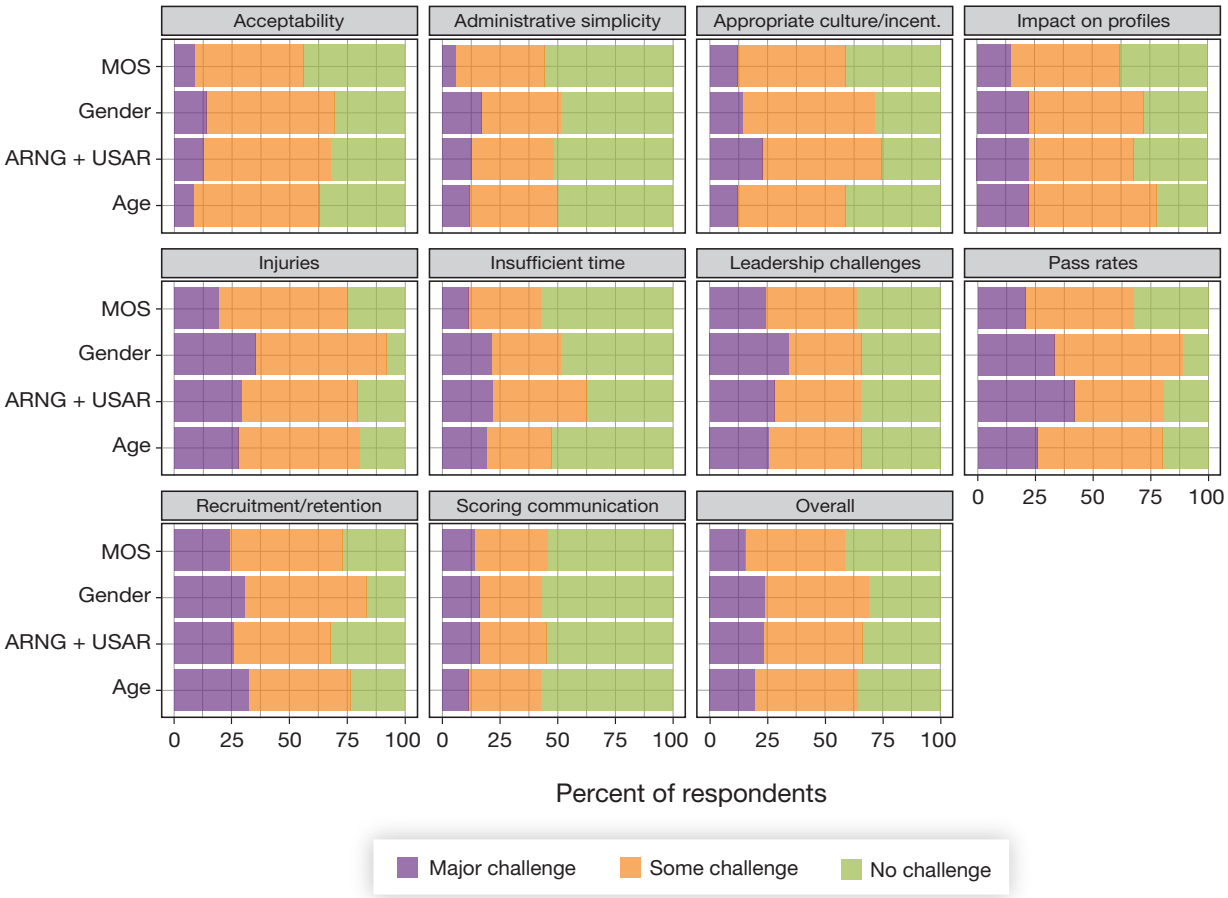
Participant ratings of the various concerns for the 450 overall points + MDL 150-pound standard revealed that, overall, concern about pass rates and concern about injuries were predominant, as Figure 4.1 illustrates. There were also concerns regarding appropriate incentives around the ACFT and instilling a culture of fitness when societal trends favor *decreasing* fitness,[1] as well as potential challenges with recruitment and retention based on standards. Scoring communication and concerns relating to insufficient time to prepare for a change were most likely to be marked as "no challenge."

Across alternative standards under consideration, some patterns held. Injuries were consistently rated as a source of concern regardless of how standards were increased (for a single event, overall ACFT, or hybrid). Pass rates were somewhat less of a concern for an option that generally increased the overall score above 360, but attaining pass rates was considered more challenging when associated with an increase in the MDL standard and when an overall score increase to 450 was specified. The fewest challenges were associated with administrative simplicity issues and scoring communication.

[1] See, for example, Regina Guthold, Gretchen A. Stevens, Leanne M. Riley, and Fiona C. Bull, "Worldwide Trends in Insufficient Physical Activity from 2001 to 2016: A Pooled Analysis of 358 Population-Based Surveys with 1.9 Million Participants," *Lancet Global Health*, Vol. 6, 2018.

FIGURE 4.1

Ratings of Concerns Related to the Practice Phase Standard for Relevant Subgroups



Workshop participants also pointed to concerns for particular subgroups. Gender subgroups (such as concern about the performance of female soldiers) were associated with injury and pass rate challenges for options that included an overall score increase to 450 points and a hybrid standard. Age subgroups (such as concern about older and younger soldiers' performance) were associated with perceived challenges regarding recruitment and retention for options that included an overall score increase to 450 points and a hybrid standard. Component subgroups (such as concern about USAR and ARNG performance) were associated with perceived challenges regarding pass rates for every option except an overall score increase above 360. Injuries were also a concern for many of the standard alternatives. MOS and rank subgroups were not associated with any particular type of concern for any of the standard alternatives.

Thus, ratings by workshop participants showed consistent concern with injuries and pass rates, and concern regarding performance in certain gender, age, and component subgroups. In particular, age was associated with recruitment and retention concerns for the more stringent standard options in the ratings. Based on ratings, there was little perceived concern relating to differences between MOS or between officers and enlisted soldiers.

Some of the most vocal concerns lay outside the scope of the study. Many participants were more concerned about the policy impact of repeated failures of the elevated close combat ACFT standard than they were about the standards themselves. While the Army will ultimately have to determine in policy the consequences of repeated failures—separation from the service, ban on reenlistment, forced reclassification, or

others—workshop participants made it clear that discussing the potential disruption of a higher close combat ACFT standard was difficult without clarity on these policies.

Facilitating Implementation of Higher Standards

After a thorough discussion of potential concerns relating to various alternative options to increase close combat MOS standards, the remaining discussion focused on developing strategies to alleviate the challenges and facilitate implementation. Table 4.1 summarizes the proposals, which are discussed in more detail in Appendix C. Goals for a standard increase must be clear, as must associated policy (such as consequences for failure); these strategies should be tempered by setting incremental goals and allowing for remedial assess-

TABLE 4.1
Summary of Proposed Facilitation Strategies

Mitigation Strategy	Brief Description	Concerns Addressed
Glide path(s)	• Soldiers need sufficient time to train and improve with minimal increases to injury risk. Two aspects are relevant: • Sufficient time is needed to allow incumbent soldiers (including USAR and ARNG) time to train up to a new standard. • New entrants (after basic combat training [BCT] or One Station Unit Training [OSUT]) to close combat MOSs need time to increase fitness and meet the new standard after arrival at their permanent duty station.	• Injuries • Insufficient training time • Pass rates for the USAR and ARNG
Clear consequences associated with standards	• Transparent and consistently enforced consequences are needed to appropriately incentivize fitness behavior and meet the new standard.	• Lack of supportive fitness culture and incentives • Lack of transparency in test-related communication
Resources to facilitate compliance	• Fitness education or access to resources such as H2F can support injury-free progress toward meeting higher standards. Even among active component soldiers, access varied: Some soldiers indicated that company commanders needed to request that fitness clinics be held at their units.	• Injuries • Training time needed for improvement • Pass rates in the ARNG and USAR
Clarification of ACFT goals	• Current ACFT messaging is inconsistent. The goals for higher ACFT standards are unclear; communication regarding higher standards needs to be consistent and transparent. Given potential confusion regarding the MDL standard, this strategy is particularly essential if the MDL included as part of the higher standards.	• Lack of transparency in test-related communication Leadership challenges for implementing new standards
Remedial assessment to show progress	• MOS-specific fitness assessments can inform a training timeline by providing information about progress toward fitness goals. A standardized, predetermined list of HPDTs for each MOS could be used to ensure transparency and consistency for soldiers who exceed the general-purpose standard but do not meet the close combat ACFT standard.	• Time to train • Injury concerns • Lack of supportive fitness culture incentives

ments where needed and providing sufficient time and other resources to allow soldiers to achieve the alternative standard safely.

Conclusion

The purpose of the workshop was to elicit potential concerns and strategies to facilitate implementation. Although the sample of soldiers was small and not fully representative, discussions highlighted that communication regarding the ACFT Practice Phase goal standard of 450 overall points + MDL 150 pounds was imprecise at best. The discussion also highlighted numerous implementation concerns that will need attention when a higher standard is put in place. Principal among these concerns were injuries and lower pass rates for different subgroups. Various facilitation strategies were proposed to address these concerns, many of which offer potential for addressing multiple concerns at once. The discussion also raised concerns surrounding related policy and practice, such as the way profile status is used, as well as challenges regarding implementation and messaging for a reality in which soldiers in close combat MOSs are in units with soldiers who in theory would *not* be subject to increased standards.

CHAPTER 5

Conclusions and Courses of Action

In this report, we have provided an assessment of the existing evidence for establishing higher fitness standards for close combat MOS soldiers, analyzed new data collected by the Army specifically for this purpose, and synthesized feedback from a stakeholder workshop. In this chapter, we conclude by summarizing the findings and providing COAs for the Army to consider regarding ACFT minimum standards for soldiers in close combat MOSs.

Summary of Main Findings

There Is Sufficient Evidence to Inform the Army's Decision on Raising ACFT Standards for Close Combat MOSs

- **The ACFT as a complete test with current administration protocols and standards has been evaluated against injury outcomes.** Passing any of the six ACFT events is associated with reduced injury risks. Higher performance levels are associated with lower injury risk for all events except the SPT. However, selecting minimum standards based on injury risk requires the Army to first specify an acceptable injury risk threshold.
- **Higher ACFT standards for select combat MOSs can be implemented as a strategy to ensure that overall fitness is maintained at the higher levels demonstrated by soldiers in close combat MOSs.** Our analysis indicated that overall ACFT scores for soldiers in close combat MOSs are, on average, 6.5 percent higher than those for soldiers in the general-purpose force based on historical performance since implementation for Regular Army soldiers on October 1, 2022.
- **For the MDL event, there is strong evidence that can inform a specific higher minimum standard for close combat MOSs based on combat task performance.** Although evidence is available to support additional individual ACFT events, the quality and quantity of evidence varies by event.

Analysis Predicts Some Soldiers Will Improve to Meet Higher Standards

- **Experience:** Prior research shows that soldiers' performance improves with experience. Prior research examining the period October 2020–October 2022 demonstrated that such gains were substantial: an average of 24.7 (male) to 38.3 (female) points on the overall ACFT score and an average of 27 (male) to 9 (female) pounds on the MDL event over the first four ACFT attempts.
- **Practice Phase:** We find evidence of performance improvement in all six ACFT events in the Practice Phase. Improvement is concentrated among the subsets of soldiers who did not meet the standard in their previous ACFT attempts and soldiers with less experience taking the ACFT. In the Regular Army, on average, relative to their prior performance, these soldiers in close combat MOSs achieved 16 more points in total and lifted 11 additional pounds on the MDL.

- **Predictive Modeling:** Our model estimates suggest that pass rates observed in the Practice Phase understate the true extent to which soldiers will continue to improve toward new standards. Adjusting for improvement, fewer subgroups, particularly those in the Regular Army, fall below a 90 percent pass rate under the COAs considered.[1]

Results Highlight Trade-Offs Between Alternative Standards

A number of higher fitness standard options satisfy important decision criteria and are reasonable COAs for the Army. In this report, we focused on eight alternative higher standards. Using our professional judgment to weigh the evidence across criteria, we summarize our analysis of these of standards in Table 5.1.

TABLE 5.1
Decision Criteria for Potential Courses of Action

Label	Standard	Defensibility	Combat MOS General Fitness	Combat Task Performance	Simplicity	Compensatory	Subgroup Concerns
1A	420 Total	●	◐	◐	●	●	●
1B	450 Total	◐	●	◐	●	●	○
2A	140 MDL	○	○	○	◐	○	●
2B	150 MDL	●	○	●	◐	○	○
3A	420 Total + 140 MDL	◐	◐	○	○	◐	◐
3B	420 Total + 150 MDL	◐	◐	●	○	◐	○
3C	450 Total + 140 MDL	◐	●	○	○	◐	○
3D	450 Total + 150 MDL	◐	●	●	○	◐	○

KEY: ● = COA has relative advantages. ◐ = COA has some relative advantage and some relative disadvantages. ○ = COA has relative disadvantages.

[1] This prediction model is based on resourcing, training, and incentives at the time of the Practice Phase. Future pass rates could change depending on how these underlying factors change.

When examining these courses of action:

- **Defensibility:** All courses of action are defensible based on research evidence and ACFT performance. However, we highlight option 2A as less likely to satisfy the FY 2024 NDAA. This is because, at 140 pounds, the standard is raised for female soldiers only, not all soldiers in close combat MOSs. We also highlight that increasing the overall number of points but maintaining the age- and gender-normed scoring may conflict with some interpretations of language provided in DoDI 1308.03.
- **Combat MOS general fitness:** Close combat MOS general fitness would be maintained by setting an overall standard of 420 overall points and would be enhanced by setting a standard of 450 overall points. While the MDL options also maintain or raise fitness levels, these options are narrowly focused on only one event and will be less likely to maintain or enhance fitness across other fitness domains.
- **Combat task performance:** In terms of combat task performance, all options raise the general fitness floor. However, options with only an increase in the overall points, which are age- and gender-normed, do not ensure that all close combat soldiers have the same capabilities. To meet that requirement, options with an age- and gender-neutral MDL standard are needed. If an option with an age- and gender-neutral MDL standard is adopted, the ACFT would have three age- and gender-neutral minimum standards—the other two being the PLK and HRP. Setting additional gender- and age-neutral standards for other events should only be set after sufficient evidence has been collected.
- **Simplicity:** Simplicity is maximized by decisions that are aligned with current ACFT objectives to promote general fitness and health. This alignment is greatest for options that increase the overall ACFT point standards. Specifying an age- and gender-neutral standard on the MDL would support a different objective to ensure minimally acceptable combat task performance. Additionally, the Army would need to spend more time messaging because soldiers may be confused as to why other events have not been selected. Without having a full understanding of why only the MDL has an age- and gender-neutral standard, soldiers may question the relevance of other events to close combat MOSs.
- **Compensatory:** This criterion is maximized when soldiers have the greatest flexibility in selecting which events to focus on to maintain or improve their fitness. Increasing the overall points allows soldiers this flexibility because they can choose how they meet higher standards.
- **Subgroup outcomes:** Another factor to consider is the prevalence of differences in subgroup pass rates that may be created by higher standards. Differences in pass rates are a negative factor unless justified by evidence. It is important to note that pass rate differences would not be expected for Regular Army age and gender subgroups when increasing the overall points standard if the ACFT scoring system was normed based on ACFT performance observed since October 2022.
  - Considering these factors, options that keep the overall points increase to 420 results in the fewest subgroup differences in pass rates. Increasing beyond 420 overall points may create disparities for certain MOSs and ARNG/USAR soldiers.
  - Younger, female soldiers in particular have lower pass rates when the MDL standard is increased. Increasing the standard for total ACFT points to 450 results in pass rates below 90 percent for several MOSs. At this standard, pass rates for soldiers in the ARNG and USAR fall below 75 percent.
  - Evidence suggests that a hybrid approach adopting an overall ACFT score minimum of 450 overall points and 150 pounds on the MDL may be too high to achieve a 95 percent overall pass rate and 90 percent pass rate for select subgroups in the short term. Soldiers in the Regular Army exhibited a 91.4 percent pass rate, and several subgroups, including female soldiers and certain age and MOS subgroups, had pass rates below 90 percent under this standard. Pass rates for soldiers in the ARNG and USAR are significantly lower under this COA. Other options for setting higher standards achieve

an overall pass rate greater than 95 percent and have less impact on Regular Army, ARNG, and USAR subgroups.

Recommendations to Facilitate Implementation of Higher Standards

1. **Supporting policy should be implemented to facilitate the movement to higher standards.** We identified five opportunities to support implementation:
   a. **Glide paths.** Provide glide paths to allow soldiers sufficient time to train and improve with minimal increases to injury risk.
   b. **Remedial assessments.** Use MOS-specific physically demanding tasks as a test for soldiers who do not meet the higher ACFT standard (these task-based tests should be standardized for each MOS).
   c. **Re-screening.** Administer the OPAT again to confirm readiness levels (e.g., during BCT).
   d. **Training resources.** Continue rolling out H2F resources to units and reserve components; invest in training for local levels (squad, company).
   e. **Clarify ACFT goals.** Promote acceptance of higher ACFT standards by establishing a clear and consistent message across all organizational levels, which includes transparent policies for failure to meet close combat ACFT standards.
2. **Consider re-norming the scoring system at the same time as adopting higher standards.** The existing age- and gender-normed scoring system was developed using diagnostic test records and should be updated to reflect current age and gender ACFT performance norms. Re-norming can address known issues where increases in raw performance do not consistently translate into additional points, particularly at the lower end of the performance scale. Finally, updating the scoring system while raising standards for the close combat force provides the Army with additional flexibility to ensure that norms are appropriately set for soldiers in both close combat and non–close combat MOSs.
3. **Collect further evidence for minimum ACFT standards through criterion-related validation studies.** Design criterion-related validation studies that examine the relationship between performance on each event in the ACFT as currently administered as a complete test battery and critical organizational outcomes, including combat task performance, fitness, retention rates, and other relevant metrics.[2] The studies should include (1) broad and diverse samples of soldiers (e.g., female and older age groups) to ensure the findings are generalizable across target MOSs and (2) a longitudinal component to assess how changes in ACFT performance over time relate to changes in the key organizational outcomes.

Final Thoughts

As currently implemented, the ACFT is designed to promote general fitness and health. Although the FY 2024 NDAA has directed the Army to increase ACFT standards for close combat MOSs, this change may further confuse soldiers on the real purpose of the test. Ideally, the Army would be able to create separate Tier II occupationally specific tests that are age- and gender neutral, as described by DoDI 1308.03 and in Chapter 1. Separating Tier I from Tier II assessments should help facilitate communication and commitment to the objectives for general health and fitness and MOS task performance. Collecting the type of data necessary to

[2] The Army has asked the University of Iowa to conduct a study to identify the minimum requirements to perform specific combat tasks. This study is currently underway.

establish evidence for these types of standards takes considerable resources and time. Therefore, the Army could prioritize MOSs that may have specific fitness requirements or for which commanders have indicated a need for more-specific assessments to guide their decisions about personnel readiness.

We also emphasize that the ACFT serves as a measure of physical readiness but is not comprehensive of all individual readiness domains. While it is a proxy for physical readiness, it does not encompass other critical components, such as nutritional, mental, spiritual, and sleep readiness, as identified in Field Manual 7-22, *Holistic Health and Fitness*. These elements collectively contribute to a soldier's ability to achieve movement lethality: "the ability to apply and sustain the right amount of strength, endurance, and speed to meet the demands of training and combat physical tasks."[3]

Unit commanders should consider using additional measures of physical fitness to ensure that their soldiers can perform the physically demanding tasks specific to their unit's missions. The ACFT is just one of many proxies for combat readiness, and its relative importance to combat performance is not yet fully understood. Therefore, further research is needed to inform broader efforts that guide soldiers on optimizing their fitness for combat, injury prevention, and general health.

Ultimately, modern conflicts, even at the close combat level, are influenced by organizational factors beyond individual capabilities. Unit commanders are well positioned to assess the close combat readiness of their formations through collective assessment measures, ensuring comprehensive preparedness for their specific operational contexts.

---

[3] Field Manual 7-22, *Holistic Health and Fitness*, Headquarters, Department of the Army, October 2020, p. 3-1.

APPENDIX A

Additional Background on Sources of ACFT Evidence

This appendix describes the key sources of evidence we evaluated to inform our assessment and initial recommendations for the Practice Phase standard. It also contains the decision criteria used in developing alternative ACFT standards for combat MOSs.

Key Sources of Evidence

We reviewed five key sources of evidence and relevant policy changes regarding the ACFT (Figure A.1):

1. **Meta-analysis:** In 2014, the Army Public Health Command's Injury Prevention Program published a meta-analysis that examined 543 correlations from primary studies to estimate relationships between physical fitness tests and military task performance.[1] Researchers provided correlation estimates for specific fitness tests (e.g., push-ups) when possible, as well as the underlying physical ability (e.g., upper body muscular strength).
2. **BSPRRS (Fort Riley):** As part of the BSPRRS, the study phase conducted at Fort Riley collected data on fitness test events and combat task performance using a combination of combat task simulations referred to as the Warrior Tasks and Battlefield Drill—Simulation Test (WTBD-ST). The WTBD-ST combined four field events: (1) Conduct a foot movement under load (movement to contact), (2) prepare a hasty fighting position, (3) move over/under/around/through obstacles, and (4) react to hand-hand contact (combatives).[2] The results from the study can be used to support the validity of certain ACFT events, but there are several limitations with data generated by the Fort Riley study:
    a. No minimally acceptable performance time on the WTBD-ST was established; therefore, identifying an acceptable cutoff score must be guided by other, less defensible factors.
    b. Only 46 female soldiers participated in the study, with none from the target combat MOSs specified in the FY 2024 NDAA.
    c. No data were collected on the Plank, which replaced the Leg Tuck after the study was completed.
    d. Event protocols changed after the study for three ACFT events: MDL, SDC, and HRP.
    e. Combat task performance was combined into a single WTBD-ST time, which prevents opportunities to determine possible relationships between fitness test events and performance on specific combat tasks.
1. **BSPRRS (Fort Benning):** Part of the BSPRRS, the study phase conducted at Fort Benning was a follow-on to the Fort Riley effort and collected data from 136 male and 16 female soldiers on eight

[1] Hauschild et al., 2014.

[2] East et al., 2019.

fitness test events[3] and four WTBD-ST events. The results from the Fort Benning study can be used to support the validity of certain ACFT events, but there are several limitations with data generated by this phase of the BSPRRS:

a. No minimally acceptable performance times on the WTBD-ST events were established; therefore, identifying an acceptable cutoff score must be guided by other factors that are less defensible.
b. Only 16 female soldiers participated in the study and, since MOS was not included in the data we received, it is unclear how many soldiers were from the target combat MOSs specified in the FY 2024 NDAA.
c. No data were collected on the Plank, which replaced the Leg Tuck after the study was completed.
d. Event protocols changed after the study for two ACFT events: SDC and HRP.

2. **RAND studies:** RAND has conducted a series of analyses in support of effective Army implementation of the ACFT. One study, focused on gaps in the evidence base, includes analyses of the role of extraneous factors (e.g., height and altitude) in biasing test scores among soldiers taking the ACFT.[4] Another RAND analysis examined the association between ACFT performance and subsequent injury risk.[5] Pass rates and performance levels for each individual ACFT event were associated with subsequent injury risk. A limitation of the injury analysis is that the sample includes all soldiers, not just soldiers in close combat MOSs. In addition, results should be understood to be associative, not causal, in nature.
3. **OPAT studies and other related studies:** To develop and evaluate OPAT, researchers examined relationships between many fitness tests and eight Criterion Measure Task Simulations:
    a. OPAT development studies: When developing the OPAT, slightly different protocols were used. For example, instead of the SDL, the Squat Lift was used, which required soldiers to lift dumbbells ranging from 25 lb to 110 lb.
    b. OPAT Evaluation Study in 2016:[6] This study involved 741 (608 male, 133 female) trainees. Male trainees were from 11B, 11C, 12B, 13B, 13F, 19D, and 19K. Female trainees were from 12B, 13B, and other high physical demand MOSs due to a limited number of females in combat MOSs at the time of the study.
    c. OPAT Longitudinal Study: This study conducted a longitudinal validation of the OPAT with 1,181 recruits, who completed the OPAT within the first two weeks of initial entry training. Of those taking the OPAT, 741 also performed physically demanding MOS tasks (PDTs) near the end of initial entry training.
    d. OPAT Injury Study in 2016: As part of the OPAT Longitudinal Study, this analysis prospectively evaluated relationships between OPAT performance and two outcome measures (i.e., injury and attrition) during initial entry training.

[3] The eight fitness tests were the Sled Drag, 2MR, Deadlift, Sled Push, Push-Ups, Power Throw, Leg Tuck, and 300-Yard Shuttle Run.

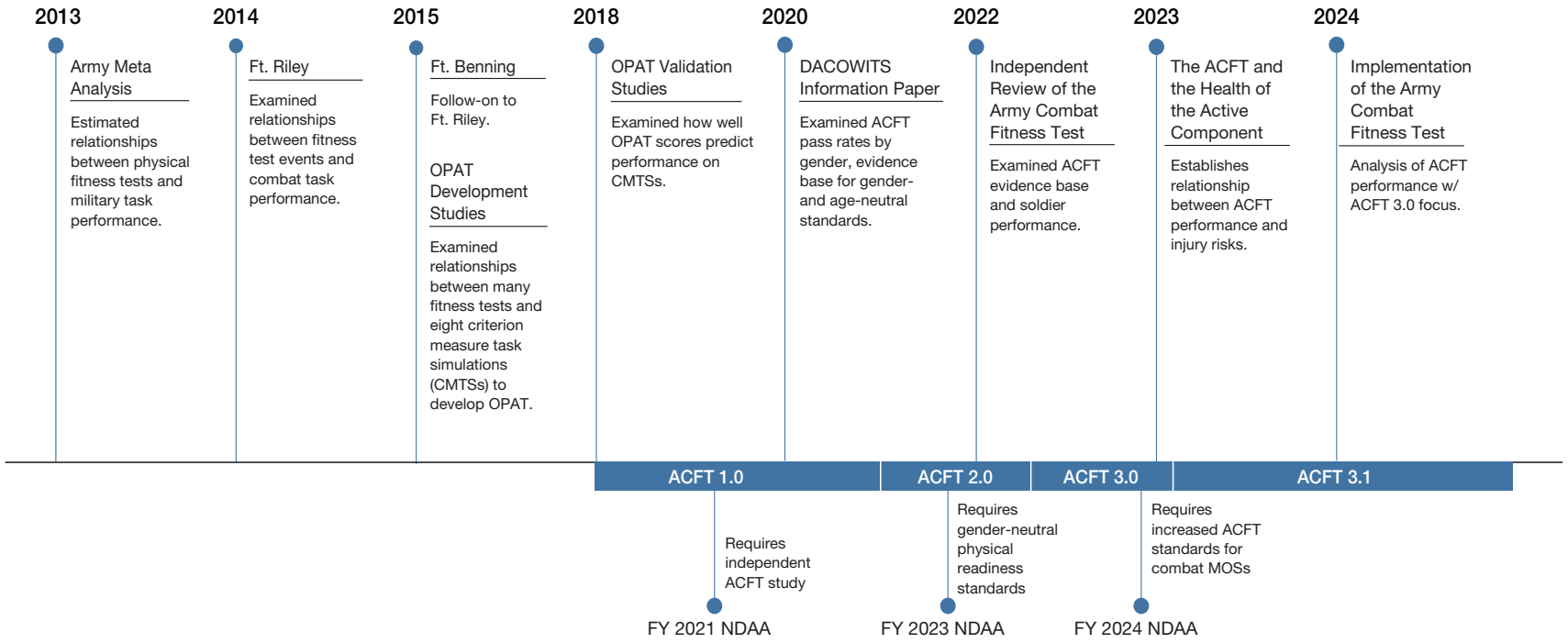
[4] See Hicks and Robson, forthcoming.

[5] Hicks et al., forthcoming.

[6] Sharp et al., 2018.

FIGURE A.1

Timeline of Major ACFT and Related Studies



NOTE: ACFT 1.0 took place from October 2018 to May 2020. The test was age- and gender-neutral during this period and was in the trial and diagnostic phase. ACFT 2.0 took place from Jun 2020 to March 2021, with a two-minute PLK backup for the Leg Tuck event. ACFT 3.0 took place from April 2021 to March 2022, with a choice between the Leg Tuck or four-minute PLK. ACFT 3.1 took place from April 2022 to the present. The test became age- and gender-normed, the Leg Tuck was replaced with the PLK. The ACFT became for-record, counting for personnel actions, in October 2022 in the Regular Army (Hicks and Robson, 2024, forthcoming).

Additional Evidence: ACFT Events and Combat Task Performance

In this section, we provide additional detail on how ACFT events have been linked to combat task performance.

Additional Evidence for the 3 Repetition Maximum Deadlift

Additional evidence for the MDL is provided from our review of the following sources:

1. **Meta-analysis:** Although previous meta-analyses have not specifically examined the MDL, other strength measures have shown moderate correlations (0.44) with combat task performance.
2. **BSPRRS (Fort Riley):** The administration protocol for the MDL varied slightly between the Fort Riley study and the current ACFT protocol. In the Fort Riley study, participants "were directed to select a weight that would allow them to execute ≤ 6 repetitions".[7] In the ACFT protocol, soldiers must select a weight that allows them to complete 3 repetitions. Although these differences could limit efforts to determine an appropriate MDL cutoff score, the positive relationship observed for both male soldiers (0.34) and female soldiers (0.45) provide supporting evidence for including the MDL as a strength measure for soldiers in combat MOSs.
3. **BSPRRS (Fort Benning):** A 1-repetition maximum deadlift was estimated based on 3-repetition maximum deadlift. Sample sizes were sufficient to determine consistent correlations between the MDL and each of the four WTBD-ST events (correlations range from 0.46 to 0.56). The number of female soldiers was too small (n = 16) to estimate whether the MDL predicts WTBD-ST equally well for male and female soldiers.
4. **OPAT studies and other related studies:** The OPAT also uses a deadlift but with a 1-repetition SDL and a standard of 160 lb for the "Heavy" physical demand category of MOSs. Studies developing and evaluating OPAT have demonstrated strong support for the deadlift. Correlations with combat task performance, such as casualty drag and casualty evacuation, are strong (i.e., correlations generally above 0.70) indicating that soldiers lifting higher weights on the deadlift can drag and evacuate a casualty a set distance in less time.

Strength Deadlift to 3 Repetition Maximum Deadlift Conversion

We estimated the ACFT MDL standard using three different formulas (see Table A.1). These formulas are imperfect but provide a range of values to consider for a potential gender- and age-neutral MDL standard.[8] We estimated a 3-repetition maximum deadlift (i.e., ACFT MDL) using a 1-repetition lift (i.e., OPAT SDL).

Additional Evidence for the Standing Power Throw

1. **Meta-analysis:** Although the SPT was not specifically examined in the meta-analysis, some indirect support for the SPT is provided by estimates using two fitness events, the Standing Broad Jump (SBJ) and Vertical Jump (VJ), that share some of the same movement patterns. The SBJ and VJ demon-

[7] East, DeGroot, and Muraca-Grabowski, 2019, p. 28.

[8] Although there is generally limited research on the accuracy of these conversion formulas, the Wathen formula provided the most accuracy in estimating one repetition maximum deadlift with 9 percent error between predicted and actual amount lifted. For more details, see Dale A. LeSuer, James H. McCormick, Jerry L. Mayhew, Ronald L. Wasserstein, and Michael D. Arnold, "The Accuracy of Prediction Equations for Estimating 1-RM Performance in the Bench Press, Squat, and Deadlift," *Journal of Strength and Conditioning Research*, Vol. 11, No. 4, November 1997.

TABLE. A.1
Conversions to Estimate a 3 Repetition Maximum Deadlift

Method	1RM Formulas	OPAT SDL Standard for Heavy MOSs (Includes Combat MOSs) 1RM = 160 lb	Alternate Higher Standard 1RM = 180 lb
Epley	1RM = MDL * (1 + 3/30)	145	164
Brzycki	1RM = MDL / (1.0278-0.0278*3)	151	170
Wathen	1RM = MDL * (1 / (0.4880+0.538$e^{-0.075*3}$)	147	165

SOURCE: Author calculations using formulas adapted from J. L. Mayhew, T. E. Ball, and J. C. Bowen, "Prediction of Bench Press Lifting Ability from Submaximal Repetitions Before and After Training," *Sports Medicine, Training, and Rehabilitation*, Vol. 3, 1992; and Terry Wood, Gianni F. Maddalozzo, and Rod A. Harter, "Accuracy of Seven Equations for Predicting 1-RM Performance of Apparently Healthy, Sedentary Older Adults," *Measurement in Physical Education and Exercise Science*, Vol. 6, No. 2, 2002.

NOTE: 1RM = 1 repetition maximum.

strate moderate to strong correlations (general range 0.5–0.7) with lift and lower tasks, stretcher carry tasks, and tasks requiring moving fast.

2. **BSPRRS (Fort Riley):** The SPT was correlated with WTDB-ST for both male soldiers (0.39) and female soldiers (0.63).
3. **BSPRRS (Fort Benning):** The sample size was large enough to determine consistent relationships between the SPT and each of the four WTBD-ST events (correlations ranged from 0.53 to 0.58).
4. **OPA studies and other related studies:** The OPAT studies demonstrated many significant correlations between related events including the Powerball Throw, Seated Power Throw, and jump tests (SBJ, VJ). Because the administration protocol is substantially different for the SPT than for these other tests, findings can only be used to provide indirect support for the SPT.

Additional Evidence for the Hand Release Push-Up

The HRP is a measure of upper body muscular endurance relevant to combat tasks that require crawling, hand-to-hand contact, and pushing obstacles. Although push-ups were less strongly correlated with combat tasks in both the Fort Riley and Fort Benning studies, previous meta-analyses provide stronger support with linkages to a range of combat tasks. Additional evidence for the HRP is provided from our review from the following sources:

1. **Meta-analysis:** Traditional push-ups were moderately to strongly correlated with lifting, carrying, quick movement, climbing, crawling, and multi-activity tasks, with the strongest correlations being 0.58 (crawl), 0.57 (repeated lift and lower), and 0.52 (move fast).[9]
2. **BSPRRS (Fort Riley):** Traditional push-up scores were provided by soldiers from their most recent APFT administration. It is possible that using the APFT push-up scores affected the correlations observed for both male soldiers (0.28) and female soldiers (0.19), which were weaker than previously reported in other studies, including the meta-analysis and the Fort Benning study.
3. **BSPRRS (Fort Benning):** Traditional push-ups were moderately correlated with each of the four WTBD-ST events. Specifically for soldiers in fighting load, the correlations ranged from 0.32 to 0.39, indicating that soldiers performing more push-ups performed better on the combat tasks.
4. **OPAT studies and other related studies:** Push-ups are not part of the OPAT, but the Army did examine relationships between 1-minute push-ups and combat tasks in multiple studies, which have shown moderate to strong correlations.

[9] Hauschild et al., 2014.

Additional Evidence for the Sprint-Drag-Carry

1. **Meta-analysis:** In a previous meta-analysis, sprint times were moderately to strongly correlated with lifting, carrying, casualty drag, and multi-activity tasks. Correlations ranged from 0.53 (casualty drag) to 0.71 (multi-activity tasks).[10]
2. **BSPRRS (Fort Riley):** An approximation of the SDC was created by combining multiple events that were administered separately: The SDC was evaluated by combined portions of the sled drag, sled push, and 300-yard shuttle run to form the 250-meter SDC test event. "The SDC is comprised of: 50m sprint, 50m sled drag, 50m sprint, 50m farmer's carry, 50m sprint. In an attempt to understand how this consolidated variable might affect the overall explained variance in the full model regression, researchers computed a composite variable based on the standardized values of the sled push, sled drag, and 300yd shuttle run."[11] The approximated event demonstrated moderate correlations with WTBD-ST for both male (0.43) and female soldiers (0.44).
3. **BSPRRS (Fort Benning):** An approximation of the SDC was used to examine relationships with WTBD-ST events. Sample sizes are large enough to determine consistent and strong relationships between the SDC approximation and each of the four WTBD-ST events (0.70 to 0.88).
4. **OPAT studies and other related studies:** Although the Army does not have an OPAT event equivalent to the SDC, the 300-meter sprint was evaluated as part of the OPAT studies and demonstrated moderate to strong correlations with a range of combat tasks. For example, among 11B soldiers, correlations were observed with foot march (0.47), sandbag carry (0.55), move under fire (0.66), casualty evac (0.55), and casualty drag (0.53).

Additional Evidence for the Plank

1. **Met-analysis:** In a previous meta-analysis, no estimates were provided for the PLK. Only indirect evidence from sit-ups can be used to support conclusions about the importance of core endurance to combat task performance. Relationships between sit-ups and combat task performance are generally weaker compared with other fitness tests but were shown to correlate with climbing and crawling tasks (0.45 and 0.48, respectively).[12]
2. **BSPRRS (Fort Riley):** The PLK was not evaluated.
3. **BSPRRS (Fort Benning):** The PLK was not evaluated.
4. **OPAT studies.** The Army did not select a core endurance measure for the OPAT but did examine relationships between 1-minute sit-ups and combat tasks in multiple studies, which revealed inconsistent and weak to moderate relationships with criterion tasks. For example, sit-ups correlated with foot march for some combat MOSs but not others.

[10] Hauschild et al., 2014.

[11] East et al., 2019, p. 32.

[12] Hauschild et al., 2014.

Additional Evidence for the Two-Mile Run

Additional evidence for the 2MR is provided from our review of the following sources:

1. **Meta-analysis:** Previous meta-analyses provide strong evidence for distance runs (1.5 and 2 miles) with correlations to combat task performance often above 0.50.[13] Additionally, these distance timed runs are very strong indicators of VO2max, which is the gold standard for measuring cardiorespiratory fitness.
2. **BSPRRS (Fort Riley):** The 2MR times were provided by soldiers from their most recent APFT administration. It is possible that using the APFT 2MR times affected the correlations observed for both male soldiers (0.31) and female soldiers (0.01), which were weaker than previously reported in other studies, including the meta-analysis and the Fort Benning study.
3. **BSPRRS (Fort Benning):** The 2MR was moderately correlated with each of the four WTBD-ST events. Specifically for soldiers in fighting load, the correlations ranged from 0.34 to 0.45, indicating that soldiers who were faster on the 2MR performed better on the combat tasks.
4. **OPAT studies and other related studies:** The OPAT also measures cardiorespiratory fitness but uses the IAR (beep test). For the "Heavy" physical demand category of MOSs, soldiers must complete 43 shuttles, which can be used to estimate a corresponding 2MR time. Across different studies, the IAR has demonstrated moderate to strong correlations with a range of combat tasks.

Example of a Method to Establish a Specific Cutoff Scores

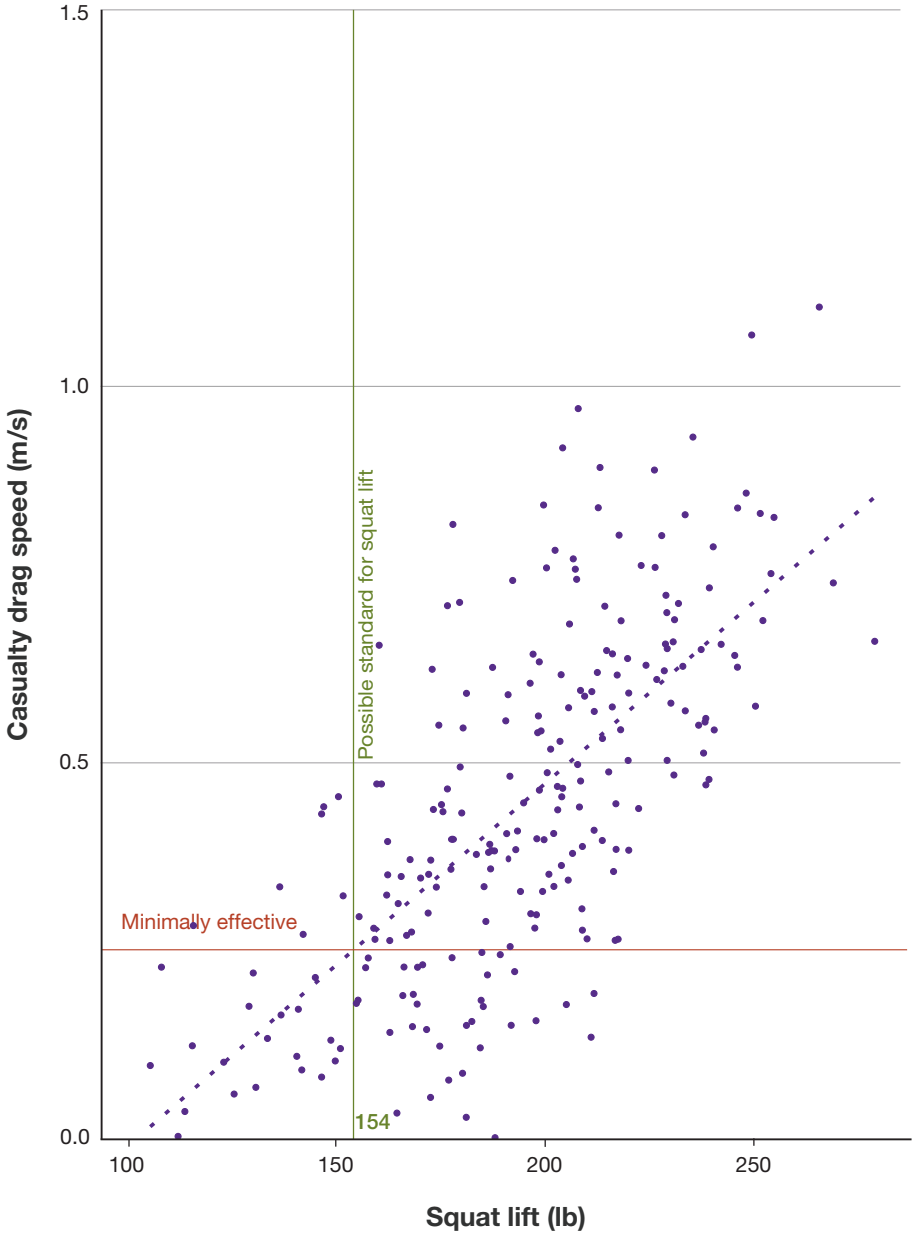
To establish evidence for the predictive validity of an ACFT test event, it is necessary to demonstrate a correlation with a specific outcome (e.g., combat task performance or injuries). However, setting a criterion-referenced event standard requires additional information, such as minimally acceptable performance on a task (e.g., completing a casualty drag task of 30 meters in less than 60 seconds). This information is needed to determine the score on the ACFT event that corresponds to the minimally acceptable level of performance on the specific outcome (e.g., casualty drag time).

We provide a hypothetical example in Figure A.2 based on simulated data for a non-ACFT fitness event—the Squat Lift—and a performance measure—the casualty drag. In this notional example, subject-matter experts set a MAPS of 0.3 meters per second on the casualty drag task. This performance standard represented by the horizontal red line differentiates soldiers who are considered successful on the combat task (above the red line) and soldiers who are considered unsuccessful (below the red line). Based on the strength of the statistical relationship between the fitness test event and combat task performance (represented by the blue regression line), we can identify a possible cutoff score on the Squat Test (green vertical line).

Other methods for setting cutoff scores, which use statistical relationships between tests and outcomes, can consider measurement error and other key organizational objectives to minimize the number of individuals who may be falsely categorized as successful or unsuccessful by the cutoff score. Cutoff scores may be drawn at other locations. There is a trade-off implicit in selecting a cutoff standard, based on a desire to minimize two types of undesirable misclassifications (Table A.2). The first, represented by observations in the top left quadrant of the figure (above the red minimally effective line and scoring below the possible standard for the Squat Lift, the green line) are soldiers who are excluded from combat because their test score is too low but could actually perform the task. The second, represented by the quadrant in the bottom right (below the red line and scoring above the green line) are soldiers who are not combat proficient in the casualty drag

[13] Hauschild et al., 2014.

FIGURE A.2
Hypothetical Relationship Between Squat Lift and Casualty Drag Speed



SOURCE. ChatGPT 4.0 generated data.

but are misclassified by the test as proficient. There is no perfect test, nor a single combat standard, so it is up to the Army to decide what constitutes effective combat performance and how much emphasis to place on minimizing each type of misclassification through standard setting.

However, even with this information, some degree of professional judgment is often required in setting a cutoff score to balance multiple and sometimes competing organizational objectives (e.g., increase physical readiness for combat versus maintain a desired end strength). Other organizational objectives that could be considered in standard setting include retention, fitness, injury risk or other health factors, time allocated to training, cultural factors, diversity, and costs of recruiting and training new soldiers.

TABLE. A.2
Outcomes Using Hypothetical Squat Lift Standard

Event	Passed or Failed	Casualty Drag Performance			
		Not Minimally Effective		Minimally Effective	
		Count	Percentage	Count	Percentage
Squat Lift Standard	Passed	7	3%	173	74%
	Failed	20	9%	35	15%

Additional Evidence for the ACFT and Injury Prevention

In this section, we provide additional event by event detail on how components of the ACFT have been linked to injury prevention in prior research:[14]

- **MDL:** Passing current MDL minimum requirements is associated with reduced injury risk for both acute trauma and overuse injuries. This rate differs slightly for male and female soldiers (35 percent and 26 percent reduction in relative injury risk, respectively). Higher performance (event score 70+ for males, 90+ for females) is associated with lower overuse injury risk. Highest performance (90+ scores) is associated with small increases in acute injury risk for male and female soldiers. These findings could support higher standards on the basis of the MDL association with injury risk reduction, particularly for male soldiers.
- **SPT:** The SPT is statistically significantly associated with *higher* injury risk for both acute trauma and overuse injuries for men, as well as for overuse injuries for women. Unlike the other events, performance in higher deciles of event scoring is not associated with reduced risk for overall injuries. These findings *do not* provide support for higher standards on the SPT based on the goal of injury risk reduction.
- **HRP:** Failure on the HRP event is statistically significantly associated with higher injury risk for both acute trauma and overuse injuries for both male and female soldiers (45 percent and 30 percent reduction in relative risk, respectively). Higher HRP performances are associated with lower injury risk and could support higher standards on the basis of the HRP association with injury risk reduction.
- **SDC:** Generally, better SDC performance is associated with lower overuse injury risk for both male and female soldiers and could support higher standards on the basis of the SDC association with injury risk reduction. However, somewhat unexpectedly, male soldiers scoring 90 and above on the SDC are at slightly higher risk for acute trauma injuries. It is unclear why injury risk is elevated for this subgroup of high performers.
- **PLK:** Stronger performance on the PLK is associated with lower future risks for overuse injury for male (scores of 70+ points) and female (scores of 80+ points) soldiers but not for acute trauma injuries. These findings could support higher standards on the basis of the PLK association with injury risk reduction for both male and female soldiers.
- **2MR:** Failure on the 2MR event is associated with higher injury risk for both acute trauma and overuse injuries. The 2MR exhibits the strongest performance injury risk gradient of any event, with higher performances associated with lower injury risk at similar rates for both male and female soldiers. Lower levels of injury risk associated with 2MR performance are from reductions in overuse injuries among these populations. These findings could support higher standards on the basis of the 2MR association with injury risk reduction.

[14] For further detail, see Hicks et al., forthcoming.

APPENDIX B

Additional Details on Methodology

In this appendix, we provide additional details about our methods for constructing the Practice Phase and historical ACFT samples, designing the Practice Phase, and models developed to predict how much soldiers can improve on the ACFT within the next year.

Sample Construction for Practice Phase and Historical ACFT Results

This project made extensive use of historical ACFT data in combination with the Practice Phase data obtained from May to August 2024. Throughout, we applied a uniform set of inclusion criteria ("business rules") to the historical and Practice Phase data to ensure that the data were clean and that they reflected our target population: close combat soldiers who would be held to a higher ACFT standard. The business rules were as follows:

1. We considered test results only from October 2022 onward. The ACFT has undergone major changes since its rollout in 2018, and different iterations of the ACFT are not strictly comparable. In addition, the ACFT was not required until October 2022; soldiers may not have been fully incentivized to do well prior to this point. Therefore, we considered results only from after October 2022.
2. We restricted to soldiers age 17–62. Although there are some soldiers who fall outside of this age range, this is the relevant target population.
3. We excluded soldiers missing gender or age. Because analyses on age- and gender- subgroups are a critical component of this project, we removed ACFT results that lacked this information.
4. We excluded soldiers on profile or who were missing test results for any of the six events. We also excluded individuals with nonstandard patterns of performance suggestive of medical profile that grants them full or partial exemption from the test (e.g., missing raw score but received 60 points on an event), refusal to participate, or out-of-range values. Because our goal is to summarize and assess "normative" performance, we removed ACFT results from soldiers who were on profile or who appeared noncompliant, or records indicating data quality issues.
5. For the historical data, we considered only for-record ACFT results. Performance is generally better on for-record ACFTs than on diagnostic ACFTs. Because higher ACFT standards would be administered on a for-record basis, we considered only for-record results.
6. We restricted to close combat MOSs only. This includes the list of close combat MOSs defined by the FY 2024 NDAA. In the Practice Phase data, we also considered MOS 13B (in addition to the FY 2024 NDAA MOS list), because the Army indicated that it was of interest.
7. Although included in the FY 2024 NDAA MOS list, we removed Special Forces MOSs. Soldiers in Special Forces MOSs have elite performance levels exceeding both existing and proposed standards and would not be representative of other close combat MOSs.
8. With the exception of analyses on changes in the performance of individual soldiers over time, we considered a single test record per soldier, selecting the most recent test. Multiple test results from the same soldier are likely more correlated than test results from different soldiers, and including

multiple results gives more weight to soldiers who have been in the Army longer.[1] Consequently, we restricted to a single test record per soldier. We chose to include the most recent test record, because there has been improvement in ACFT performance over time and thus the most recent test is on average most similar to future performance.

9. For the analyses of changes in individual performance over time, we considered multiple tests per soldier, and included all soldiers who had at least two ACFT results that met criteria 1–7.
10. When considering subgroup analyses by age, gender, and/or MOS, we considered any subgroup that had a sample size of at least ten soldiers.

Figure B.1 provides information on how these business rules constructed the final sample. After business rules were applied, we had 115,979 ACFT results (box A in Figure B.1). For analyses on individual changes in performance over time, we had 197,803 historical ACFT results from 132,288 soldiers (box B in Figure B.1). After business rules were applied to the Practice Phase data, we had 44,293 Practice Phase ACFT results (box C in Figure B.1). Of those Practice Phase results, we could connect 37,487 to a historical ACFT record (box D). To give some examples of how each sample was used:

- Box A, historical ACFT data, most recent test only: used to estimate historical pass rates, identify subgroups that were historically at risk, calculate sample size requirements for the Practice Phase, develop provisional Practice Phase standard
- Box B, historical ACFT data, multiple tests per soldier: used to develop models of how ACFT performance by a given soldier changes over time
- Box C, Practice Phase ACFT data: used to estimate pass rates during the Practice Phase under the provisional Practice Phase standard and identify subgroups that were at-risk under the provisional Practice Phase standard
- Box D, Practice Phase ACFT data connected to historical records, multiple tests per soldier: used to predict ACFT performance in future and assess how the Practice Phase affected ACFT performance for a given soldier.

Practice Phase Design and Implementation

To better understand how soldiers in close combat MOSs may perform against higher standards, the Army administered a Practice Phase on the ACFT. In this Practice Phase, a sample of soldiers took the ACFT under a higher set of minimum standards. ACFT scores collected during this phase allowed us to estimate the impact of higher standards on pass rates.

To determine how many and which soldiers should participate in the Practice Phase, we conducted a statistical power analysis to determine the sample sizes and composition required to obtain an estimate of the pass rates for different subgroups of soldiers. We describe the power analysis, the final recommendations we delivered to the Army for the Practice Phase design, and how the Practice Phase implementation compared with our recommendations.

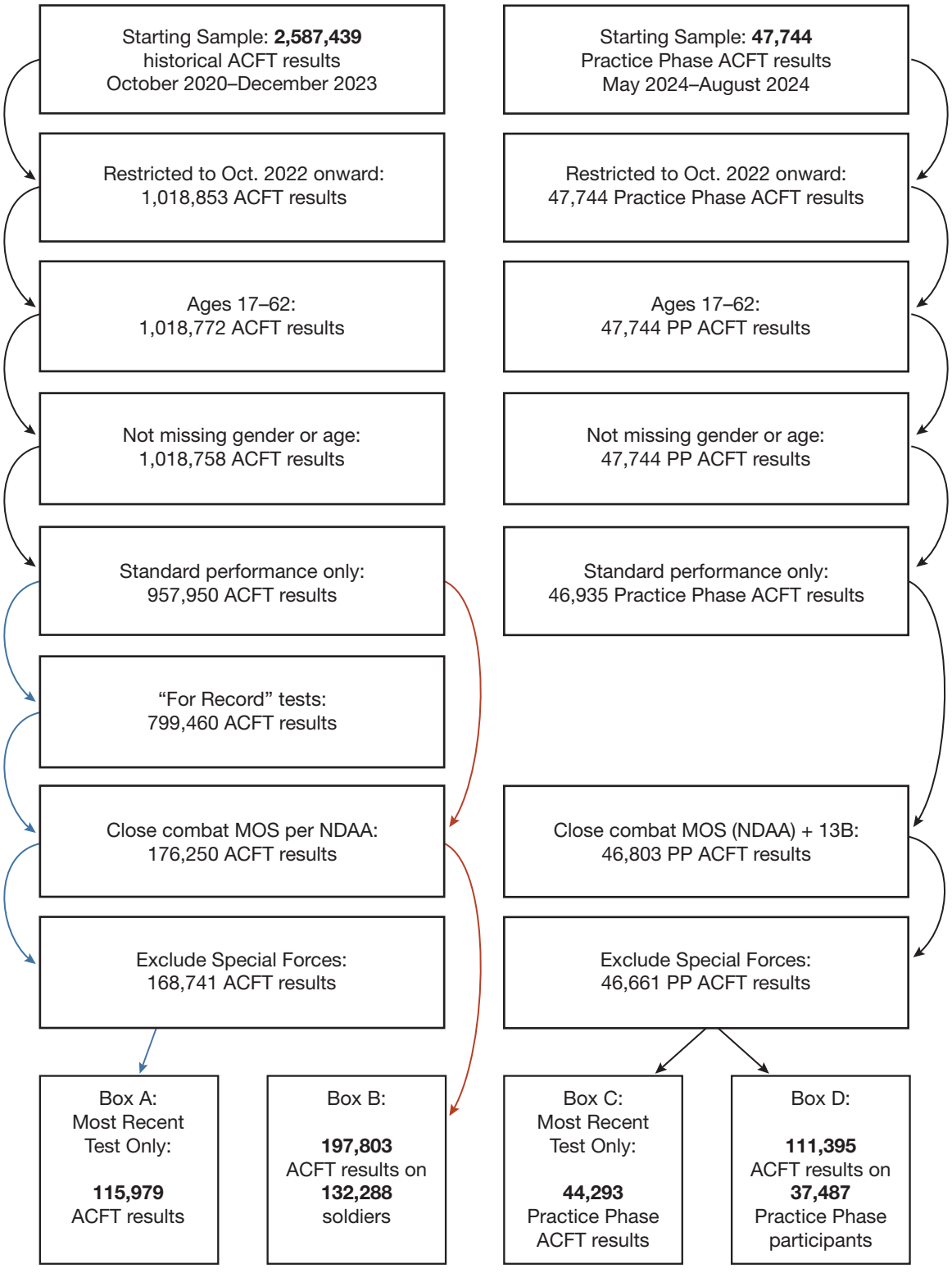
Approach to Statistical Power Analysis

The goal of the Practice Phase was to estimate pass rates on the ACFT under a higher proposed standard among soldiers in close combat MOSs. Unlike a conventional survey, in which estimates of the *overall* pass

[1] Garrett M. Fitzmaurice, Nan M. Laird, and James H. Ware, *Applied Longitudinal Analysis*, John Wiley & Sons, 2011.

FIGURE B.1

Business Rules Used to Construct Historical and Practice Phase ACFT Samples



NOTE: For constructing box B, the historical ACFT data with multiple test results per soldier, diagnostic tests were permitted, and Special Forces soldiers were retained.

rate would be the primary interest, estimates of the pass rates among key subgroups were arguably of equivalent or greater interest than the overall pass rate. As a result, the Practice Phase needed to be powered to provide high-quality estimates *within subgroups* (meaning that the number of soldiers in each subgroup would be large enough to make statistical inferences from the scores). The subgroups of interest were gender, age group, MOS, and component.

In an ideal world, the Practice Phase would provide high-quality estimates in all subgroups at a granular level—for example, the pass rate for Regular Army female soldiers age 17–21 in MOS 11B. However, this highly granular analysis required an immense sample size. To reduce sample size requirements, we developed a set of demographic and MOS clusters within which performance seemed fairly homogenous, to reduce the number of subgroups under consideration. These clusters were created by fitting (1) a linear regression of total ACFT scores on age group, gender, MOS, component, and an interaction between age group and gender, and (2) a logistic regression for pass/fail at the provisional standard used for sample size calculations (discussed below) on age group, gender, MOS, component, and an interaction between age group and gender. Both models were fit in the historical ACFT data. MOSs with similar coefficients (approximately no more than a 5-point difference on total score and no more than a 5 percent relative difference in pass rate) on both models were grouped, and similarly for age/gender subgroups. That resulted in the following age/gender clusters:

- males: 17–36, 37–46, 47–51, 52+
- females: 17–21, 22–31, 32–41, 42+.

And the following eight MOS clusters:

- 11A, 11B/11C, 11Z, 12A/19A, 12B/19D, 13A/19C/19Z, 13F, 19K.

Having established our list of subgroups (age/gender cluster × MOS cluster × component), we used standard statistical theory to generate sample size recommendations.[2] To estimate the pass rate (a proportion) within a subgroup of a finite population at some desired level of precision, the necessary sample size is:

$$n = \frac{N z^2 p(1-p)}{\delta^2(N-1) + z^2 p(1-p)}$$

where:

- N: the total size of the subgroup—for example, how many Regular Army MOS 11B female soldiers ages 17–21 are there in the Army? We estimated N in each subgroup using September 2022 Army headcounts provided in the Total Army Personnel Database.
- z: the Normal distribution critical value for the confidence interval size; we wanted 95 percent confidence intervals, hence $z = 1.96$.
- p: the expected pass rate in this subgroup—for example, what proportion of Regular Army MOS 11B female soldiers age 17–21 will pass the proposed ACFT standard?
- δ: the desired margin of error; we aim for a ±3 percent margin of error, so $\delta = 0.03$.

Of these quantities, the only one we were missing was p, the estimated pass rate in each subgroup. Note that sample size demands increase as p gets closer to 50 percent; therefore, the conservative approach is to

[2] Herman Burstein, "Finite Population Correction for Binomial Confidence Limits," *Journal of the American Statistical Association*, Vol. 70, No. 349, 1975.

set p closer to 50 percent. Because we expected p to be greater than 90 percent in most subgroups under the preexisting 360/60 standard, we wished to estimate expected pass rates under a higher, more challenging standard, thus pushing the expected pass rate closer to 50 percent. Note that the standard we used for estimating sample sizes does *not* reflect RAND recommendations for a higher provisional standard; rather, it was a conservative choice meant to ensure that sample sizes were adequate. For the purpose of calculating sample sizes, we required the following to pass:

- MDL of at least 140 lb
- SPT of at least 4 meters
- HRP of at least 20 repetitions
- SDC of no more than 2 minutes and 45 seconds
- PLK of at least 90 seconds
- 2MR of no more than 21 minutes.

In each subgroup in the historical ACFT data, we calculated the percentage of soldiers who would meet this standard and used that as our estimate of p. In some subgroups, p was estimated to be either 0 percent or 100 percent (usually because of small sample sizes). In those cases, we followed the recommendations of Agresti and Coull (1998) and added two passes and two failures to the subgroup to stabilize the estimated passing rate.[3]

With these quantities in place, we calculated sample size recommendations under the above formula for each subgroup of interest.

Recommended Approach

Based on our calculations, we recommended that the Army sample approximately 21,000 soldiers in total. We provided the Army with a detailed breakdown of sample size recommendations in each age/gender cluster × MOS cluster × component subgroup; the component × gender breakdown is summarized in Table B.1. These sample size calculations were created under the assumption of random sampling, in which soldiers in each subgroup would be randomly sampled for inclusion into the Practice Phase.

We recommended that the Army randomly sample the recommended number of soldiers from each subgroup and inform these soldiers of the requirement to take a Practice Phase ACFT from May to August 2024, with the goal of meeting the higher provisional standard that the Army would select in April 2024.

TABLE B.1
Recommended Practice Phase Sample Sizes, by Component and Gender

Component	Female Soldiers	Male Soldiers
Regular Army	2,345	5,088
Army National Guard	1,559	7,638
Army Reserve	572	3,989

NOTE: These by component and gender sample sizes are further broken down into MOS and age clusters. The full breakouts are available in a corresponding Excel spreadsheet, available upon request.

[3] Alan Agresti and Brent A. Coull, "Approximate Is Better than 'Exact' for Interval Estimation of Binomial Proportions," *American Statistician*, Vol. 52, No. 2, 1998.

Practice Phase Implementation

In February 2024, the Army began identifying Regular Army and ARNG units for participation in the Practice Phase. Because the Army ultimately selected a standard of 450 overall points + 150 MDL as the provisional Practice Phase standard—for some subgroups a higher standard than the standard we used for constructing sample size recommendations—RAND's sample size recommendations were too small to achieve the desired +/–3 percent level of precision for some subgroups. In addition, because of Army operational constraints, random sampling was infeasible. As a result, Practice Phase estimates of the pass rates may be subject to sampling bias. However, we do not expect this bias to be large, because the Army sampled a high percentage of all soldiers in the combat MOSs of interest and did not, to our knowledge, intentionally target particularly high- or low-performing units for inclusion.

From May 1, 2024, to August 8, 2024, the Army sampled 44,293 soldiers (after applying the business rules described above). Of these, 42,779 were male soldiers and 1,514 were female soldiers. The Army vastly exceeded RAND's recommended sample size for male soldiers but did not obtain the recommended number of female soldiers. Some subgroups of close combat soldiers have small populations within the Army, and even a complete sampling of these groups would not lead to high confidence in understanding potential future pass rates. In general, sampling recommendations were met for all MOS clusters and for male soldiers age 17–46, whereas ARNG soldiers, USAR soldiers, older subgroups, and female soldiers were not sampled at the target level, limiting our ability to estimate pass rates in these groups.

ACFT Improvement Models

To better understand how individual performance on the ACFT changes over time, we modeled ACFT performance of individual soldiers over time. We used these models to generate predicted ACFT performance in September 2025 (one year post-Practice Phase) for Practice Phase participants, to provide the Army with an estimate for what pass rates might look like when soldiers have more time to train to the higher standard.[4]

Improvement Model

We fit two mixed effects models—one for raw performance on the MDL, the second for total points—in the historical ACFT longitudinal sample (multiple test records per soldier). As predictors, we used

- information on baseline ACFT performance: date of first ACFT, age at first ACFT, performance (either total score or raw MDL) on first ACFT
- information on experience with the ACFT: passage of time since first ACFT (modeled quadratically), how many ACFTs had been taken (modeled quadratically), and whether each ACFT was for-record or diagnostic
- gender, MOS, and component
- interactions between gender and age, passage of time since first ACFT and age, passage of time since first ACFT and gender, and passage of time since first ACFT and performance on first ACFT.

To model the correlation between repeated ACFTs from the same soldier, we included a random intercept for each soldier.[5]

[4] Full model results available upon request.

[5] Fitzmaurice, Laird, and Ware, 2011.

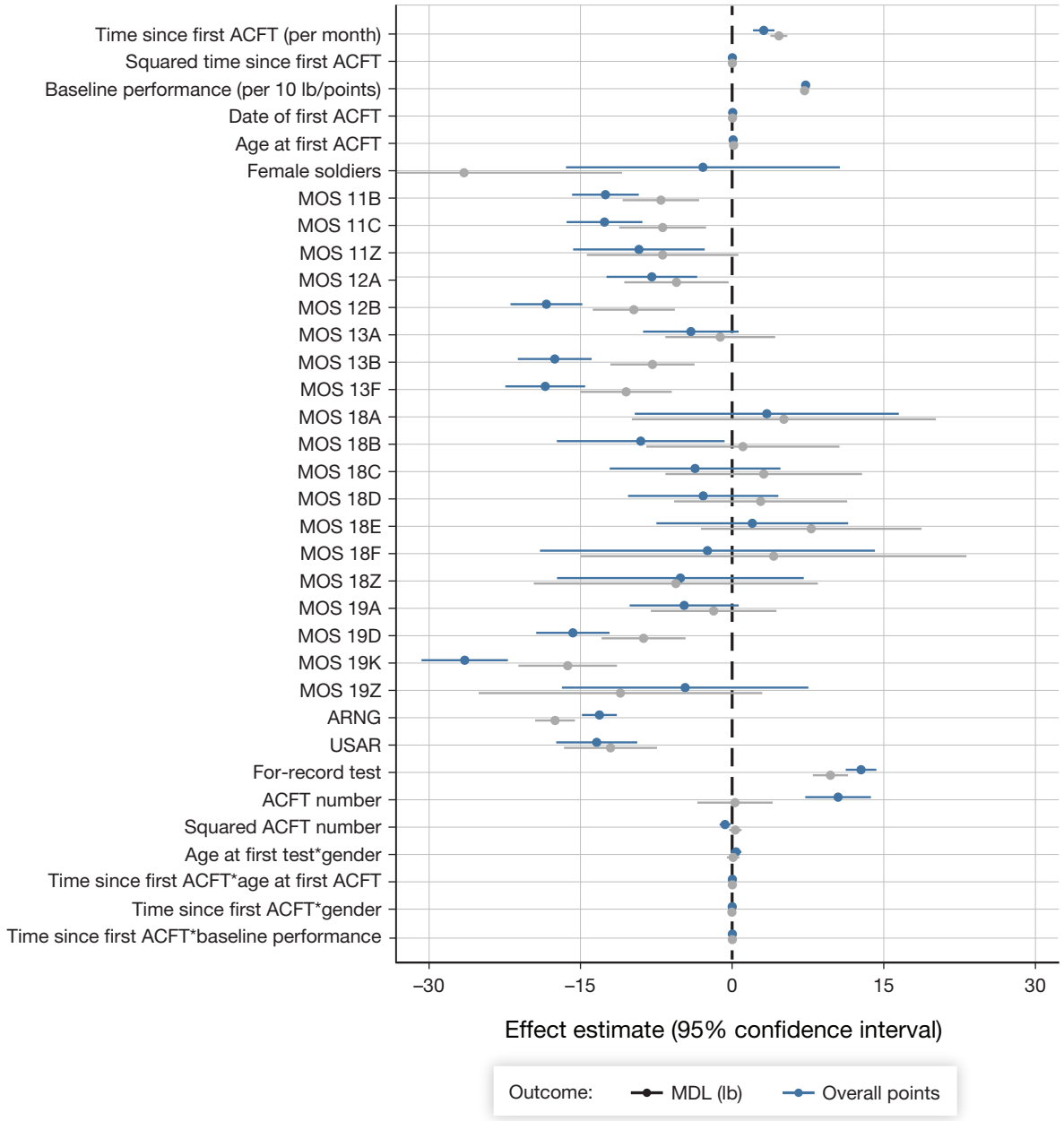
The results from these two models are shown in Figure B.2. Results are fairly consistent across the two outcomes. In both cases, increasing time since first ACFT and higher baseline performance corresponds to higher ACFT performance. The reference group for MOS (11A) generally does better on the ACFT than most other MOSs, except for most of the 18 (Special Forces) MOSs. Female soldiers have lower performance than male soldiers, an effect that is particularly pronounced for raw MDL performance, and similarly for USAR and ARNG soldiers relative to the Regular Army. Performance is higher on for-record ACFTs than on diagnostic ACFTs. In the fitted model, the soldier random intercepts were sampled from a normal distribution with standard deviation of 23 points for total score and 26 pounds for MDL, suggesting a high degree of individual-to-individual heterogeneity in performance.

Generating Predicted September 2025 Performance

Using the improvement models described, we generated a predicted ACFT score for each soldier who participated in the Practice Phase. For soldiers who appeared in the Practice Phase and could be linked to a historical ACFT record, we identified their baseline performance in the historical data, age at first ACFT, date of first ACFT, and number of previous ACFTs. For soldiers who could not be linked to a historical ACFT record (the 6,806 soldiers who appeared in Box C but not Box D of Figure B.1), we used their Practice Phase performance as their baseline performance and the start of their ACFT record. For every soldier, we calculated how much time would elapse between their Practice Phase test and September 1, 2025. We then input these predictors into each of the two models described above. To account for individual variation, for soldiers who appeared in the longitudinal sample (and thus had a personalized random intercept fitted from the improvement model), we used that random intercept when generating predictions for September 1, 2025. For soldiers with no pre–Practice Phase record, we randomly sampled a random intercept for them from the fitted random intercept distribution produced by the improvement models. We then identified which soldiers were predicted to pass a given total points/raw MDL standard by September 1, 2025.

FIGURE B.2

Effect Estimates from the Two ACFT Improvement Models



NOTE: Effect estimates appearing to the right of the vertical dashed line are associated with improved ACFT performance; effect estimates to the left of the vertical dashed line are associated with reduced ACFT performance.

APPENDIX C

Additional Details on Workshop Methodology and Results

After analyzing the results from the Practice Phase (Chapter 3, Appendix B), we conducted a stakeholder workshop to review results, identify concerns, and develop mitigation strategies to facilitate implementation of higher ACFT standards for the close combat MOSs. A total of 38 soldiers participated in the one-day workshop. Participants were selected to represent key subgroups, including gender, age group, MOS, and component. Additional details on the design of the workshop are presented this appendix, and results are presented in Chapter 4.

Higher fitness standards on the ACFT as called for by the FY 2024 NDAA are a matter of policy, but implementation would benefit from information similar to what might be collected as part of a formal standard-setting process.[1] Thus, we used a similar process to obtain stakeholder feedback regarding concerns about potential standards and how they can best be implemented. To solicit this information systematically, we conducted a workshop. The purpose of the workshop was to consider already-proposed and empirically driven standards and surface concerns regarding those standards, from soldiers who would be affected (i.e., relevant stakeholders). We also wanted to obtain detailed information on strategies to mitigate these concerns. Specifically, the main objectives of the workshop were to

- identify challenges related to various potential COAs for the implementation of various higher standard options
- identify potential policy mitigation strategies.

Workshop Demographics

Selection of attendees is an essential component of the workshop process to represent the perspectives of the full range of stakeholders. Stakeholders primarily include those for whom those standards would be applied: personnel in close combat MOSs. Within that constraint, various constituencies are important: personnel in different close combat MOSs mentioned in the FY 2024 NDAA (and 13Bs), who are subject to different physical requirements as part of their job but who all must exemplify fitness on the ACFT; different age- and gender-normed groups, who are held to equivalent levels of fitness by ACFT scoring requirements; different ranks, who may be subject to greater or lesser flexibility in their ability to prepare for the ACFT but nonetheless are held to similar requirements; and different components, who, while also subject to the ACFT, have

[1] A long history supports the use of considered judgments in the process of examining and setting standards for employment and educational tests, and how best to do so. See, for example, Ronald K. Hambleton, Mary J. Pitoniak, and Jenna M. Copella, "Essential Steps in Setting Performance Standards on Educational Tests and Strategies for Assessing the Reliability of Results," in Gregory J. Cizek, ed., *Setting Performance Standards: Foundations, Methods, and Innovations*, 2nd ed., Routledge, 2012.

different access to equipment and time to prepare, as well as, historically, different ACFT implementation timelines. Thus, relevant groups of stakeholders echo distinctions already made in policy and implementation of the ACFT. Soliciting views of stakeholders of a variety of close combat MOSs, age and gender groups, ranks, and components enables consideration of the standard in light of the concerns of those who would be held to its requirements.

Although efforts were made to ensure representation, some stakeholder groups had relatively few representatives present in the workshop. For example, the ARNG and USAR were represented by only four soldiers, combined. No soldiers represented the 18 CMF (Special Forces). Some soldiers reported that they were or had been drill sergeants or otherwise had responsibility for designing physical fitness programs for soldiers in their units; however, none were experts in exercise science per se and so represented perspectives that would be typical of the soldiers expected to implement any new standard in the field. Potential shortcomings were not insurmountable: For example, the data collection and research team included National Guard and Reserve members. The exclusion of Special Forces soldiers from consideration was because our analysis showed that soldiers within the Special Forces community have uniformly high ACFT scores and these MOSs are unique compared with other close combat MOSs: It is less likely that their perspective would be required to surface challenges with implementation for increased ACFT standards. Finally, we incorporated notable expertise in exercise science on the study team to be able to speak to the feasibility of any specific training program ideas that would potentially be surfaced as part of the discussion of how best to facilitate higher standards.

When asked directly, no participants indicated they had participated in the Practice Phase. However, most participants did indicate that they had taken a "Practice ACFT" earlier in the summer, although it was *not associated with a particular score or performance standard*. Thus, all participants had recent experience with the ACFT to base their judgments on, although the answers also suggest that communication regarding the Practice Phase was substantially lacking during data collection. One additional note is that workshop participants scored, on average, far higher on the ACFT (556) than their peers in close combat MOSs in the total force (519) .

Workshop Process

The one-day workshop was held at RAND's Northern Virginia offices, in person, and included four sessions. Sessions included a mix of presentations from RAND facilitators, full-room discussions, and small-group discussions. For all sessions, participants were seated in their small groups, which were designed to include stakeholders from a variety of age, gender, and MOS perspectives. All small groups had access to binders containing current ACFT scoring guides, Practice Phase performance data for various subgroups, and case studies of soldiers. These case studies described how soldiers with a variety of scores on the ACFT as implemented could reach various increased standards, using different strategies, such as achieving an increased score across ACFT events versus selecting a few events to target for particular score gains. Information was presented both in terms of point increases and raw performance (e.g., the time decrease needed on the 2MR to achieve the targeted score point increase). These case studies were developed to enable stakeholders to better understand the performance gains that would be required to meet the various ACFT standard alternatives.

In the first session, workshop participants were informed of the nature of the project and the objectives of the workshop itself. RAND researchers and Army leadership presented information on the importance of the workshop to facilitate productive dialogue. In the second session, data were provided from the Practice Phase, including information on the performance of various stakeholder groups. This included a brief discussion of information available throughout the day to stakeholders in binders, including the case studies. After presenting this information, a large group discussion was used to solicit potential concerns that might

pertain to increased standards on the ACFT. Prior to the workshop, the research team had generated several concerns, but these were supplemented by those generated by the stakeholders to ensure that the breadth of relevant concerns were considered.

Stakeholder perceptions of the challenges faced for the variety of standards under consideration were solicited using a rating form that assessed views on the current 360-point (6 x 60) status quo, a 140 or 150 MDL requirement, a 420 or 450 overall point requirement, and combinations of the above, including the Practice Phase standard of 450 overall points + MDL 150 pounds. They were asked to rate whether any of the generated list of ten concerns presented challenges for each of the standard options.

In the third session, stakeholders were presented with the most frequently chosen concerns across standards, and implementation facilitation strategies were subsequently generated in small-group discussion format. (See discussion of the rating results in Chapter 4.) After brainstorming mitigation strategies, small groups presented their top choice of facilitation strategy to the larger group and were guided in discussion to generate concerns and nuance with implementation of any given facilitation strategies. Although not all facilitation strategies were appropriate for all concerns, several raised in the discussion alleviated multiple concerns. The fourth and final session focused on unit-specific concerns such as, inter alia, how higher standards for close combat MOSs would affect training plans for units with a mix of close combat and non–close combat MOSs.

Strategies to Facilitate Implementation of Higher Standards

After a thorough discussion of potential challenges relating to various options to increase close combat MOS standards, the discussion focused on developing facilitation strategies to alleviate the challenges. A variety of specific facilitation strategies were discussed in small groups and proposed and discussed at the larger workshop group level, but in general they revolved around a few themes.

Glide Paths

The first facilitation strategy theme addressed glide paths: providing adequate time to train up to new standards, which would address some of the challenges relating to injuries, insufficient training time, and pass rates for the USAR and ARNG. Participants nearly unanimously expressed concerns about increased injury risk for soldiers attempting to immediately meet higher standards that they have not been training for, and this was evident in both the group discussions and the ratings of concerns. However, time for seasoned close combat soldiers to meet a higher standard was not the only nuance. Discussion also raised the issue of new soldiers needing time to train and develop fitness, especially as broader U.S. fitness declines. Facilitation strategies included variants on glide paths relevant to the two related issues of time.

Implementation Timeline

For seasoned soldiers in particular, an "implementation timeline" (the time between when the Army announces a specific higher close combat ACFT standard and when soldiers face as-yet-to-be-determined policy repercussions for failing to meet that standard) was discussed that included an extended rollout for the ARNG and USAR. Importantly, it was noted that the career repercussions of failure should be announced at the same time as the higher standards themselves, or else many soldiers will assume that the new standards are merely guidelines with "no teeth."

An implementation timeline would be a one-time grace period for current soldiers to meet the new close combat standards. Participants favored an implementation timeline similar to that of ACFT 3.0: The Army

should announce the new close combat standard (and the impact of failing), and then wait six months for active duty soldiers and one year for the ARNG and USAR soldiers before applying those sanctions to failures.[2]

ARNG and USAR service members will have a mobilization period that can be used to meet active component close combat ACFT standards prior to deployment, although there are many competing priorities that must be met during the time, and additional resources may be required to accomplish this goal. Given this window, however, and difficulties associated with access to equipment and time to train for ARNG and USAR service members in Troop Program Unit status, an intermediate standard above the general fitness requirement but below the close combat requirement(s) might meet the combat readiness objectives of the FY 2024 NDAA while reducing disruption to the ARNG and USAR.

Compliance Timeline

For new soldiers in particular, a glide path was suggested to allow meeting particular benchmarks on the path to the final standard. Rather than a one-time "implementation timeline," this might be considered a "compliance timeline" that might entail a 6-month and 12-month benchmark, with each benchmark closer to the new, higher standard. These glide paths would depend on the specifics of the higher standards themselves.

For new entrants, one aspect included re-administration of the OPAT, possibly during BCT. This would confirm that entrants were meeting their initial fitness requirements; discussion suggested that during the recruitment phase the incentives might encourage providing "aspirational" rather than achieved OPAT scores. Participants recommended monthly (or quarterly for ARNG/USAR) diagnostic tests to ensure that soldiers were continuing to make progress on their ACFT scores. Participants noted the higher availability of H2F resources at permanent duty stations compared with BCT and OSUT and said that new soldiers would benefit from access to those resources for at least some period of time before facing negative career repercussions for failing to meet the close combat ACFT standards. Time in the unit also alleviated challenges associated with training timelines that differed by MOS.

Clear Consequences Associated with Standards

Discussion raised the need for clear consequences for not meeting higher standards (e.g., reclassification, incentive payback) that would be clearly communicated and enforced using a transparent timeline. This also included ensuring that OPAT scores were sufficient to enter a career field. This would address some of the concerns regarding appropriate supportive culture and incentives, as well as perceived lack of transparency in test-related communication.

Stakeholders asserted that without such consequences, uniformly imposed, higher standards for close combat MOSs would not achieve the ultimate goal of encouraging higher overall fitness but instead the standards would be seen as toothless and adding to the confusion around the changes in the ACFT and challenges maintaining fitness culture owing to the COVID-19 pandemic. Clarity was discussed for related issues, including consequences for repeated ACFT failure. The appropriate policies and expectations around going on profile, and for going off profile, were also discussed. Discussion indicated that the expectations around going on and off profile were inconsistent and arbitrary.

[2] Some participants recommended the immediate implementation of the higher close combat standards for schools. This provoked some debate between those who thought higher standards were appropriate—especially for badge-producing schools, such as Airborne—and those (generally more senior noncommissioned officers) who were concerned about the disruption this could cause for soldiers currently slotted. It also raised fairness concerns, as soldiers in non–close combat MOSs would not face the same restriction.

Resources to Facilitate Compliance

Several facilitation strategies related to increased fitness education or access to other resources in one way or another. This facilitation strategy would address issues relating to risk of injuries, time needed for improvement, and challenges related to pass rates in the ARNG and USAR. There was general agreement that H2F resources could be of great utility, especially as they roll out in the reserve and National Guard. However, even among active component soldiers, there was concern that access to these resources varied: For example, some soldiers indicated that company commanders needed to request that fitness clinics be held at their units, and that otherwise the individuals with expertise in H2F were not effectively accessible to soldiers. This might require better informing soldiers about these resources and ensuring greater outreach to the company and squad level. Additional fitness-related content was proposed for trainees in BCT and for soldiers in more advanced professional military education, although there was general agreement that this would be extremely difficult to implement.

Clarify ACFT Goals

Extended discussion of standards revealed the importance of communication during implementation of higher standards for close combat MOSs, although this challenge was not evident in stakeholders' ratings. Implementation of higher standards needs to take into account that soldiers need to understand the purpose of the ACFT and the need for higher standards. This addresses concerns noted with scoring communication.

Stakeholders said that messaging around the ACFT was inconsistent and expressed confusion regarding the need for higher standards. They also expressed confusion regarding the goals of the test, which as currently implemented are for general health and fitness. This nuance is not well understood, in part because the word *combat* remains in the test's name, and initial marketing of the ACFT played up its combat-relevance (while actual validation evidence was insufficient to substantiate these associations). Additional communication concerns were related to potential challenges to unit leadership: Mixed MOSs within a unit mean that some soldiers not in the close combat MOSs targeted by the FY 2024 NDAA may also be held to a higher combat standard, for reasons of pure administrative simplicity within a unit. Messaging to address this aspect of likely implementation in practice may be beneficial.

Discussion also revealed that the need for a clear communication strategy would be particularly acute should the policy determination include an MDL increase. Stakeholders did not understand why the MDL, specifically, was chosen as the validation exercise when many other components of the ACFT, such as the SDC, seemed more relevant to close combat tasks. Many participants expressed surprise that the MDL was the only ACFT component with sufficient validation to link to close combat tasks—and this provoked an extended, primarily negative, discussion of the ACFT in general. Participants also expressed concerns about injury risk in training their soldiers for the MDL and noted that it is the one exercise that they are most concerned about leading to injuries.[3] A final communication difficulty centered on the amount of the standard increases under consideration. A 140-pound minimum MDL standard would have no effect—and a 150-pound minimum MDL would have almost no effect—on male soldiers, who are the majority of the close combat force. Instead, the higher MDL standards would primarily affect female soldiers in close combat MOSs.

[3] This concern is not without merit. Evidence suggests that performance in the highest-scoring groups for both men and women is associated with an increased risk of acute trauma injuries (Hicks et al., forthcoming).

Remedial Assessment

A final facilitation strategy was the use of some form of remedial assessment. This would address such challenges as time to train and injury concerns and would potentially support fitness culture and incentives surrounding ACFT use. Stakeholders highlighted several ways in which MOS-specific combat fitness was assessed beyond the ACFT HPDTs, Expert Physical Fitness Assessment, and other unit-specific assessments. Given challenges in setting up ACFT testing and time requirements, such assessments might be used to demonstrate incremental improvements for soldiers struggling to meet new ACFT standards and provide targets along an implementation timeline, and so could best be used for soldiers whose ACFT scores do not meet the combat standard but still exceed the general-purpose force standard. Such a strategy would provide commanders additional assurance that a given soldier was capable of executing combat tasks. However, it would be essential to standardize these assessments to ensure that their use is transparent and consistent. For example, a standardized, predetermined list of HPDTs for a given MOS could be developed and used as the assessment.

APPENDIX D

Pass Rates by Subgroups for Each Alternative Standard

In this appendix, we show the pass rates and confidence intervals from Practice Phase scores for gender and age subgroups and MOS subgroups for the Regular Army in Figures D.1 and D.2, respectively. We also present the specific pass rates for these subgroups with adjusted pass rates to account for prior ACFT performance and for potential improvement (Tables D.1–D.8). In these tables, which were presented to workshop participants, we shade all rows where the subgroup Practice Phase pass rate or lower-bound confidence interval is below 90 percent. The same data are shown for the ARNG and USAR in Figures D.3–D.4 and Tables D.9–D.16.

To aid interpretation, we highlight, using red italics, observed pass rates below 90 percent. We also highlight lower-bound estimates below 90 percent when the observed pass rates may be 90 percent or higher. The data in the last two columns make the following adjustments:

- **+ Crediting Prior Performance:** Presents adjusted pass rates that assume that soldiers who did not meet the standard during the Practice Phase but met the standard in a previous for-record test would pass again in the future.
- **+ Possible Improvements:** Presents adjusted pass rates that credit prior performance and improvements expected by soldiers who are predicted to be within range of meeting the standard within one year.[1]

[1] See Appendix B for a detailed discussion of the methodology employed to estimate predicted improvement over 12 months.

FIGURE D.1

Pass Rates and Confidence Intervals for Regular Army, by Gender and Age Group

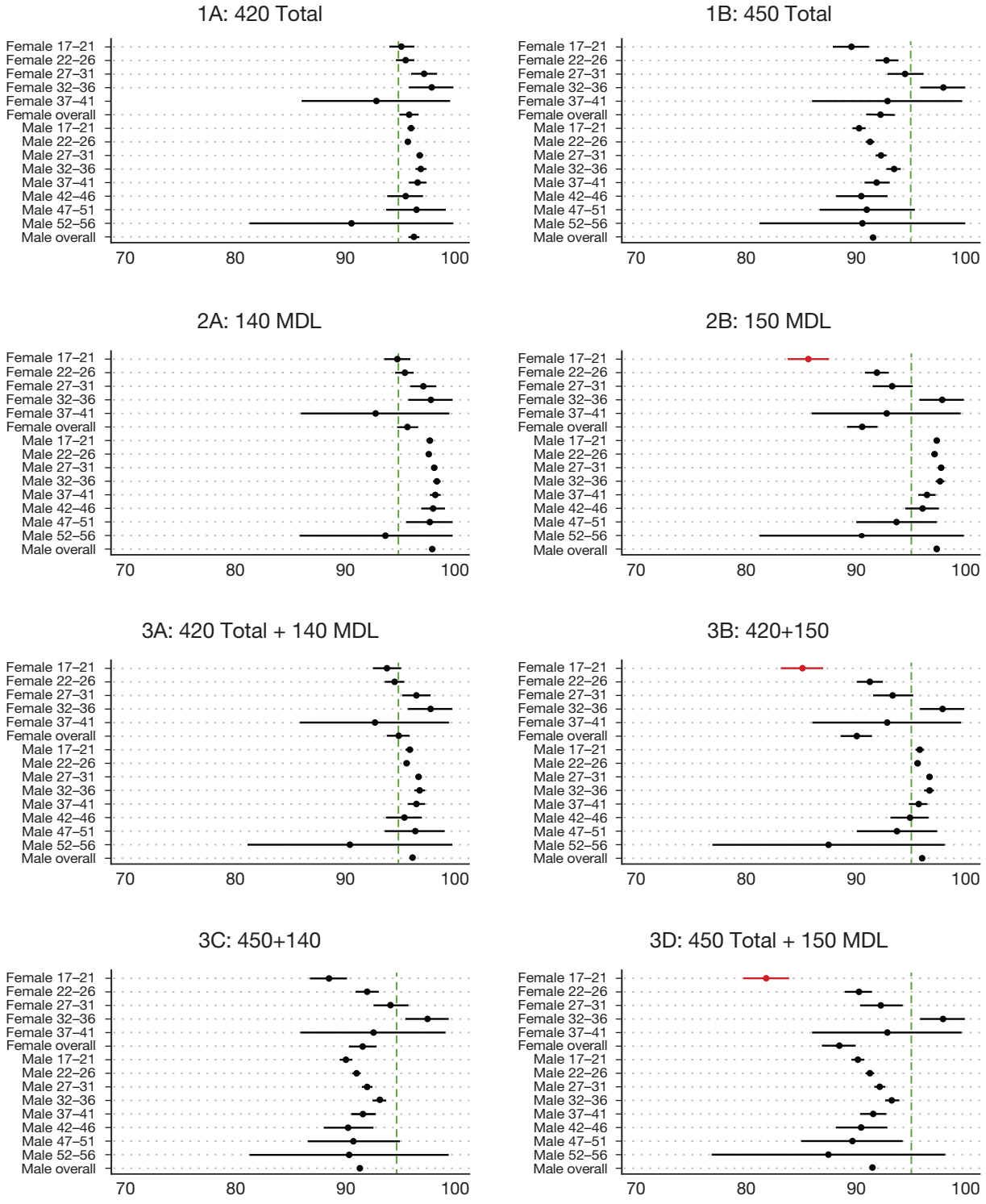


FIGURE D.2

Pass Rates and Confidence Intervals for Regular Army, by MOS

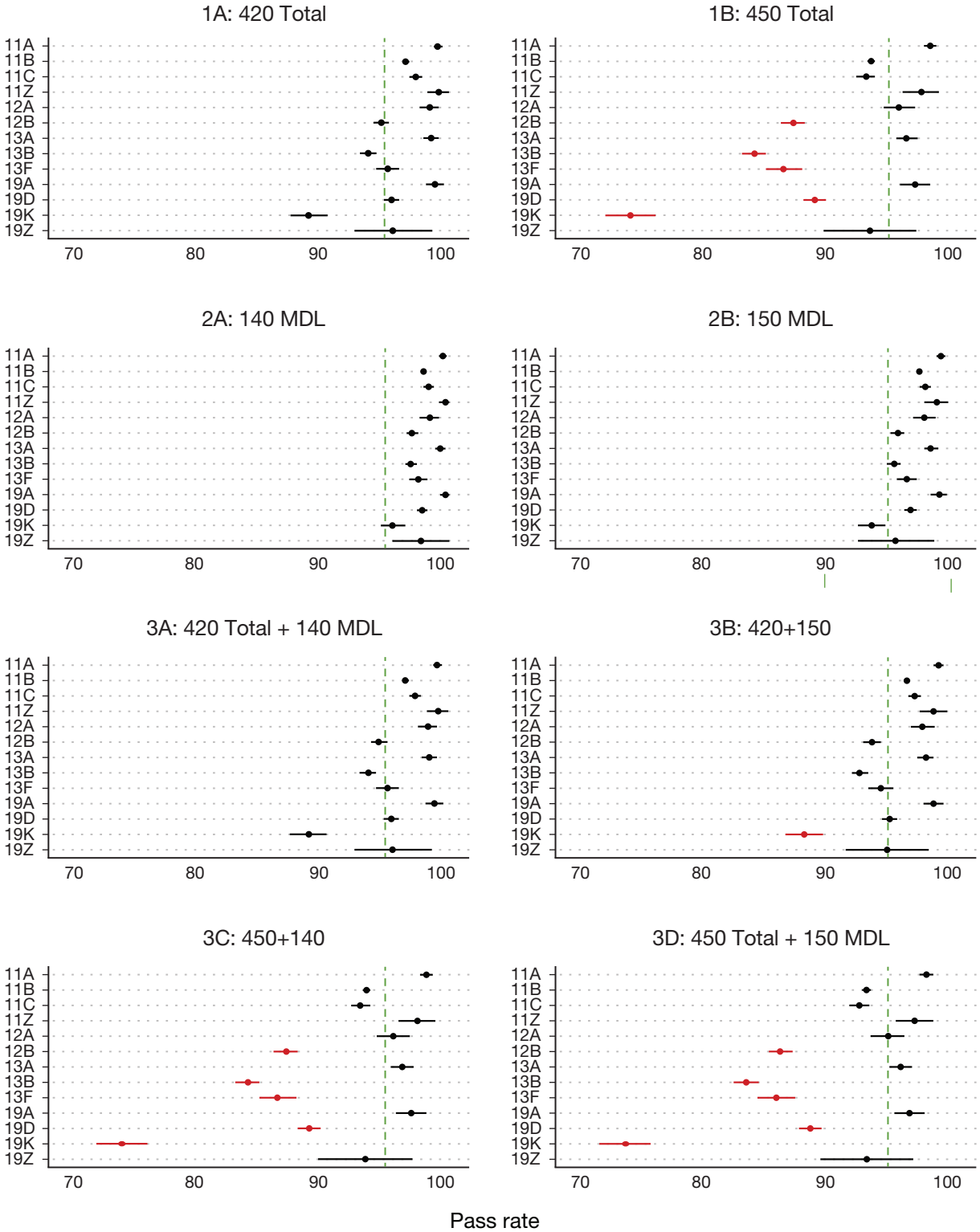


TABLE D.1

Pass Rates and Confidence Intervals for Regular Army—Standard 1A: 420 Total

	Observed Pass Rates			Pass Rates with Assumptions	
Subgroup	Practice Phase	95% CI - Lower	95% CI - Upper	+ Credit Previously Met Standard	+ Improvement Within Range
Overall					
Overall	96.3	96.1	96.4	98.7	99.2
Female Overall	95.8	94.8	96.8	98.5	98.6
Male Overall	96.3	96.1	96.4	98.8	99.2
Gender-Age					
Female 17–21	95.2	94.1	96.4	98.2	98.2
Female 22–26	95.6	94.7	96.4	98.8	99.0
Female 27–31	97.3	96.1	98.5	98.4	98.9
Female 32–36	98.0	95.9	100.0	98.0	98.0
Female 37–41	92.9	*86.0*	99.7	92.9	92.9
Male 17–21	96.1	95.7	96.4	98.5	99.0
Male 22–26	95.8	95.6	96.1	98.8	99.2
Male 27–31	96.9	96.6	97.2	99.0	99.4
Male 32–36	97.0	96.5	97.5	98.8	99.4
Male 37–41	96.7	95.9	97.5	98.5	99.2
Male 42–46	95.6	93.9	97.2	97.8	99.0
Male 47–51	96.6	93.8	99.3	97.2	97.2
Male 52–56	90.6	*81.2*	100.0	90.6	93.8
MOS					
11A	99.1	98.8	99.5	99.9	100.0
11B	96.6	96.4	96.9	99.0	99.3
11C	97.4	96.9	97.9	99.3	99.5
11Z	99.2	98.3	100.0	99.2	99.4
12A	98.5	97.7	99.2	99.1	99.2
12B	94.7	94.1	95.3	98.3	99.0
13A	98.6	98.0	99.2	99.3	99.7
13B	93.7	93.0	94.3	96.9	97.8
13F	95.2	94.3	96.1	97.5	98.2
19A	98.9	98.2	99.6	99.5	99.7
19D	95.5	94.9	96.1	98.6	99.3
19K	*89.0*	87.6	90.5	96.2	98.3
19Z	95.6	92.6	98.7	99.3	100.0

TABLE D.2

Pass Rates and Confidence Intervals for Regular Army—Standard 1B: 450 Total

	Observed Pass Rates			Pass Rates with Assumptions	
Subgroup	Practice Phase	95% CI - Lower	95% CI - Upper	+ Credit Previously Met Standard	+ Improvement Within Range
Overall					
Overall	91.6	91.4	91.8	96.0	97.6
Female Overall	92.3	91.0	93.6	96.2	97.2
Male Overall	91.6	91.3	91.8	96.0	97.6
Gender-Age					
Female 17–21	*89.6*	87.9	91.2	94.0	96.4
Female 22–26	92.8	91.8	93.9	97.1	97.4
Female 27–31	94.5	92.9	96.2	96.7	97.8
Female 32–36	98.0	95.9	100.0	98.0	98.0
Female 37–41	92.9	*86.0*	99.7	92.9	92.9
Male 17–21	90.3	*89.7*	90.9	94.8	97.1
Male 22–26	91.3	90.9	91.7	96.3	97.6
Male 27–31	92.3	91.8	92.8	96.5	97.8
Male 32–36	93.5	92.8	94.1	96.6	98.2
Male 37–41	91.9	90.8	93.1	96.1	98.0
Male 42–46	90.5	*88.2*	92.9	94.8	97.8
Male 47–51	91.0	*86.7*	95.4	95.2	96.6
Male 52–56	90.6	*81.2*	100.0	90.6	93.8
MOS					
11A	98.2	97.7	98.7	99.6	99.9
11B	93.5	93.2	93.8	97.4	98.5
11C	93.1	92.3	93.8	96.8	98.1
11Z	97.5	96.0	98.9	98.6	99.2
12A	95.7	94.5	97.0	98.1	98.9
12B	*87.3*	86.3	88.2	94.4	96.2
13A	96.3	95.5	97.2	98.1	99.3
13B	*84.2*	83.2	85.1	90.2	93.3
13F	*86.5*	85.1	88.0	92.3	95.4
19A	97.0	95.8	98.2	98.0	99.5
19D	*89.0*	88.1	89.9	95.0	97.1
19K	*74.3*	72.3	76.3	83.7	90.2
19Z	93.4	*89.7*	97.1	98.5	99.3

TABLE D.3

Pass Rates and Confidence Intervals for Regular Army—Standard 2A: 140 MDL

	Observed Pass Rates			Pass Rates with Assumptions	
Subgroup	Practice Phase	95% CI - Lower	95% CI - Upper	+ Credit Previously Met Standard	+ Improvement Within Range
Overall					
Overall	98.0	97.8	98.1	99.4	99.4
Female Overall	95.7	94.8	96.7	98.1	98.3
Male Overall	98.0	97.9	98.2	99.5	99.5
Gender-Age					
Female 17–21	94.9	93.7	96.1	97.6	97.9
Female 22–26	95.6	94.7	96.4	98.3	98.5
Female 27–31	97.3	96.1	98.5	98.9	98.9
Female 32–36	98.0	95.9	100.0	98.0	98.0
Female 37–41	92.9	*86.0*	99.7	92.9	92.9
Male 17–21	97.9	97.6	98.2	99.3	99.3
Male 22–26	97.8	97.6	98.0	99.5	99.5
Male 27–31	98.3	98.1	98.6	99.6	99.6
Male 32–36	98.6	98.2	98.9	99.6	99.6
Male 37–41	98.4	97.9	98.9	99.4	99.5
Male 42–46	98.2	97.1	99.3	99.4	99.4
Male 47–51	97.9	95.7	100.0	97.9	97.9
Male 52–56	93.8	*85.9*	100.0	93.8	93.8
MOS					
11A	99.5	99.2	99.8	100.0	100.0
11B	98.0	97.8	98.1	99.4	99.4
11C	98.4	98.0	98.8	99.6	99.7
11Z	99.7	99.2	100.0	99.7	99.7
12A	98.5	97.7	99.2	99.2	99.2
12B	97.1	96.7	97.6	99.2	99.3
13A	99.3	98.9	99.7	99.8	99.8
13B	97.0	96.6	97.5	98.9	98.9
13F	97.6	96.9	98.3	99.1	99.1
19A	99.7	99.3	100.0	99.7	99.7
19D	97.9	97.5	98.3	99.5	99.5
19K	95.6	94.7	96.6	99.0	99.0
19Z	97.8	95.6	100.0	100.0	100.0

TABLE D.4

Pass Rates and Confidence Intervals for Regular Army—Standard 2B: 150 MDL

	Observed Pass Rates			Pass Rates with Assumptions	
Subgroup	Practice Phase	95% CI - Lower	95% CI - Upper	+ Credit Previously Met Standard	+ Improvement Within Range
Overall					
Overall	97.2	97.1	97.4	99.1	99.3
Female Overall	90.6	*89.2*	92.0	95.0	96.4
Male Overall	97.4	97.3	97.6	99.3	99.4
Gender-Age					
Female 17–21	*85.7*	83.8	87.6	91.6	94.3
Female 22–26	92.0	90.9	93.1	96.2	97.1
Female 27–31	93.4	91.6	95.3	96.2	97.8
Female 32–36	98.0	95.9	100.0	98.0	98.0
Female 37–41	92.9	*86.0*	99.7	92.9	92.9
Male 17–21	97.5	97.2	97.8	99.3	99.3
Male 22–26	97.3	97.0	97.5	99.3	99.4
Male 27–31	97.9	97.6	98.2	99.5	99.6
Male 32–36	97.8	97.4	98.2	99.3	99.6
Male 37–41	96.6	95.8	97.4	98.6	99.1
Male 42–46	96.2	94.6	97.7	98.2	99.4
Male 47–51	93.8	90.1	97.5	95.9	97.2
Male 52–56	90.6	*81.2*	100.0	90.6	90.6
MOS					
11A	99.2	98.9	99.6	99.8	100.0
11B	97.5	97.3	97.7	99.3	99.4
11C	98.0	97.5	98.4	99.4	99.6
11Z	98.9	97.9	99.8	99.4	99.7
12A	97.9	97.0	98.8	98.9	99.1
12B	95.8	95.2	96.3	98.7	98.9
13A	98.4	97.9	99.0	99.1	99.5
13B	95.5	94.9	96.0	98.2	98.5
13F	96.5	95.7	97.3	98.7	98.9
19A	99.1	98.4	99.7	99.4	99.7
19D	96.8	96.3	97.3	99.2	99.4
19K	93.7	92.6	94.8	97.9	98.5
19Z	95.6	92.6	98.7	100.0	100.0

TABLE D.5

Pass Rates and Confidence Intervals for Regular Army—Standard 3A: 420 Total + 140 MDL

	Observed Pass Rates			Pass Rates with Assumptions	
Subgroup	Practice Phase	95% CI - Lower	95% CI - Upper	+ Credit Previously Met Standard	+ Improvement Within Range
Overall					
Overall	96.2	96.1	96.4	98.7	99.2
Female Overall	95.0	93.9	96.0	97.8	98.0
Male Overall	96.3	96.1	96.4	98.8	99.2
Gender-Age					
Female 17–21	94.0	92.7	95.3	97.3	97.6
Female 22–26	94.7	93.8	95.6	98.0	98.1
Female 27–31	96.7	95.4	98.0	98.4	98.9
Female 32–36	98.0	95.9	100.0	98.0	98.0
Female 37–41	92.9	*86.0*	99.7	92.9	92.9
Male 17–21	96.1	95.7	96.4	98.5	99.0
Male 22–26	95.8	95.6	96.1	98.8	99.2
Male 27–31	96.9	96.6	97.2	99.0	99.4
Male 32–36	97.0	96.5	97.5	98.8	99.4
Male 37–41	96.7	95.9	97.5	98.5	99.2
Male 42–46	95.6	93.9	97.2	97.8	99.0
Male 47–51	96.6	93.8	99.3	97.2	97.2
Male 52–56	90.6	*81.2*	100.0	90.6	93.8
MOS					
11A	99.1	98.8	99.5	99.9	100.0
11B	96.6	96.4	96.9	99.0	99.3
11C	97.4	96.9	97.8	99.2	99.5
11Z	99.2	98.3	100.0	99.2	99.4
12A	98.4	97.6	99.1	99.0	99.1
12B	94.5	93.9	95.2	98.2	98.9
13A	98.5	97.9	99.1	99.2	99.6
13B	93.7	93.0	94.3	96.9	97.8
13F	95.2	94.3	96.1	97.5	98.2
19A	98.9	98.2	99.6	99.5	99.7
19D	95.5	94.9	96.1	98.6	99.3
19K	*89.0*	87.5	90.4	96.2	98.2
19Z	95.6	92.6	98.7	99.3	100.0

TABLE D.6

Pass Rates and Confidence Intervals for Regular Army—Standard 3B: 420 Total + 150 MDL

	Observed Pass Rates			Pass Rates with Assumptions	
Subgroup	Practice Phase	95% CI - Lower	95% CI - Upper	+ Credit Previously Met Standard	+ Improvement Within Range
Overall					
Overall	95.9	95.7	96.0	98.5	99.1
Female Overall	90.1	*88.6*	91.5	94.6	96.1
Male Overall	96.1	95.9	96.2	98.7	99.2
Gender-Age					
Female 17–21	*85.1*	83.1	87.0	91.0	93.7
Female 22–26	91.3	90.1	92.5	95.9	96.8
Female 27–31	93.4	91.6	95.3	96.2	97.8
Female 32–36	98.0	95.9	100.0	98.0	98.0
Female 37–41	92.9	*86.0*	99.7	92.9	92.9
Male 17–21	95.9	95.5	96.3	98.5	99.0
Male 22–26	95.7	95.4	95.9	98.8	99.2
Male 27–31	96.8	96.5	97.1	98.9	99.4
Male 32–36	96.8	96.3	97.2	98.6	99.4
Male 37–41	95.8	94.9	96.6	98.1	99.0
Male 42–46	95.0	93.2	96.7	97.6	99.0
Male 47–51	93.8	90.1	97.5	95.9	97.2
Male 52–56	*87.5*	76.8	98.2	87.5	90.6
MOS					
11A	99.0	98.6	99.4	99.8	100.0
11B	96.5	96.2	96.7	98.9	99.2
11C	97.1	96.6	97.6	99.1	99.5
11Z	98.6	97.5	99.7	99.2	99.4
12A	97.7	96.8	98.7	98.6	99.0
12B	93.7	93.0	94.4	97.8	98.5
13A	98.0	97.3	98.6	98.9	99.5
13B	92.7	92.1	93.4	96.5	97.5
13F	94.4	93.4	95.4	97.2	98.1
19A	98.6	97.8	99.4	99.2	99.7
19D	95.1	94.5	95.7	98.4	99.2
19K	*88.3*	86.8	89.8	95.5	97.6
19Z	94.9	91.6	98.2	99.3	100.0

TABLE D.7

Pass Rates and Confidence Intervals for Regular Army—Standard 3C: 450 Total + 140 MDL

	Observed Pass Rates			Pass Rates with Assumptions	
Subgroup	Practice Phase	95% CI - Lower	95% CI - Upper	+ Credit Previously Met Standard	+ Improvement Within Range
Overall					
Overall	91.6	91.3	91.8	96.0	97.6
Female Overall	91.8	90.5	93.1	95.6	96.6
Male Overall	91.6	91.3	91.8	96.0	97.6
Gender-Age					
Female 17–21	*88.7*	86.9	90.4	93.1	95.5
Female 22–26	92.3	91.2	93.4	96.4	96.8
Female 27–31	94.5	92.9	96.2	96.7	97.8
Female 32–36	98.0	95.9	100.0	98.0	98.0
Female 37–41	92.9	*86.0*	99.7	92.9	92.9
Male 17–21	90.3	*89.7*	90.9	94.8	97.1
Male 22–26	91.3	90.9	91.7	96.3	97.6
Male 27–31	92.3	91.8	92.8	96.5	97.8
Male 32–36	93.5	92.8	94.1	96.6	98.2
Male 37–41	91.9	90.8	93.1	96.1	98.0
Male 42–46	90.5	*88.2*	92.9	94.8	97.8
Male 47–51	91.0	*86.7*	95.4	95.2	96.6
Male 52–56	90.6	*81.2*	100.0	90.6	93.8
MOS					
11A	98.2	97.7	98.7	99.6	99.9
11B	93.5	93.2	93.8	97.4	98.5
11C	93.0	92.3	93.8	96.8	98.1
11Z	97.5	96.0	98.9	98.6	99.2
12A	95.6	94.3	96.9	98.0	98.7
12B	*87.2*	86.2	88.1	94.3	96.1
13A	96.3	95.4	97.2	98.1	99.2
13B	*84.2*	83.2	85.1	90.2	93.2
13F	*86.5*	85.1	88.0	92.3	95.4
19A	97.0	95.8	98.2	98.0	99.5
19D	*89.0*	88.1	89.9	95.0	97.1
19K	*74.3*	72.3	76.3	83.7	90.2
19Z	93.4	*89.7*	97.1	98.5	99.3

TABLE D.8

Pass Rates and Confidence Intervals for Regular Army—Standard 3D: 450 Total + 150 MDL

	Observed Pass Rates			Pass Rates with Assumptions	
Subgroup	Practice Phase	95% CI - Lower	95% CI - Upper	+ Credit Previously Met Standard	+ Improvement Within Range
Overall					
Overall	91.4	91.1	91.6	95.9	97.5
Female Overall	*88.5*	86.9	90.0	93.3	95.0
Male Overall	91.5	91.2	91.7	96.0	97.6
Gender-Age					
Female 17–21	*81.8*	79.7	83.9	88.4	91.9
Female 22–26	90.3	*89.0*	91.5	95.1	95.9
Female 27–31	92.3	90.4	94.3	95.6	97.3
Female 32–36	98.0	95.9	100.0	98.0	98.0
Female 37–41	92.9	*86.0*	99.7	92.9	92.9
Male 17–21	90.2	*89.6*	90.8	94.8	97.1
Male 22–26	91.3	90.9	91.7	96.3	97.6
Male 27–31	92.2	91.7	92.7	96.4	97.8
Male 32–36	93.3	92.7	94.0	96.6	98.2
Male 37–41	91.6	90.4	92.8	95.9	97.7
Male 42–46	90.5	*88.2*	92.9	94.8	97.8
Male 47–51	*89.7*	85.0	94.3	95.2	96.6
Male 52–56	*87.5*	76.8	98.2	87.5	90.6
MOS					
11A	98.2	97.6	98.7	99.5	99.9
11B	93.4	93.0	93.7	97.3	98.4
11C	92.8	92.0	93.6	96.6	98.1
11Z	97.2	95.7	98.7	98.6	99.2
12A	95.1	93.7	96.4	97.6	98.6
12B	*86.5*	85.6	87.5	93.9	95.8
13A	96.1	95.2	97.0	98.0	99.2
13B	*83.8*	82.8	84.8	90.1	93.1
13F	*86.2*	84.7	87.7	92.1	95.2
19A	96.8	95.6	98.0	97.9	99.5
19D	*88.9*	88.0	89.8	94.9	97.1
19K	*74.2*	72.1	76.2	83.6	90.2
19Z	93.4	*89.7*	97.1	98.5	99.3

FIGURE D.3

Pass Rates and Confidence Intervals for ARNG and USAR, by Gender and Age Group

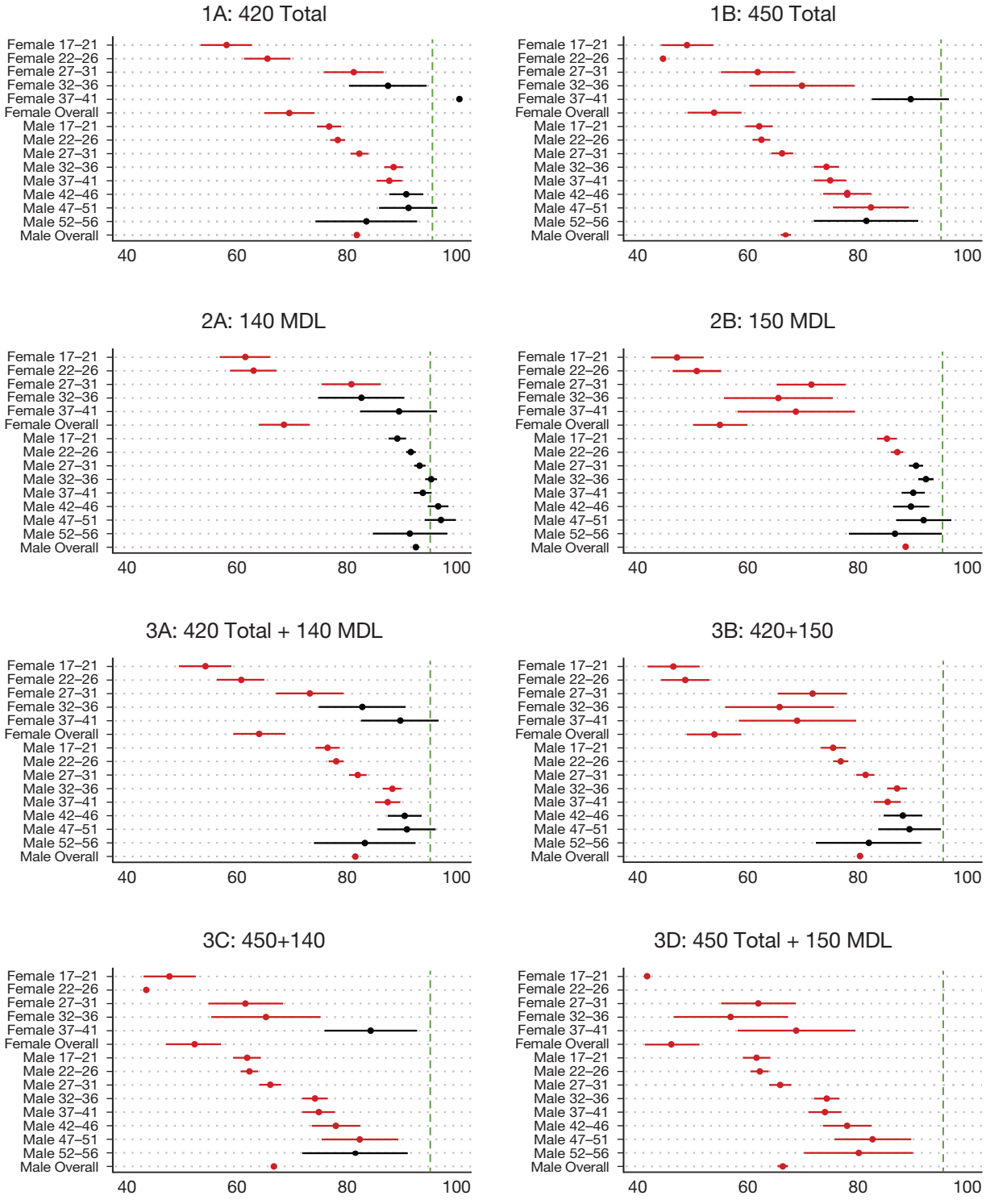


FIGURE D.4

Pass Rates and Confidence Intervals for ARNG and USAR, by MOS

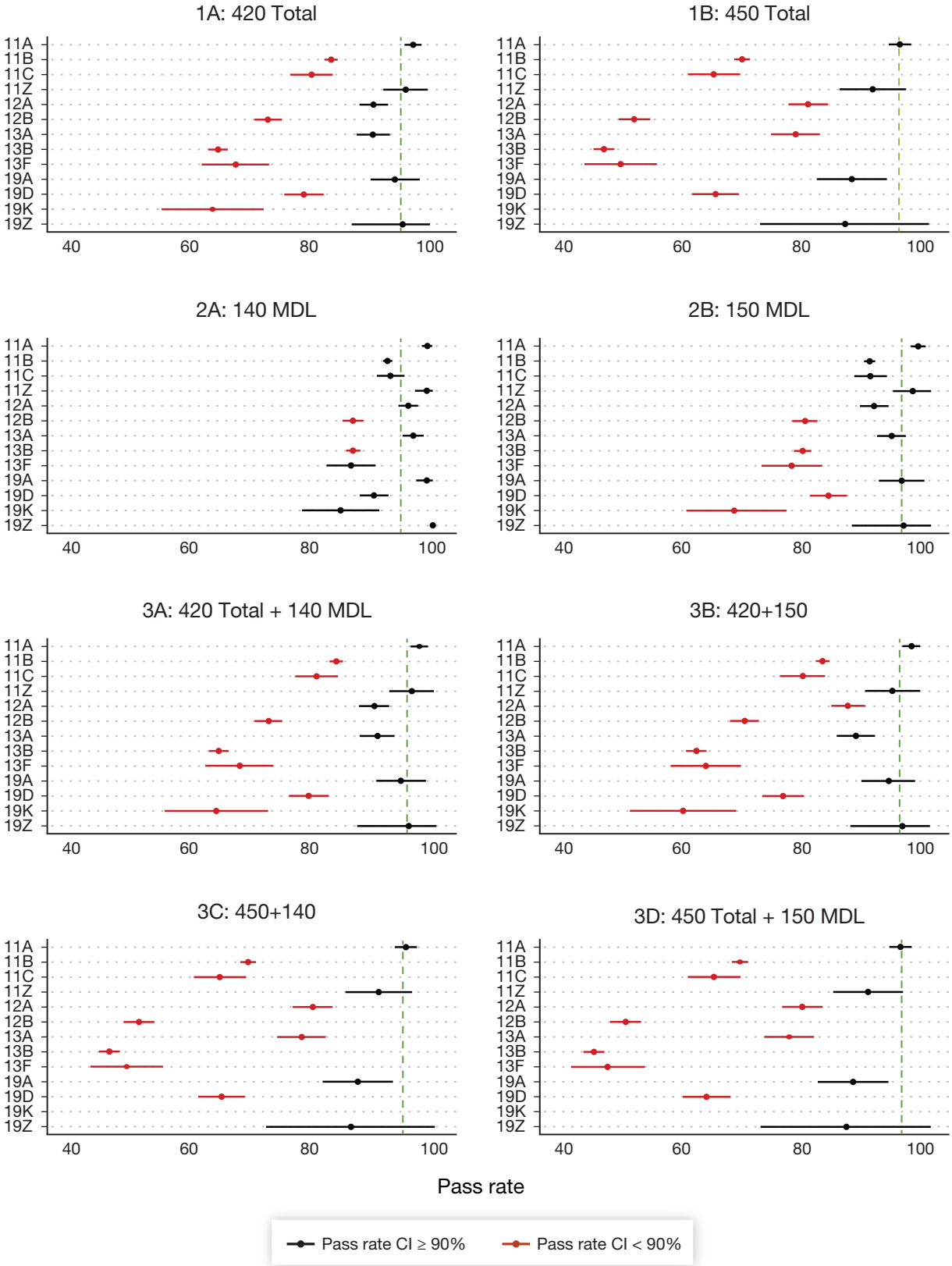


TABLE D.9

Pass Rates and Confidence Intervals for ARNG + USAR—Standard 1A: 420 Total

	Observed Pass Rates			Pass Rates with Assumptions	
Subgroup	Practice Phase	95% CI - Lower	95% CI - Upper	+ Credit Previously Met Standard	+ Improvement Within Range
Overall					
Overall	*80.8*	80.1	81.6	87.5	91.4
Female Overall	*69.1*	64.5	73.7	78.1	81.0
Male Overall	*81.3*	80.5	82.0	87.9	91.8
Gender-Age					
Female 17–21	*57.7*	53.0	62.3	70.3	70.3
Female 22–26	*65.1*	60.9	69.3	73.6	79.1
Female 27–31	*80.8*	75.3	86.2	88.5	90.4
Female 32–36	*87.0*	79.9	94.0	95.7	100.0
Female 37–41	100.0	100.0	100.0	100.0	100.0
Male 17–21	*76.3*	74.1	78.5	84.5	89.1
Male 22–26	*77.9*	76.5	79.2	86.1	90.0
Male 27–31	*81.8*	80.2	83.4	87.8	92.5
Male 32–36	*88.0*	86.3	89.7	92.7	95.5
Male 37–41	*87.2*	84.9	89.5	91.6	94.4
Male 42–46	90.3	*87.2*	93.4	93.0	95.0
Male 47–51	90.7	*85.4*	95.9	91.6	97.2
Male 52–56	*83.1*	73.8	92.3	86.4	88.1
MOS					
11A	97.2	95.8	98.6	99.1	99.8
11B	*83.7*	82.6	84.8	89.6	93.4
11C	*80.5*	77.0	84.0	86.0	92.4
11Z	96.0	92.3	99.6	99.0	99.0
12A	90.7	*88.4*	93.1	94.1	96.0
12B	*73.3*	71.1	75.6	82.1	86.7
13A	90.6	*87.9*	93.4	93.0	95.9
13B	*65.1*	63.5	66.7	77.1	83.2
13F	*68.0*	62.4	73.5	79.7	84.8
19A	94.2	90.2	98.3	96.2	98.1
19D	*79.2*	76.0	82.5	86.3	88.3
19K	*64.2*	55.8	72.6	79.2	84.2
19Z	95.5	*87.1*	100.0	95.5	95.5

TABLE D.10

Pass Rates and Confidence Intervals for ARNG + USAR—Standard 1B: 450 Total

	Observed Pass Rates			Pass Rates with Assumptions	
Subgroup	Practice Phase	95% CI - Lower	95% CI - Upper	+ Credit Previously Met Standard	+ Improvement Within Range
Overall					
Overall	*66.1*	65.2	67.0	74.0	80.9
Female Overall	*53.6*	48.7	58.6	63.0	68.8
Male Overall	*66.6*	65.6	67.5	74.4	81.4
Gender-Age					
Female 17–21	*48.6*	43.9	53.4	57.7	60.4
Female 22–26	*44.2*	39.8	48.6	55.8	64.3
Female 27–31	*61.5*	54.8	68.3	71.2	78.8
Female 32–36	*69.6*	60.0	79.2	73.9	78.3
Female 37–41	*89.5*	82.4	96.5	94.7	94.7
Male 17–21	*61.8*	59.3	64.3	69.7	78.3
Male 22–26	*62.2*	60.6	63.8	72.0	78.8
Male 27–31	*66.0*	64.0	68.0	73.7	80.7
Male 32–36	*74.1*	71.8	76.4	80.4	86.6
Male 37–41	*74.8*	71.8	77.7	80.3	86.9
Male 42–46	*77.9*	73.5	82.3	80.9	86.3
Male 47–51	*82.2*	75.3	89.2	85.0	96.3
Male 52–56	*81.4*	71.8	90.9	83.1	83.1
MOS					
11A	95.3	93.5	97.1	96.7	97.2
11B	*69.7*	68.4	71.0	77.5	84.6
11C	*65.1*	60.9	69.4	72.5	82.8
11Z	90.9	*85.5*	96.3	93.9	96.0
12A	*80.4*	77.2	83.6	85.0	89.9
12B	*52.2*	49.7	54.8	64.0	71.1
13A	*78.4*	74.4	82.3	83.0	88.9
13B	*47.3*	45.6	49.0	57.7	65.4
13F	*50.0*	44.1	55.9	59.4	69.5
19A	*87.5*	81.8	93.2	90.4	95.2
19D	*65.4*	61.6	69.2	71.4	79.0
19K	*38.3*	29.8	46.9	50.8	60.8
19Z	*86.4*	72.6	100.0	86.4	95.5

TABLE D.11

Pass Rates and Confidence Intervals for ARNG + USAR—Standard 2A: 140 MDL

	Observed Pass Rates			Pass Rates with Assumptions	
Subgroup	Practice Phase	95% CI - Lower	95% CI - Upper	+ Credit Previously Met Standard	+ Improvement Within Range
Overall					
Overall	91.7	91.1	92.2	95.8	95.9
Female Overall	*68.2*	63.6	72.9	78.4	81.6
Male Overall	92.6	92.1	93.1	96.5	96.5
Gender-Age					
Female 17–21	*61.3*	56.6	65.9	74.8	75.7
Female 22–26	*62.8*	58.5	67.0	73.6	77.5
Female 27–31	*80.8*	75.3	86.2	88.5	92.3
Female 32–36	*82.6*	74.7	90.5	87.0	95.7
Female 37–41	*89.5*	82.4	96.5	94.7	94.7
Male 17–21	*89.2*	87.6	90.8	94.8	94.8
Male 22–26	91.7	90.8	92.6	95.9	95.9
Male 27–31	93.3	92.3	94.4	97.3	97.4
Male 32–36	95.4	94.3	96.5	98.1	98.1
Male 37–41	93.9	92.2	95.5	96.7	96.7
Male 42–46	96.7	94.8	98.6	97.0	97.0
Male 47–51	97.2	94.2	100.0	97.2	97.2
Male 52–56	91.5	*84.7*	98.4	91.5	91.5
MOS					
11A	99.1	98.2	99.9	100.0	100.0
11B	92.6	91.9	93.4	96.5	96.5
11C	93.1	90.9	95.4	96.3	96.6
11Z	99.0	97.1	100.0	100.0	100.0
12A	96.0	94.4	97.6	97.0	97.5
12B	*87.0*	85.3	88.7	92.8	93.2
13A	96.8	95.1	98.5	98.0	98.2
13B	*87.0*	85.9	88.2	94.5	94.9
13F	*86.7*	82.7	90.7	91.8	91.8
19A	99.0	97.3	100.0	99.0	99.0
19D	90.4	*88.1*	92.8	94.0	94.0
19K	*85.0*	78.7	91.3	95.0	95.0
19Z	100.0	100.0	100.0	100.0	100.0

TABLE D.12

Pass Rates and Confidence Intervals for ARNG + USAR—Standard 2B: 150 MDL

	Observed Pass Rates			Pass Rates with Assumptions	
Subgroup	Practice Phase	95% CI - Lower	95% CI - Upper	+ Credit Previously Met Standard	+ Improvement Within Range
Overall					
Overall	*87.1*	86.4	87.7	93.0	94.6
Female Overall	*54.5*	49.6	59.5	64.4	66.5
Male Overall	*88.3*	87.7	89.0	94.1	95.7
Gender-Age					
Female 17–21	*46.8*	42.1	51.6	60.4	63.1
Female 22–26	*50.4*	46.0	54.8	59.7	61.2
Female 27–31	*71.2*	64.9	77.4	75.0	78.8
Female 32–36	*65.2*	55.3	75.1	87.0	87.0
Female 37–41	*68.4*	57.8	79.1	68.4	68.4
Male 17–21	*84.9*	83.1	86.7	91.8	93.4
Male 22–26	*86.8*	85.6	87.9	93.5	95.2
Male 27–31	90.2	*88.9*	91.5	95.5	96.8
Male 32–36	92.0	90.6	93.4	96.3	97.4
Male 37–41	*89.7*	87.6	91.8	94.7	96.1
Male 42–46	*89.3*	86.0	92.6	93.0	95.3
Male 47–51	91.6	*86.6*	96.6	93.5	96.3
Male 52–56	*86.4*	78.0	94.9	91.5	91.5
MOS					
11A	97.9	96.7	99.1	99.8	100.0
11B	90.0	*89.1*	90.9	95.1	96.0
11C	90.1	*87.5*	92.8	94.7	96.1
11Z	97.0	93.8	100.0	100.0	100.0
12A	90.7	*88.4*	93.1	94.1	95.4
12B	*79.5*	77.4	81.5	87.0	89.5
13A	93.6	91.2	95.9	95.9	96.5
13B	*79.1*	77.7	80.5	89.4	92.6
13F	*77.3*	72.4	82.3	86.3	88.7
19A	95.2	91.5	98.9	98.1	99.0
19D	*83.3*	80.3	86.3	90.9	93.4
19K	*68.3*	60.2	76.5	85.8	91.7
19Z	95.5	*87.1*	100.0	100.0	100.0

TABLE D.13

Pass Rates and Confidence Intervals for ARNG + USAR—Standard 3A: 420 Total + 140 MDL

	Observed Pass Rates			Pass Rates with Assumptions	
Subgroup	Practice Phase	95% CI - Lower	95% CI - Upper	+ Credit Previously Met Standard	+ Improvement Within Range
Overall					
Overall	*80.6*	79.9	81.4	87.4	91.3
Female Overall	*63.8*	59.1	68.6	74.3	78.1
Male Overall	*81.3*	80.5	82.0	87.9	91.8
Gender-Age					
Female 17–21	*54.1*	49.3	58.8	69.4	70.3
Female 22–26	*60.5*	56.2	64.8	69.8	75.2
Female 27–31	*73.1*	66.9	79.2	82.7	86.5
Female 32–36	*82.6*	74.7	90.5	87.0	95.7
Female 37–41	*89.5*	82.4	96.5	94.7	94.7
Male 17–21	*76.3*	74.1	78.5	84.5	89.1
Male 22–26	*77.9*	76.5	79.2	86.1	90.0
Male 27–31	*81.8*	80.2	83.4	87.8	92.5
Male 32–36	*88.0*	86.3	89.7	92.7	95.5
Male 37–41	*87.2*	84.9	89.5	91.6	94.4
Male 42–46	90.3	*87.2*	93.4	93.0	95.0
Male 47–51	90.7	*85.4*	95.9	91.6	97.2
Male 52–56	*83.1*	73.8	92.3	86.4	88.1
MOS					
11A	97.2	95.8	98.6	99.1	99.8
11B	*83.7*	82.6	84.8	89.6	93.4
11C	*80.5*	77.0	84.0	86.0	92.4
11Z	96.0	92.3	99.6	99.0	99.0
12A	*89.9*	87.4	92.3	93.2	95.4
12B	*72.7*	70.4	74.9	81.6	86.3
13A	90.4	*87.5*	93.2	92.7	95.6
13B	*64.6*	63.0	66.2	77.0	83.2
13F	*68.0*	62.4	73.5	79.7	84.8
19A	94.2	90.2	98.3	96.2	98.1
19D	*79.2*	76.0	82.5	86.3	88.3
19K	*64.2*	55.8	72.6	79.2	83.3
19Z	95.5	*87.1*	100.0	95.5	95.5

TABLE D.14

Pass Rates and Confidence Intervals for ARNG + USAR—Standard 3B: 420 Total + 150 MDL

	Observed Pass Rates			Pass Rates with Assumptions	
Subgroup	Practice Phase	95% CI - Lower	95% CI - Upper	+ Credit Previously Met Standard	+ Improvement Within Range
Overall					
Overall	*78.8*	78.0	79.6	86.0	90.4
Female Overall	*53.4*	48.4	58.3	63.3	65.9
Male Overall	*79.8*	79.0	80.5	86.9	91.3
Gender-Age					
Female 17–21	*45.9*	41.2	50.7	58.6	61.3
Female 22–26	*48.1*	43.7	52.5	58.1	61.2
Female 27–31	*71.2*	64.9	77.4	75.0	78.8
Female 32–36	*65.2*	55.3	75.1	87.0	87.0
Female 37–41	*68.4*	57.8	79.1	68.4	68.4
Male 17–21	*74.9*	72.7	77.2	83.6	88.8
Male 22–26	*76.3*	74.9	77.7	85.0	89.5
Male 27–31	*80.8*	79.1	82.4	87.1	92.0
Male 32–36	*86.5*	84.7	88.3	91.7	95.1
Male 37–41	*84.8*	82.3	87.2	90.7	94.3
Male 42–46	*87.6*	84.1	91.1	90.6	93.6
Male 47–51	*88.8*	83.1	94.5	90.7	96.3
Male 52–56	*81.4*	71.8	90.9	86.4	88.1
MOS					
11A	97.0	95.5	98.4	99.1	99.5
11B	*82.6*	81.5	83.7	88.9	93.0
11C	*79.4*	75.7	83.0	85.3	92.2
11Z	93.9	*89.5*	98.4	99.0	99.0
12A	*86.7*	84.0	89.5	91.4	93.7
12B	*70.0*	67.6	72.3	79.1	84.0
13A	*88.0*	84.9	91.1	91.2	94.4
13B	*62.2*	60.5	63.8	74.9	81.8
13F	*63.7*	58.0	69.4	77.0	82.8
19A	93.3	*88.9*	97.6	96.2	98.1
19D	*76.2*	72.8	79.6	84.0	87.7
19K	*60.0*	51.4	68.6	73.3	81.7
19Z	95.5	*87.1*	100.0	95.5	95.5

TABLE D.15

Pass Rates and Confidence Intervals for ARNG + USAR – Standard 3C: 450 Total + 140 MDL

	Observed Pass Rates			Pass Rates with Assumptions	
Subgroup	Practice Phase	95% CI - Lower	95% CI - Upper	+ Credit Previously Met Standard	+ Improvement Within Range
Overall					
Overall	*66.0*	65.1	66.9	73.9	80.8
Female Overall	*52.2*	47.2	57.2	61.8	67.3
Male Overall	*66.6*	65.6	67.5	74.4	81.4
Gender-Age					
Female 17–21	*47.7*	43.0	52.5	57.7	60.4
Female 22–26	*43.4*	39.0	47.8	55.8	63.6
Female 27–31	*61.5*	54.8	68.3	69.2	75.0
Female 32–36	*65.2*	55.3	75.1	69.6	78.3
Female 37–41	*84.2*	75.8	92.6	89.5	89.5
Male 17–21	*61.8*	59.3	64.3	69.7	78.3
Male 22–26	*62.2*	60.6	63.8	72.0	78.8
Male 27–31	*66.0*	64.0	68.0	73.7	80.7
Male 32–36	*74.1*	71.8	76.4	80.4	86.5
Male 37–41	*74.8*	71.8	77.7	80.3	86.9
Male 42–46	*77.9*	73.5	82.3	80.9	86.3
Male 47–51	*82.2*	75.3	89.2	85.0	96.3
Male 52–56	*81.4*	71.8	90.9	83.1	83.1
MOS					
11A	95.3	93.5	97.1	96.7	97.2
11B	*69.7*	68.4	71.0	77.5	84.6
11C	*65.1*	60.9	69.4	72.5	82.6
11Z	90.9	*85.5*	96.3	93.9	96.0
12A	*80.2*	76.9	83.4	84.6	89.5
12B	*52.0*	49.5	54.5	63.8	71.0
13A	*78.4*	74.4	82.3	83.0	88.9
13B	*47.2*	45.5	48.9	57.7	65.4
13F	*50.0*	44.1	55.9	59.4	69.1
19A	*87.5*	81.8	93.2	90.4	95.2
19D	*65.4*	61.6	69.2	71.4	79.0
19K	*38.3*	29.8	46.9	50.8	60.8
19Z	*86.4*	72.6	100.0	86.4	95.5

TABLE D.16

Pass Rates and Confidence Intervals for ARNG + USAR – Standard 3D: 450 Total +150 MDL

	Observed Pass Rates			Pass Rates with Assumptions	
Subgroup	Practice Phase	95% CI - Lower	95% CI - Upper	+ Credit Previously Met Standard	+ Improvement Within Range
Overall					
Overall	*65.3*	64.4	66.2	73.3	80.4
Female Overall	*45.8*	40.8	50.7	55.1	59.2
Male Overall	*66.1*	65.1	67.0	74.0	81.2
Gender-Age					
Female 17–21	*41.4*	36.8	46.1	51.4	54.1
Female 22–26	*37.2*	33.0	41.5	49.6	55.0
Female 27–31	*61.5*	54.8	68.3	65.4	71.2
Female 32–36	*56.5*	46.2	66.9	69.6	73.9
Female 37–41	*68.4*	57.8	79.1	68.4	68.4
Male 17–21	*61.2*	58.7	63.7	69.1	78.1
Male 22–26	*61.8*	60.1	63.4	71.6	78.6
Male 27–31	*65.5*	63.5	67.5	73.5	80.6
Male 32–36	*73.9*	71.6	76.2	80.1	86.3
Male 37–41	*73.6*	70.6	76.6	79.6	86.5
Male 42–46	*77.6*	73.2	82.0	80.9	86.3
Male 47–51	*82.2*	75.3	89.2	85.0	96.3
Male 52–56	*79.7*	69.8	89.5	83.1	83.1
MOS					
11A	95.1	93.3	96.9	96.7	97.2
11B	*69.3*	68.0	70.6	77.2	84.4
11C	*65.1*	60.9	69.4	72.5	82.6
11Z	*89.9*	84.3	95.5	92.9	94.9
12A	*79.3*	76.1	82.6	84.0	88.6
12B	*50.9*	48.4	53.4	62.8	70.1
13A	*77.2*	73.2	81.2	82.2	88.0
13B	*45.8*	44.1	47.5	56.3	64.3
13F	*48.0*	42.1	54.0	57.8	67.6
19A	*87.5*	81.8	93.2	90.4	95.2
19D	*63.9*	60.1	67.8	70.3	78.7
19K	*37.5*	29.0	46.0	50.0	60.8
19Z	*86.4*	72.6	100.0	86.4	95.5

Abbreviations

2MR	Two-Mile Run
ACFT	Army Combat Fitness Test
APFT	Army Physical Fitness Test
ARNG	Army National Guard
BCT	basic combat training
BSPRRS	Baseline Soldier Physical Readiness Requirements Study
CI	confidence interval
COA	course of action
DoDI	Department of Defense Instruction
FY	fiscal year
H2F	Holistic Health and Fitness
HPDT	high physical demand tasks
HRP	Hand Release Push-Up
IAR	Interval Aerobic Run
MAPS	minimally acceptable performance standard
MDL	3 Repetition Maximum Deadlift
MOS	military occupational specialty
NDAA	National Defense Authorization Act
OPAT	Occupational Physical Assessment Test
OSUT	One Station Unit Training
PLK	Plank
SDC	Sprint-Drag-Carry
SDL	Strength Deadlift
SPT	Standing Power Throw
USAR	U.S. Army Reserve
WTBD-ST	Warrior Task and Battle Drill—Simulation Test

References

Agresti, Alan, and Brent A. Coull, "Approximate Is Better than 'Exact' for Interval Estimation of Binomial Proportions," *American Statistician*, Vol. 52, No. 2, 1998. American Education Research Association, American Psychological Association, and National Council on Measure in Education, *Standards for Educational and Psychological Testing*, American Education Research Association, 2014.

American Psychological Association, *Principles for the Validation and Use of Personnel Selection Procedures*, 5th ed., August 2018.

Burstein, Herman, "Finite Population Correction for Binomial Confidence Limits," *Journal of the American Statistical Association*, Vol. 70, No. 349, 1975.

Canino, Maria C., Bruce S. Cohen, Jan E. Redmond, Marilyn A. Sharp, Edward J. Zambraski, and Stephen A. Foulis, "The Relationship Between Soldier Performance on the Two-Mile Run and the 20-m Shuttle Run Test," *Military Medicine*, Vol. 183, No. 5/6, 2018.

Department of Defense Instruction 1308.03, *DoD Physical Fitness/Body Composition Program*, Office of the Under Secretary of Defense for Personnel and Readiness, U.S. Department of Defense, March 10, 2022.

Department of the Army Pamphlet 611-21 (Smartbook), *Military Occupational Classification and Structure*, Headquarters Department of the Army, November 2023.

East, Whitfield B., *Fit to Serve: A History of US Army Physical Readiness*, Army University Press, U.S. Army Combined Arms Center, 2024.

East, Whitfield B., David DeGroot, and Stephanie Muraca-Grabowski, *Baseline Soldier Physical Readiness Requirements Study*, Research and Analysis Division, U.S. Army Center for Initial Military Training, Technical Report T19.041-13.1, November 2019.

Field Manual 7-22, *Holistic Health and Fitness*, Headquarters, Department of the Army, October 2020.

Fitzmaurice, Garrett M., Nan M. Laird, and James H. Ware, *Applied Longitudinal Analysis*, John Wiley & Sons, 2011

Foulis, Stephen A., Jan E. Redmond, Bradley J. Warr, Edward J. Zambraski, Peter N. Frykman, and Marilyn A. Sharp, *Development of the Occupational Physical Assessment Test (OPAT) for Combat Arms Soldiers*, U.S. Army Research Institute of Environmental Medicine, USARIEM Technical Report T16-2, 2015.

FY 2024 NDAA—*See* Public Law 118-31.

Guthold, Regina, Gretchen A. Stevens, Leanne M. Riley, and Fiona C. Bull, "Worldwide Trends in Insufficient Physical Activity from 2001 to 2016: A Pooled Analysis of 358 Population-based Surveys with 1.9 Million Participants," *Lancet Global Health,* Vol. 6, 2018.

Hambleton, Ronald K., Mary J. Pitoniak, and Jenna M. Copella, "Essential Steps in Setting Performance Standards on Educational Tests and Strategies for Assessing the Reliability of Results," in Gregory J. Cizek, ed., *Setting Performance Standards: Foundations, Methods, and Innovations*, 2nd ed., Routledge, 2012.

Hardison, Chaitra M., Paul W. Mayberry, Heather Krull, Claude Messan Setodji, Christina Panis, Rodger Madison, Mark Simpson, Mary Avriette, Mark E. Totten, and Jacqueline Wong, *Independent Review of the Army Combat Fitness Test: Summary of Key Findings and Recommendations*, RAND Corporation, RR-A1825-1, 2022. As of September 17, 2024:
https://www.rand.org/pubs/research_reports/RRA1825-1.html

Hauschild, Veronique, David DeGroot, Shane Hall, Karen Deaver, Keith Hauret, Tyson Grier, and Bruce Jones, *Correlations Between Physical Fitness Tests and Performance of Military Tasks: A Systematic Review and Meta-Analyses*, U.S. Army Public Health Command, PHR No. 12-02-0614, June 2014.

Headquarters Department of the Army, EXORD 134-24, "Practice Physical Fitness Test," April 30, 2024.

Hicks, Daniel, and Sean Robson, *Implementation of the Army Combat Fitness Test: An Updated Analysis and Future Directions*, RAND Corporation, RR-A2232-1, forthcoming.

Hicks, Daniel, Carra S. Sims, Mary Avriette, Max Steiner, and Sarah Baker, *The Army Combat Fitness Test (ACFT) and the Health of the Active Component: Understanding the Link Between the ACFT and Personnel Health and Injuries,* RAND Corporation, RR-A2005-1, forthcoming.

LeSuer, Dale A., James H. McCormick, Jerry L. Mayhew, Ronald L. Wasserstein, and Michael D. Arnold, "The Accuracy of Prediction Equations for Estimating 1-RM Performance in the Bench Press, Squat, and Deadlift," *Journal of Strength and Conditioning Research*, Vol. 11, No. 4, November 1997.

Mayhew, J. L., T. E. Ball, and J. C. Bowen, "Prediction of Bench Press Lifting Ability from Submaximal Repetitions Before and After Training," *Sports Medicine, Training, and Rehabilitation*, Vol. 3, 1992.

Nevill, Alan M., Roger Ramsbottom, Gavin Sandercock, Carlos Eduardo Bocachica-González, Robinson Ramírez-Vélez, and Grant Tomkinson, "Developing a New Curvilinear Allometric Model to Improve the Fit and Validity of the 20-m Shuttle Run Test as a Predictor of Cardiorespiratory Fitness in Adults and Youth," *Sports Medicine*, Vol. 51, 2021.

Public Law 116-283, William M. (Mac) Thornberry National Defense Authorization Act for Fiscal Year 2021, Section 598, Limitations on Implementation of Army Combat Fitness Test, January 1, 2021.

Public Law 118-31, National Defense Authorization Act for Fiscal Year 2024, December 22, 2023.

Sax van der Weyden, Megan, Michael Toczko, Marcie Fyock-Martin, and Joel Martin, "Relationship Between a Maximum Plank Assessment and Fitness, Health Behaviors, and Moods in Tactical Athletes: An Exploratory Study," *International Journal of Environmental Research and Public Health*, Vol. 19, 2022.

Sharp, Marilyn A., Stephen A. Foulis, Jan E. Redmond, Maria C. Canino, Bruce S. Cohen, Keith Hauret, Peter N. Frykman, Joseph R. Pierce Jr., Richard B. Westrick, Brooke M. Pacheco, Deborah L. Gebhardt, and Edward J. Zambraski, *Longitudinal Validation of the Occupational Physical Assessment Test (OPAT)*, U.S. Army Institute of Environmental Medicine, USARIEM Technical Report T18-05, 2018.

Society for Industrial and Organizational Psychology, *Principles for the Validation and Use of Personnel Selection Procedures*, 5th ed., August 2018.

U.S. Army, "Army Combat Fitness Test," webpage, undated. As of September 21, 2024: https://www.army.mil/acft/

Taylor-Clark, Tanekkia M., Lori A. Loan, Pauline A. Swiger, Larry R. Hearld, Peng Li, and Patricia A. Patrician, "Predictors of Temporary Profile Days Among U.S. Army Active Duty Soldiers," *Military Medicine*, Vol. 188, No. 5-6, May–June 2023.

U.S. Army, *2022 Health of the Force*, 2022.

Withrow, Kevin L., Daniela A. Rubin, J. Jay Dawes, Robin M. Orr, Scott K. Lynn, and Robert G. Lockie, "Army Combat Fitness Test Relationships to Tactical Foot March Performance in Reserve Officers' Training Corps Cadets," *Biology*, Vol. 12, 2023.

Wood, Terry, Gianni F. Maddalozzo, and Rod A. Harter, "Accuracy of Seven Equations for Predicting 1-RM Performance of Apparently Healthy, Sedentary Older Adults," *Measurement in Physical Education and Exercise Science*, Vol. 6, No. 2, 2002.

RAND ARROYO CENTER

The National Defense Authorization Act for Fiscal Year 2024 calls for higher minimum fitness standards for soldiers in close combat military occupational specialties (MOSs). This report details options for increasing Army Combat Fitness Test (ACFT) standards for such soldiers.

After reviewing the evidence for setting higher fitness standards, the RAND team identified eight options for the Army to pilot in a "Practice Phase" among a sample of soldiers in close combat MOSs. The Army chose to pilot an increase in the total ACFT point standard to 450 points overall, along with a requirement of 150 pounds on the 3 Repetition Maximum Deadlift. Evidence from the Practice Phase suggests trade-offs across alternative standards and that the Practice Phase standard may be too high to achieve the Army's desired pass rates for various subgroups in the short term, particularly Guard and Reserve soldiers.

Soldiers' performance improves with experience, so the pass rates observed in the Practice Phase may understate the extent to which soldiers will continue to improve toward new standards. Several options for higher fitness standards satisfy important decision criteria and are reasonable courses of action for the Army. Options that raise the overall points requirement may be simplest to implement while promoting fitness among all close combat soldiers.

$39.50

ISBN-10 1-9774-1443-5
ISBN-13 978-1-9774-1443-4

53950

9 781977 414434

www.rand.org

RR-A3092-1