Westat

A Comparison of Questionnaire Evaluation Methods to Assess the Airborne Hazards and Open Burn Pit Registry

Darby Steiger and April Fales, Westat

Presentation to FCSM
October 25, 2022

Background

- Many questionnaire evaluation methods are used to improve data quality
  - Expert reviews (Olson, 2010)
  - Cognitive interviews (Willis, 2004)
  - Usability testing (Couper, 1999)
  - Respondent debriefings (Martin, 2006)
- Ongoing question:
  - Do different methods produce different results? (Presser & Blair, 1994; Hughes, 2004; Rothgeb, 2007; Yan et al, 2012; Maitland and Presser, 2018)

AHOBPR – Background

- Veterans Health Administration Office of Public Health
- Registry launched in 2014 for service members to document exposures to toxic chemicals and fumes from burn pits and other airborne hazards during wars in Iraq and Afghanistan
- >300,000 responses to the registry (about 10% of eligible Veterans)
- National Academies Panel in 2017 identified flaws in the questionnaire and recommended engaging survey experts to guide improvements

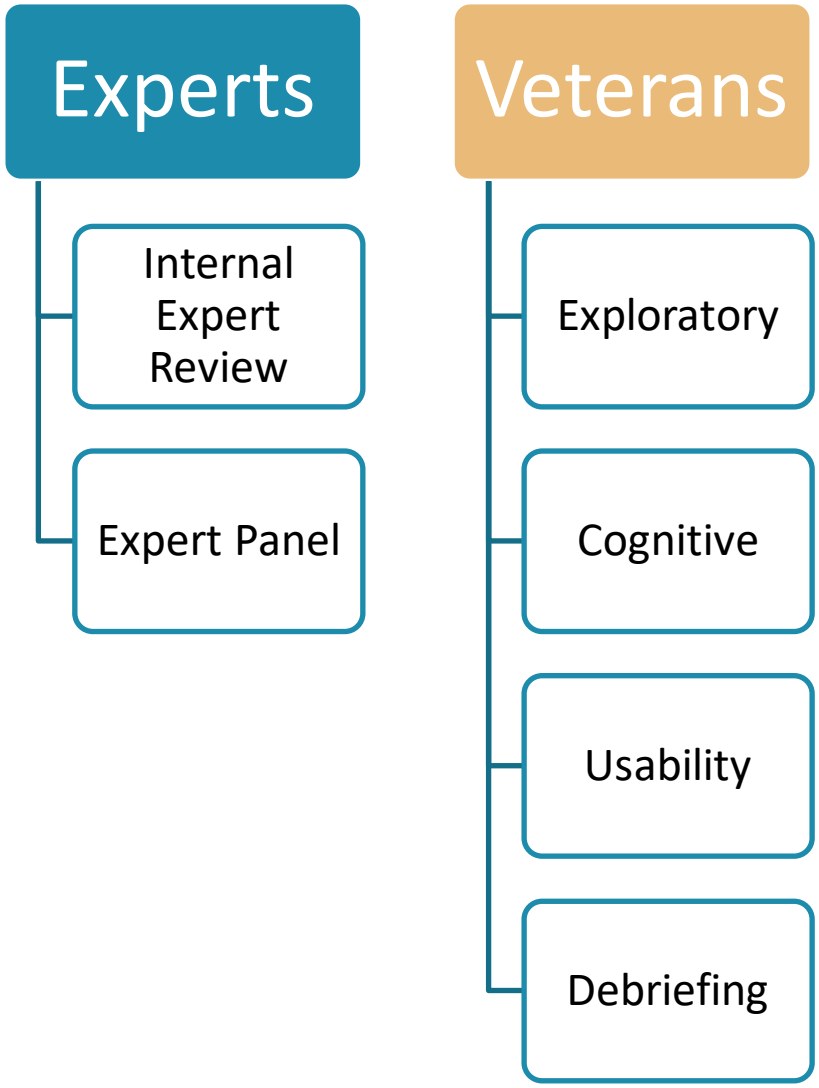
- GOAL: To provide a comprehensive review of the registry questionnaire and make recommendations to improve the registry's ability to meet VA's objectives

What Are Airborne Exposures?

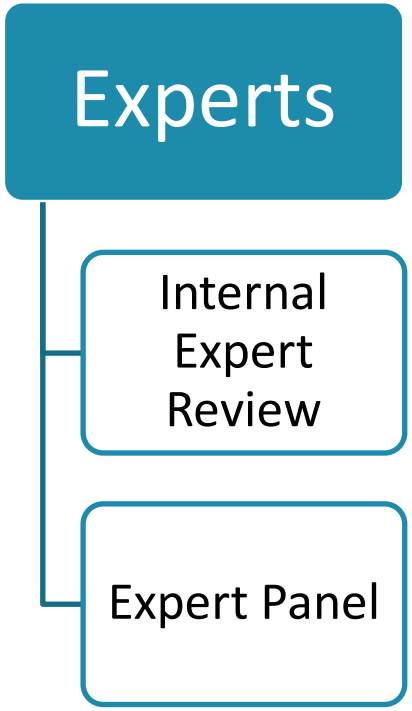
- Gulf war oil well fires (early 1990s)
- War-related airborne exposures (OIF/OEF/OND)
  - Burn pits, sewage ponds, IEDs
  - Heavy smoke from warfare
  - Exhaust from convoys/military vehicles
  - Fumes from refueling vehicles, engine maintenance, construction
- Environmental exposures
  - Local heavy pollution, pesticides
  - Dust storms

Westat

Methodology



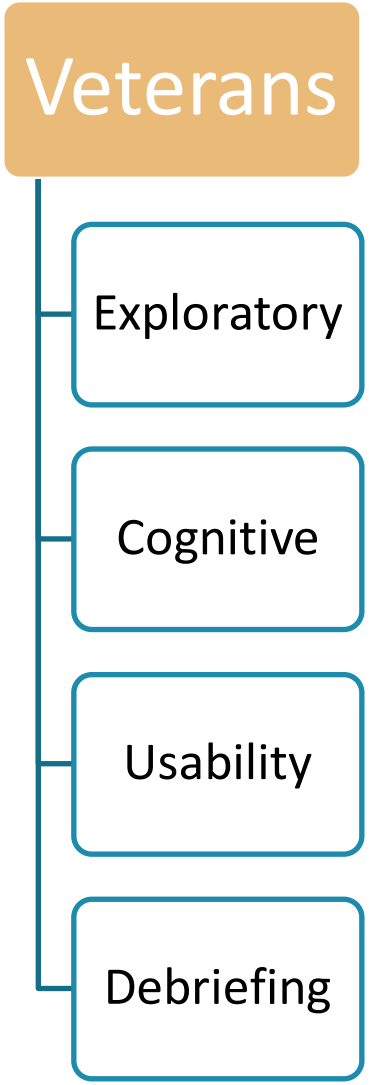
Westat



- 2 Senior Methodologists
- Reviewed each question independently
- Meeting to adjudicate findings

- 6 Methodologists, 1 Pulmonologist, 1 Epidemiologist
- Reviewed each question independently
- Met as a panel with VHA to discuss common themes and make additional recommendations

Westat



- Goal to complete up to 9 interviews in each study
- Joint recruitment effort for all 4 studies
  - Snowballing
  - Friends & family of staff
- Screener criteria
  - Deployed to Iraq, Afghanistan or other SW Asia theaters
  - Have not completed the AHOBPR registry
  - Exposed to airborne hazards during a deployment
  - Eligible to access registry (debriefing group only)
  - Mix of branches, demographics to extent possible
  - Willing to use Zoom
- $60 Electronic gift card upon completion

Westat

Key Features of Each Veteran Study

Exploratory (60 min Zoom, audio only) N=9	• Can Veterans recall details of deployments, including airborne exposures? • Why have they not yet participated in registry? • What do they expect is covered in the registry?
Cognitive (60 min Zoom, audio + video) N=9	• Feedback on a subset of 25 items identified by expert panel as problematic. • Items are displayed on screen, one at a time • Concurrent probes
Usability (60 min Zoom, audio + video) N=4	• Using a test account, observe Veteran do the registry • Focus is on navigation, screen design, instructions
Debriefing (30 min Zoom, audio only) N=8	• Veteran completes registry on their own • 30-minute debriefing call with interviewer to gather their feedback on their experiences completing it

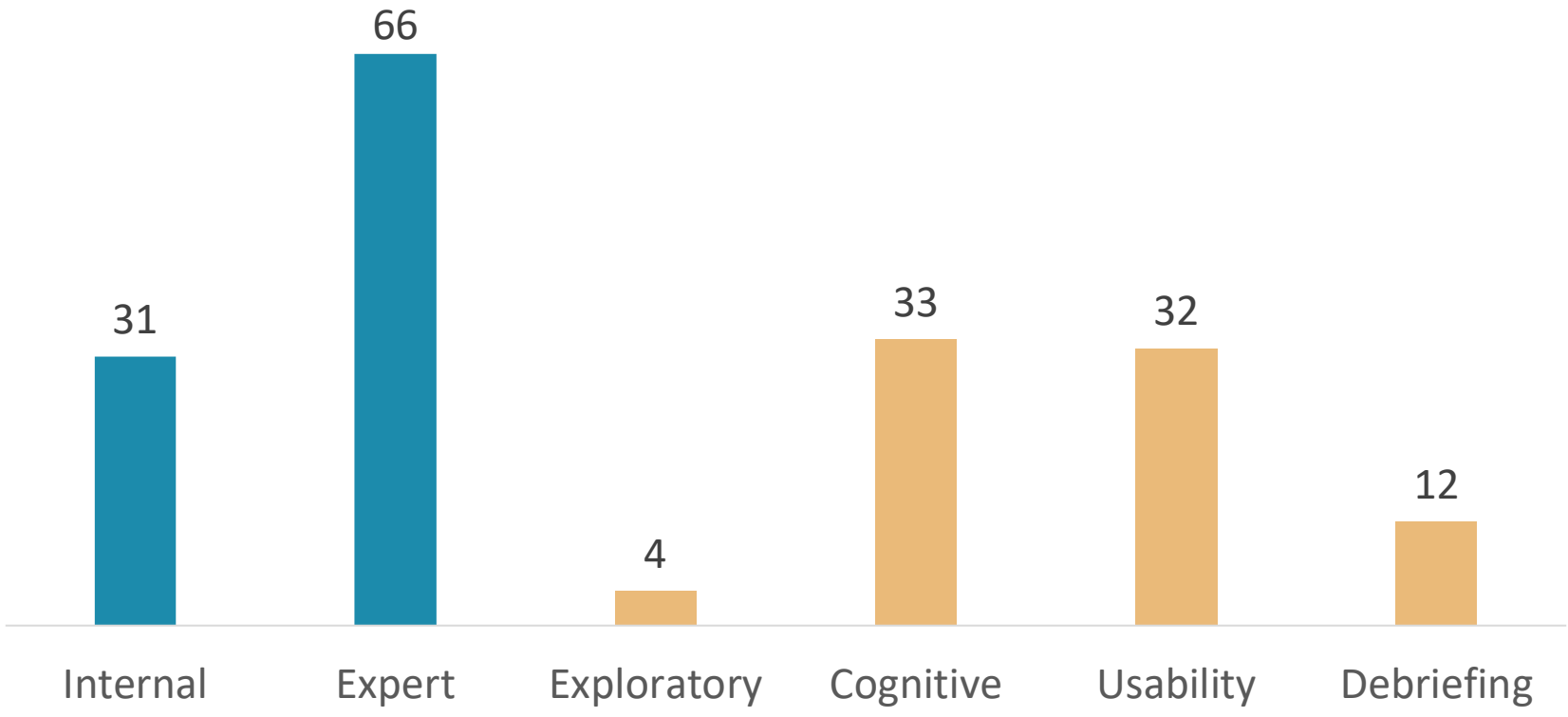
Westat

Analysis Approach

- Focus on subset of 24 items pertaining to airborne and environmental exposures
- Each source analyzed independently
- All results brought into NVivo 12
- Coding scheme developed
- Queries to explore codes by source, question problems by source

Number of Issues Identified by Source

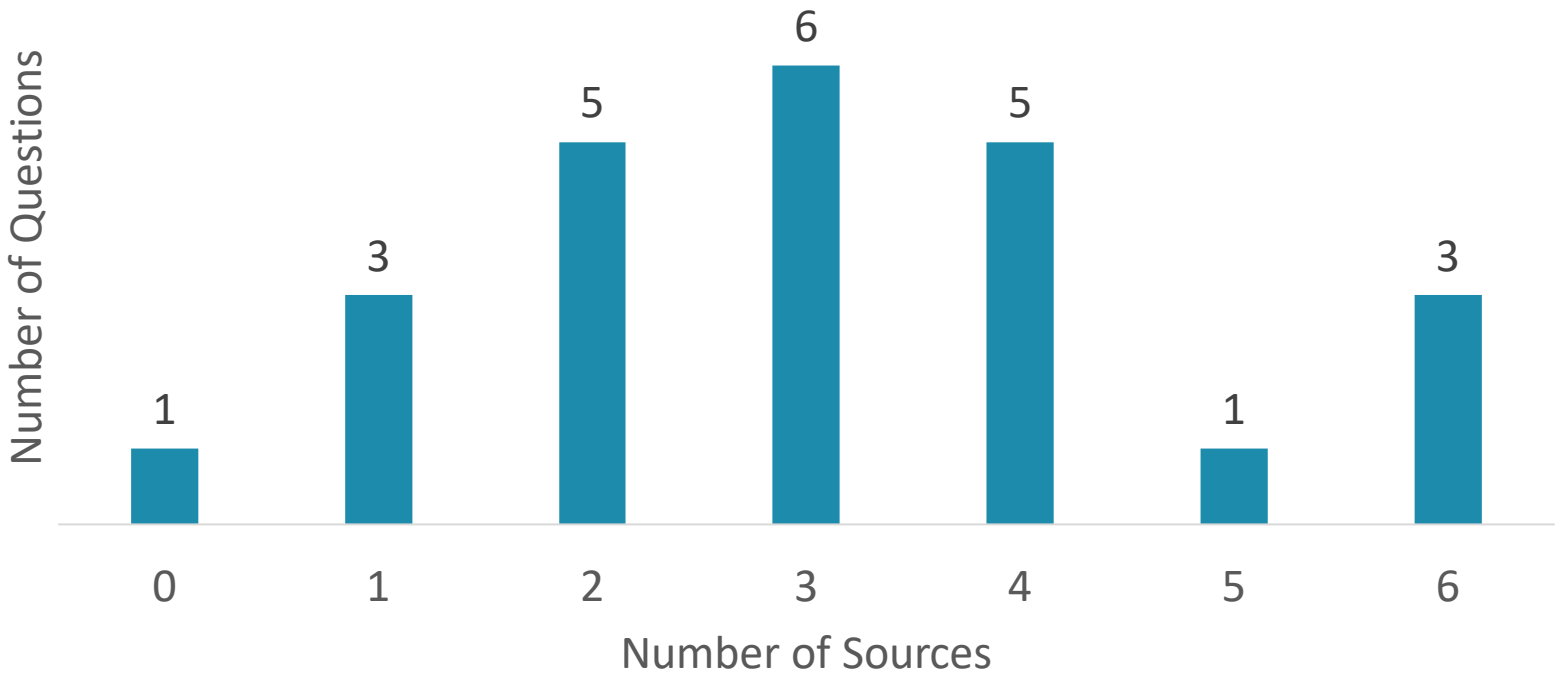
- Total of 177 issues identified across all sources
- The expert panel identified twice as many issues as any other source.
- Exploratory interviews tended not to identify as many issues.
- Significant difference in mean # of problems identified by sources (F<.001)



Westat

Overall Cohesion: Number of Sources Identifying a Problem With Each Question

- Only 3 questions were identified by all 6 sources as problematic. (Deployment table, burn pit exposure, hours of burn pit exposure).
- Only one question was not identified by any source as problematic



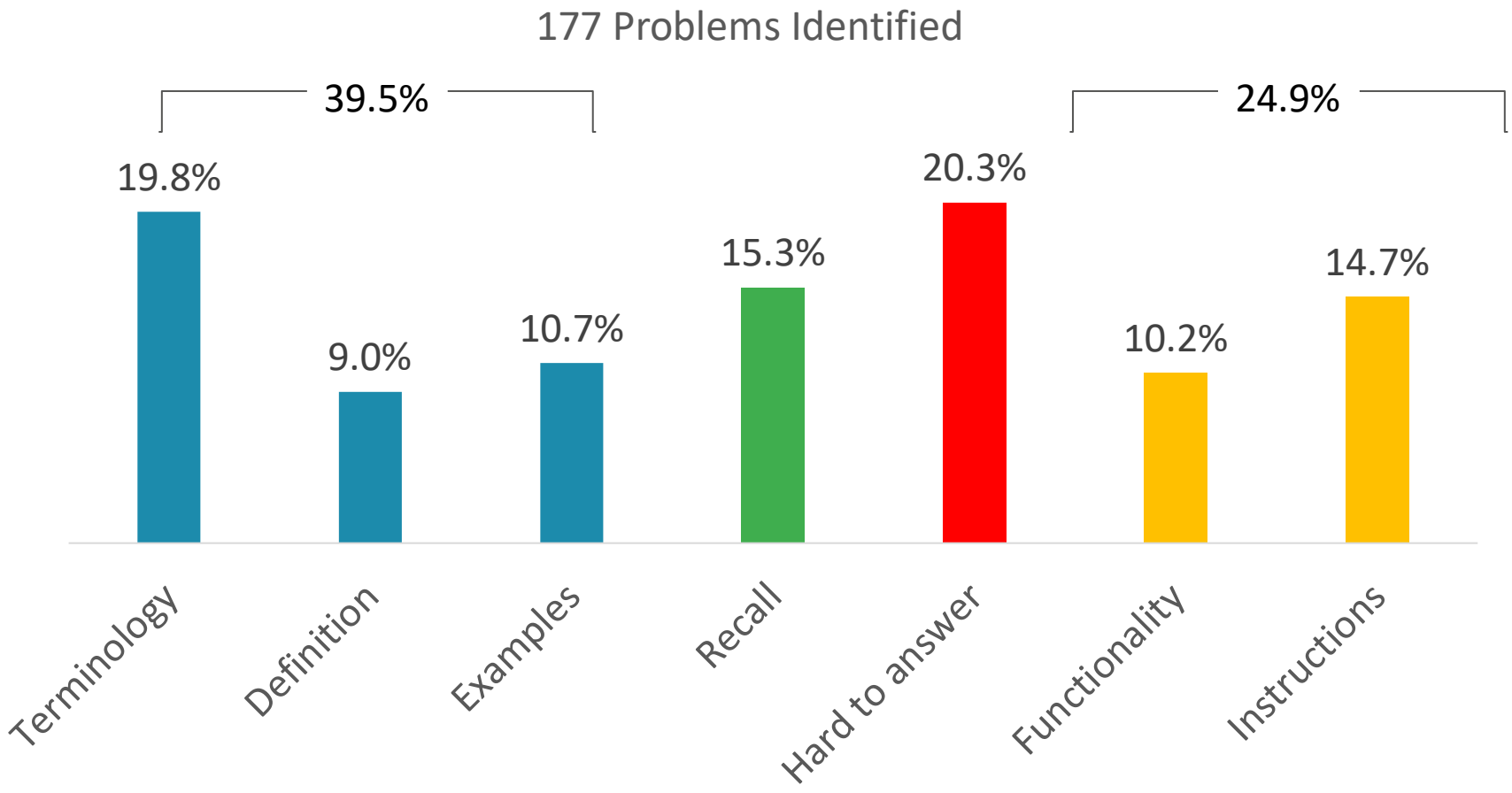
Westat

Coding Scheme (Tourangeau, Rips, Rasinski, 2000)

- Comprehension
  - Terminology
  - Definitions
  - Examples & parentheticals
- Recall
  - Ability to recall details
- Judgment
  - Ability to assess memories/Question would be hard to answer
- Response
  - Instructions
  - Functionality

Problem Codes Identified Across All Sources

- Overall, the most common problems identified across sources were comprehension (39.5%) and response mapping (24.9%).

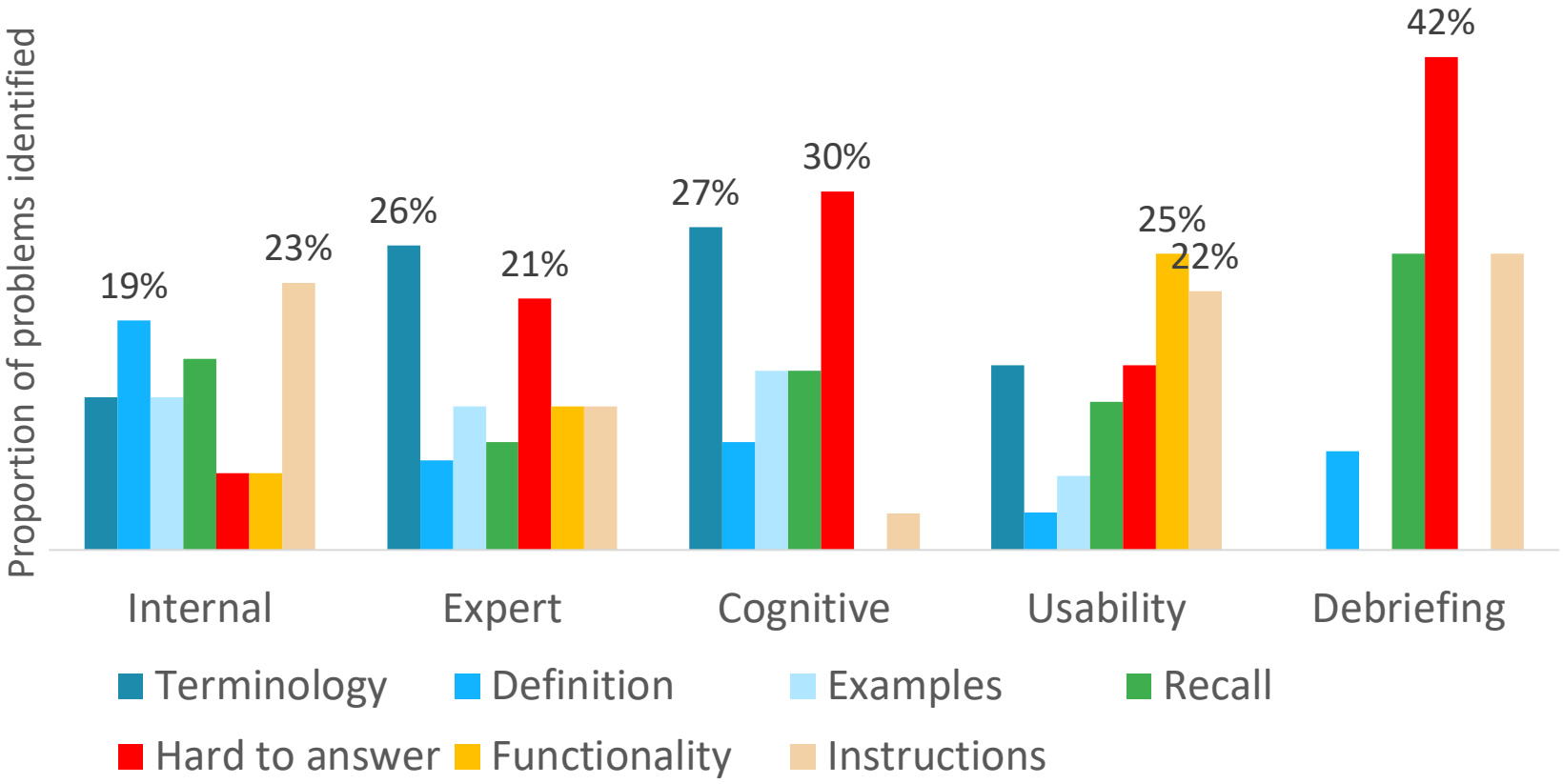


Westat

Comprehension | Recall | Judgment | Response

Types of Problems Identified by Source

- Different types of problems were identified by each source.
- Even within experts vs. Veterans, types of problems differed.
- Correlations were generally low; highest between usability & debriefing.



Westat

Comprehension | Recall | Judgment | Response

Example #1: Deployment Verification

IE	EP	Exp	Cog	UX	Deb

Correct	Branch	Begin Date	End Date	Conflict	Location	Opt Out
○ Yes ◉ No	Navy	08/14/1995	01/04/1996		Arabian Sea	○ Yes ◉ No
◉ Yes ○ No	Navy	06/14/1999	11/02/2001	Gulf	Persian Gulf (USS Michelle)	○ Yes ◉ No
◉ Yes ○ No	Navy	06/14/1999	11/02/2000		Persian Gulf	○ Yes ◉ No
◉ Yes ○ No	Army	05/01/2005	09/01/2005	oef	Uzbekistan (Karshi-Khanabad (K2))	○ Yes ◉ No

If deployments are missing from the list above please use the search provided below to find and add deployments in the Southwest Asia theater, Djibouti or Afghanistan.

You do not need to add deployments outside the Southwest Asia theater, Djibouti, or Afghanistan.

Need help adding your deployments to the Deployment History Table?

Search base, Country, or body of water:

Search | Reset

Westat®

Comprehension	Recall	Judgment	Response

Example #2: Burn Pit Exposure

IE	EP	Exp	Cog	UX	Deb

D. Were you near a burn pit during these dates (on the base or close enough to the base for you to see the smoke)? (Answer Required)

Please select an answer

Yes

G. On a typical day, how many hours did smoke or fumes from the burn pit enter your **work site or housing**? (Answer Required)

- ○ Never
- ○ Enter 1 to 24 hours.
- ○ I do not wish to answer

Westat

Comprehension	Recall	Judgment	Response

Example #3: Days Near Heavy Smoke in a Typical Month

IE	EP	Exp	Cog	UX	Deb

Across all of your deployments,

B. In a typical month, how many days were you near heavy smoke from weapons, signal smoke, markers or other combat items? (Answer Required)

- ○ Never
- ◉ Enter 1 to 31 days. 25
- ○ I do not wish to answer
- ○ Don't know

Westat

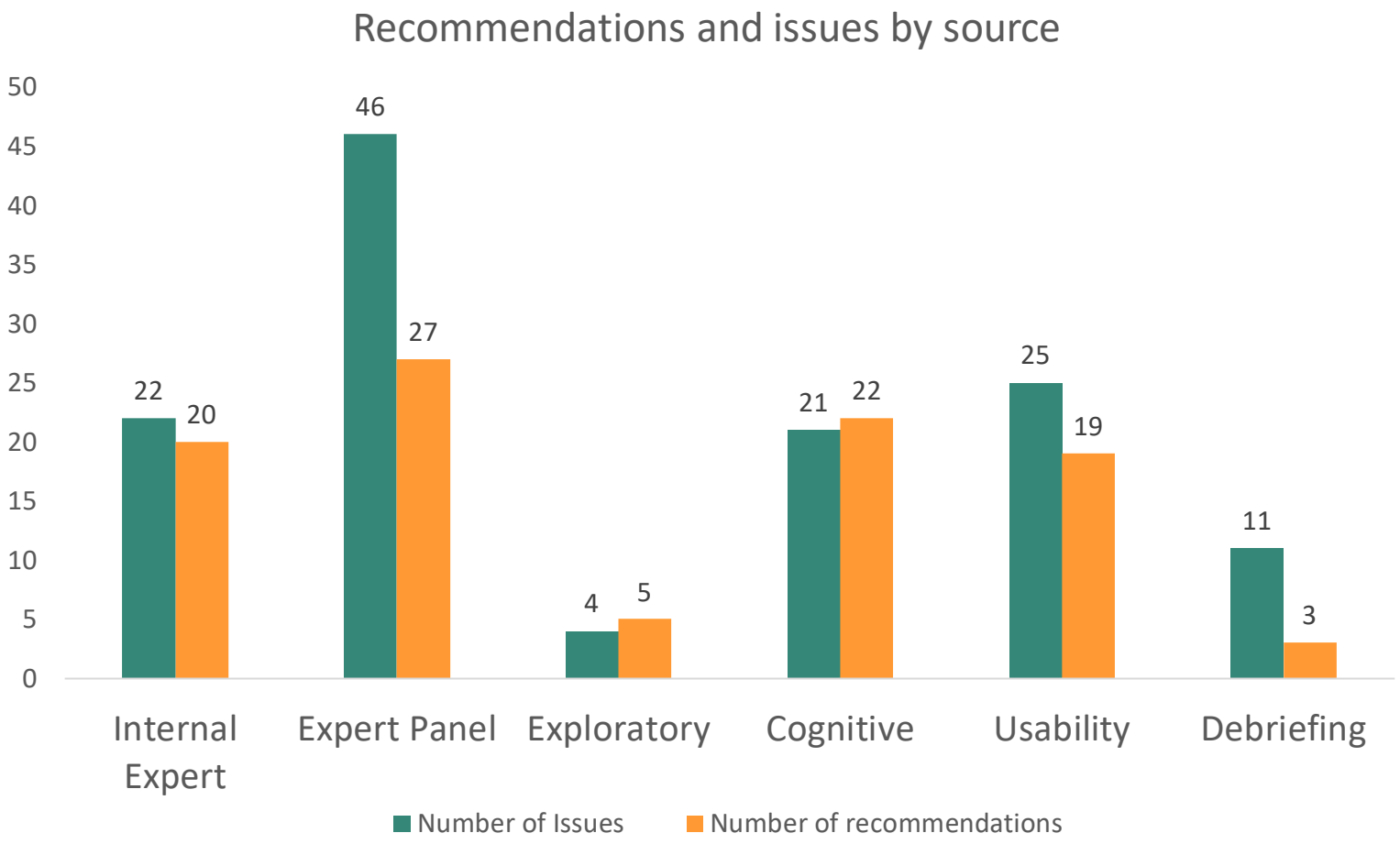
Comprehension	Recall	Jugdment	Response

Did the Issues Identified Result in Recommendations?

- Analyzed 16 of the 24 items
  - The eight we omitted were completely overhauled
  - Examined findings from each source
  - Looked for common themes
  - Used our methodological expertise
- Most items had between three and five recommendations for improvements
  - One item had zero recommendations
  - One item had 20 recommendations
  - Each item had recommendations from between two and five sources

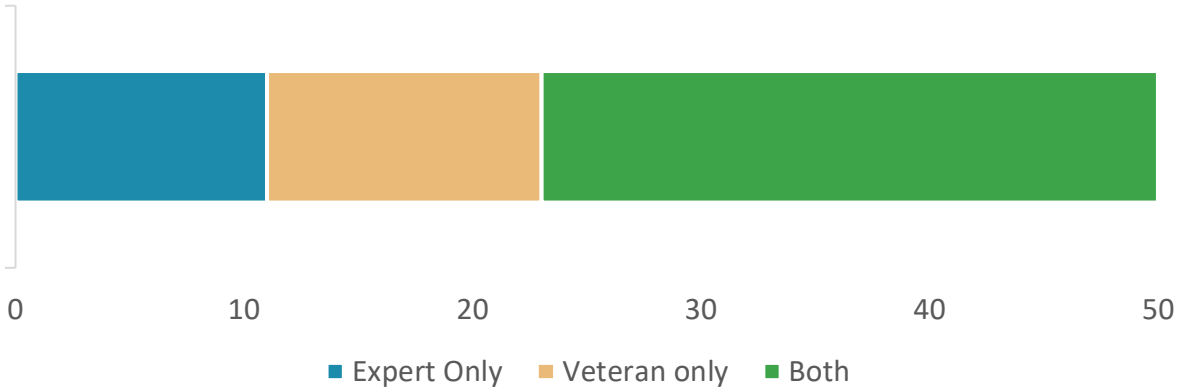
Recommendations and Issues by Source

- Expert Panel and Debriefing had lower proportion of issues translate into recommendations



Westat

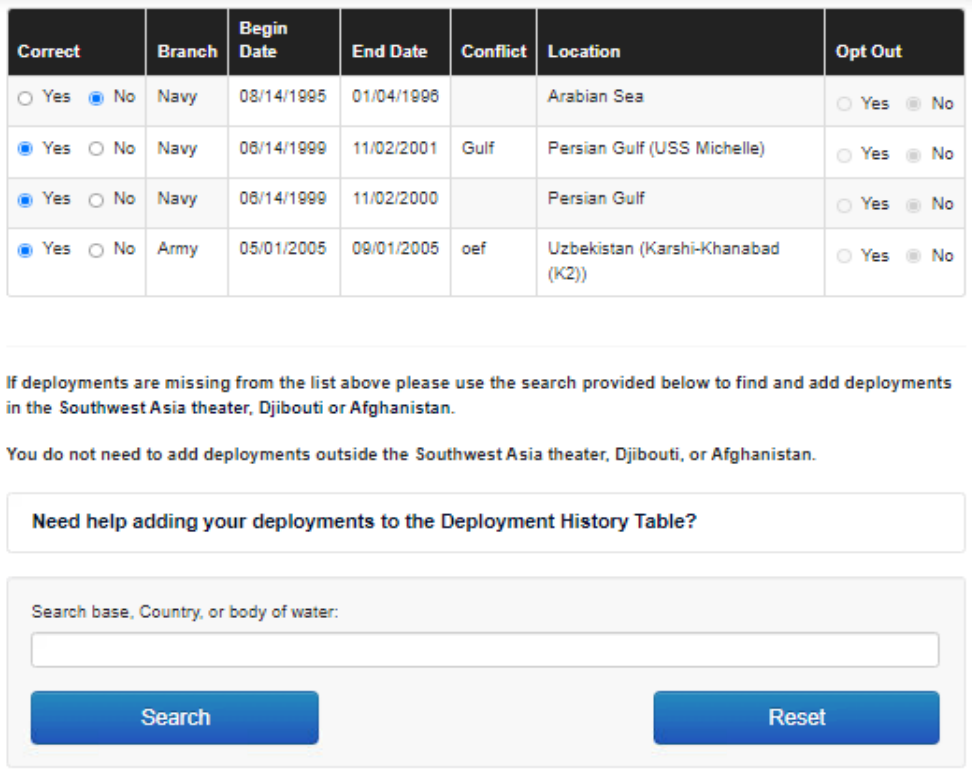
Concordance of Sources by Recommendations



- 50 total recommendations
  - 11 Expert Only recommendations focused on instructions, definitions, and examples
  - 12 Veteran Only recommendations focused on comprehension, functionality, terminology, and instructions

Westat

Example #1: Deployment Verification



Correct	Branch	Begin Date	End Date	Conflict	Location	Opt Out
○ Yes ◉ No	Navy	08/14/1995	01/04/1996		Arabian Sea	○ Yes ◉ No
◉ Yes ○ No	Navy	06/14/1999	11/02/2001	Gulf	Persian Gulf (USS Michelle)	○ Yes ◉ No
◉ Yes ○ No	Navy	06/14/1999	11/02/2000		Persian Gulf	○ Yes ◉ No
◉ Yes ○ No	Army	05/01/2005	09/01/2005	oef	Uzbekistan (Karshi-Khanabad (K2))	○ Yes ◉ No

If deployments are missing from the list above please use the search provided below to find and add deployments in the Southwest Asia theater, Djibouti or Afghanistan.

You do not need to add deployments outside the Southwest Asia theater, Djibouti, or Afghanistan.

Need help adding your deployments to the Deployment History Table?

Search base, Country, or body of water:

Search | Reset

- Allow users to edit information directly in the deployment table, rather than having to re-enter the information as a new deployment. (I, U, D)
- Only display the opt-out option for deployments less than 14 days. (I, E, C, D)
- For adding deployments, provide examples right above the search bar of how to search for a base (e.g. do not enter the word “Camp”). (E, U, D)

Westat®

Example #2

F. Did your duties during these dates include the burn pit (examples include trash burning, hauling trash to the burn pit, burn pit security, trash sorting at the burn pit)? (Answer Required)

Please select an answer

- Acknowledge that burn pit duties may have been infrequent or irregular. (I, E)
- Remove parentheses and incorporate examples into question wording. (E)

Westat

Summary

- Six testing methods were used to assess the AHOBPR
- Different methods found different issues and experts and Veterans often found different types of problems
  - I, E, C tended to focus on comprehension
  - U tended to focus on functionality
  - D tended to focus on recall
  - Exploratory didn't produce many specifics
- All sources agreed that certain items were problematic
  - Deployment verification, burn pit exposure, typical month
- Nearly half of the final recommendations came from either only experts or only Veterans, suggesting the value of both sources of review.

Implications

- Similar to findings in literature, different testing techniques produced different findings
  - Experts identified the most problems (Presser & Blair)
  - Low consistency across sources (Yan et al)
  - Use of multiple methods was effective (Maitland & Presser)
- Testing a questionnaire 8 years after it was launched and after 300K+ have already responded raises thorny issues
  - How to interpret data already collected on items found to be most problematic
  - How will changes impact trends

Westat®

Thank you!

DarbySteiger@Westat.com
AprilFales@Westat.com

Additional Info

Background

- An ongoing question is whether these methods produce similar findings (Rothgeb et al, 2007)
  - Based on a review of 83 items from 3 federal telephone surveys
  - Compared findings from expert review, QAS forms appraisal, and cognitive interviews
  - Created a summary score for each item to assess how many sources identified it as problematic
    - Analyzed differences between expert reviewers
  - Coded nature of problems and compared problems by technique
  - Determined QAS to be most productive in identifying problems; high consistency between experts; and that the different techniques identified somewhat different issues.

AHOBPR Goals

Clinicians

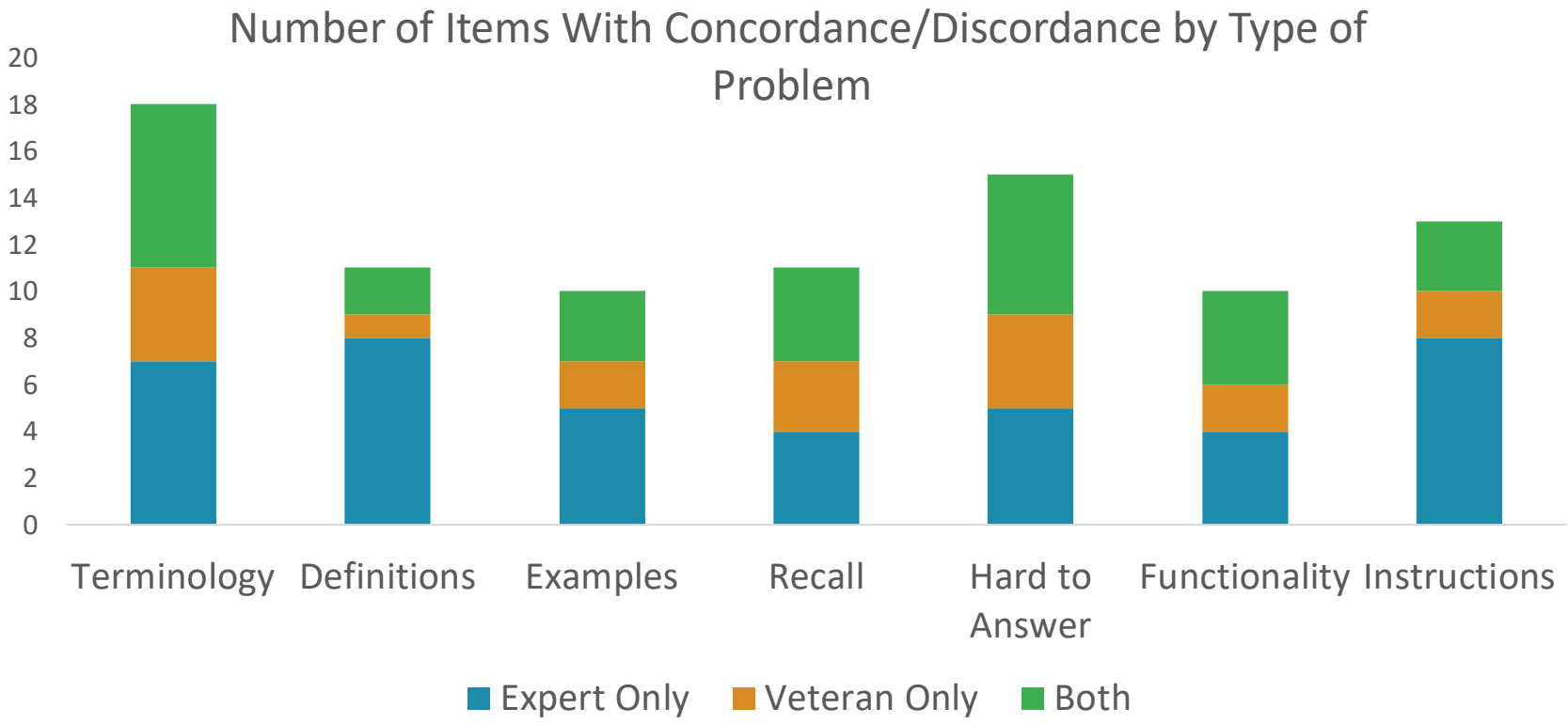
- Gather data to help facilitate the clinical follow-up visit
- Long delays between registry & clinical visit
  - Less interest in current health status
  - Less interest in exposure at each deployment/segment

Researchers

- Gather data to explain nature of exposures and impact on health
  - Value data about each deployment
  - Value data on current health outcomes

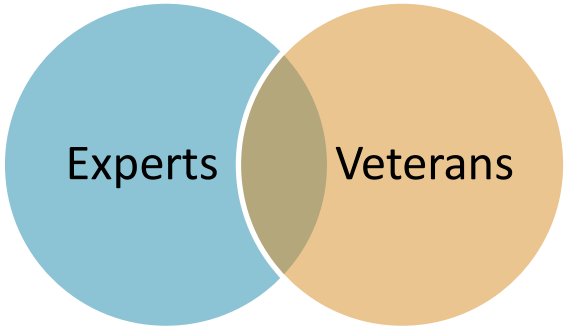
Concordance of Sources By Item

- Experts and Veterans found the most agreement on items with terminology problems and items hard to answer.
- Experts much more likely to find problems with definitions, instructions, and examples.



Westat

Examples of Discordance/Concordance of Experts and Veterans



- Expert issues not identified by Veterans
  - Use of examples in parentheses
  - Use of dropdown response format
  - Instruction that answer is required on every item
- Veteran issues not identified by Experts
  - Functionality of entering new deployments
  - Applicability of some items to Air Force, Navy veterans
- Issues identified by both Veterans and Experts
  - Meaning of opting out of a deployment
  - Keeping track of deployments
  - Definition of burn pit, exposure, sewage pond
  - Calculating hours/days of exposure, “typical” month

Example #1: Deployment Verification

EXPERTS	**Internal Expert (Comprehension)** In what order are deployments presented? Is there a maximum number of deployments that are asked about? Can a respondent skip a deployment if they want to?	**Expert (Hard to Answer)** This process could be extremely time consuming if the member had multiple, short deployments.
VETERANS	**Exploratory (Recall)** Veterans sometimes began their narrative with the caveat that their deployment "was xx years ago," and they were not sure how accurate their answers would be.	**Cognitive (Comprehension)** Participants found the information to be redundant and needing to be streamlined.
	Usability (Functionality) One respondent would have preferred to be able to make updates within the table rather than adding a whole new deployment.	**Debriefing (Functionality)** Three participants were missing deployments from their list. Two of those participants did not see a place to add the missing deployments, and the other participant chose not to add the additional deployment.

Example #2: Burn Pit Exposure

EXPERTS	**Internal Expert (Definitions)** Consider revisions to remove parentheticals, clarify which deployment is being asked about, and to expand definition of exposure to include smell.	**Expert (Terminology)** What counts as "near"? That needs to be defined. Does the respondent know for sure what counts as a "burn pit"?
VETERANS	**Exploratory (No problems)** All Veterans were able to provide detailed answers to the kinds of airborne hazards and burn pits they encountered during deployments.	**Cognitive (Definitions)** Participants did not have a consistent definition for what a burn pit is.
	Usability (Recall) One respondent thought it was a cognitively demanding task to try to recall the number of hours and thought people might need to review their memories before answering some of these questions.	**Debriefing (Definitions)** Two participants wanted a definition for burn pits, and one wanted a definition for deployments.

Example #3: Days Near Heavy Smoke

EXPERTS	**Internal Expert (Hard to Answer)** Is there such a thing as a typical month, especially if R is supposed to be thinking across all deployments?	**Expert (Hard to Answer)** This question may be trying to get more precision that people can adequately provide.
VETERANS	**Exploratory (Not mentioned)**	**Cognitive (Hard to Answer)** Participants thought this would be a difficult question to answer, because there is no "typical month" when it comes to these types of events. As one said, "every day is totally different. Some months are totally lopsided."
	Usability (Hard to Answer) A respondent was confused about which deployment was being asked about "in a typical month." He thought about typical months within the deployments in which he performed those duties, rather than averaging across all deployments.	**Debriefing (Hard to Answer)** Participants who had different experiences across deployments found these hard to answer – some said that they tried to take an average, while others said they reported based on the times they had more exposures.

Key Findings

- Other than exploratory, which did not actually see the instrument, recommendations from all sources were used to make final recommendations for the deployment data screen.
- Experts had the most recommendations, but only 3 were unique specific recommendations that were not captured by Veterans or Internal reviewers.
- Few recommendations came from debrief despite that being the "real life" group, that is, the eligible population completing the registry in full.
- The only item that did not have any recommendations based on Veteran interviews was the question about burn pit duties - some commented on other duties that could be included in the list, but did not think that the question needed to be changed.