

Considerations for flight testing of search and rescue (SAR) innovations

Christopher J Bennett, PhD¹, Jane Hodgkinson, PhD², Jim Nixon, PhD^{2,3}

¹ Department of Aerospace Engineering, Embry-Riddle Aeronautical University, Prescott, AZ, USA

² Cranfield University, Cranfield, Bedfordshire, UK (at the time of collaboration)

³ Health & Safety Executive, UK

Email bennec28@erau.edu

<https://doi.org/10.61618/SITW4437>

Abstract

Flight testing of new aircraft, equipment and systems is mandatory in the aviation industry to establish and verify the airworthiness of any new airborne product as determined by the appropriate authority, for example the Federal Aviation Administration (FAA) in the United States or the European Union Aviation Safety Agency (EASA) in Europe. Testing procedures for civilian and military category products are well documented and strictly regulated (see, for example the FAA flight test guide for certification of normal, utility, acrobatic, and commuter category aircraft (FAA, 2011) and often use traditional techniques. However, there are no published guidelines for flight testing new developments in the search and rescue (SAR) domain.

This article summarizes, from experience, recommendations for a successful SAR flight test campaign. This includes considerations for effective and efficient trial design, best practices for data capture across a range of environmental conditions, and methods to minimize sources of error. A detailed example of a search procedure with sequential steps is also provided. The principles discussed herein target quality data capture via flight test, both efficiently and safely, to improve and evolve the vital work that SAR operations conduct worldwide.

KEY WORDS Flight-Test design, innovations, review, case study

Introduction

Background

Airborne Search and Rescue (SAR) is a key element of emergency services worldwide. The objective of the SAR operation is to locate a casualty in the fastest possible time in order to complete a safe and successful rescue. When searching for casualties in challenging environments (rough seas or rugged terrain), SAR divisions typically deploy a helicopter, usually with a crew of four: Captain, First Officer (FO), rear search support crew, and a medically trained winch-man. Twin-engine, medium-lift helicopters are usually the aircraft of choice for such operations due to their ability to track low and slow over a search area, and also for their capability to hover and winch for the rescue phase. However, fixed wing aircraft are also commonly used to search much wider areas (or for larger targets) where faster, higher altitude flying is appropriate.

In the US, maritime SAR is typically conducted by the US Coast Guard who operate over 200 aircraft (both rotary-wing and fixed-wing) (U.S. Department of Defense, 2019). Inland SAR is typically coordinated by the US Air Force Rescue Coordination Center (AFRCC) and can leverage USAF, state, local or civilian resources, equipment and personnel (U.S. Department of Defense, 2023a). Deployment of resources, in either case, depends upon the severity/urgency of the distress call, availability of personnel and equipment, geographic location, terrain, and weather conditions. Between 2007 and 2017 the US Coast Guard conducted an average of 21,336 sorties per year, saving a total of 48,522 lives (Bureau of Transportation Statistics, 2017), and between 2013 and 2023 the AFRCC coordinated on average 750 missions per year, saving a total of 3430 lives (U.S. Department of Defense, 2023b).

The authors of this article previously completed a full-scale evaluation, including in-flight trials, of new conspicuity aids for casualties at sea for an EASA funded project (Bennett, Hodgkinson, & Nixon, 2019), and Bennett is currently involved in a SAR product design project which will soon be ready for ground and in-flight testing. This article has been motivated by the many discussions, issues that have arisen, and resolutions that have been devised in previous and on-going research campaigns.

Introduction of technology in SAR

SAR is a highly skilled, technology driven industry, continuously evolving to update equipment, processes, and procedures with the aim of improving the probability of detection, and therefore survival rates. Much of the equipment currently utilized by SAR operations has been developed specifically for this application. Furthermore, procedures such as search patterns, altitudes/speeds, and SAR crew resource management (CRM) have been developed to optimize the effectiveness of any given mission. However, there are a number of key technologies currently used by SAR operations which have filtered through other industries such as commercial aviation and military applications, for example with the introduction of the forward looking infrared (FLIR) system and night vision goggles (NVGs).

Clearly, the specific processes and procedures employed by any given SAR crew should be derived based on the equipment they have at their disposal. For example, if a new FLIR system has a wider

swath than a previous model, the search pattern should be modified to take advantage of the improved capability and hence reduce the time taken to survey a given area. Therefore, as new SAR equipment is introduced, methods pertaining to the search procedure requires continual reassessment to fully implement the equipment effectively. This process obviously requires full-scale, in-flight testing and training. Ultimately, an evaluation as to whether a new innovation is worthwhile to implement from a cost and/or operational standpoint should be made. For example, a governing body is unlikely to invest hundreds of millions of dollars implementing a new technology on its fleet of aircraft if it is projected to save only a handful of lives over its operational lifespan. This is the unfortunate reality. However, if a new piece of equipment can be rolled out relatively cost effectively, and is projected to significantly improve the probability of detection and therefore the survival rate of a missing person, then the case for implementation is much stronger.

Aims and Organization of this article

This document details important considerations for designing and conducting a realistic and representative trial in the SAR domain, and also describes how many of the factors described in Figure 1 can be utilized to assess new technologies or working practices. The principles are applicable to research, governing bodies and private companies needing to assess the benefit, and airworthiness, of new innovations in a SAR context.

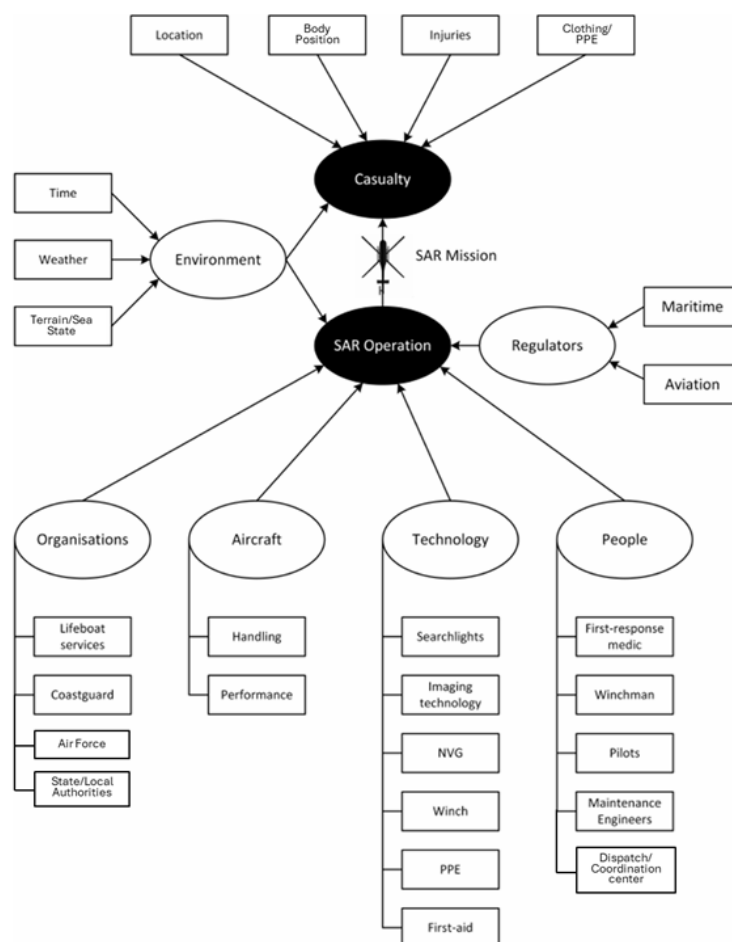


Figure 1: The SAR mission as a system of systems.

Aspects of the SAR mission considered in this article include:

- The performance and handling of the aircraft in support of the SAR mission.
- Imaging hardware and software including thermal imaging or the use of computer vision in SAR (Gotovac, Papic, & Marzusic, 2016), (Lygouras *et al.*, 2019).
- The color, intensity and type of aircraft search light.
- The rescue equipment and procedures used in a mission (for example the winch).
- The use of NVGs as an aid for the search and the optimal ways in which to use them.
- The effects of casualty clothing (including immersion suits, life jacket, and helmets on the mission.)

One key challenge when conducting research in this field is that any new design must be tested and validated in a real environment to ensure it is both safe to use, and also improves the SAR operation in a measurable and meaningful way. A comprehensive program of research is therefore required to support any hypotheses or claims made by equipment manufacturers. Typically, such research campaigns are conducted in the following sequence:

- Research into new SAR innovation: to understand the new capabilities, how it could be of benefit, and if it is possible to integrate the new equipment with the current operation.
- Research into current SAR operation: to understand current equipment capabilities, methods, procedures, and if the new SAR innovation is compatible with the current operation. This also establishes a benchmark for the new innovation to be tested against.
- Lab/ground testing and training: to test the new innovation against the benchmark in a controlled and safe environment. Measurements are taken and supporting data is gathered. The host SAR crew should also be trained on the new equipment in this phase, and feedback on its usage should be collected.
- Full-scale/flight-testing: to test the new innovation in a realistic environment. Measurements are taken and supporting data is gathered.
- Present findings: to justify the upgrade (or not) and outline the projected benefits and potential improvements to the SAR operation. Summarize the data and provide evidence such that the governing body can make an informed decision based on a projected improvement to cost ratio.

The protocol presented in this paper is derived from the authors prior experience, in particular, recommendations made to EASA (Bennett, Hodgkinson, & Nixon, 2019) currently under review/revision for an ongoing SAR product design project. Since no published guidelines, checklists, or instructions existed to aid with this specific type of project, the authors developed bespoke requirements. On reflection, this article represents a reference guide the authors hoped was available to consult at the start of this body of research. The lessons learned, trial design revisions, and protocol updates, as well as considerations of sources of error, have been distilled into the work presented in this article. The

findings are generalized such that they may be applied to assess any type of new equipment or operational method/procedure in a SAR context.

Here, an outline of principles and guidelines for the design and implementation SAR flight trials, are written to encompass a broad range of scenarios. The first topic of discussion is the design of the trials/tests. This will include a summary of best practices. Next, the myriad of factors which can harm the reliability and validity of the trials will be outlined, followed by a description of ways in which they may be resolved, or at least minimized. Finally, a step-by-step procedure that could be applied to any SAR based trial is provided as an example.

Approaches to SAR trials – A review of previous studies

Different approaches have been taken by researchers in various fields to the problem of assessing the efficacy of new innovations in SAR. These primarily focus on maritime SAR. The type of testing, and the design of the trials, is highly dependent on the type of product in question, for example, aircraft hardware, new imaging technology, drift models, algorithms for defining and programming the aircraft's search pattern, and unmanned autonomous vehicles (UAVs).

For a complex application such as SAR target detection, a full-scale test is required to validate the operation of the technical systems and the interaction with human operators and observers. However, to do this thoroughly, a sufficient number of realistic SAR scenarios would be required to enable statistically significant comparisons between different options to account for system reproducibility across a number of variables: lighting conditions, weather and visibility, sea states, orientation of target object, plus human factors such the level of alertness of operators and their ability to discriminate the target from clutter in the field of view. Some of these may be chosen but still have variability: the weather and lighting might be broadly similar but can still vary between one test and the next. There are additional purely stochastic factors such as whether the crew were looking in a particular direction at one time or another, or whether a mission's success was due to happenstance rather than by systematic design and execution. Clearly, the number of missions that would be required to control, or average, over such factors is unfeasibly large if conducted under realistic conditions, at scale. It is also difficult to run many repeated scenarios in the same location without the same human operators drawing on knowledge of previous runs: where to look and what the target looked like under the prevailing condition. Despite these limitations, well-considered trials and comparisons can be usefully made. Different types of testing provide complementary information, as discussed below.

Specification testing

For certain technical subsystems, it is possible to define a set of operating requirements and specifications that any given new innovation must fulfill. Hence, the new technology may be tested quantitatively against the corresponding requirement. This also includes testing of models and algorithms used in planning missions. The key here is that a single part of the whole SAR system is

evaluated in situations that have reduced complexity. Subsystems that have been tested in this manner include control systems for UAVs (Almeshal & Alenezi, 2018), where it is possible to identify the ideal flight path and measure deviations from that ideal.

The advantage of such testing is that it may be readily controlled and replicated, which is essential to the development of technical solutions particularly in the early stages. The limitation is that it can only be applied to subsystems that are separable and well understood. In (Bennett, Hodgkinson, & Nixon, 2019), for example, performance of retro-reflective materials that may be applied to crew survival suits were characterized optically in the laboratory. Although performance standards exist for these materials (International Maritime Organization, 1989), (American Society for Testing and Materials, 2013), they relate to one type of lamp and one standardized spectral response of the detector only. Therefore, additional tests looking purely at optical performance, in ground-based trials of detectability, were required.

Detectability

Measurements of target detectability can be made during trials on sea or land using the same or similar detection equipment to that deployed on SAR aircraft. For this type of trial, the location of the target should be known, but the assessment is of the strength of the detection signature under different circumstances. 'Strength of the signal', here, may refer to the distance at which the target is observable by the SAR crew, the contrast of the target within its environment, or both. See, for example, the tests conducted in (Burciu, Abramowicz-Gerigk, Przybyl, Plebankiewicz, & Januszko, 2020) aimed at enhancing heat signatures of life rafts, the tests of a novel detection system based on sensor fusion conducted in (Burciu, Abramowicz-Gerigk, Przybyl, Plebankiewicz, & Januszko, 2020), and the investigations into immersion suit conspicuity performed in (Hodgkinson, Nixon, Bennett, & Tatum, 2020) and (Nixon, Hodgkinson, & Bennett, 2020). In all cases, a stronger signal should lead to a higher probability of detection within a given time.

Drift models

Drift is the term used to describe the motion of floating objects due to wind and sea conditions. By knowing the location at which the target entered the water, and the weather and sea conditions, mathematical models can be applied to predict the change in position of the target over time. Numerous field trials have been performed to assess drift models that aim to identify the likely location of targets and thereby narrow the search area for the SAR crew. Trials are typically conducted with standard buoys as well as real targets in different seas, to validate the reliability of the mathematical models and meteorological inputs to handle the local conditions, for example the tests conducted in (Cho et al., 2014). Trajectories are compared over large datasets to analyze accuracy and reproducibility, in a similar way to weather forecasts, for example. In (Coppini et al., 2016), known initial and final recovery locations of drifting objects such as ships, boats, and people lost overboard were used to assess

common drift models. However, good quality, reliable data is often too sparse to permit statistically significant evaluations, as discussed in (Breivik & Allen, 2008).

Realistic field exercises

Exercises should be designed to simulate a real SAR event, in order to test the function of a whole system or at least the part of the system involving SAR operators. In the context of maritime SAR, researchers often use a weighted raft or SAR training dummy/manikin as a search target rather than human participants, for reasons of safety. This cannot account for differences in behavior of live casualties, for example waving to an approaching aircraft to attract attention. Furthermore, resources to conduct trials with 'live' SAR crews are limited, often requiring volunteers and/or integration with training exercises. Due to the practical and operational limitations, the amount of raw data which can be collected in any given trial is severely limited. Hence, there are very few published trials which can be considered to test all or a large part of the SAR system in a realistic way.

Despite this, there have been several key studies involving full scale SAR trials which have led to meaningful conclusions and, as a result, operational changes. For example, in (Donderi, 1994) a total of 116 Coast Guard crew members on-board two ships were tasked with identifying 23 tethered life rafts using the unaided eye, binoculars and NVGs for a period of one hour. From such a large-scale realistic test within a realistic environment, it was possible to capture a wealth of statistically significant data. Other full-scale trials have involved investigating UAVs for SAR in a post-tsunami situation (Ferreira et al., 2018), life jacket tracking devices (Lilja et al., 2013), (Miano et al., 2019), and helicopter visualization systems (Miller, Kelly, & Ehler, 1999). It is interesting to note the differing approaches to 'measuring' the success of each trial. The performance metric needs to be carefully considered so that the results can be ultimately related to the probability of detection. This will be discussed in more detail later in this document.

It is therefore evident there is a lack of published research using a full-scale SAR operation, and hence forms the motivation for this article.

SAR trial design principles

Trial goals

Any given trial should be aimed at improving one of the overall goals of the SAR operation:

- **Prioritize:** using local knowledge and computer software (to calculate drift, for example), to identify the search area.
- **Search - Detect:** improve the probability of detection of objects.
- **Search - Identify:** recognize and categorize whether the detected object is a target of interest.
- **Rescue:** improve the efficiency of the rescue of a casualty.
- **Safety:** improve safety for the crew and/or the casualty.

The goal should be well defined, specific, and quantifiable to determine if the goal was achieved.

Trial management

The research team are responsible for planning the trials and managing the project. Clearly, they should work in close collaboration with the other parties involved to ensure that the objectives are:

- achievable given the specific equipment available,
- reasonable given the integration with a live SAR operation,
- safe for all parties involved.

Therefore, the research team personnel should be selected to include experience and expertise in the subject area. This may include, human factors/CRM, flight operations, flight test, planning/scheduling, communications, flight mechanics, sensor technology, software, optics, etc. This will gain the project credibility and allow the best use of limited flight time available. The synergy of the research team, whether large or small, is key. They should have a broad range of skills, expertise and experience with regards to the scope of the project. Interaction and effective communication between the research team and the SAR crew is also paramount to disseminate goals, ideas and findings at each step of the project. The research team are also responsible for detailed and professional mission briefings for the SAR crew.

Participants

It is important for the research tasks to be performed by trained individuals, familiar with the current methods and procedures. It is important that participants are recruited from a genuine population of working SAR crew, to provide a realistic test. Crew should also not be required to operate in unfamiliar roles. Again, these considerations are aimed at obtaining as realistic conditions as possible for the tests, and maintaining scientific validity of the processes and procedures. This will ensure that the best quality data is captured from the trial.

Performance metric

The chosen performance metric should allow results to be measured scientifically, and lead to quantitative conclusions. It is often difficult to isolate specific factors from a real-world event and hence the trials should be designed in a way to eliminate variability as much as possible to capture consistent data.

There are typically two types of performance metric which can be measured in this context: a simple binary measure of success or failure (for example, target identified or not), or a continuous numerical value (for example, distance or time). The binary measure requires a significant number of tests to be completed to satisfactorily assess the difference in performance between the benchmark verses the new innovation. In contrast, measuring a continuous variable (for example, the distance to the target at which the object is first detected) provides more granular data, and hence differences between variables

can be derived from a reduced number of tests. This metric can be directly related to the time required to search a specified area (since the longer the distance at which the target may be detected, the wider the search swath of the aircraft) and thereby the probability of survival of casualties [12]. In the search phase, the success of a trial may be measured by the time taken to locate the casualty, and is a parameter considered in trial planning software.

However, for a limited number of flight trials the time taken to detect the target may be strongly influenced by luck: the target might happen to be located early or late in the chosen flight trajectory. The problem is avoided if the distance at which the target is first detected is measured, rather than the time taken to identify. This can be achieved conveniently by 'dropping' coordinates using the aircraft's GPS system, for example. The distance and time parameters are easily linked; the greater the detection distance, the wider the swath of the aircraft sensors and the shorter the overall flying time needed to cover a given area (Burciu, Abramowicz-Gerigk, Przybyl, Plebankiewicz, & Januszko, 2020).

Data gathering

Having decided upon a suitable performance metric which will enable direct comparisons with the control condition, the question becomes how to accurately capture the data with the resources available. Typically, in-flight research uses sensors and instrumentation to record information in real time. This alleviates workload on the pilot/crew on-board the aircraft. Clearly, it would not be appropriate to expect a pilot to manually record data while in control of an aircraft for safety reasons. Furthermore, the operation should be conducted in as close to realistic conditions as possible, i.e. without added pressures, tasks, or distraction. Regarding the aircraft itself, it may be possible to add additional equipment for the purpose of the flight test, or modify the aircraft in some way to help capture the necessary data. However, depending on the type and intrusiveness of the modification, airworthiness assessment and certification may be required. Hence, the design of the flight test should take advantage of any existing equipment/sensors wherever possible.

An example of an efficient strategy for data capture during a SAR trial is to record the FLIR video of the entire search with audio from the aircrew intercom (IC) overlaid. Throughout the US and Europe SAR aircraft are typically fitted with the FLIR system and the video/audio recording feature should be available as standard. The FLIR video recording contains live flight information including ground speed, flight path and GPS location. In combination with the IC, GPS coordinates at the point when the target was first identified, and the GPS coordinates of the target itself, could be recovered following the trials via analysis of the video. From experience, this is a convenient method for retrospectively calculating the identification distances following the trials and was minimally obtrusive to the crew, allowing them to operate in a natural manner. It must be noted that FLIR video/audio recordings may be proprietary and/or protected against release to the public. This should be respected by the research team when presenting the findings.

Due to the significant variability (planned or unplanned) that could occur during a SAR trial, particularly regarding the environment in which the search takes place, there is also important supplementary data that should be recorded (see the lists below). Checklists are recommended to ensure integrity of the data in a complex trial where details could easily be forgotten. This additional data should be captured either at base before take-off, or by the non-flying pilot during the test. Again, considerations must be made to ensure that the requests for data logging by the crew are reasonable, and do not interfere with the basic operation during flight.

The following details should be recorded (see (Bennett, Hodgkinson, & Nixon, 2019) for an example test card) by a member of the aircrew **prior to take-off**:

- Date and time (if retrospective information regarding weather/sea conditions is required)
- Aircraft Type and Registration (for information on performance and equipment)
- Crew (for analysis of experience levels - hours, aircraft types, training, military/commercial background etc.)

The following details should be recorded (see (Bennett, Hodgkinson, & Nixon, 2019) for an example test card) by the non-flying pilot **prior to the start of each search pass**:

- Trial Number and configuration (for correlation with the data collected)
- Conditions (for analysis of effects on the success of a trial) including: air temperature, ambient pressure, visibility, sea temperature, wind speed and direction, weather, sea state, light meter reading (this can be communicated from a ground-based member of the support team).

Raw data should also be supplemented with comments and feedback from the aircrew in a debrief back at base. It is beneficial for the researchers to attend the flight debrief (where possible) and interview the crew to document their experience immediately following the trial. A combination of qualitative and quantitative data allows the researchers to conclude whether the innovation being tested offers a benefit in comparison to the current standard.

It is also important to prioritize the data which is sought so that critical data is collected first, followed by any supporting/repeated data if the resources and opportunities allow. It is a case of managing expectations and being realistic with what can be achieved given the resources available. It is common for trials to go awry when integrating research with a live operation. For example during the testing phase for (Bennett, Hodgkinson, & Nixon, 2019), while integrating a research trial with the SAR crew's scheduled training exercise, the crew received a real distress call and had to abandon the test mid-trial. This led to the loss of usable comparative data with respect to the control condition which had already been completed, and also the loss of the training manakin which was fortunately recovered the following day by the local lifeboat. As a further example, in a later trial for (Bennett, Hodgkinson, & Nixon, 2019), rough sea conditions prohibited the targets from being deployed in the search area originally planned for the test. This meant that the SAR crew, already in the air and ready to search, were briefed incorrectly. The trial, and subsequent data, was recovered by repositioning the targets into a more

sheltered area of coastline. The SAR crew were updated via radio of the new search procedure and were able to successfully complete the trial in full, albeit under conditions which could be considered marginally less representative of a real SAR event.

To ensure maximum validity of the data, the SAR crew must employ methods and equipment as defined by the most up-to-date International Aeronautical and Maritime Search and Rescue (IAMSAR) guidelines (IMO Publishing, 2019).

Controls

It is important to test any new innovation directly against a control to quantify any potential improvements and ultimately make recommendations. The most valid approach in most cases is to use the current industry standard as a baseline comparison. It should be tested at the same time and in the same conditions in order to provide a meaningful conclusion based on the performance metric (discussed above). For example, in a previous project by the authors (Bennett, Hodgkinson, & Nixon, 2019) regarding survival suit conspicuity, a standard unmodified immersion suit was chosen as the control measure.

Based on a review of the literature in the field of SAR flight testing, it appears that many published studies have not explicitly evaluated a direct comparison of the control, rather concentrating on validating the efficacy of the new innovation in isolation. As a result, it is difficult to make solid evidential recommendations. However in some cases, controls are not possible, for example when testing location beacons Lilja et al., 2013).

Blinding

Another consideration when conducting multiple trials of the same type, with the same participants, and in the same location is 'blinding'. In this context, blinding concerns eliminating the influence of previous trials on the current one, for example, the participants gaining experience of the experimental set-up. This is somewhat unavoidable when testing against a control as discussed above. Hence, it is important to design the trials in such a way to minimize this effect. Furthermore, 'double blinding', which concerns eliminating the influence of the scientists designing the trials on the participants of the trial, should also be considered. Clearly, there is an implication of blinding on the repeatability and reproducibility of any given trial, which is discussed further in the next section.

Again drawing upon experience, the key here is to design the tests in such that the SAR operation gains minimal transferable knowledge and experience from one test to the next. For the authors previous project regarding immersion suit conspicuity (Bennett, Hodgkinson, & Nixon, 2019), the trials were designed such that the aircraft's flight path would always traverse directly over the search area, and that the target would always be within the scanning arc of the pilots and FLIR system. Therefore, the metric for success (the distance at which the target was first identified) was intentionally not dependent on how the aircraft was flown. It was specified that, having completed a run, the aircraft was to continue

past the target to a landmark sufficiently far from the search area such that it would be impossible to identify the target, before turning around to make the next pass. There was, however, some variation in the location of the target within the search area due to tidal drift between passes, and also due to random nature of how the support boat deployed new targets within the search area. It was found that in most of the trials conducted, the visibility in the prevailing conditions was so limited that the crew were unable to make much use of knowledge gained on prior runs. The results show that the crew's performance on repeated runs was remarkably similar and did not noticeably improve as more trials were conducted. This is the goal in maximizing blinding.

Despite these successes in (Bennett, Hodgkinson, & Nixon, 2019), double blinding was difficult to achieve fully, as the research team were responsible for placing the target object on the sea surface and therefore knew its location. However, there are restrictions forbidding untrained personnel from being on-board the SAR aircraft, particularly when flying over water. Only SAR crew were allowed to be present and therefore scientists running the trial were explicitly not present when the helicopter was in the air. Therefore the research team are unable to directly or indirectly influence the crew's actions during the search phase. Radio communication between the team on the surface and the air crew was limited to events that took place after target detection occurred, or (in some cases) the search was abandoned.

Blinding in trial design is not frequently discussed, though it may have been implemented; for example in (Donderi, 1994) the measures taken to ensure that participants were unaware of object locations is not discussed. However, it is also important to ensure that a trial remains tractable and easy to conduct within a reasonable time window, therefore the target location must be somewhat constrained.

Repeatability and reproducibility

In a realistic field trial, it is difficult to assess the repeatability and reproducibility of a measurement while also ensuring blinding of participants and a realistic participant population. For example, it is typically not practical or efficient to change the search area, or to significantly change the position of the target within the search area, for each run due to the limitations of resources and time constraints. However, efforts should be made to blind the SAR crew as best possible from run to run as described above. Furthermore, there is typically not enough potential participants within the population to send out a new crew for each run.

Repeatability and reproducibility is also difficult to achieve regarding environmental factors such as weather, light conditions, visibility, sea state, sun position with respect to the direction of flight, etc. The key is to minimize the time required for each run so that the change in prevailing conditions is minimized from run to run. In (Bennett, Hodgkinson, & Nixon, 2019), the average trial time was 6-8 minutes and so any change in conditions could be considered negligible. Therefore, the dependency on the conditions can be removed from the comparison of the data.

It is recommended that if the trials are likely to take greater than 10-15 minutes, weather and sea conditions should be recorded by the SAR crew prior to each trial. This can be completed based on the weather radar at base, and the Automatic Terminal Information Service (ATIS) at the station airport. This should include sea state and tidal phase, lighting (cloud cover, position of sun or moon, etc.) and prevalence of rain, cloud or mist. For longer duration trials, any significant change in the conditions listed above during the test should also be noted, or otherwise considered constant.

Best practices for maximizing repeatability and reproducibility as employed for the maritime SAR trials in (Bennett, Hodgkinson, & Nixon, 2019) include choosing the trial time to be within a 2 hour window of slack water (high or low tide) to minimize the change in sea state and drift, and when the weather conditions were not predicted to significantly change. Differences in lighting conditions (ie. position of the sun/moon, and glare from nearby towns/cities) and the physical orientation of the target in the water should be mitigated by performing two runs in opposite directions. To assess reproducibility over a range of environmental variables, trials should be conducted in different locations and at different times of the year, in different sea states, wherever possible.

Sources of inconsistencies, interference and bias

There are a number of factors which could introduce inconsistencies, interference and systematic bias to the results of full-scale SAR simulations, confounding the validity of the trial. To allow comparison of results from multiple flight trials, it is important that these effects were mitigated or minimized as best possible. Below is a summary of typical sources of error, type, and best practices for mitigation/minimization. This is by no means a comprehensive list. The highly specific nature of any test will inevitably lead to a new and complex series of potential sources of inconsistencies, interference and/or bias. The subsections below summarize the most common and potentially adverse examples when conducting tests specifically in the SAR flight test domain.

Visual scanning range of the SAR crew

This is considered a potential inconsistency. This can be mitigated/minimized, for example, by maintaining the following scanning ranges constant throughout the testing: FO 10-1 o'clock, Captain 11-2 o'clock, rear crew 1-5 o'clock, FLIR operator 9-3 o'clock.

Position and scanning motion of the searchlight

This is considered a potential inconsistency. This can be mitigated/minimized, for example, by maintaining the searchlight fixed at 12 o'clock, fully de-focused.

Multiple targets

This is considered a potential source of interference. This can be mitigated/minimized by testing the new innovation and current standard separately, in the same environment, in the same conditions.

Search crew becoming familiar with the test procedure

This is considered a potential source of bias. This can be minimized by employing different search crews and a range of experience levels wherever practically possible.

Search crew becoming familiar with the search area

This is considered a potential source of bias. This can be minimized by scheduling a number of search trials in different locations wherever practically possible.

Randomness of success in identifying targets

This is considered a potential inconsistency. This can be minimized by conducting multiple tests of the same configuration.

Pretesting, efficient use of resources, and financial constraints

It is clear that from the considerations discussed above that there is a delicate balance to be found between scientific rigor and the practicality of integrating trials with a live SAR operation. Any such tests obviously puts a strain on resources, relies upon cooperation (and often volunteers), and can be very costly. Utilizing crewed helicopters as the SAR platform costs approximately \$10,000 per hour (excluding crew time). From a research perspective, it is key to realize that the day-to-day operation will always be prioritized at any base, potentially limiting the time window during which tests can be conducted.

Flight testing should therefore be the absolute final stage for any study of this kind. By its very nature, flight testing is expensive and comes with risks. Time spent in the air gathering data must be fully utilized, and the trials should be meticulously designed in such a way that the maximum amount of scientific data can be extracted from a minimal amount of flight hours. Time is critical and therefore it is not realistic to expect a large number of repetitions. Only once all constraints have been considered can a test plan be formulated.

Therefore, prior to any flight test planning, all necessary laboratory and ground testing should be completed to limit the number of variables that need to be tested at full scale. This should include, in particular, compatibility checks of all equipment with the proposed innovation. The goal is to clearly justify and constrain the flight tests by using the data from the lab/ground testing as evidence. This will minimize the number of discrete tests, and repetitions of tests, to be conducted at full scale.

The probability of success of a trial should also be considered. For example, is it likely that the data/evidence gathered from any given test will substantiate the claims/goals of the project? If the answer is yes, then the full-scale trials can be justified in terms of the resources required. If no, then the flight tests should be redesigned to improve the probability of success within the given constraints.

To alleviate (or absorb) some of the costs associated with flight testing, there may be operational opportunities to exploit. For example, SAR crews regularly undertake mandatory training exercises and so the research team may be able to integrate the field trials with these objectives. This was the case for the tests conducted in (Donderi, 1994) and (Bennett, Hodgkinson, & Nixon, 2019).

Trial location

The location of any trial should be chosen based on a number of factors: safety, representative conditions, financial, convenience, availability of equipment and personnel.

In the authors previous project (Bennett, Hodgkinson, & Nixon, 2019), ground and flight trials were conducted at 2 different SAR bases in the UK. It is important to recognize that SAR operations will always take precedence over any research related trials and therefore the research team should be prepared to spend long periods waiting for access to operational aircraft or crew members. A private research office, if available, can be used as a base, allowing the team to work effectively and also to minimize disruption to the SAR operation. For safety reasons, SAR officers are required peace and quiet during rest periods and equipment maintenance takes place in a no distraction zone. Ultimately, it is important that the research team respects normal SAR working practices.

Locations for the flight tests themselves should be chosen to be safe and un-congested in terms of both airspace and shipping. The search area should be large enough so that the exact location of the target is not trivially obvious to the SAR crew, but should be confined to the defined scanning ranges of the crew when following the predetermined flight track. Any landmarks should also be sufficiently far from the search area so that the target was not trivially visible from the turning points.

Local authorities should be informed such that the search trial is not mistaken for a genuine SAR mission. If the flight trial is to be conducted near a shore, it is a good practice to utilize a member of the research team to reassure and inform bystanders. These events are likely to attract attention and potentially cause distress to the public if there is a lack of information. Also, there is potential for outside personnel to interfere with the test in progress. For example in (Bennett, Hodgkinson, & Nixon, 2019), during one of the night-time trials, a member of the public parked their car on the promenade and directed the main headlight beam toward the beach from where the lifeboat crew were supporting the trial. Although this person was clearly trying to assist with a potential genuine rescue attempt, this type of occurrence would have been detrimental to the validity of the trial data, particularly in the case of evaluating the conspicuity of a casualty at sea from the air. In this case, the ground-based member of the research team was able to reassure the individual, with the result that they turned off their headlights.

Trial briefing

A briefing between the parties (for example, SAR crew, support boat crew, and the researchers/observers) should take place before each trial. A date and time for the trial should be decided, taking into consideration the requirement for a range of conditions as described above.

It is important for all involved to remain flexible and adaptable in the case that the trial does not go as planned. Appropriate contingency planning should ensure that time in the air is not wasted. A plan B should be formulated and briefed before each flight trial to give the commander the ability to change the objective of the trial mid-flight if it was deemed not safe to continue with plan A. This decision lies ultimately with the aircrew in this situation, as they are most at risk, and should be respected. A working radio frequency should be established so that the SAR crew can communicate with the research team (and support boat, for example, if applicable). Clear and efficient radio communication is key for a successful trial.

As an example of adaptability while on trial, during the second series of flight tests in Port Talbot, Wales in (Bennett, Hodgkinson, & Nixon, 2019), the plan initially was to conduct a search several miles out to sea, near to a tethered buoy which was used to signpost a shipping lane. Due to severe sea conditions and bad weather, deemed unsafe for the support boat, the search area was changed to a sheltered bay. As a result, the SAR crew had to change the aircraft's flight path to track parallel with the coast rather than in and out to sea. This decision allowed the boat crew to operate more safely, while maintaining continuity and validity of the trial.

Trial procedure and execution

While each SAR trial is unique depending on the overall objective and the data sought, the following example flight test methodology highlights the level of detail required. It is important to link the procedures of all parties involved with focus on communication. For example, the search helicopter must be aware of the support boat status at all times and maintain live communication via radio. The research team on the ground should also monitor these radio communications. This is necessary to ensure safety and efficient completion of the flight test sequence.

Effective and efficient teamwork, as well as clear and decisive communication is essential. Figure 2 shows a helicopter and support boat working together at sea. Conducting a flight trial is not without risk, so it should be performed exactly as planned, where every person involved understands their role and is confident in performing under pressure. Clearly the aircrew are experienced in operating under these conditions, however the ground support and research team may not be. The key message is that meticulous planning can ultimately achieve the research goals while saving time, effort, money, and also minimizing risk for all involved.



Figure 2: Effective teamwork between the SAR helicopter and the support boat during a flight trial, Bristol Channel, UK. Note that the image has been edited to disidentify the aircraft and the winchman sitting in the bow of the support boat.

Figure 3 shows a schematic for reference with Table 1 which sequentially lists the procedure for one of the flight tests conducted in (Bennett, Hodgkinson, & Nixon, 2019), as an example. This procedure applies specifically to the testing of new equipment in flight. However, if minor changes are made to the methodology, this strategy may also be used to test more conceptual SAR processes such as pilot scanning ranges, aircraft search patterns, and pilot training techniques.

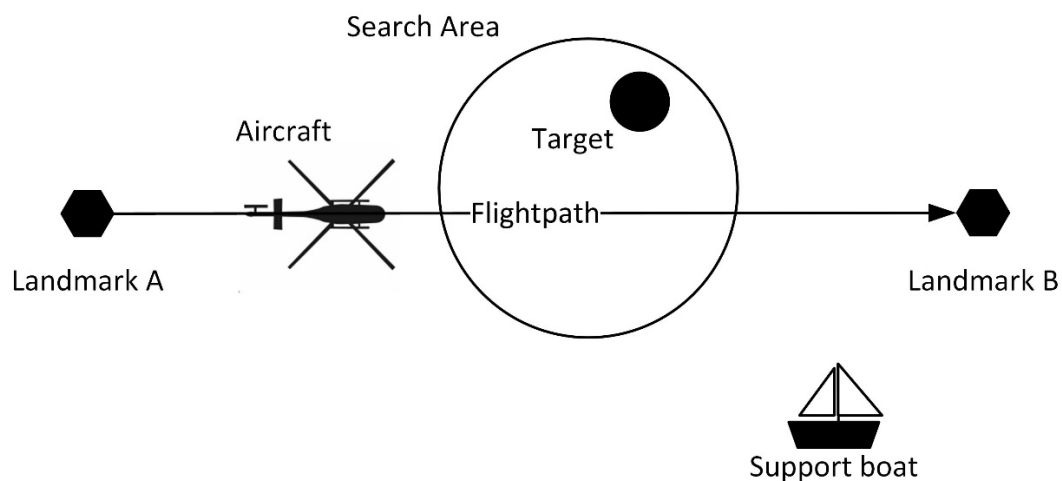


Figure 3: Schematic of search for the example given in Table 1.

1	Boat launches and transits to search area.
2	Boat deploys first target with PLB attached.
3	Boat transits to stand-off area and makes a radio call to the aircraft 'ready for search'.
4	Pilot completes the first portion of the test card.
5	Aircraft takes off and transits to landmark A.
6	FLIR operator starts the video recording.
7	Non-flying pilot completes the test card for trial 1.
8	Aircraft makes first pass from landmark A to landmark B searching for the target. On identifying, the crew member who spots says 'spotted from cockpit/on FLIR'.
9	Aircraft transits to directly above the target and the pilots says 'on top'.
10	Aircraft continues to landmark B and turns around.
11	Aircraft makes second pass from landmark B to landmark A searching for the target. On identifying, the crew member who spots says 'spotted from cockpit/on FLIR'.
12	Aircraft transits to directly above the target and the pilots says 'on top'.
13	Aircraft makes radio call to support boat 'second pass complete'.
14	If the target was not found, the SAR crew may use the PLB to locate and aid the recovery by boat.
15	Support boat retrieves the first target and deploys the second.
16	Aircraft continues to landmark A and turns around.
17	Support boat transits to stand-off area and makes radio call to aircraft 'ready for search'.
18	Non-flying pilot completes the test card for trial 2.
19	Aircraft repeats the two search passes.
20	The trial procedure repeats as necessary.
21	On completion of the final pass the FLIR operator stops the recording.
22	Support boat retrieves the target following the final pass and returns to dock.
23	The aircraft transits back to base, FLIR video is downloaded, and the crew debrief as normal recording any further details pertinent to the trial.
24	Research team performs the necessary analysis of the FLIR video recording.

Table 1: Example search procedure.**Data analysis, results and recommendations**

Assuming that the flight tests are successful and sufficient data has been collected, the research team can proceed to perform the required analysis. The first step is to quantify the results in context of the control condition. This includes assessing the validity of the data across the expected full range of conditions in a real SAR scenario. Both quantitative results (based on the chosen metric) and qualitative/anecdotal results (provided by the SAR crew in debriefs) should be presented.

Based on the analysis of the results, realistic and practical recommendations should be made. For example, the financial implications of introducing a new standard, in terms of the cost of both the new equipment itself and any additional crew and maintenance training (initial and recurrence) required, should be considered. Generally, it is not the role of a research team to recommend that SAR operations

change, but to provide an evidence base that can be used to inform decisions that may, for example, balance implementation costs and quality of response.

As an example, in the authors previous project (Bennett, Hodgkinson, & Nixon, 2019) focusing on crew immersion suit conspicuity, it was recommended that additional retro-reflective tape, positioned strategically on the survival suit, should be employed. The recommendation represents excellent cost effectiveness: the material itself is relatively inexpensive, the suits do not require radical redesign, no additional maintenance is required, and it was demonstrated that the modifications are compatible with existing SAR equipment on-board the aircraft. It was also shown through additional ground testing that the modified suit design did not lead to unwanted reflections, distraction, or discomfort within the cockpit. Therefore, the project and overall findings were considered successful in completing its objective by EASA, with the recommendation that regulators, immersion suit manufacturers, and SAR should consider the proposed design modifications to aid conspicuity at sea.

Conclusions and final remarks

In light of no published guidelines specifically for SAR flight-testing, and also with the noted lack of published articles related to full-scale SAR testing, this article presents the key considerations for designing and conducting a successful campaign, drawing upon prior experiences and lessons learned.

In the first half of this article, a review of the literature is provided to highlight the differing approaches to SAR testing. While a range of platforms, targets, locations and methods have previously been studied, published articles in this field focus mainly on a single component of the SAR operation, and often not at full scale or representative conditions. In addition, no previous studies consider a full-scale SAR flight trial using a helicopter (the most widely used method) probably due to the expense and availability.

In the second half, each SAR trial design principle is discussed to highlight the constraints, as well as methods to improve the validity of the trials, and ultimately the data required to solidify recommendations. For each section, examples are provided based on published studies or prior experience of the authors, primarily in the context of maritime SAR. However, these basic principles are relevant to any SAR flight test campaign, and it is the intention of the authors that this article can be used as a reference guide for future research projects in this field. It is hoped that the article can be used to ensure that the very best possible data can be captured in the most efficient and safe way. This article has been motivated by the many discussions, issues that have arisen, and resolutions that have been devised in previous research campaigns. The authors hope that the findings are of interest to the flight test community and most importantly can be used to improve and evolve the vital work that SAR operations conduct across the world.

References

- Federal Aviation Administration (FAA) (2011). Flight Test Guide for Certification of Part 23 Airplanes. US Department of Transport, No. 23-8C.
- Bennett, C. Hodgkinson, J. Nixon, J. (2019). Crew Immersion Suits Conspicuity. European Union Aviation Safety Agency (EASA), No. EASA_REP_RESEA_2017_2.
- U.S. Department of Defense (2019). Factsheet. Online:
<https://www.uscg.mil/datasheet/display/Article/1547982/aircraft/>, accessed June 26 2025.
- U.S. Department of Defense (2023a). Air Force Rescue Coordination Center – Mission. Online:
<https://www.1af.acc.af.mil/>, accessed June 26 2025.
- Bureau of Transportation Statistics (2017). U.S. Coast Guard Search and Rescue Statistics. Online:
<https://www.bts.gov/content/us-coast-guard-search-and-rescue-statistics-fiscal-year>,
accessed June 26 2025.
- U.S. Department of Defense (2023b). AFRCC 2023 Annual Report. Online:
<https://www.1af.acc.af.mil/Units/AFRCC/Annual-Reports/>, accessed June 26 2025.
- Gotovac, S. Papic, V. Marzusic, Z. (2016). Analysis of saliency object detection algorithms for search and rescue operations. 24th International Conference on Software, Telecommunications and Computer Networks (SoftCOM), No. 16522518.
- [8] Lygouras, E. Santavas, N. Taitzoglou, A. Tarchanidis, K. Mitropoulos, A. Gasteratos, A. (2019). Unsupervised Human Detection with an Embedded Vision System on a Fully Autonomous UAV for Search and Rescue Operations. *Sensors*, Vol. 19(16), No. 3542.
- Almeshal, A. Alenezi, M (2018). A Vision-Based Neural Network Controller for the Autonomous Landing of a Quadrotor on Moving Targets. *Robotics*, Vol. 7, No. 71.
- International Maritime Organization (1989). Use and fitting of retro-reflective materials on life-saving appliances. IMO, Resolution A.658(16).
- American Society for Testing and Materials (2013). Standard Practice for Measuring Photometric Characteristics of Retroreflectors. ASTM International, Active Standard ASTM E809.
- Burciu, Z. Abramowicz-Gerigk, T. Przybyl, W. Plebankiewicz, I. Januszko, A. (2020). The impact of the improved search object detection on the SAR action success probability in maritime transport. *Sensors (MDPI)*, Vol. 20, No. 3962.
- Hodgkinson, J. Nixon, J. Bennett, C. Tatum, R. (2020). Field investigation of retroreflective materials for enhanced target detection in maritime search and rescue. *Proceedings - Optical Sensing and Detection VI*, Vol. 11354, No. 61.

- Nixon, J. Hodgkinson, J. Bennett, C. (2020). Modified immersion suits for helicopter aircrew: Evidence for improved conspicuity from sea trials. *Safety Science*, Vol. 130, No. 104903.
- Cho, K. Li, Y. Wang, H. Park, K. Choi, J. Shin, K. Kwon, J. (2014). Development and validation of an operational search and rescue modeling system for the Yellow Sea and the East and South China Seas. *Journal of Atmospheric and Oceanic Technology*, Vol. 31, Pg. 197-214.
- Coppini, G. Jansen, E. Turrisi, G. Creti, S. Shchekinova, E. Pinardi, N. Lecci, R. Carluccio, I. Kumkar, Y. D'Anca, A. Mannarini, G. Martinelli, S. Marra, P. Capodiferro T. Gismondi, T. (2016). A new search-and-rescue service in the Mediterranean Sea: a demonstration of the operational capability and an evaluation of its performance using real case scenarios. *Natural Hazards and Earth System Sciences (NHESS)*, Vol. 16, Pg. 2713–2727.
- Breivik, O. Allen, A. A. (2008). An operational search and rescue model for the Norwegian Sea and the North Sea. *Journal of Marine Systems*, Vol. 69, Pg. 99–113.
- Donderi, D. (1994). Visual acuity, color vision and visual search performance at sea. *Human Factors*, Vol. 36, No. 1, Pg. 129–144.
- Ferreira, F. Ferri, G. Petillot, Y. Liu, X. Franco, M. Matteucci, M. Pérez-Grau, F. Winfield, A. (2018). Scoring robotic competitions: balancing judging promptness and meaningful performance evaluation. *IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC)*, Torres Vedras, Portugal, Pg. 179-185,
- Lilja, J. Pynttari, V. Kaija, T. Mäkinen, R. Halonen, E. Sillanpää, H. Heikkinen, J. Mantysalo, M. Salonen, P. de Maag, P. (2013). Body-worn antennas making a splash: lifejacket-integrated antennas for global search and rescue satellite system. *IEEE Antennas and Propagation Magazine*, Vol. 55, No. 2, Pg. 324-341.
- Miano, J. Empig, E. Gaw, A. Mendoza, O. Adlaon, D. Canedo, S. Dangcal, R. Sumalpong. A. (2019). Microcontroller-based vessel passenger tracker using GSM system: an aid for search and rescue operations. *International Journal of Advanced Computer Science and Applications*, Vol. 10, No. 9, Pg. 261-268.
- Miller, J. Kelly, J. Ehler, J. (1999). Flight test and performance of a nongated active television system. *AeroSense '99 - Infrared Technology and Applications XXV*, Orlando, Florida, USA.
- IMO Publishing (2019). *International Aeronautical and Maritime Search and Rescue Manual*. International Civil Aviation Organization (ICAO) and the International Maritime Organization (IMO), Ed. 2019, London, UK.