Safety Science 164 (2023) 106064

Contents lists available at ScienceDirect

Safety Science

journal homepage: www.elsevier.com/locate/safety

Factors influencing the time required to don thermal protective immersion suits correctly

Hooshyar Azizpour[a,*], Edwin R. Galea[a,c], Sveinung Erland[a], Bjørn-Morten Batalden[a,b], Steven Deere[c], Helle Oltedal[a]

[a] Department of Maritime Studies, Western Norway University of Applied Science (HVL), Norway
[b] Department of Technology and Safety, The Arctic University of Norway (UiT), Norway
[c] Fire Safety Engineering Group, University of Greenwich, United Kingdom

ARTICLE INFO

Keywords:
Survival suit
Donning time
Corrcetness
Evacuation
Polar waters
Passenger ship

ABSTRACT

The cold environment in polar regions introduces additional challenges when abandoning passenger vessels and offshore facilities. The International Maritime Organization Polar Code requires vessels operating in polar regions to be equipped with approved thermal protective immersion suits (TPIS) that can be donned unassisted within 120 s. As time is critical during an evacuation, quantifying the Net Donning Time (NDT) is important as this may need to be factored into passenger ship evacuation analysis. Furthermore, an incorrectly donned TPIS may be ineffective in providing the required thermal protection, so in addition to NDT, it is important to understand the factors that impact donning correctness. In this study, we present the results of a series of trials that quantified participants' performance while donning a TPIS with integrated buoyancy. Analysis of data from 108 participants revealed that NDT ranged from 65 to 341 s, with over 90 % requiring a total donning time of greater than 120 s. The mean NDT was dependent on a complex relationship between, age (increases by 6.6 % for each 10 years), gender (increases by 33 % if female), experience (decreases by 17 % with experience), method of instruction (increases by 21 % with video instruction) and failure to remove shoes (increases by 26 %). Furthermore, the method of instruction significantly impacted the number of donning errors, with instruction by video producing an average of 1.5 errors while written instruction producing 2.3. Finally, a donning time distribution is suggested for use in evacuation modelling analysis.

1. Introduction

Decreasing sea ice coverage in polar regions in recent times has resulted in a growth in the popularity of adventure cruises involving large passenger ships sailing in polar waters (Luck et al., 2010; Maher, 2017). The increase in ship traffic inevitably results in a higher probability of accidents or incidents involving these vessels in these challenging conditions (Khan et al., 2020). In light of this, and acknowledging the inadequacy of existing safety provisions for passenger ships operating in polar waters, the International Maritime Organization (IMO) introduced the Polar Code in 2017 (Polar Code, 2017). The Polar Code requires that passenger ships operating within polar waters are required where appropriate, to provide thermal protective clothing and insulated immersion suits (referred here as Thermal Protective Immersion Suit (TPIS)), for each person on board.

The unpredictability and speed at which maritime emergencies may occur make time a critical factor (Andreassen et al., 2020), whether it be associated with the time required to gather the passengers in the assembly stations, the time required by passengers to don their TPIS, or the time available to move passengers from the assembly station to the life safety apparatus (LSA) and consequently abandon the vessel. Given that emergencies may occur on passenger ships in polar waters, and that passengers and crew are likely to be encumbered by TPIS it is important to know how the TPIS is likely to impact time-critical procedures and operations. In particular, how long does it take to don TPIS, and how does the wearing of TPIS impact the movement rates of passengers and crew? An essential design requirement of TPIS is that they can be quickly donned during an emergency. According to IMO (SOLAS) and International Organization of Standardization (ISO) requirements, TPIS must be unpacked, properly donned and secured without assistance within 120 s in ambient temperatures of $20 \pm 2^{\circ} C$ (ISO, 2012; SOLAS, 1998). Within this paper, the total time required to don the TPIS is referred to as the

* Corresponding author.
E-mail address: Azizpour.h@gmail.com (H. Azizpour).

https://doi.org/10.1016/j.ssci.2023.106064
Received 7 April 2022; Received in revised form 23 November 2022; Accepted 4 January 2023
Available online 10 May 2023

Total Donning Time (TDT). This includes the time associated with opening all packaging, removing the suit from the packaging and donning the suit. It is further noted that it is essential that the TPIS is correctly donned. A TPIS that is incorrectly donned can impact the effectiveness of the thermal protection and buoyancy offered by the suit and may also adversely impact the ability of the wearer to walk efficiently.

In most cases, apart from anecdotal information, or information from marketing materials, a rigorous evidence base characterising the impact of TPIS on human performance within maritime environments does not exist. Furthermore, quantifying TPIS donning time is critical for three reasons: (1) developing achievable evacuation procedures for passenger ships operating in polar waters, (2) enhancing the design of TPIS, and (3) modelling evacuation performance using ship-based evacuation models (Galea et al., 2013; Gwynne et al., 2003; Vassalos et al., 2002; Pradillon, 2004).

Since 2002 the IMO has published a set of guidelines for evacuation modelling associated with new and existing passenger ships (IMO-MSC/Circ., 1033). The human performance data specified in these guidelines are based on research associated with land-based scenarios such as walking speed data collected from the built environment, such as rail stations. As such, the human performance data implemented within maritime evacuation models do not incorporate the performance of individuals under conditions associated with maritime emergencies such as adverse vessel orientation (heel or trim) or conditions associated with extreme weather such as may be encountered in polar regions. However, within the guidelines the IMO invited the Member States to collect and submit information and data resulting from research and development activities on human behaviour associated with ship evacuation (IMO-MSC/Circ., 1533).

Implicit within the intent of the IMO Polar Code (Polar Code, 2017) and the associated ISO standards (ISO, 2012) is the requirement that the TPIS should not adversely impact passenger ship evacuation. This is reflected by the requirement that the TPIS can be donned within 120 s and that it does not adversely impact walking speeds of individuals by more than 25 %, compared with normal walking speeds (ISO, 2012). However, thus far there is a lack of a substantive evidence base quantifying the impact of various types of TPIS on these parameters, and there is little understanding of the impact that the TPIS may have on evacuation performance within a maritime environment.

From the mid-1990s, the first ship evacuation models started to appear in the literature (Vassalos et al., 2002; Galea and Owen, 1994; Galea et al., 1998; Galea, 2000), and these publications highlighted the need for the collection of maritime specific human performance data, such as walking rates in maritime environments involving adverse vessel orientation, the impact of protective clothing, such as lifejackets on walking speeds and passenger response times (Galea et al.). Interest in quantifying the performance of people resulted in two significant land-based studies with a major focus on the impact of the maritime environment on walking speeds. Both studies attempted to reproduce key aspects of the maritime environment using land-based simulators. Both studies occurred independently and at around the same time, one in the Netherlands at the Dutch Research Institute (TNO) (Bles et al., 2002) and the other at an industrial research facility in Canada (Glen, et al., 2003). The TNO study made use of a modified shipping container on hydraulics to represent a ship corridor at various angles of heel and trim while the Canadian study made use of a purpose-built facility called SHEBA (Ship Evacuation Behaviour Assessment Facility) that could be heeled at various angles.

The SHEBA facility allowed measurements of human performance and behaviour in a typical ship passageway and stairway. Tests were conducted with participants with and without life jackets. While the SHEBA trials involved participants wearing lifejackets and collected data on donning of lifejackets (Glen, et al., 2003), none of the studies considered the impact of TPIS on the performance of individuals. More recently, several other lifejackets donning trials have been reported providing useful data concerning lifejacket donning times for infants and adults in full-scale studies (Brown et al., 2008; MacDonald et al., 2011).

One of the few studies concerning the donning of TPIS was conducted by Mallam et al. (2012). Their trials involved 32 test subjects (18 male and 14 female) with an average age of 22.9 ± 2.0 years, donning two different types of TPIS in both static and dynamic environments. The dynamic environment was created using an electric motion platform ($2m \times 2m$) with six degrees of freedom. The two types of TPIS were randomly distributed among both genders such that each type of suite was donned by nine males and seven females. Each participant repeated the trials seven times using the same type of suit. Each participant received verbal instruction on how to don the TPIS and was also allowed 300 s to read the instruction sheet prior to attempting to don the suit. Participants could also read the instructions again during the rest period between each subsequent donning trial. Thus, the participants can be expected to be well briefed as to the donning procedures prior to the start of the trial. The average donning time, determined by analysis of video recordings, was found to vary between 90.1 s and 115.9 s depending on the type of suit. A key finding of this work was that there was a significant learning effect associated with repeated donning of the TPIS. However, unfortunately, the donning times and the correctness of donning for the first donning attempt of each suit was not reported, and so it is uncertain how long a time was required for the first donning of each suit or what level of correctness that was achieved. Furthermore, there are a number of other limitations associated with these trials that reduce the usefulness of the findings. For example, all the participants were in the early twenties and so unrepresentative of the broad cross-section of the population that may need to utilise the TPIS, the sample size was very small, participants were instructed to tie back long hair prior to the trial so as not to interfere with the donning process and finally, the TPIS used in the trials were not representative of the type of survival suit approved for evacuation in polar waters (Mallam et al., 2012; Mallam et al., 2014).

In another donning study, the effect of learning and training on the correctness of donning survival gear (immersion suits) was investigated using 536 seafarers (290 officers and 246 ratings). Less than 1 % of the donning trials involved an error relating to the correctness of donning. In this experiment, all participants had received the necessary safety training required to serve at sea (Sanli et al., 2019). As the participants in this study were trained professional seafarers, the results do not shed light on the performance of typical cruise or adventure cruise passengers.

To address this lack of maritime relevant data and amass an evidence base that can be used to assess the impact of TPIS on evacuation performance in polar regions, Wester Norway University of Applied Science (HVL) and The Arctic University of Norway (UiT) embarked on the ARCtic EVACuation (ARCEVAC) project. As part of the AREVAC project, two different types of TPIS were used in a series of experiments to assess their impact on walking speeds and quantify donning times and the factors that influenced donning times. The two TPIS differed significantly, one was a lightweight survival suit produced by Hansen Protection (Sea Pass passenger suit) (Brünig et al., 2021) and the other was an immersion suit with fully integrated buoyancy and thermal insulation produced by Viking (Yousafe Blizzard PS5002). The impact of TPIS on walking speeds of individuals along a corridor at four different angles of the heel ($0^{\circ}, 10^{\circ}, 15^{\circ}$ and 20°) has recently been reported in Azizpour et al. (2022).

The aim of this paper is to systematically explore TPIS donning time and correctness and the factors that influence these parameters. To explore these issues a series of donning trials are conducted with over 100 volunteers donning a buoyancy integrated immersion suit produced by Viking (Viking TPIS, see Fig. 1). In addition, the paper provides a quantification of the TDT for the Viking TPIS that can be used in evacuation modelling analysis for passenger ships in polar conditions.

(a) Viking TPIS (b) TPIS carry bag with instructions placard (c) TPIS in sealed plastic bag

Fig. 1. The Viking TPIS and its packaging.

2. Experimental procedure and data collection

To identify the factors influencing donning times and the correctness of donning and to quantify the TDT for the Viking TPIS a series of trials were conducted with volunteers recruited from the local community. This section describes the recruitment of participants, the TPIS used in the trials and the experimental procedures employed.

2.1. Trial participants

Trial participants were recruited through the local media, social media, word of mouth etc. Recruited participants were asked to complete a pre- and post-trial questionnaire that included questions related to demographical information and potential previous experience in donning TPIS. In total 108 volunteers (71 male and 37 female) aged between 18 and 72 years of age were recruited. Older participants were not allowed in order to reduce the risk of injuries if slipping. Other recruitment criteria were that participants should be in good health without any serious condition that could impair their movement or vision. The total number of participants in different age groups for males and females is presented in Table 1. Of the 108 participants, 59 stated that they did not have prior experience of donning TPIS. The rest of the participants (49 people) either claimed to have a previous experience with donning a survival suit or had donned another type of survival suit (Hansen TPIS (Brünig et al., 2021) prior to the trial (see Table 1). The average height and weight of the male group were 1.83 m (Standard Deviation (SD) = 0.06 m) and 84.1 kg (SD = 11.9 kg), respectively. The female group had an average height and weight of 1.67 m (SD = 0.05 m) and 68.3 kg (SD = 10.5 kg), respectively.

2.2. TPIS

The TPIS used in this study was supplied by Viking and is a buoyancy integrated immersion suit equipped with a thick layer of thermal insulation, satisfying the thermal requirements of the Polar Code (2017). This TPIS is a one size fits all suit, accommodating a wide array of body types and heights. The suit consists of integral foot coverings and a hood with non-integral but attached gloves. Rubber seals around the face and wrists are intended to prevent water ingress into the suit. The foot covering was equipped with rubber soles requiring the suit to be worn without shoes (Fig. 1.a). A total of 25 Viking TPIS were used in the trials, of which 14 were new (previously unused) and 11 were previously used, at least once. Each TPIS was stored in a zipped carry bag provided with the suit (see Fig. 1.b). The zipper of the carry bag extended over three sides of the carry bag. In addition, unused suits are sealed within a plastic bag within the carry bag (see Fig. 1.c). Once used by the first participants, suits were folded and placed inside the carry bags without the sealed plastic bag, ready for the next group of participants. We define the Net Donning Time (NDT) to be the time required to open the carry bag, extract the TPIS and don the suit (see Eq. (4) and Eq. (5) for details).

2.3. Trial procedures

The donning trials were conducted at two shore-based facilities, the ARCOS safety centre in Tromsø) and the ResQ safety centre in Haugesund. In total, 84 volunteers participated at the ARCOS safety centre and

Table 1
Arithmetic mean, minimum, and maximum NDT (s) for the different age groups of participants given their previous donning experience and method of instruction (number of participants shown in brackets).

		Mean, Min-Max and (Number of participants)						
Method of Instruction	Experience	18 – 19 Years of age		30 – 50 Years of age		≥51 Years of age		Total number
		Male	Female	Male	Female	Male	Female	
Written Instruction (WI)	No Experience (NE)	149.1 90.6–208.4 (25)	202.9 135 –335 (12)	221.4 164.4–341.1 (4)	185.6 142.6–210.6 (3)	201.7 125.4–278.1 (2)	187.2 187.2–187.2 (1)	47
	Experience (E)	122.7 64.7–265.6 (14)	145.7 110.6–224.6 (4)	139.4 77.4–249.3 (13)	198.4 106.9–235.4 (7)	109.4 93.2–125.6 (2)	209.7 167.9–251.5 (2)	42
Video Instruction (VI)	No Experience (NE)	156 101.9–221.3 (7)	315.5 315.5–315.5 (1)	N/A	177.5 172.4–182.6 (2)	165.1 165.1–165.1 (1)	365 365–365 (1)	12
	Experience (E)	N/A	179.8 179.8–179.8 (1)	170.3 146.7–193.9 (2)	185.8 111.8–259.7 (2)	112.3 112.3–112.3 (1)	201.7 201.7–201.7 (1)	7
Total number		46	18	19	11	6	8	108

24 at the ResQ safety centre. Full details concerning the trial procedures can be found in the Supplementary Materials (see *Sec*. S1), here we present a summary of the key details. Upon the arrival of the participants at the trial location, participants went through a registration process which included completing the pre-trial questionnaire and consent form, and participants were then given a group safety briefing. Participants were also instructed to remove coats and jackets and to leave all personal belongings behind prior to being escorted to the trial area.

The TPIS within its carry bag was placed on the floor in front of each participant. Participants were instructed to imagine that they were at sea on board a passenger ship sailing in polar waters and the evacuation alarm had just been sounded. The participants were told that they had to don the suit as quickly and as correctly as possible so that they would be ready to safely evacuate the vessel. The task would start once the instructor yelled "GO" and the end point was defined as the time that the participant raised their arms above their head.

Prior to starting the trials, a sub-group of randomly selected participants were shown a two-minute instructional video demonstrating the correct donning procedure. In total 19 participants were shown the video demonstration. This sub-group consisted of 10 male and 9 female participants aged between 18 and 72 years.

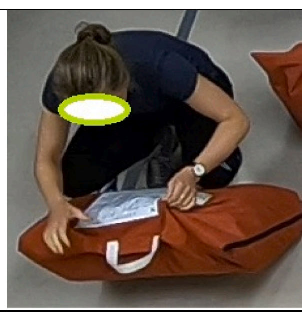
In addition, written instructions (provided by the manufacturer) were available to all participants through a laminated sheet located prominently on the suit carrying cover (Fig. 1.b and Supplementary Material *Sec*. S2). Participants were not permitted to read the instructions prior to the start of the trial. The participants' donning performance was recorded throughout the donning trial using two GoPro Hero cameras (frame rate of 25 FPS). A range of quantitative and qualitative data was collected during the trials through video footage and questionnaires (see Supplementary Material *Sec*. S3). Quantitative data concerning donning correctness and speed of donning was collected through analysis of the video footage.

Presented in Fig. 2 are example frames extracted from the trial video footage highlighting important behaviours noted during the donning trials (additional information can be found in Supplementary Material Figure S2). The images demonstrate examples of participant behaviour as they read the instructions (Fig. 2.1), unpack the TPIS (Fig. 2.2 to Fig. 2.4) and attempt to don the suit (Fig. 2.5 to Fig. 2.7).

Prior to the start of the experiments, an application for ethical approval for the research was sent to the Norwegian Centre for Research Data (NSD). All appropriate measures were taken to ensure the safety and anonymity of participants. Participation in the trials was completely voluntary and the participants could withdraw from the trials at any time.

3. Results

In this section the main results from the data collection are summarised. This consists of data extracted from the video analysis supported by data extracted from trial questionnaires.



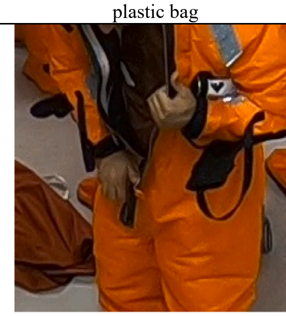
2.1. Reading the instructions

2.2. Start of the donning process with participant touching the zipper

2.3. Carry bag is opened, the start of the process to remove TPIS from the sealed plastic bag

2.4. End of removing TPIS from the plastic bag

2.5. Removable neoprene gloves



2.6. Pulling up the zipper

2.7. End of donning process

Fig. 2. Examples of key participant behaviours during donning trials.

3.1. Definition of variables

The controlled variables of primary interest consisted of:

- Demographic: primarily age and gender
- Experience: no previous donning experience (NE) or previous donning experience (E)
- Instruction method: written instruction only (WI) or video plus written instruction (VI)

Information relating to Demographic and Experience variables were quantified through analysis of the pre- and post-trial questionnaires. For the Experience variable, participants were asked if they had previously donned a TPIS. If they answered yes, they were considered experienced, irrespective of how long ago or how often they had undertaken the task. Participants were placed in one of the two Instruction categories at the start of the trial. As a main purpose of the trial was to establish a donning time distribution that could be used in evacuation modelling analysis, it was decided to focus on the minimum instructional method associated with just written instructions, and so most of the participants were placed in this category.

Observational parameters derived from the video analysis included; net donning times (NDT), preparation time (PT), extraction time (XT) and donning errors. These terms are further defined in this and subsequent sections. The process by which observational parameters were reliably and consistently extracted from video footage relied on the specification of a data dictionary and the precise definition of key parameters that were to be quantified. The definitions of key timed events as defined in the data dictionary are as follows:

(1) **Trial Start/End:**
(a) **Trial start time:** Time at which 'GO' command is heard on audio track (t_s).
(b) **Trial end time:** Time at which the participant has raised both hands to the highest level they could reach, indicating they were finished (see Fig. 2.7) (t_e).
(2) **Preparation phase:**
(a) **Start preparation:** Trial start time (see Fig. 2.1) (t_{sp}, note by definition, $t_{sp} = t_s$).
(b) **End preparation:** Time at which the participant disengaged from the preparation phase and engaged in the donning process by touching the zipper tracker in order to open the cover (see Fig. 2.2) (t_{ep}).
(3) **Extraction phase:**
(a) **Start extraction process:** Time at which the participant first touches the plastic bag with the intent to open it (once the cover is opened). (See Fig. 2.3) (t_{xs}).
(b) **End extraction process:** Time at which the participant has fully extracted the TPIS from the plastic bag (see Fig. 2.4) (t_{xe}).

Having defined these parameters, it is possible to determine the time required by each participant to undertake various tasks. These are defined as follows:

During the preparation phase, it is anticipated that participants will take some time to read the donning instructions which are available on the face of the package in the form of a large placard (see Fig. 1.b and Supplementary Material *Sec.* S2) prior to attempting to don the TPIS. The time spent during the preparation phase (PT) for each participant is defined as the time interval from the start of the trial (i.e., 1a) to the end of the participants preparation process (i.e., 2b), see Eq. (1).

$$PT = t_{ep} - t_s \tag{1}$$

The time required to extract the TPIS from the plastic bag (XT) for each participant is defined as the time interval from the start of the extraction process (i.e., 3a) to the end of the extraction process (i.e., 3b), see Eq. (2).

$$XT = t_{xe} - t_{xs} \tag{2}$$

The NDT for a participant with a used TPIS (NDT_{used}), i.e., suit not in a sealed plastic bag, is defined as the time interval from the end of their preparation phase (i.e., 2b) to their trial end time (i.e., 1b). The NDT_{used} can be determined for 94 participants and is given by Eq. (3).

$$NDT_{used} = t_e - t_{ep} \tag{3}$$

The NDT for a participant with a new TPIS (NDT_{new}), i.e., suit in a sealed plastic bag, is defined as the time interval from the end of their preparation phase (i.e., 2b) to their trial end time (i.e., 1b) less the XT time. The NDT_{new} can be determined for 14 participants and is given by Eq (4).

$$NDT_{new} = \left(t_e - t_{ep}\right) - XT \tag{4}$$

The Total Donning Time (TDT) for a participant with a used TPIS (TDT_{used}) is defined as the time interval from the trial start time (i.e., 1a) to their trial end time (i.e., 1b) plus the XT time. The TDT_{used} can be determined for 94 participants and is given by Eq. (5).

$$TDT_u = (t_e - t_s) + XT \tag{5}$$

Thus, the TDT includes the preparation time and a representation of the extraction time. Depending on the nature of the intended application, the XT can be represented by the mean XT, maximum XT or the XT distribution within Eq. (5).

The Total Donning Time (TDT) for a participant with a new TPIS (TDT_{new}) is defined as the time interval from the trial start time (i.e., 1a) to their trial end time (i.e., 1b). The TDT_{new} can be determined for 14 participants and is given by Eq. (6).

$$TDT_n = (t_e - t_s) \tag{6}$$

As the TDT_{new} inherently includes a measure of the actual extraction time achieved by each participant, there is no need to add the XT term to Eq. (6).

Finally, it is important to note that the NDT is a combined measure of two parameters, the time required to extract the TPIS from the zippered carry bag and the time required to don the TPIS. Thus, the NDT does not simply measure the inherent ease or difficulty associated with donning the TPIS. This is particularly important to keep in mind when attempting to compare the inherent donning performance of the TPIS described in this analysis with another TPIS design.

Throughout the video analysis, other behavioural data such as donning errors were quantified and recorded in the form of binary variables (Yes = 1, No = 0) (see *Sec.* 3.4). A randomly selected set of footage was analysed by two raters using the definition of variables provided in the data dictionary. The analysis was undertaken independently by the two raters to quantify key observational parameters (e.g., donning times, instruction times, opening times, etc.) and behavioural parameters. As part of the inter-rater assessment, video footage for 20 participants was analysed by the two raters and the results were compared using inter-rater analysis methods (McGraw and Wong, 1996; McHugh, 2012). Interclass Correlation Coefficient (ICC) was used to compare the measurements of durations and Kappa statistics were used for comparison of quantified behavioural variables. Results showed excellent agreement between raters with an average ICC value of 0.99 and a Kappa value of 0.85, respectively, for duration measurement and behavioural data. Analysis of the video footage for the Viking TPIS required approximately 63 person-hours of effort, 5076 data points were collected.

3.2. Net donning time

The donning data from 108 participants were collected from two different locations (see *Sec.* S1) thus, the possible influence of location on the NDT was assessed using a Mann-Whitney test. Results from the test did not indicate that the location of trials influenced the NDT (P-value = 0.23). Therefore, data from the two different locations were merged into one dataset. Descriptive statistics for NDT according to the different age groups, gender, experience and method of instruction are presented in Table 1.

Across all categories, the NDT for males varied from 64.7 s to 341.1 s (see Table 1), with an overall mean of 147.5 s while for females the NDT varied from 106.9 s to 364.9 s (see Table 1) with an overall mean of 198.9 s. Taken across all categories, this suggests that males were on average quicker in removing the TPIS from the zippered carry bag and donning the TPIS.

A distribution identification test, based on the Anderson-Darling goodness-of-fit test (Stephens et al., 1986), suggests that the NDT for both males and females was best represented by log-normal distributions. The Anderson-Darling test gave P-values of 0.42 and 0.64 for the male and the female group, respectively. As presented in Fig. 3, NDT can be represented by log-normal distributions with location (μ) and scale (σ) of 4.94 and 0.334 respectively for males and 5.25 and 0.297 respectively for females. The influence of age, gender, experience, and method of instruction on NDT is examined in detail in Section 4.3.

3.3. Preparation and extraction times

By definition, all participants spent some time during the preparation phase as defined by Eq. (1). This is the time interval between the trial start time and the participant purposefully touching the zipper to open the carry bag. Among all participants who did not receive the video instruction (i.e., the 89 participants in the WI group, see Table 1) the average time spent in the preparation phase prior to beginning to open the carry bag was 2.5 s (SD = 5.2s) with a range from approximately 1 s to a maximum of approximately 35 s.

For the 89 participants in the WI group, it is reasonable to assume that some or all of the 'preparation time' is spent reading the donning instructions. While there is a large amount of text on the instructions placard, the actual donning instructions consist of eight short bullet points and associated pictograms (see Supplementary Material *Sec*. S2). Thus, once the appropriate text is identified on the placard, the donning instructions would not require much time to read but may require more time to correctly interpret. The time spent in the preparation phase (PT) for the 89 participants in the WI group is distributed as shown in Fig. 4.a. A Mann-Whitney test did not show significant differences in PT time based on gender (P-value = 0.62).

As can be seen in Fig. 4.a. some 14 (16 %) participants had an extended preparation time (i.e., greater than 2 s). Given this large tail, it is difficult to represent the PT distribution using a continuous mathematical expression. Thus, two expressions are used to describe PT distribution. The long tail of the PT distribution is reasonably well described using a log-normal distribution (P-value = 0.75) with the location (μ) and scale (σ) respectively 2.35 and 0.56.

By taking an average preparation time of 1 s for participants without extended preparation time and assuming that the additional time for those with extended preparation time (16 %) follows a log-normal distribution, the PT distribution for the 89 participants in the WI group can be approximated by Eq. (7).

$$PT = 1 + U * X \tag{7}$$

where $U \sim \text{Bernoulli}(0.16)$ and $X \sim \text{Log-normal}(2.35, 0.56)$

Using this formulation, each person is allocated a 1 s PT and 16 % have an additional PT derived randomly from the log-normal distribution (X).

Finally, as 14 new suits were available during the trial, 14 (eight males and six females) participants were engaged in the extraction phase, where the TPIS had to be removed from the sealed plastic bag. The duration of the extraction phase (i.e., XT as defined by Eq. (2)) varied from 9.8 s to 31.5 s, with a mean of 19.4 s (SD = 7 s). A Mann-Whitney test did not show that the mean XT was significantly different between the male and female groups (P-value = 0.33). The XT was not significantly different from the log-normal distribution (P-value = 0.06) and should be a reasonable choice for modelling purposes (though a distribution fit is in any case deemed to be uncertain with such small sample size). The time spent in the extraction phase (XT) by the 14 participants who had new suits is distributed as shown in Fig. 5. The XT distribution can be approximated by a log-normal distribution and is given by Eq. (8).

$$XT = Log - normal(2.9, 0.39) \tag{8}$$

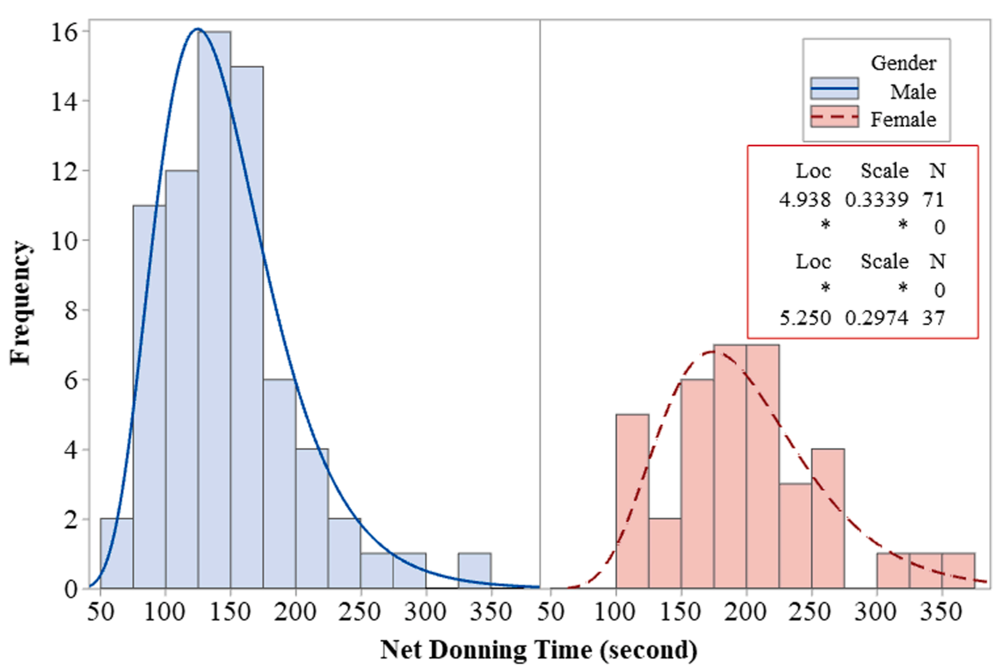
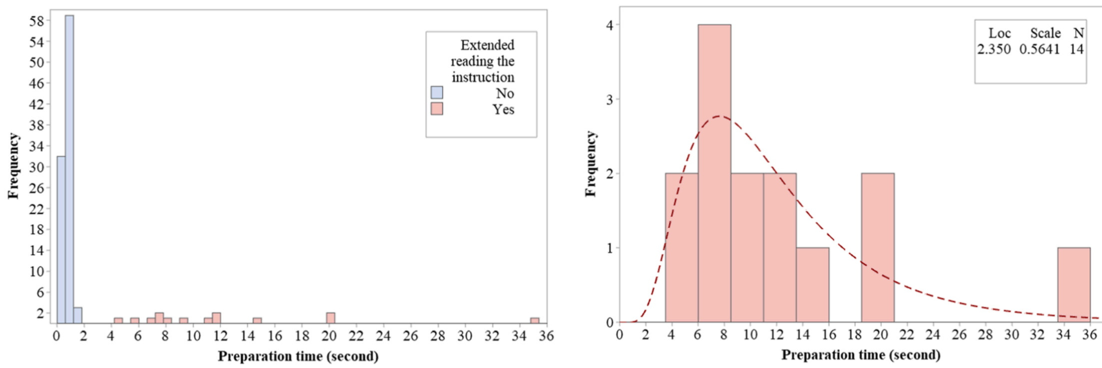


Fig. 3. Distribution of NDT for both male and female groups.



(a) PT distribution (89 participants)

(b) PT distribution for participants with extended preparation times (14 participants)

Fig. 4. Preparation of all (a) and preparation with extended reading (b) time distribution.

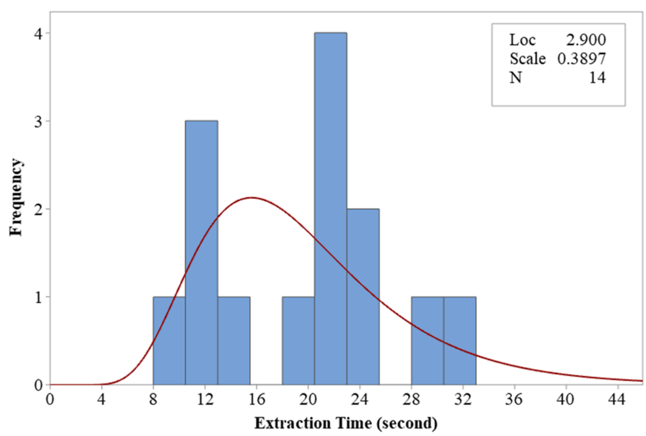


Fig. 5. Extraction time (XT) distribution (14 participants).

3.4. Donning errors

The number of donning errors incurred by each participant was also evaluated. A donning error was defined as a key feature of the donning process that was not correctly completed by the participant. These features were based on the checklist of features identified by the TPIS manufacturer and indicated on the laminated instructions appearing on the suit carry case (see Fig. 1.b and Supplementary Material *Sec.* S2). Donning errors are associated with seven key donning features as shown in Table 2. To be classed as a donning error, the donning feature must be in a final state other than the correct final state identified in Table 2.

Throughout the video analysis, correctness of donning of all items listed in Table 2 was checked/quantified for each participant using a binary variable (Incorrect/No = 0, Correct/Yes = 1). The Error Count (number) for each participant was defined as a metric to investigate the correctness of donning. An 'Error Count' of zero indicates that the participant donned the TPIS correctly with no errors, while an 'Error Count' of seven denotes that the participant made seven errors during donning. The error metric did not distinguish between the different types of errors listed in Table 2.

Table 2
List of key donning features and their correct final state.

#	Key donning feature	Definition of correct state
1	Shoes	Shoes should be removed prior to donning suit
2	Hood (see Fig. 1.a and Fig. 2.7)	Hood should be pulled over the head covering the entire head
3	Ankle straps (see Fig. 1.a)	Both ankle straps should be securely fastened
4	Interior length straps	Both interior length straps should be adjusting to ensure suit legs are not to baggy
5	Zipper (see Fig. 2.6)	The zipper should be pulled up all the way past the chin
6	Chest buckle	The chest buckle must be fastened
7	Gloves (see Fig. 2.5)	Both gloves should be worn

As shown in Table 3, the 108 participants incurred a total of 234 donning errors, 158 by male (of which there are 71) and 76 by female (of which there are 37) participants. On average, males incurred 2.2 donning errors while females incurred 2.0, and experience and video instruction decreased the mean number for donning errors. Presented in Fig. 6 is a pie chart showing the frequency of each donning error.

The least number of donning errors was associated with the hood and gloves, representing only 0.4 % (1) and 0.9 % (2), respectively, of the total number of errors (see Fig. 6). Only 0.9 % and 1.8 %, respectively, of the population donned these suit components incorrectly. Participants with very long hair appeared to require longer time to ensure that their hair was within the hood, and this was confirmed by participants in their post-trial questionnaire (see Supplementary Material *Sec.* S3).

The next most frequent donning error concerned the fastening of the ankle strap, representing 6 % (14) of the total number of errors (see Fig. 6). Given the universal size (one size fits all) of the TPIS, the ankle strap is necessary to secure that the shoe of the TPIS remains in place on the wearers foot. In total 13 % (9) of men and 14 % (5) of women failed to fasten the ankle straps. The next most frequent donning error, representing 8.1 % of the total number of errors, involved the zipper (see Fig. 6). When the zipper is correctly pulled up over the chin it creates a waterproof seal, however participants struggled with pulling the zipper above the neck and chin. Approximately 11 % (8) of males and 30 % of females (11) struggled with this task.

The third most frequent donning error, representing 22.6 % of the total errors (see Fig. 6) concerned the lifting harness (chest buckle). This is designed to aid the retrieval of an individual from water. If the lifting harness is not buckled up it is more difficult to rescue passengers that have fallen into the water. Only 46 % (33) of males and 59 % (22) of females had buckled up their lifting harness when donning was completed. The second most common donning error, representing 23.1 % of the total errors (see Fig. 6), was associated with failing to remove shoes prior to donning. More than half the males (54 % i.e., 38) and 43 % (16) of females failed to remove their shoes. The most frequent donning error, representing 38.9 % of the total errors (see Fig. 6), was failure to adjust the length straps. These straps are intended to compensate for the universal size of the TPIS. The length straps, one located on each side, adjust the length of the TPIS, lifting the gusset and bunching up excess fabric in the legs and upper body area. Some 88 % (63) of men and 76 % (28) of women failed to adjust these.

Presented in Fig. 7 is the distribution of the number of donning errors incurred by participants from the male and female groups. For each

Table 3
Descriptive statistics for donning errors according to different methods of instruction and experience.

Method of Instruction	Experience	(Number of errors) / (Number of people)	Average number of errors / person	Mode of errors per person	Number of occurrence of mode	Min number of errors per person	Max number of errors per person	Number of people
Written Instruction (WI)	No Experience (NE)	118/47	2.5	3	18	0	5	47
	Experience (E)	87/42	2.1	2	17	0	4	42
Video Instruction (VI)	NoExperience (NE)	19/12	1.6	2	7	1	2	12
	Experience (E)	10/7	1.4	1	4	1	2	7
Total		**234/108**	**2.2**	**2**	**44**	**0**	**5**	**108**

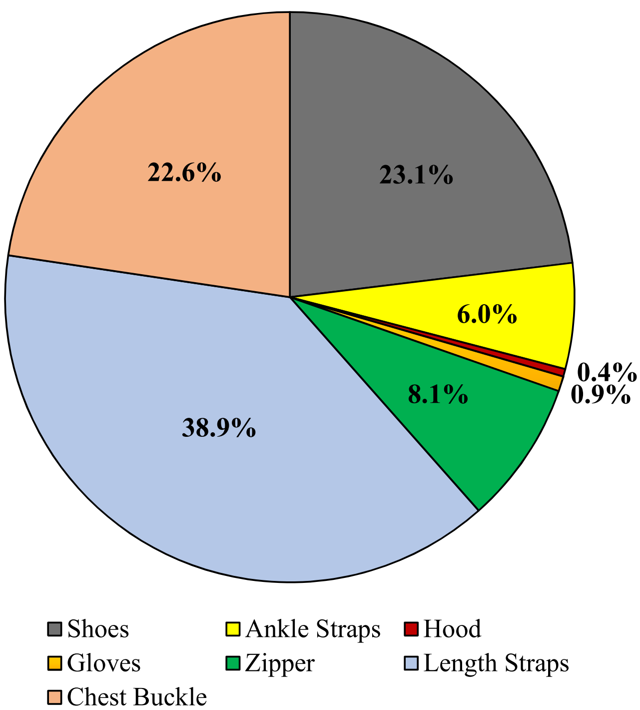


Fig. 6. Frequency for each type of donning error.

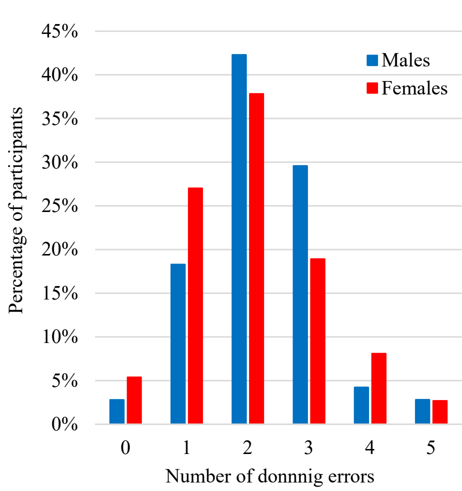


Fig. 7. Percentage of male and female participants committing a given number of donning errors.

participant the number of donning errors varies from 0 to 5 for both males and females, with the most common number of errors being 2 for both males and females.

4. Identification of factors impacting number of donning errors and NDT using regression modelling

In this section a regression model is developed to explore which parameters are most influential in impacting NDT and the number of donning errors and the nature of the interaction. All the regression analysis was performed using Minitab (version 19.2) and a significance level of 0.05 is used in all statistical inferences.

4.1. Regression models

4.1.1. Poisson regression

Poisson regression (Hoffmann, 2016) can be used to predict a dependent variable that are counts (e.g., number of donning errors) following a Poisson distribution given one or more independent variables or predictors. Let $Y \sim \text{Poisson}(\mu_Y)$ denote the number of donning errors following a Poisson distribution with expected number of donning errors given by μ_Y. In order to avoid negative values of μ_Y, one assumes that there is a log-linear relationship between μ_Y and the predictors x_i, $i = 1, \ldots, n$, in the following way:

$$Ln(\mu_Y) = a_0 + a_1x_1 + a_2x_2 + \cdots + a_nx_n \quad (9)$$

By exponentiation of Eq. (9) we have:

$$\mu_Y = e^{a_0} * e^{a_1x_1} * e^{a_2x_2} * \cdots * e^{a_nx_n} = A_0 * A_1^{x_1} * A_2^{x_2} * \cdots * A_n^{x_n} \quad (10)$$

In the Poisson regression model, each 1-unit increase in the predictor x_i multiplies the expected value of Y by $e^{a_i} = A_i$. Here A_i can be interpreted as a growth factor, and $A_i - 1$ gives the relative increase in expected number of donning errors per unit increase of x_i (all other factors being kept constant).

4.1.2. Log-linear regression

The potential impact of various predictors on the NDT was investigated using a log-linear regression model by log transforming the NDT (as a response factor) in a general linear regression model (Levine et al., 2001). If the response variable (i.e., NDT) is log-transformed, the effect of any predictor in a linear regression model would be a percentage-wise reduction or increase in the NDT. A log-linear multiple regression model for response variable Z (i.e., NDT) and predictors x_i can generically be represented as follows:

$$Ln(Z) = b_0 + b_1x_1 + b_2x_2 + \cdots + \varepsilon, \varepsilon \sim Normal(0, \sigma) \quad (11)$$

By exponentiation of Eq. (11) we have:

$$Z = e^{b_0} * e^{b_1 x_1} * e^{b_2 x_2} * \cdots * e^{\varepsilon} = B_0 * B_1^{x_1} * B_2^{x_2} * \cdots * \widetilde{\varepsilon},$$
$$\widetilde{\varepsilon} \sim Log - normal(0, \sigma) \quad (12)$$

In the log-linear regression model, each 1-unit increase in the predictor x_i multiplies the expected value of Z by $e^{b_i} = B_i$. Here B_i can be interpreted as a growth factor, and $B_i - 1$ gives the relative increase in NDT per unit increase of x_i (all other factors being kept constant).

4.2. Factors influencing the number of donning errors

A Poisson regression analysis (see *Sec.* 4.1.1) was undertaken to explore the impact of predictor parameters on the expected average number of donning errors (ADE $= \mu_Y$). The predictor parameters explored in the regression analysis are preparation time (x_1), method of instruction (x_2), experience (x_3), gender (x_4) and age (x_5). The definition and state of the predictor variables are presented in Table 4. Note that video instruction (VI) is an abbreviation of video with written instruction, i.e., these participants had access to both forms of instruction.

In addition to the predictor parameters identified in Table 4, potential interactions between the parameters were also considered through the introduction of interaction terms such as method of instruction and donning experience ($x_2 \times x_3$). According to the results of the stepwise Poisson regression, only the x_1, x_2 and x_3 predictor parameters were found to be significant, while none of the interaction terms turned out to be significant ($R^2 = 25.4$ %). As a result, the expected average number of donning errors can be estimated by:

$$ADE = 2.81 * 0.96^{x_1} * 0.59^{x_2} * 0.73^{x_3} \quad (13)$$

Presented in Table 5 are the coefficients, standard errors and P-values for the significant predictor parameters in Eq. (13). Also presented is the expected change in predicted ADE per unit increase in predictor parameter.

From Table 5, 'method of instruction' (x_2) is predicted to have the greatest impact on the ADE, followed by 'experience' and then 'PT'. From Eq. (13), a group of people exposed to written instruction only ($x_2 = 0$), without previous donning experience ($x_3 = 0$) and a PT of 0 s (i.e., no time to read instructions) are expected to incur an average of 2.8 donning errors (ADE $= 2.81 * 1 = 2.81$ from Eq. (13)). However, if the same group has a PT of 10 s (i.e., has more available time to read the instructions), then they are expected to incur an average of 1.9 donning errors (ADE $= 2.81 * 0.96^{10} = 1.9$). Thus, for inexperienced persons exposed only to written instructions, the average number of donning errors is predicted to decrease by approximately 33 % for every 10 s of preparation time.

Presented in Fig. 8 is a plot of the expected ADE (Eq. (13)) for participants with different types of instruction as a function of preparation time. However, it should be noted that only a single participant had a PT as high as 35 s, but any test intended to confirm how representative this data point is would be very uncertain due to the small sample size. If this PT is considered an outlier and excluded from the analysis, then the P-value for PT increases from 0.02 (i.e., a significant result) to 0.09 (i.e., not a significant result). Thus, while it may be argued that intuitively it would be expected that PT would exert a significant impact on the expected number of donning errors, the analysis presented here is not conclusive and would benefit from additional data.

Table 4
Definition of predictor parameters examined in the regression analysis.

Predictor Parameter	Definition	State	
x_1	Preparation Time (s)	0 – 35 s	
x_2	Method of Instruction	1 = Video Instruction (VI)	0 = Written Instruction (WI)
x_3	Experience	1 = Yes (E)	0 = No (NE)
x_4	Gender	1 = Female	0 = Male
x_5	Age (years)	18 to 72 years	

Table 5
Contributing factors and change in the ADE given one unit increase in each of the influencing variables (when all other variables are fixed).

Variable	Coefficient	Standard Error of Coefficient	$A_i = e^{a_i x_i}$	Change in the ADE per unit increase of x_i	P-value
x_1	−0.044	0.02	0.96	Approximately −4% per second preparation time	0.023
x_2	−0.52	0.2	0.59	−41 % with receiving video instruction	0.009
x_3	−0.32	0.15	0.73	−27 % with having previous donning experience	0.03

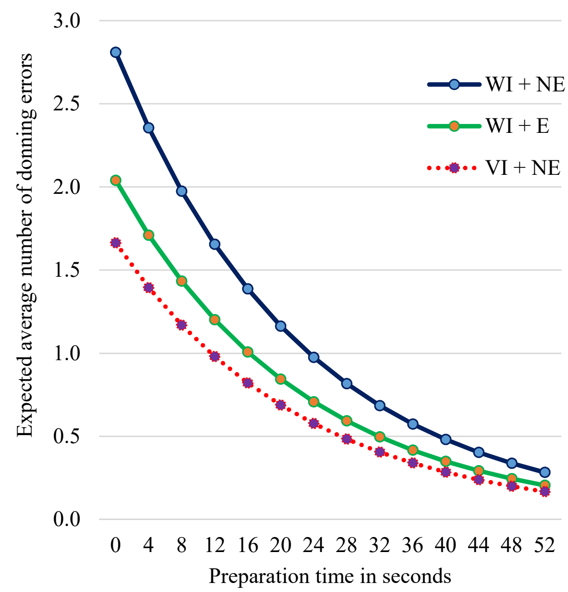


Fig. 8. Predicted average number of donning errors as a function of preparation time for various methods of instruction and experience.

As can be seen from Fig. 8, for a given preparation time those with WI and no experience always produce more errors on average than those with VI and no experience. The WI and experience group fall between the two. To produce an average of one donning error, those with WI and no experience require a preparation time of 23 s, those with WI and experience require on average a preparation time of 16 s, while those with VI and no experience require a mean preparation time of 12 s. When interpreting Fig. 8, it should be noted that the maximum recorded preparation time was approximately 35 s. Furthermore, for the VI group none of the participants has a preparation time of greater than 2 s and so the curve is essentially a model extrapolation (hence shown as a dashed line).

4.3. Factors influencing NDT

As with the donning errors, a variety of parameters such as age, gender, method of instruction, etc may influence donning performance of individuals. Here, we explore the potential impact of the control variables (i.e., age, gender, height, weight, previous donning experience, video instruction) (see *Sec.* 3.1) and a selection of the observed variables (such as taking the shoes off prior to donning, number of donning errors including and excluding the error associated with the taking off the shoes at the beginning, duration of preparation time) on the NDT (response variable) using a log-linear regression model. The purpose of the modelling in this section is not to predict or quantify the NDT for modelling application, but to analyse the postulated impact of all aforementioned factors on donning performance of individuals. More generally, the regression analysis is used to understand the interrelation between the different variables. A recommended expression to predict NDT for modelling application is presented in Section 5.4.

As in the donning errors analysis (see *Sec*. 4.2), the five predictor parameters of preparation time (x_1), method of instruction (x_2), experience (x_3), age (x_4), and gender (x_5) were considered (see Table 4). In addition, the observational variable, 'failure to remove shoes prior to donning' (x_6) was included. This parameter was included in the analysis as it was noted from the video analysis that if participants did not remove their shoes prior to donning the suit, they tended to struggle with their inserted foot getting stuck in the suit's thigh, thereby extending their donning time. If $x_6 = 1$, i.e., YES, then shoes were not removed and if $x_6 = 0$, i.e., NO, then shoes were removed prior to donning.

First, the donning data from the participants without previous experience (59 people) was considered within the regression model. Thus, the parameters investigated were those identified above, excluding the x_3 parameter, and we call this Model 1. As in the donning error analysis, potential interactions between the parameters were also considered.

The regression analysis suggests that gender (x_4) strongly impacts the average NDT, with being female increasing the expected average NDT by 29 %. However, apart from gender (x_4), none of the parameters were found to significantly influence the average NDT (see Table 6) producing a model with an R^2 of 25.5 %. While not significant, age (x_5), preparation time (x_1) and video instruction ($x_2 = 1$) were found to potentially increase the average NDT. The result concerning preparation time and video instruction may appear surprising as it is expected that having more preparation time and having video instruction would better equip the participants to don the suit faster and so we could expect that the average NDT would decrease rather than increase. However, as shown in *Sec*. 4.2, both preparation time and video instruction tend to decrease the number of donning errors. If participants perform fewer donning errors, it is possible that the NDT increases (as suggested by the regression models) as participants correctly undertake all the tasks required to correctly don the suit. The one exception is the donning error associated with (not) removing shoes prior to donning. As stated previously, from analysis of the video footage, participants who did not remove their shoes prior to donning struggled with the donning process increasing their NDT. Thus, as noted in the regression model if $x_6 = 1$ (shoes not removed), the average NDT increases by approximately 16 % (see Table 6). The positive correlation between age (x_5) and expected average NDT is to be expected. The regression analysis suggests that expected average NDT increases by 6 % for every 10-year increase in age. Donning the suit is a strenuous physical activity requiring a certain amount of flexibility and so it is expected that donning time will generally increase with age.

To increase the power of the analysis, the data associated with the experience parameter (x_3) was included in the regression model, which we call Model 2. This increased the number of data points from 59 to 108. With the increased data set, all the parameters, with the exception of preparation time (x_1) are now found to significantly impact the average NDT (see Table 7). Furthermore, the impact of each variable on the expected average NDT has also increased. However, the results show that only experience (x_3) significantly reduces the expected average NDT. Having previous donning experience reduced the expected average NDT by about 17 %. Thus, these results are consistent with earlier findings (Mallam et al., 2012) that suggest that experience tends to reduce the expected donning time. However, unlike the previous studies, the current study has quantified the potential impact of experience on donning time.

All other parameters tended to increase the expected average NDT, being female ($x_4 = 1$) by 33 %, failure to remove shoes ($x_6 = 1$) by 26 %, video instruction ($x_2 = 1$) by 21 %, preparation time (x_1) by 11 % for every 10 s and age (x_5) by 6.6 % for each 10 years increase in age. However, it is noted that preparation time was not found to be significant, with a P-value of 0.07 (see Table 7). Other factors such as height, weight, and cross-product terms representing potential interactions between the parameters were not found to have a significant impact on the NDT.

The resultant log-linear regression model describing the NDT is presented in Eq. (14) and can predict approximately 37 % of the variation in the NDT ($R^2 = 37.0$ %). The parameters in Eq. (14) are defined in Table 7.

$$NDT = 105.2 * 1.011^{x_1} * 1.21^{x_2} * 0.83^{x_3} * 1.33^{x_4} * 1.0064^{x_5} * 1.26^{x_6} * \varepsilon^{\sim} \tag{14}$$

where $\tilde{\varepsilon} \sim Log\text{-}normal.(0, 0.29)$

5. Discussion

The analysis presented in *Sec*. 4 identified the main factors influencing the number of donning errors (ADE, see *Sec*. 4.2) and the net donning time (NDT, see *Sec*. 4.3). In this section the impact of these relations is discussed.

5.1. Factors influencing the number of donning errors

Donning errors (see *Sec*. 3.4) can have a range of detrimental effects on the safety of the person wearing the TPIS. Some donning errors (related to hood, zipper and gloves) can reduce the effectiveness of the thermal protection provided by the suit, reducing the survival time offered by the TPIS. Some donning errors (related to ankle and interior straps) may make it more difficult to walk, increasing the time required to reach a place of safety or potentially causing trips and falls. Some donning errors (related to shoes) may make it more difficult to don the TPIS, reducing the time available to reach a place of safety. Thus, multiple donning errors have an accumulative effect on reducing safety and so should ideally be eliminated completely, or at the very least, reduced in frequency by the population and absolute number incurred by individuals.

Table 6
Model 1: Definition of contributing factors and change in the NDT of the inexperienced participants, given one unit increase in each of the influencing variables (when all other variables are fixed).

Variable	Definition (Unit)	Coefficient	Standard Error of Coefficient	$A_i = e^{a_i}$	Change in the mean NDT per unit increase of x_i	P-value
x_1	Preparation time (seconds)	0.0054	0.0073	1.005	About + 5 % for every 10 s of preparation	0.46
x_2	Method of instruction $x_2 \in \{\text{VI} = 1, \text{WI} = 0\}$	0.13	0.11	1.14	About + 14 % with receiving VI	0.23
x_4	Gender $x_4 \in \{\text{Male} = 0, \text{Female} = 1\}$	0.25	0.076	1.29	About + 29 % longer donning time for females	0.002
x_5	Age $x_5 \in (18 - 72\,\text{year old})$	0.0060	0.0032	1.0060	+6% per every 10-year increase in age	0.07
x_6	Failure to remove shoes prior to donning $x_6 \in \{\text{Yes} = 1, \text{No} = 0\}$	0.15	0.092	1.16	About + 16 % increase in donning time with NOT removing shoes prior to donning	0.11

Table 7
Model 2: Definition of contributing factors and change in the NDT given one unit increase in each of the influencing variables (when all other variables are fixed).

Variable	Definition (Unit)	Coefficient	Standard Error of Coefficient	$A_i = e^{a_i}$	Change in the mean NDT per unit increase of x_i	P-value
x_1	Preparation time (seconds)	0.011	0.0061	1.011	About + 11 % for every 10 s of preparation	0.07
x_2	Method of instruction $x_2 \in \{\mathrm{VI} = 1, \mathrm{WI} = 0\}$	0.19	0.085	1.21	About + 21 % with receiving VI	0.03
x_3	Experience $x_3 \in \{\mathrm{Yes} = 1, \mathrm{No} = 0\}$	−0.19	0.060	0.83	About −17 % reduction in donning time with having previous experience	0.002
x_4	Gender $x_4 \in \{\mathrm{Male} = 0, \mathrm{Female} = 1\}$	0.29	0.060	1.33	About + 33 % longer donning time for females	<0.001
x_5	Age $x_5 \in (18 - 72\ \text{year old})$	0.0064	0.0025	1.0064	+6.6 % per every 10-year increase in age	0.01
x_6	Failure to remove shoes prior to donning $x_6 \in \{\mathrm{Yes} = 1, \mathrm{No} = 0\}$	0.23	0.066	1.26	About + 26 % increase in donning time with NOT removing shoes prior to donning	0.001

From Section 3.4 the number of donning errors incurred by an individual ranged from 1 to 5 with an average of 2.2 for the 108 participants. Clearly, it is desirable to reduce the average number of errors committed during the donning process, and so it is essential to determine the factors that influence donning errors. A Poisson regression model was used to investigate the potential influence of all the background (demographic, experience), randomized (instruction method), and observed (NDT, instruction reading times and extraction times) variables on donning errors.

Results of the Poisson regression suggest that of all the variables considered, three appeared to significantly impact the number of donning errors. These were, preparation time, method of instruction (see *Sec.* 4.3) and previous donning experience, producing P-values of 0.02, 0.009 and 0.03, respectively. Fig. 9 presents the histogram of donning errors for four groups of participants according to the type of instruction they received and their previous experience of donning. As shown in Fig. 9 and supported by Fig. 8, VI appears to be the most effective instruction methodology, producing an average of 1.5 donning errors amongst the 19 participants who had VI, compared to an average of 2.3 donning errors for the 89 participants who had WI. Furthermore, participants who received VI made a maximum of two donning errors while those that received WI made up to five errors (see Fig. 9). Of secondary importance, but still of significance is experience (E). As seen in Fig. 9 and again supported by Fig. 8, E also appears to reduce the number of errors. The average number of donning errors for those with no experience (NE), irrespective of instruction methodology, is 2.3 (59 participants) compared with 1.98 (49 participants) for those with E. Experience has a smaller impact on the propensity to generate errors than method of instruction. For those with WI, the maximum number of errors for those with NE and E decreases from 5 to 4 while the average decreases from 2.5 to 2.1. Similarly, for VI, the maximum remains unchanged at 2 and the average changes from 1.58 to 1.42 for NE and E. Thus, as with the expected donning time (see *Sec.* 4.3) these results are consistent with earlier findings (Mallam et al., 2012; Sanli et al., 2019) that suggest that experience tends to reduce the expected number of donning errors. However, unlike the previous studies, the current study has quantified the potential impact of experience on donning correctness.

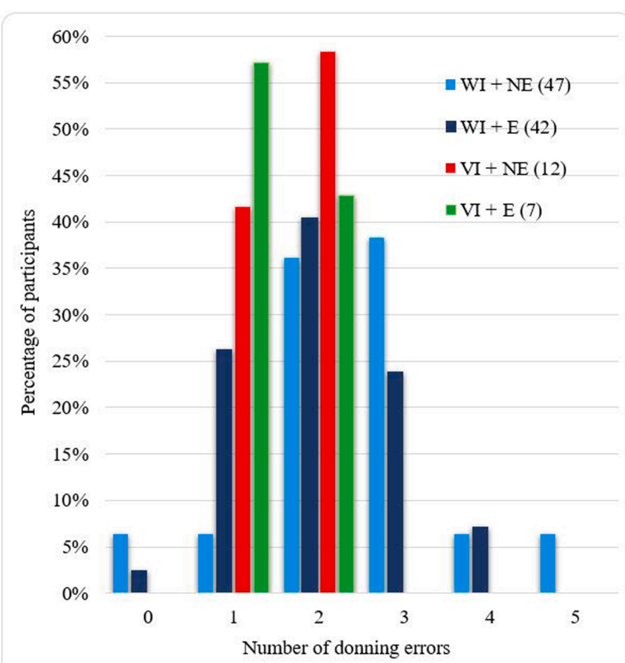


Fig. 9. Histogram of number of donning errors for participants according to instruction method and experience.

It is noted that the impact of experience in this study may be masked by how the experience was defined and measured. The quality, frequency and how recent the experience was gained is not represented in the current study. So, the experience claimed by participants could have been donning a similar TPIS, once 20 years ago, or once in the previous week, or every day throughout a person's sea going career. Furthermore, within this study, all would have been considered equivalent. These factors are likely to influence the effectiveness of the experience but are not considered within this study.

Thus, the impact of experience identified in this study can be considered indicative at best. In reality, experience may have a more profound effect depending on the nature of the experience. This is particularly important when considering utilising the presented data and correlations to represent the performance of mariners/crew regularly trained in donning TPIS. However, one of the prime motivations of this study was to determine the factors that impact donning correctness and time, and clearly experience is an important influential factor for both. However, another motivation of this study was to quantify the expected donning time for the TPIS, and this is further described in Section 5.5.2.

5.2. Importance of the donning error associated with shoe removal

The NDT is a key parameter of interest as the time required to don the TPIS may directly impact the amount of time available for passengers to reach a place of safety. In time critical evacuation situations, the longer it takes to don the TPIS, potentially the shorter is the time available to reach a place of safety. Thus, factors that tend to increase the NDT should be avoided and their number minimised. While most donning errors tend to decrease the NDT – as they usually result in some key donning function not being completed – neglecting to remove shoes prior to donning tends to increase the NDT due to the inherent difficulty of the resulting donning process while wearing shoes. From the video analysis it is known that 50 % of the participants (54 participants out of total 108) failed to remove their shoes prior to the first donning attempt (see Fig. 6). Furthermore, as shown in *Sec.* 4.3, failing to remove shoes prior to donning increases the average NDT by approximately 26 %.

The primary factors that influence this particular donning error are expected to be method of instruction and experience. Analysis of the donning video footage reveals that 100 % (19) of participants with VI and 52 % (22 out of 42) of those with previous experience in the WI group removed their shoes prior to the donning process. However, just approximately 30 % (14 out of 47) of those in the WI group with NE removed their shoes prior to the donning process. It is also expected that the duration of the preparation time is likely to impact whether or not the shoes are removed prior to donning, however, this cannot be determined easily from the basic frequencies. Nevertheless, it is important to identify procedural measures that can reduce the frequency of this donning error, in particular when dealing with those with NE and when relying on written instructions of the type associated with the tested TPIS.

To quantify the impact of background and randomised variables (see Table 4) on the probability of removing shoes (PRS), binary logistic regression (Hoffmann, 2016) was used (see Supplementary Material *Sec.* S4). The analysis reveals that only the preparation time, PT (x_1), method of instruction (x_2) and experience (x_3) were found to be significant. Furthermore, as previously suggested, VI had the most significant influence on PRS while E was the second most significant variable.

The PRS for the WI + NE group is considerably smaller than that for the VI + NE group with PT of up to 20 s. The PRS for the VI + NE group, even with PT = 0 s is 91 %, while that for the WI + NE group is just 22 %. If the PT is increased to 20 s, the PRS for the WI + NE group is just 85 %. While having experience improves the PRS for the WI group, the improvement is small and decreases as PT increases, furthermore, the impact of experience is even smaller for the VI group. Analysis also suggests that to achieve a 95 % probability of removing their shoes, the WI + NE group require a PT of at least 28 s, whereas the VI + NE group only require a PT of 4 s (for further details see Supplementary Material *Sec.* S4).

While the average time spent in the preparation phase may be considered short (2.3 s, see section 4.3), the actual donning instructions are rather short, consisting of only eight short bullet points and associated pictograms (see Fig. 1.b and Supplementary Material *Sec.* S2). It could therefore be argued that it should not require much time to read the instructions. However, the donning instructions are somewhat lost in a large amount of text, consisting of irrelevant text associated with care of the TPIS and various language options. Thus, it can take some time to actually locate the necessary information. Furthermore, the text font size is rather small, making it difficult for many to read. Indeed, based on the responses to the post-trial questionnaire (see Supplementary Material *Sec.* S5), many participants encounter these difficulties and state that they could not read the instructions. The 'extended preparation time' group (see Fig. 4.b) was identified as a sub-set of participants that spent more time in the preparation phase (approximately 4 s to 35 s) and so were more likely to have read the instructions, and thus more likely to note the requirement to remove shoes prior to donning. 71 % (10 out of 14) of the participants in the 'extended preparation time' sub-group took their shoes off prior to donning. This result is almost as good as the VI group, that achieved all 19 participants removing their shoes prior to donning.

The findings suggest that WI can also be an effective approach to providing donning instructions. However, it is essential that the instructions are short, accompanied by clear pictorials, written in large fonts, not combined with 'care' instructions, and simply focus on the essential items. According to the SOLAS (IMO-SOLAS, 2014), ship passengers must undergo a safety drill including assembling at lifeboat stations prior to or immediately following departure. It is likely that as part of the assembly drill passengers will be shown a video of the donning procedures, but it is unlikely that passengers will remember the correct donning procedures when required during an emergency. It is thus essential that clear, simple, short and unambiguous donning instructions are provided with the TPIS packaging.

5.3. NDT and donning errors

As stated in Section 4, NDT and donning errors are key parameters of interest as they each directly impact passenger survivability albeit in different ways. The longer the NDT, the less time is available for passengers to reach a place of safety, and donning errors have a cumulative impact on reducing overall safety by compromising thermal protection and possibly buoyancy. But how is NDT related to the number of donning errors? Presented in Fig. 10 is a graph of mean NDT across all participant control groups (108 participants) grouped according to the number of donning errors. It is clear that a simple direct correlation between number of donning errors and mean NDT does not exist. Furthermore, a Kruskal-Wallis test was used to investigate the impact of donning errors on NDT. The test did not find that NDT was significantly influenced by donning correctness (P-value = 0.49).

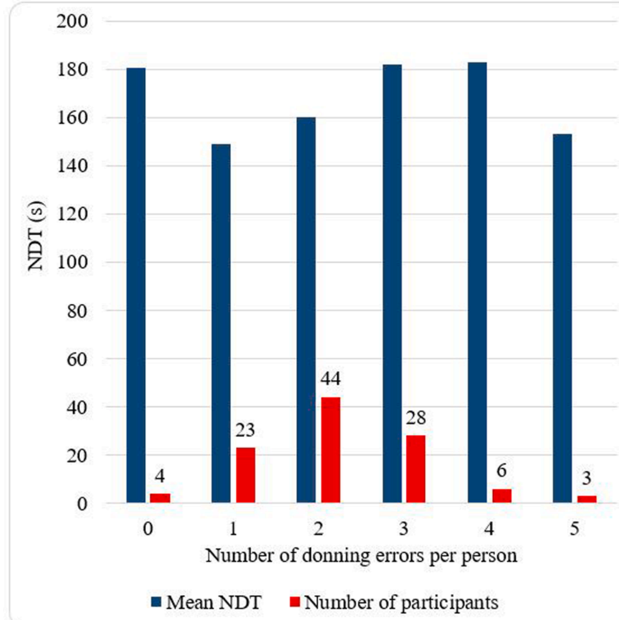


Fig. 10. Mean NDT as a function of number of donning errors.

Intuitively, this result may appear strange. It could be argued that most donning errors tend to decrease NDT through omission, i.e., neglecting to undertake an essential task. However, it is also possible that some participants may struggle with a particular essential task, such as pulling the zipper over the chin, only to eventually give up incurring a donning error while also having wasted time in the attempt, increasing their NDT. Thus, some donning errors may either increase or decrease average donning times, depending on the nature of the individual involved. In contrast, the donning error associated with the failure to remove shoes consistently increases donning time through the increased difficulty incurred in donning while wearing shoes. The donning error associated with shoe removal was the most common error, representing approximately 40 % of all errors committed and was committed by 50 % (54 out of 108) of the participants across all the trials (see *Sec.* 3.4). Thus, the relationship between the number of donning errors and NDT is inherently complex, with some donning errors increasing NDT for some participants and decreasing it for others, while other types of donning error tending to consistently increase NDT. This complex relationship explains the lack of correlation observed in Fig. 10.

5.4. Quantification of donning time for regulatory purposes

Here the quantification of the TPIS donning time is provided to assess regulatory compliance and for proposed use in evacuation simulation analysis used to demonstrate that proposed vessel design and procedures are appropriate.

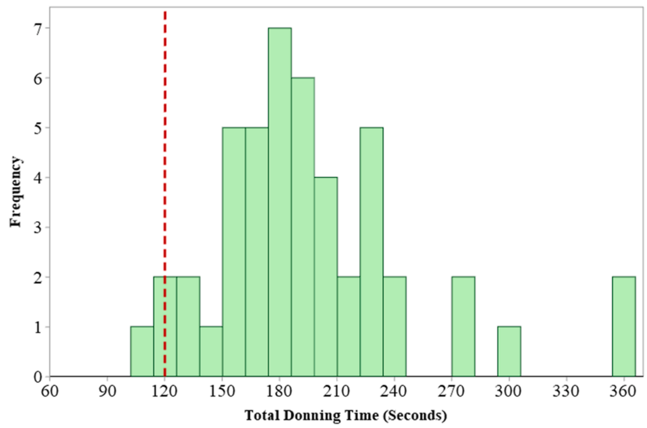


Fig. 11. TDT distribution (WI + NE group) highlighting the 120 s regulatory requirement.

5.4.1. Regulatory compliance

As discussed previously (see *Sec.* 1) it is an IMO regulatory requirement that TPIS can be unpacked and donned without assistance within 120 s (ISO, 2012; IMO-SOLAS, 1998). For the TPIS investigated we compare the TDT (derived from Eq. (5) and Eq. (6)) for inexperienced participants (59 participants) with the requirement. For the XT component in Eq. (5), we use the mean XT (19.4 s) derived from the trials involving the 14 new TPIS. Presented in Fig. 11 is a frequency plot of the 47 TDT's based on the data derived from the trials involving participants exposed to WI and who were inexperienced with the critical 120 s time indicated. As can be seen from Fig. 11, 95.7 % of the participants have a TDT in excess of the maximum permitted donning time of 120 s.

Even if only the NDT is considered (i.e., excluding the preparation time and the extraction time) 89.4 % of the participants fail to don the TPIS within 120 s. Clearly, the TPIS used in this study is not easy to don. This is also supported by the participants responses to the post-trial questionnaire (see Supplementary Material *Sec.* S5) where the majority of female participant responses (38 %) found the TPIS very difficult or difficult to don, while 18 % of the males found it very difficult or difficult to don. However, 80 % of the participants suggested that it would have been easier if there was a live visual demonstration during the donning process while 50 % said it would have been easier had there been some physical assistance during the donning process. While these types of interventions are not permitted during the regulatory assessment of the TPIS, these observations have important implications for the procedures adopted onboard vessels.

5.4.2. Evacuation modelling

The guidelines for evacuation analysis of passenger ships (IMO-MSC/Circ.,1533) specify population parameters that must be used in the evacuation analysis. These include parameters such as passenger response times, passenger deck walking speeds and passenger stair walking speeds. The walking speed data is provided as a function of age and gender. Furthermore, for evacuation modelling applications involving vessels operating in polar waters it may be appropriate to include the time required by passengers to don a TPIS. This could be significant in evacuation analysis as the donning times for TPIS can be up to 120 s (as required by IMO regulation) or more (see Fig. 15). However, currently there are no formulations characterising the donning times for TPIS that can be used in agent-based evacuation modelling applications, apart from simply assuming a uniform 120 s regulatory compliant donning time. To address this limitation, we use the data generated from the donning trials to specify a donning time relationship that can be used in agent-based evacuation modelling.

We define the Total Donning Time ($TDT_{modelling}$) for modelling applications by combining Eq. (1), (2) and (5), to produce,

$$TDT_{modelling} = PT + XT + NDT_{modelling} \tag{15}$$

where $PT = 1 + U * X,\ XT \sim Log - normal(2.9,\ 0.39)$

and $U \sim Bernoulli\ (0.16),\ X \sim Log - normal(2.35,\ 0.56)$

In Eq. (15), $NDT_{modelling}$ is defined by the log-linear regression model derived from the data-set for the group with WI, i.e., involving 89 participants. The data for the participants with VI are excluded in order for the NDT to be representative of the most conservative and likely situation on-board the vessel. The same type of log-linear regression model as in *Sec.* 4.3 is applied, except that the variables method of instruction and preparation time are excluded from the regression model. In this analysis, the background variables, age, gender and previous donning experience appear to have a significant influence on the NDT. Thus, $NDT_{modelling}$ is determined using these predicting factors in a log-linear regression model. The resulting model, as defined by Eq. (16), can predict $R^2 = 27.5\%$ of the expected variance of the NDT,

$$NDT_{modelling} = 130.3 * 1.006^{Age} * 1.32^{Gender} * 0.79^{Experience} * \tilde{\varepsilon}; \tag{16}$$

Where: $\tilde{\varepsilon} \sim Log - normal(0,\ 0.3)$

$$Age \in (18 - 72), Gender \in \{Male = 0, Female = 1\}$$

$$Experience \in \{People\ without\ donning\ experience = 0, People\ with\ donning\ experience = 1\}$$

As can be seen in Eq. (16), previous donning experience is one of the factors that can have a significant impact on the donning time. People with experience can perform approximately 21 % faster than their inexperienced counterparts. However, as noted in Section 5.1, given the vagueness of the definition of experience used in this study, it is suggested that the quantification of donning time for the inexperienced is more reliable and representative of expected performance than the quantification for the experienced. Thus, the predicted NDT for experienced should be used with care as it is likely to underestimate the performance of highly trained personnel such as crew.

Presented in Fig. 12 is the mean TDT as a function of age and gender. The mean TDT is calculated using Eq. (15), setting Experience to zero (excluding the standard deviations) in Eq. (16) and using the mean values for PT and XT. The mean donning time for the male group ranges between 166 s and 218 s, while the female mean donning time ranges between 211 s and 278 s. The donning time of both genders increases approximately 5.7 % for a 10-year increase in age. Furthermore, females

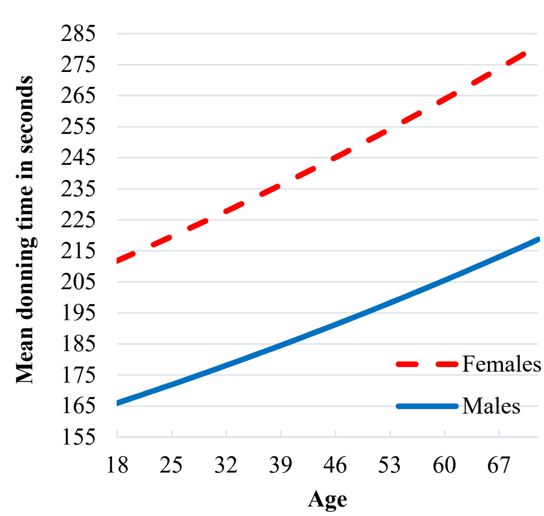


Fig. 12. Mean TDT for males and females as a function of age for people without experience (NE) using Eqs. (15) and (16).

at any age require on average over 32 % longer to don than their male counterparts of the same age. Nevertheless, there will always be a number of females who manage to don the TPIS faster than their male counterparts due to natural variations within each group (i.e., $\widetilde{\varepsilon}$ in the regression model of NDT in Eq. (16) and the σ parameter in the log-normal distributions of PT and XT).

In addition to age, gender and experience, the type and ergonomics of the TPIS design and its packaging will also have an influence on the donning performance of individuals. Thus, the TDT presented in this paper is only intended to be representative of the particular TPIS examined. Furthermore, from analysis of the behaviour of participants during the donning trials and reflecting on their comments in the questionnaires, certain aspects of the TPIS design could be modified to improve donning performance. These aspects are discussed in the Supplementary Material (see Supplementary Material *Sec.* S5).

6. Limitations

As with any experimental study involving human test subjects, there are limitations associated with this work which should be considered when reviewing the results. The limitations of the current study are identified as follows:

- In order to conduct the research in an ethical manner and to reduce the risk of injury to the participants, the experiment was conducted in a controlled environment and experimental protocol eliminated some factors, such as stress, darkness, slippery surfaces, dynamic motion, adverse deck orientation, etc. which could have a detrimental impact on donning performance of individuals. Indeed, in the post-trial questionnaire, 95 % of the participants suggested that their donning times would be adversely impacted by dynamic motions or adverse deck orientation, with 48 % suggesting that their donning times could be doubled under such circumstances (see Supplementary Material *Sec.* S5).
- The TPIS used in the trials were new or in as good as new condition and perfect working order. In actual applications, it is assumed that the TPIS used by passengers will be well maintained and in good working condition as required by international regulations.
- While the physical space available to the participants during the trials was representative of the floor area per passenger required by international regulation, it is possible that in actual emergency situations, passengers may be in environments with less physical space. This may make donning more difficult.
- Prior to the start of the trial, participants were instructed to remove excessive clothing such as winter jackets, scarves or heavy jumpers. In reality, such extra warm clothing may be worn by passengers in real situations, making donning of the TPIS more difficult. Furthermore, in real situations, passengers may be instructed to remove such clothing prior to donning of the TPIS, increasing the number of preparation tasks and hence increasing the preparation time (PT) and hence the total donning time (TDT).
- All trial participants (who were aged from 18 to 72 years old) were in good health and physical condition. Almost 60 % of the participants were under the age of 20 years, with just 13 % of the participants being over the age of 51 years. Furthermore, the average Body Mass Index (BMI) for male and female participants was 25 (SD = 4) and 24 (SD = 3), respectively. The majority of participants in the trial were within the normal BMI range with none of the participants in the obese category. It is noted that in the UK and USA 27 % and 38 %, respectively, of the population are classified as obese (Gallagher et al., 2000). Thus, the sample population used in the trials may not be considered fully representative of the target population. While further research is required to include a wider cross-section of the public, the donning times measured in these trials may be considered to be optimistic.
- The video instruction was shown to the participants immediately prior to their participation in the trial. In real life situations, passengers may have viewed the video or undertaken an emergency assembly drill hours or days before the time of the actual emergency. Furthermore, during an actual emergency, passengers may not patiently concentrate and watch an instructional video prior to the donning. Further investigation is required to understand the impact of the duration of the time interval between receiving video instruction and actual donning on donning performance.
- For the participants in the written instruction group (WI), it is assumed that participants devote some time during the preparation phase to reading the donning instructions. However, in a real situation, it is possible that passengers may spend less (or more) time reading the instructions and so this may impact the NDT and the TDT.
- Concerning the validity of the statistical analyses, the participants performed the donning procedure in groups of up to 15 persons at the same time. Hence the individual donning errors (and donning time) could potentially have been influenced by the performance of the other participants in the same group. The statistical tests that have been performed is usually underpinned by that the individual samples are independent, and some caution should therefore be taken when considering the analysis of the factors that influence donning errors in particular.

7. Conclusion

Thermal Protection Immersion Suits (TPIS) are required by the IMO for all vessels operating in polar waters and must be designed so that they can be donned, unaided within 120 s. To meet this requirement, TPIS are typically designed as a universal one-size fits all. The one-size fits all approach has the advantage of reducing the time required to distribute the TPIS and the inevitable disadvantage of impacting the donning time, walking performance and general manoeuvrability of those individuals who are either very large or small in stature. The aim of this study was to explore the factors influencing the donning speed and correctness through an experimental trial involving 108 volunteers (71 males and 37 females) aged between 18 and 72 years old.

A key finding of this work is that the mean net donning time (NDT) was dependent on a complex relationship between, age (increases by 6.6 % for each 10 years), gender (increases by 33 % if female), experience (decreases by 17 % with experience), method of instruction (increases by 21 % with video instruction) and failure to remove shoes prior to commencing the donning process (increases by 26 %). The study is unique in that it identifies and quantifies, for the first time, the factors that influence donning time for the type of TPIS used in this study. This is important to ship operators as unnecessarily prolonging the time involved in donning the TPIS may mean that less time is available to safely abandon the vessel. With the insight that this information provides, ship operators can develop procedures to minimise the time required to don the TPIS. The information is also important to TPIS designers and manufacturers, as it identifies design issues that make it difficult to quickly don the TPIS.

Perhaps of greater importance than the donning speed is the donning correctness and the factors that influence correctness. Clearly, a TPIS that is incorrectly donned will impact life critical issues such as thermal protection and buoyancy. Thus, a further unique aspect of this study is that it identified that the number of donning errors is significantly impacted by the method of instruction, with video instruction (VI) producing an average of 1.5 errors while written instruction (WI) producing 2.3. This finding is again important to both ship operators and TPIS designers and manufacturers. For example, as part of the ship abandonment procedures, showing a live donning demonstration or playing a video of the donning process during an actual emergency may be more effective at reducing donning errors than relying on passengers to read the donning instructions. Nevertheless, project findings also suggest that WI can also be an effective approach to reducing the number of donning errors. However, it is essential that suit manufacturers provide instructions that are short, accompanied by clear pictorials, written in large fonts, not combined with 'care' instructions, and simply focus on the essential items.

In reviewing these results, it is important to note the limitations associated with the study. In particular, the study focuses on only a single type of TPIS, other suits may have different characteristics. This aspect is currently being examined in a related study involving a different type of TPIS. The majority (60 %) of the participants where under the age of 20 years, with just 13 % being over the age of 51 years and none of the participants being classed as obese. Furthermore, the trials were conducted in ideal environmental conditions. These limitations were a combination of practical and ethical considerations, the latter being intended to reduce the risk of injury to the participants.

Furthermore, it is interesting to note that the preparation time was not found to have a significant influence on NDT and a significant but weak influence on the expected number of donning errors. The later relationship was further weakened, to the point of insignificance, if the maximum preparation time data point was considered an outlier and removed from the analysis. Thus, it is suggested that the important relationship between preparation time and both NDT and number of donning errors requires further analysis.

The final key result, addressing an important aim of the paper was the specification of a donning time distribution that can be used in agent-based passenger ship evacuation analysis. Passenger ship evacuation analysis using modelling techniques, as required by IMO for all new builds, currently does not represent the time required by passengers to don the TPIS. As the time required to don the TPIS is a critical factor identified in the IMO (SOLAS), it is reasonable to assume that it may also be an important factor in evacuation analysis for passenger ships intended to operate in polar waters. The donning time distribution suitable for modelling analysis defined in this work allows engineers to assess whether the time required to don the TPIS critically impacts the evacuation process, and if it does, enables them to refine procedures to reduce the impact. This latter point is currently being pursued by the authors in a continuing study.

CRediT authorship contribution statement

Hooshyar Azizpour: Methodology, Investigation, Resources, Data Curation, Formal Analysis, Writing - original draft, writing - review & editing. **Edwin R. Galea:** Supervision, Methodology, Investigation, Formal analysis, Conceptualization, Writing - review & editing. **Sveinung Erland:** Supervision, Formal analysis, Writing - review & editing. **Bjørn-Morten Batalden:** Supervision, Resources, Project administration, Methodology, Investigation, Writing - review & editing. **Steven Deere:** Methodology, Investigation, Formal analysis. **Helle Oltedal:** Supervision, Investigation, Funding acquisition, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

The authors would like to express their deepest appreciation to all those who contributed to this study. This work could never have been accomplished without the financial support from MARKOM-2020 (T92). Special thanks also goes to Viking Life-Saving Equipment for providing the survival suits used in the study. Further, we acknowledge with great appreciation, the invaluable support of the ARCOS and ResQ safety centres who provided access to invaluable facilities to conduct the experiments and to the staff of HVL and UiT who assisted with the safe and efficient running of the experiments. Finally, we are indebted to the 108 volunteers who freely gave their time to improve maritime safety.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ssci.2023.106064.

References

Andreassen, N., Borch, O.J., Sydnes, A.K., 2020. Information sharing and emergency response coordination. Safety Science 130.

Azizpour, H., Galea, E.R, Erland, S., Batalden, B.M., Deere, S., Oltedal, H., 2022. An experimental analysis of the impact of thermal protective immersion suit and angle of heel on individual walking speeds. Safety Science 152, 105621.

Bles, W., Nooij, S., Boer, L., Sharma, S.S., 2002. Influence of ship listing and ship motion on walking speed. In: International Conference on Pedestrian and Evacuation Dynamics 2001. Springer, pp. 437–452.

Brown, R., Boone, J., Small, G., MacKinnon, S., Igloliorte, G., Carran, A., 2008. Understanding passenger ship evacuation through full-scale human performance trials. International Conference on Offshore Mechanics and Arctic Engineering 48197, 645–650.

Brünig, R., Galea, E.R., Batalden, B.M., Oltedal, H.A., 2021. A methodology for collecting donning times of thermal protective immersion suits intended to be worn by passengers on vessels operating in cold environments vol. 1201, no. 1.

Galea, E.R., 2000. Safer by design: Using computer simulation to predict the evacuation performance of passenger ships. The Institute of Marine Engineers 112 (2), 7–16.

Galea, E.R., Deere, S., Brown, R., Filippidis, L., 2013. Two Evacuation Model Validation Data-sets for Large Passenger Ships. Journal of Ship Research 57 (3), 155–170.

Galea, E.R., Filippidis, L., Lawrence, P., Owen, M., "An evacuation demonstration of a typical high speed craft using the EXODUS evacuation model," Report prepared by FSEG for RINA, Ref G/DG/1998/000116, London, 1998.

Galea, E.R., Filippidis, L., Gwynne, S.L., Lawrence, P., Sharp, G., Blackshields, D., Glen, L., October 2002. The development of an advanced ship evacuation simulation software product and associated large scale testing facility for the collection of human shipboard behaviour data," The Royal Institution of Naval Architects (RINA). International Conference. Human Factors in Ship Design and Operation 2-3.

Galea, E.R., Owen, M., 1994. Predicting the evacuation performance of mass transport vehicles using computer models. IMarE Conference 106, no. 2.

Gallagher, D., Heymsfield, S.B., Heo, M., Jebb, S.A., Murgatroyd, P.R., Sakamoto, Y., 2000. Healthy percentage body fat ranges: an approach for developing guidelines based on body mass index. The American Journal of Clinical Nutrition 72 (3), 694–701.

Glen, L., Lgloliorte, G., Galea, E.R., Gautier, C., 2003. Experimental determination of passenger behaviour in ship evacuations in support of advanced evacuation simulation. In: International Conference on Passenger Ship Safety. Royal Institution of Naval Architects (RINA, London, pp. 129–138.

Gwynne, S., Galea, E.R., Lyster, C., Glen, I., 2003. Analysing the evacuation procedures employed on a Thames passenger boat using the maritimeEXODUS evacuation model. Fire Technology 39 (3), 225–246.

Hoffmann, J.P., 2016. Regression models for categorical, count, and related variables: An applied approach. Univ of California Press.

15027-3-Immersion Suits Test Methods, ISO, 2012.

Interim guidelines for evacuation analyses for new and existing passenger ships (MSC/Circ. 1033), IMO, London, 2002.

International Convention for the Safety of Life at Sea (SOLAS): Revised recommendations on testing of Life-Saving Appliences , ANNEX 6, Resolution MSC.81(70), IMO, London, 1998.

International Convention for the Safety of Life at Sea (SOLAS), IMO (amendment), 2014.

Khan, B., Khan, F., Veitch, B., 2020. A Dynamic Bayesian Network model for ship-ice collision risk in the Arctic waters. Safety Science 130.

Levine, D.M., Ramsey, P.P., Smidt, R.K., 2001. Applied statistics for engineers and scientists: using Microsoft Excel and Minitab. Pearson.

Luck, M., Maher, P.T., Stewart, E.J., 2010. Cruise tourism in polar regions: promoting environmental and social sustainability?. Routledge.

MacDonald, C.V., Brooks, C., Kozey, J., Habib, A., 2011. An ergonomic evaluation of infant life jackets: Donning time & donning accuracy. Applied Ergonomics 42 (2), 314–320.

Maher, P.T., 2017. Tourism futures in the Arctic. In: The Interconnected Arctic Congress. Springer, Cham, pp. 213–220.

Mallam, S.C., Small, G., MacKinnon, S., 2012. Donning time of marine abandonment immersion suits under simulated evacuation conditions. The Journal of Ocean Technology 7 (3), 45–59.

Mallam, S.C., Small, G., MacKinnon, S., 2014. "Immersion suit donning in dynamic environments: Implications for design, construction & use," TransNav: International Journal on Marine Navigation and Safety of Sea. Transportation 8 (3), 429–437.

McGraw, K.O., Wong, S.P., 1996. Forming inferences about some intraclass correlation coefficients. Psychological methods 1 (1), 30.

McHugh, M.L., 2012. Interrater reliability: the kappa statistic. Biochemia Medica 22 (3), 276–282.

Polar Code: International code for ships operating in polar waters, MEPC 68/21/Add.1 Annex 10, 2017, IMO, 2017.

Pradillon, J., 2004. ODIGO-modelling and simulating crowd movement onboard ships. In: 3rd International Conference on Computer and IT Applications in the Maritime Industries. Siguenza, Spain, pp. 278–289.

Revised guidelines on evacuation analysis for new and existing passenger ships (MSC/ Circ. 1533), IMO, London, London 2016.

Sanli, E., Ennis, K.A., Brown, R., Eickmeier, C., Carnahan, H., 2019. Forgetting of marine emergency duties tasks: Predictors of relearning. Safety Science 120, 492–497.

Stephens, M.A., 1986. Tests based on EDF statistics. In: D'Agostino, R.B., Stephens, M.A. (Eds.), Goodness-of Fit Techniques. Marcel Dekker Inc., pp. 97–194

Vassalos, D., Kim, H., Christiansen, G., Majumder, J. "A mesoscopic model for passenger evacuation in a virtual ship-sea environment and performance-based evaluation," 2002.