Supplementary Material for: Passenger donning time and donning correctness for a non-insulated immersion suit – an experimental study

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

[a] Department of mechanical engineering and maritime studies, Western Norway University of Applied Sciences (HVL), Norway

[b] Department of Technology and Safety, UiT The Arctic University of Norway, Norway

[c] Fire Safety Engineering Group, University of Greenwich, United Kingdom

*Corresponding author: Ria Brünig (ria.bruenig@hvl.no)

List of Abbreviations

Abbreviation	Meaning
AG	Age Group
CF	Correction Factor
DE	Donning Error
DP	Donning Phase
DT	Donning Time
Dt_{new}	DT for participants with new suit
Dt_{used}	DT for participants with used suit
E	Experienced
IMO	International Maritime Organization
ISO	International Organization for Standardization
LSA	Life-saving appliances
NE	Not Experienced
NI	No Instruction
PDF	Probability density function
PP	Preparation Phase
PT	Preparation Time
PXP	Combined Preparation and Extraction Phase
PXT	Combined Preparation and Extraction Time
SM	Supplementary Material
SOLAS	Safety of Life at Sea
t_{de}	End donning process
t_{ds}	Start donning process
TDT	Total Donning Time
t_e	End trial time

Abbreviation	Meaning
t_{e_bz1}	End trial time for broken zipper 1
t_{e_bz2}	End trial time for broken zipper 2
t_{pe}	End preparation process
TPIS	Thermal Protective Immersion Suit
t_{ps}	Start preparation process
t_{pxe}	End combined preparation and extraction time
t_{pxs}	Start combined preparation and extraction time
t_s	Start trial time
t_{xe}	End extraction process
t_{xs}	Start extraction process
VI	Video Instruction
WI	Written Instruction
XP	Extraction Phase
XT	Extraction Time

This document presents supplementary material (SM) for Bruenig, et al. [1], relating to the experimental donning trials concerning the experimental methodology employed in the study, the statistical analysis, the response to the participant questionnaire, and a comparison of trial methodology and results with the related work by Azizpour et al. [2] for a different type of Thermal Protective Immersion Suit (TPIS).

The donning trials discussed in the SM were conducted at three shore-based facilities in Norway: at the ARCOS safety centre in Tromsø, and at the Faculty of Science at UiT The Arctic University of Norway in Tromsø from 18/08/2018 to 23/08/2018, and at the ResQ safety centre in Haugesund from 10/04/2019 to 23/08/2019. The experimental methodology employed in the collection of the TPIS donning data has previously been described and discussed in detail [3] and briefly in [1]. The TPIS used in these trials was a lightweight, non-insulated immersion suit without integrated buoyancy, produced by Hansen Protection (Sea Pass passenger suit) [1], [3]. The suit is complete with integrated hood, gloves, and foot coverings. It is designed for use onboard passenger vessels or recreational vessels, and it is approved as a non-insulated immersion suit in accordance with SOLAS' LSA Code Med/1.5a [4] to be used with additional approved lifejacket.

S1: Supporting data for experimental methodology

S1.1 Inter-rater analysis

To support consistency in data interpretation, a data dictionary was developed outlining key definitions, such as start and end times of various phases (see Sec. S2). To ensure homogeneity, the inter-rater reliability was tested. Twenty participants from varying cohorts, that displayed a broad variety of behaviour and performance variables, were chosen and their data extracted. The same twenty participants were then given to another researcher to extract the data based on the same definitions. The Intraclass Correlation Coefficient (ICC) and Cohen's Kappa were calculated for interval and ratio (e.g. study count, total duration, etc.), and nominal (e.g. stance, shoes status, ankle strap fastened, etc.) measurements, respectively. ICC estimates and their 95% confident intervals were calculated using SPSS based on absolute agreement, 2-way mixed-effects model. An excellent level of agreement on average (ICC = 0.918) was found. Cohen's Kappa was calculated using SPSS for the nominal measurements. The interrater reliability for the raters was found to be strong (Kappa = 0.830) on average.

S1.2 Independence of trial locations

Donning data for participants in the NI group were collected in three different locations (see [3] for details). A Kruskal-Wallis test was performed to assess a possible influence of location on the donning time (DT) collected for participants using a used suit (DT_{used}). Similarly, a Kruskal-Wallis test was used to assess a possible influence of location on number of donning error (DE) for participants that had no instructions available (NI group). There was no indication from any of the two tests that the trial location had an influence on DT_{used} ($H(2,61)=0.339$, $p=0.84$), nor DE ($H(2,49)=3.421$, $p=0.18$). Thus, the data from all three locations were merged into a single dataset.

S.1.3 Testing environment and participant setup

As mentioned in [1] and described in [3], the donning trials were part of the ARCEVAC project, which collected performance data of two different types of TPIS, a heavy TPIS with integrated buoyancy and a light-weight TPIS (see [2], [3], respectively for more details). The trials were executed at three different locations in Norway: at the ARCOS safety centre in Tromsø, at the Faculty of Science at UiT The Arctic University of Norway in Tromsø, and at the ResQ safety centre in Haugesund. The donning trials at the two locations in Tromsø took place simultaneously over a duration of 6 days. Some participants were at both locations and donned one of the two TPIS, that were used in this project, at each location. It was carefully recorded, which TPIS had been donned, so that the participants would don each type of TPIS only once.

The donning took place inside on a smooth flat floor. A square of 3m x 3m was marked on the floor. Figure SM 1 shows a picture of the three trial locations. The size of the various cohorts varied from 2 to 15 participants and so the minimum floor space available for participants to don the TPIS was $0.6m^2$. For comparison, the minimum required deck space for assembly areas is $0.35m^2$pp ([5], Chapter III, Part B, Regulation 25.2).

Figure SM 1 Trial locations from left to right: ARCOS centre, Faculty of Science at UiT, and ResQ centre

Figure SM 2 illustrates the distribution of cohorts by the available space per participant in the cohort and the number of participants per cohort. Within each cohort, participants were assigned to either the lightweight or heavy TPIS. This allocation was based on multiple factors, including whether participants had previously taken part in the trials at another location, as well as their age group and gender, to maintain balance across sub-groups.

Participants were able to observe each other during the donning task, which may have influenced individual performance. Additionally, variations in the space available to each cohort could also have impacted performance. However, both the ability to observe others and the varying spatial conditions reflect realistic conditions that may occur during actual evacuations on board ships. For this reason, these factors were not controlled for in the study.

Furthermore, some trials included mixed-instruction groups due to logistical constraints related to suit availability and scheduling. Specifically, two cohorts comprised a total of ten participants who received different types of instructions within the same trial. One of these trials included seven participants distributed across the NI (no instruction), WI (written instruction), and VI (video instruction) conditions, while the other involved three participants who received either WI or VI. While these mixed groupings could potentially introduce variability in group dynamics, the number of affected participants was small, and no systematic effects were observed.

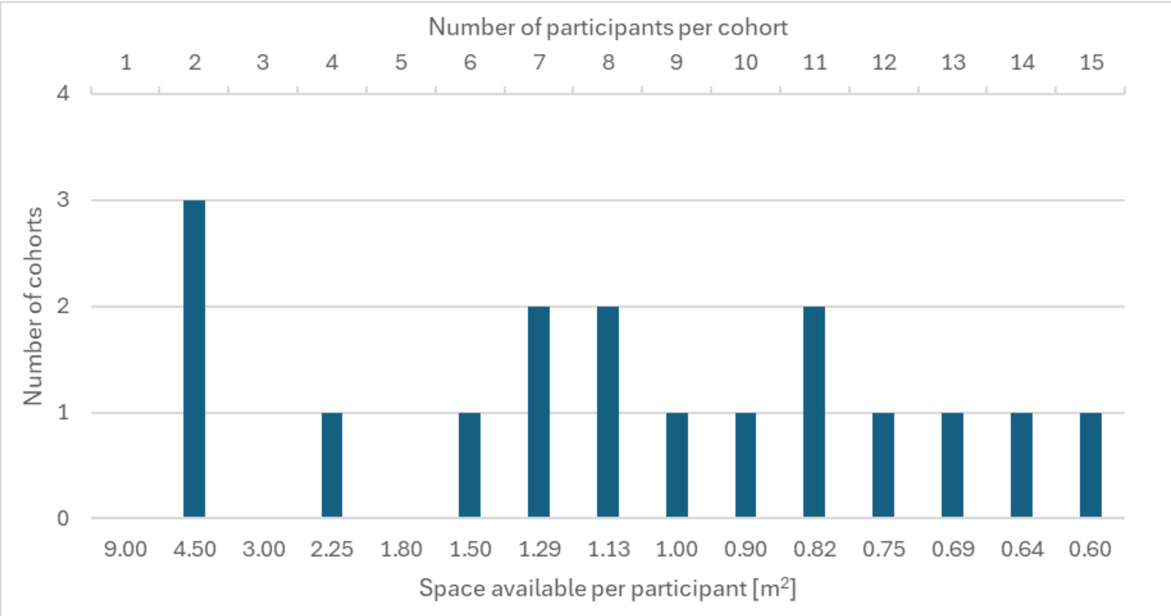


Figure SM 2 Distribution of cohorts by the available space per participant and the number of participants per cohort

S2: Definition of key time intervals (Data Dictionary)

To determine the key time intervals defining the donning process, the timing of the key events is derived from the video analysis. The process by which the timings of these events were reliably and consistently extracted from video footage relied on the specification of a data dictionary that precisely defined these key events. The precise definition of key timed events and parameters as defined in the data dictionary are as follows:

(1) Start/End Trial:

Start trial time t_s: Time at which the 'GO' command is heard on audio track. This is the time for event 'a' in Figure 2 in [1] (see Start Trial in Table 2 in [1]).

End trial time t_e: Time at which the participant raised both hands above their head. This is precisely defined when the hand reached the highest level and comes to a stop. If participants did not raise their arms, another meaningful end-time was determined, and its explanation noted. This is the time for event 'd' in Figure 2 in [1] (see End Trial in Table 2 in [1]).

(2) Preparation phase (PP):

This is the time period between the start of the trial and the start of the XP. During the PP the participant may undertake a number of activities such as pick up the TPIS package, study the instructions on the package, etc. The PP is only applicable for participants with a new suit. Furthermore, some participants with a new suit began extracting the suit while undertaking preparation tasks and so a distinct PP could not be defined (see method 2 in Figure 2 in [1] and Table 2 in [1]).

Start preparation process t_{ps}: The start was marked by the "GO" command. By definition, $t_{ps} = t_s$. This is the time for event 'a' in Figure 2 in [1] (see Start Trial in Table 2 in [1]).

End preparation process t_{pe}**:** Time at which the participant disengaged from preparation activities and begins to attempt to tear open the packaging. This is the time for event 'b' in Figure 2 in [1] (depicted by End PP in Table 2 in [1]).

(3) Extraction phase (XP):

This is the period during which the TPIS is removed from the packaging and is ready to be donned. During the XP participants do not stop to study the instructions or undertake any other task. As with the PP, the XP is only applicable for participants with new suit. However, some participants with new suit began extracting the suit while undertaking preparation activities and so a distinct XP could not be defined (see method 2 in Figure 2 in [1] and Table 2 in [1]).

Start extraction process t_{xs}: Time at which the participant begins to attempt to tear open the packaging. By definition, $t_{xs} = t_{pe}$. This is the time for event 'b' in Figure 2 in [1] (depicted by End PP in Table 2 in [1]).

End extraction process t_{xe}**:** Time at which the participant has fully extracted the TPIS from the packaging. This is the event 'c' in Figure 2 in [1] (depicted by End XP in Table 2 in [1]).

(4) Combined preparation and extraction phase (PXP):

Some participants began to open the packaging while still engaged in preparation activities. Thus, for these participants it is not possible to define a distinct PP and XP (see method 2 in Figure 2 in [1] and Table 2 in [1]). Thus, the PXP phase is a combination of the PP and the XP and can be defined for all participants with new suits.

Start combined preparation and extraction process t_{pxs}: The start is marked by the "GO" command. By definition, $t_{pxs} = t_s$. This is the time for event 'a' in Figure 2 in [1] (see Start Trial in Table 2 in [1]).

End combined preparation and extraction process t_{pxe}**:** Time at which the participant has fully extracted the TPIS from the packaging. This is the event 'c' in Figure 2 in [1] (depicted by End XP in Table 2 in [1]).

(5) Donning phase (DP):

This is the period during which the TPIS is donned. While the end of the DP is the same for all participants, and is marked by the participants raising their arms, the start of the DP is marked by different events for those with new or used suits.

For participants donning a new suit:

Start donning process t_{ds}: The start is marked by the end of the XP. By definition, $t_{ds} = t_{xe}$ for method 1 and $t_{ds} = t_{pxe}$ for method 2 This is the time for event 'c' in Figure 2 in [1] (see End XP in Table 2 in [1]).

End donning process t_{de}: The end is marked by participants raising their arms signifying the end of the trial. By definition, $t_{de} = t_e$. This is the time for event 'd' in Figure 2 in [1] (see End DP in Table 2 in [1]).

For participants donning a used suit:

Start donning process t_{ds}: The start is marked by the time at which the 'GO' command is heard on audio track. By definition, $t_{ds} = t_s$. This is the time for event 'a' in Figure 2 in [1] (see Start Trial in Table 2 in [1]).

End donning process t_{de}: The end is marked by participants raising their arms signifying the end of the trial. By definition, $t_{de} = t_e$. This is the time for event 'd' in Figure 2 in [1] (see End DP in Table 2 in [1]).

Having defined these parameters, it is possible to determine the time required by each participant to undertake each of the various phases. During the PP, 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 prior to attempting to don the TPIS (see Figure 1 in [1] and [3] for details). For each participant, the duration of the PP, identified as the preparation time (PT), is given by,

$$PT = t_{pe} - t_s \tag{S1}$$

The time required to extract the TPIS from the packaging (XT) for each participant is defined as the time interval from the start of the extraction phase to the end of the extraction phase and is given by,

$$XT = t_{xe} - t_{pe} \tag{S2}$$

The duration of the combined preparation and extraction phase (PXT) is defined as the time interval from the start of the trial to the end of the extraction phase and is given by,

$$PXT = t_{xe} - t_s \tag{S3}$$

The donning time (DT) for a participant with a used suit (i.e., a TPIS not in packaging) (DT_{used}) is defined as the time interval from the trial start time (t_s) to their trial end time (t_e) and is given by,

$$DT_{used} = t_e - t_s \tag{S4}$$

The DT for a participant with a new suit, i.e., a TPIS in packaging (DT_{new}) is defined as the time interval from the end of their extraction phase (t_{xe}) to their trial end time (t_e) and is given by,

$$DT_{new} = t_e - t_{xe} \tag{S5}$$

The DT_{used} can be determined for 61 participants and DT_{new} can be determined for 22 participants. It is noted that while there were 68 participants with a used suit, 19 of these had a TPIS with a broken zipper and as a result the data from seven of these participants could not be used. Similarly, while there were 23 participants with a new suit, DT_{new} data from one could not be used due to a broken zipper. Furthermore, it is noted that there is a small inconsistency in the definition of DT_{new} and DT_{used}. The start of the DT for a new suit is measured from when the TPIS is extracted and in the hands of the participant (i.e., event 'c' in Figure 2 in [1]) whereas for the used suit, it is measured from the 'GO' time (i.e., event 'a' in Figure 2 in [1]). Thus, for participants with a used suit, there is the additional time incorporated into the DT equivalent to the time required for the participant to bend down and pick up the folded suit at the start of the donning process. This time has been measured for several participants (N=6) of differing age and gender and was found to be less than 1.6 s and so is considered negligible.

The DT is a measure of the time required to don the TPIS, and so is an inherent measure of the ease or difficulty associated with putting on the suit. This is a useful measure as it allows for the comparison of one TPIS design with another. Furthermore, the XT is a measure of the ease of opening the packaging and so enables the comparison of various packaging designs, while the PT is a measure of the time participants spend in preparing to don the TPIS, which can include studying the instructions.

However, for regulatory applications, the Total Donning Time (TDT) is required. The TDT is a measure of the time required for the entire process of preparation, extraction, and donning. Thus, the TDT is defined as,

$$TDT = PXT + DT \tag{S6}$$

The DT is given by Eq. S4 and Eq. S5 for used and new suits, respectively. PXT is given by Eq. S3 and can be derived for the 23 participants donning a new TPIS. However, a PXT could not be measured for participants donning a used TPIS, and so to determine the TDT for a used suit an approximation for PXT must be used. Depending on the nature of the intended application, PXT can be represented by the mean of the measured PXT distribution for the 23 new suits, or the maximum PXT or values can be randomly selected from the PXT distribution.

In addition to the timed events associated with donning the TPIS, behavioural data such as DEs was quantified and recorded in the form of binary and nominal variables (see Sec. 2.2 in [1], and [3]).

S3: Number of usable data points and donning time correction factor

A total of 25 new TPIS in vacuum sealed packaging were available at the start of the experiment. One of these was used in the preparation of the donning demonstration video, reducing the number of available TPIS for the new suit trials to 24. Furthermore, a video malfunction during the new suit trials reduced this to 23 participants, that were recorded donning a new TPIS. That participant was part of the group with written instructions (WI group). In addition, another 72 participants donned a used suit, of which four participants did not clearly specify their level of prior donning experience. Due to these essential missing data, they were excluded from the DT and DE analysis. Thus, a total of 96 participants donned the TPIS during the trials, of which five were excluded in the DT and DE analysis due to missing data, resulting in 91 participants for the DT and DE analysis. Data from the five excluded participants is included in the analysis of the post-trial questionnaires, where appropriate.

Furthermore, it became apparent during the trials that the zipper was of poor quality, and a substantial number of zippers malfunctioned and broke (see Sec. 2.6 in [1]). However, given that the number of TPIS were limited (25 available in total), it was necessary to make use of the data from the suits with broken zippers. In total 10 zippers (representing 40% of the available TPIS) broke during the donning process and a further 11 participants were required to don a TPIS with a broken zipper. Thus, of the 91 participants that donned the TPIS, 21 (23%) involved a broken zipper.

The reaction of individual participants when the zipper broke during the donning process or when donning a TPIS with a broken zipper varied widely. For example, some participants tried to fix and close the zipper for some time until this was either interrupted by the trial manager or the participant gave up and moved on themselves. Other participants rapidly made eye contact with the trial manager, who then instructed the participant to ignore the broken zipper and to continue with the donning process. However, in order to not lose the data points, it was judged appropriate to include participants with a broken zipper in the DT analysis, that had performed all other donning tasks prior the attempt to close the zipper, and to add an appropriate correction factor to represent the time required to close the zip. In total 13 (62%) of the 21 participants with broken zippers were included in the DT analysis. The correction factor was determined from the analysis of the video footage. However, all participants with broken zippers were excluded from the analysis of donning correctness as the zipper end-status for these participants could not be determined. Thus, a dataset comprising of 83 data points is available for the DT analysis while a dataset comprising of 70 data points is available for the DE analysis.

The failure of the zipper was related to the slider and consisted of several failure types, including slider completely missing, slider attached to one row of teeth, and slider attached to both rows of teeth but not closing when pulled up. In all cases, the reaction of the participants to the broken zipper varied widely, ranging from trying to fix and close the zipper for some time, to contacting a researcher and receiving instruction to ignore it or to end the DP, or the participant moved on themselves after some time. The varying reactions influenced the determination of the end time t_e. To make the time measurement more coherent, two new end times for the DP were defined:

End trial time for broken zipper 1 t_{e_bz1}: Applicable only for participants that donned a TPIS with a zipper that broke during the DP. It is defined as the time at which the zipper breaks, with the proviso that all other donning tasks were performed prior the attempt to close the zipper. Note, this does not mean that the participant achieved a correct final stage of all key donning features, but it means that the participant did not engage in any other donning task afterwards. Participants engaging in donning tasks after attempting to close the zipper were excluded from the dataset – in total, donning data from three participants were therefore rejected as a result of the broken zipper. It should be kept in mind that closing the zipper is often one of the last tasks when completing all donning tasks. End trial time for broken zipper 1 could be determined for seven participants.

End trial time for broken zipper 2 t_{e_bz2}: Applicable only for participants that donned a TPIS with a zipper that was broken before the trial start time t_s. It is defined as the time at which the participant touches the zipper in the attempt to close it, with the proviso that the participants did not perform any other donning tasks afterwards – in total, donning data from five participants were therefore rejected as a result of the broken zipper. Participants, that engaged in a donning task (e.g., putting hood on) after noticing that the zipper was broken, were excluded from the analysis. End trial time for broken zipper 2 could be determined for six participants.

The TPIS Donning Time (DT_{used} and DT_{new}) is defined using equation (1) and (2) in [1] as (t_e - t_s) and (t_e - t_{xe}), respectively, and this assumes an intact zipper. However, it is noted that for participants with broken zippers, t_{e_bz1} and t_{e_bz2} are not comparable to t_e. The use of t_{e_bz1} for calculating DT does not account for the time required by the participant to raise their arms above their head to indicate the end moment of the trial. Similarly, t_{e_bz2} does not account for the time required by the participant to close the zipper and raise their arms above their head. Depicted in Figure SM 1 is a graphical representation of the differences in trial end times. To account for the irregularity, two correction factors were established:

Correction factor arms up, CF_{arms_up}: This correction factor (CF) is used to account for participants with a zipper that fails during the DP. It is determined from data collected from participants with an intact zipper that did not perform any other donning tasks after closing the zipper. The CF_{arms_up} is determined from the time it takes the participant after the zipper was completely closed (event b* in Figure SM *3*) until trial end time t_e. The moment the participant completely closes the zipper is defined from the video footage as the moment that the hand of the participant pulling the zipper up is observed to stop moving upwards, and no further attempt is made to pull the zipper further. The time is then divided by DT of the particular participant to characterize CF_{arms_up} as a proportion of DT. This CF was determined from video data of 51 participants, with the average value used in the determination of DT. The average CF_{arms_up} was determined to be 0.035 or 3.5%.

Correction factor zipper up, CF_{zipper_up}: This CF is used to account for participants with a broken zipper at the start of the trial. It is determined from data collected from participants with an intact zipper that did not perform any other donning tasks after closing the zipper. The CF_{zipper_up} is determined from the time it takes the participant to touch the zipper with the purpose to close it (event a* in Figure SM 3) until trial end time t_e. Again, this time is then divided by DT of the particular participant to characterize CF_{zipper_up} as a proportion of DT. This CF was determined from video data of 20 participants, with the average value used in the determination of DT. The average CF_{zipper_up} was determined to be 0.14 or 14.1%.

Thus, the DT for participants that had a zipper that failed during DP (i.e., DT_{cf_bz1}), and for participants that had a zipper that was broken prior to the trial (i.e., DT_{cf_bz2}), are defined using equations (S7) and (S8), respectively:

$$DT_{cf\_bz1} = \frac{t_{e\_bz1} - t_s}{1 - CF_{arms\_up}} \quad \text{(S7)}$$

$$DT_{cf\_bz2} = \frac{t_{e\_bz2} - t_s}{1 - CF_{zipper\_up}} \quad \text{(S8)}$$

Here t_s should be replaced by t_{xe} in the two formulas if the suit is new. It is noted that DT_{cf_bz1} does not include the time to completely close the zipper, as it breaks before the task is fully accomplished. However, the error margin is judged to be insignificantly small.

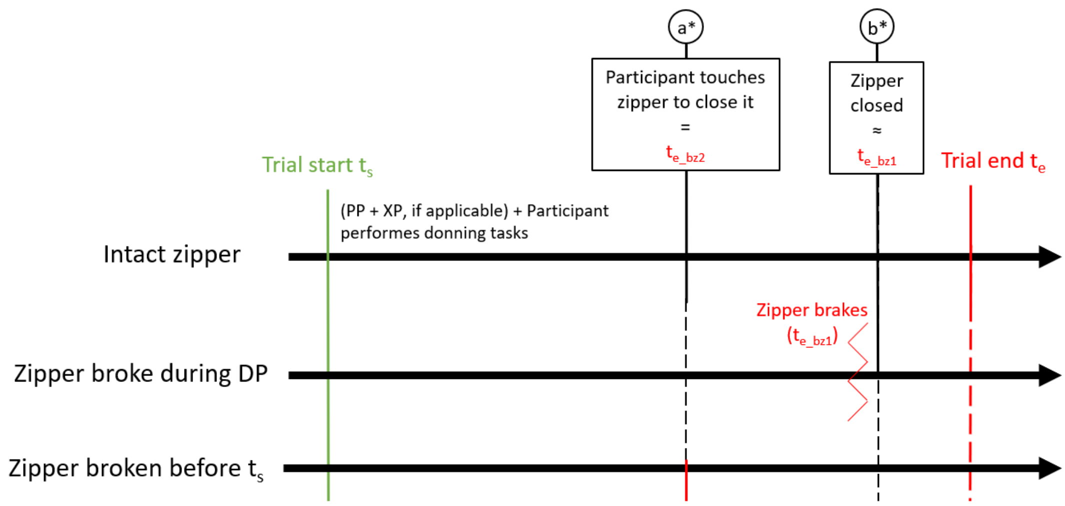


Figure SM 3 Timeline for donning of TPIS for participants with intact zipper and broken zipper

The DT distributions for the participants with the intact zipper (70 data points) and failed/broken zippers (13 data points) with CFs are depicted in Figure SM *4*.

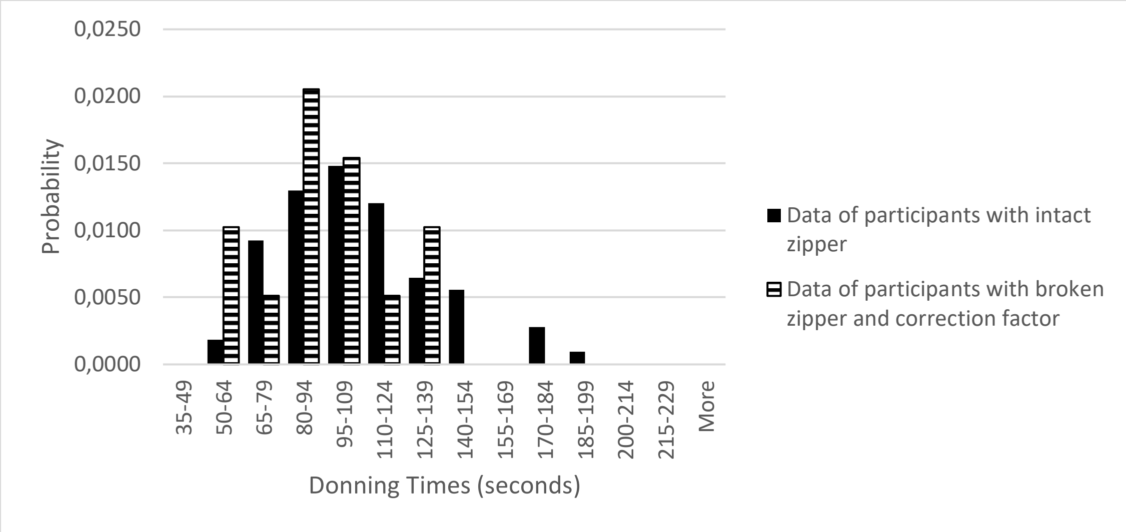


Figure SM 4 Distribution of DT and DT_{cf}

To evaluate if DT data for participants with failed/broken zippers (DT_{cf}) can be merged with the data set representing participants with an intact zipper (DT), non-parametric testing was conducted comparing DT, DT_{cf_bz1}, and DT_{cf_bz2}, and DT with DT_{cf} (combined DT_{cf_bz1}, and DT_{cf_bz2}). An independent-samples Kruskal-Wallis Test was performed with pairwise comparisons of DTs across all correction factor categories. The test showed no statistically significant difference between DT, DT_{cf_bz1}, and DT_{cf_bz2}; $H(2)=1.794$, $p=0.408$. A Mann-Whitney U Test was conducted to assess if the participants with a correction factor (combined DT_{cf_bz1}, and DT_{cf}_bz2) have a significant effect on the DTs of participants having an intact zipper. Results showed no statistically significant difference in DT (Median=93.8 s, Mean=95.4 s, SD=24.4 s) and DT_{cf} (Median=101.3 s, Mean=107.6 s, SD=29.6 s); $U(DT=70, DT_{cf}=13)=352$, $p=0.197$). Thus, the DT data points with correction factors (13 data points) were merged with the DT data set of participants having an intact zipper (70 data points), producing a DT data set comprising of 83 data points in total.

S4: Statistical analysis of quantitative data – additional information

S4.1 Fit of donning time distribution

To support the determination whether the DT data can be defined by a normal or log-normal distribution, Shapiro-Wilk tests were performed, as described in Sec. 4 in [1]. Table SM 1 shows the results (p-values) of the performed tests. A p-value below 0.05 indicates that the null hypothesis is rejected.

Table SM 1 Results of the Shapiro-Wilk tests to assess the fit of the data, referring to Sec. 3.5 in [1]

Null hypothesis	Dependent variable	Number of data points	p-value
The sample of DT for males comes from a normal distribution.	Donning time (DT)	N=54	0.041
The sample of DT for females comes from a normal distribution.	Donning time (DT)	N=29	0.038
The sample of LN(DT) for males comes from a normal distribution.	log-transformed DT	N=54	0.781
The sample of LN(DT) for females comes from a normal distribution.	log-transformed DT	N=29	0.591
The sample of DT for AG1 comes from a normal distribution.	Donning time (DT)	N=30	0.670
The sample of DT for AG2 comes from a normal distribution.	Donning time (DT)	N=31	0.094
The sample of DT for AG3 comes from a normal distribution.	Donning time (DT)	N=22	0.092
The sample of LN(DT) for AG1 comes from a normal distribution.	log-transformed DT	N=30	0.971
The sample of LN(DT) for AG2 comes from a normal distribution.	log-transformed DT	N=31	0.825
The sample of LN(DT) for AG3 comes from a normal distribution.	log-transformed DT	N=22	0.387

S4.2 Impact of age on donning correctness considering male participants in the NE+NI group

For each of the three male age groups (AG), the mean, median and mode number of DEs are AG1 (N=5), 1.0, 1 and 0,1, respectively, while for AG2 (N=7), 0.71, 1, 1 respectively and AG3 (N=11), 1.9, 2, 2 respectively. A Kruskal-Wallis test with post-hoc Dunn's tests was performed to assess the influence of age on the number of DEs. The results suggest a weak correlation between age and the number of DEs ($H(2)=6.016$, $p=0.049$), with AG1 and AG2 making fewer DEs than AG3. When the Dunn's test using Bonferroni correction for multiple tests was used to compare all pairs of groups, the difference in the number of DEs between the pairs of AGs was not found to be significant after Bonferroni adjustment (all $p>0.071$). Table SM 2 shows the results (p-values) of the performed tests. A p-value below 0.05 indicates that the null hypothesis is rejected.

However, given the small number of data points, the Kruskal-Wallis has little power and so to strengthen the statistical power, the AGs were re-defined into two groups, where the partition is represented by the midpoint of the overall age range. With the lower and upper limits set to 18 and 72 years of age, respectively, the new age groups are defined as, AG A with ages ranging from 18 to 45 years and AG B with ages from 46 to 72 years. Previously, the age limit between AG 2 and AG 3 was 50 years of age, which is relatively close to the age limit of the new AGs of 45 years of age. As there was no statistically significant difference in the number of errors between the two younger

AGs (i.e. AG1 and AG2), it is suggested that the new age group definition will not bias the results but provide a stronger statistical power. Furthermore, the number of male participants within each age group is now approximately equal, being 11 in AG A and 12 in AG B.

A Mann-Whitney test for male participants was conducted to assess the impact of AG A and AG B on number of DEs. Results are presented in Sec. 4.2 in [1], in Table SM2, and repeated here as follows: The test indicated a statistically significant difference, U(AG A=11, AG B=12)=30,p=0.02). Males in AG A (i.e., the younger age group) made fewer DEs than males in AG B (i.e., the older age group). This result supports the indicative results above (comparing AG1/AG2/AG3), that age has a significant impact on the number of DEs made, and that the older male population makes substantially more DEs than the younger male population. Unfortunately, the number of females in the AG B was insufficient to undertake a similar analysis.

Table SM 2 Results of the statistical tests to assess the impact of age on donning correctness considering male participants in the NE+NI group, referring to Sec. 4.2 in [1]

Test	Null hypothesis	Dependent variable	Independent variable	Number of data points	p-value
Kruskal-Wallis test with post-hoc Dunn's test	The distribution of DE for males is the same across categories of Age Group (NE+NI group)	Donning error (DE)	Age Group (AG1, AG2, AG3)	AG1: N=5, AG2: N=7, AG3: N=11	0.049
Dunn's test	The distribution of DE for males is the same across AG1 and AG2 (NE+NI group)	Donning error (DE)	Age Group (AG1, AG2)	AG1: N=5, AG2: N=7	1.000 with Bonferroni adjustment
Dunn's test	The distribution of DE for males is the same across AG1 and AG3 (NE+NI group)	Donning error (DE)	Age Group (AG1, AG3)	AG1: N=5, AG3: N=11	0.283 with Bonferroni adjustment
Dunn's test	The distribution of DE for males is the same across AG2 and AG3 (NE+NI group)	Donning error (DE)	Age Group (AG2, AG3)	AG2: N=7, AG3: N=11	0.071 with Bonferroni adjustment
Mann-Whitney U test	The distribution of DE for males is the same across AG A and AG B (NE+NI group).	Donning error (DE)	Age Group (AG A, AG B)	AG A: N= 11, AG B: N=12	0.020

S4.3 Impact of experience on donning time considering the NI group

Sec. 4.4 in [1] reports the performed statistical tests for AG 1 and AG 3. Unfortunately, AG2 (30-50 years) had too few data points in the experienced group (N=3) for a meaningful analysis; however, it is noted that AG1 and AG3 cover both ends of the age range. Furthermore, the data set was divided into two age groups AG A (18-45 years) and AG B (46-72 years) to increase the sub-sample sizes and strengthen the statistical power. Again, a Mann-Whitney test was conducted to assess the influence of level of experience on DT for AG A (N=37) and AG B (N=24). No statistically significant difference in DT between experienced and inexperienced participants was found (AG A: U(NE=25, E=12)=148, p=0.95 ; AG B: U(NE=18, E=6)=45, p=0.55). This supports the finding reported in [1], that for the TPIS investigated in this study, experience was not a significant factor in determining average DT.

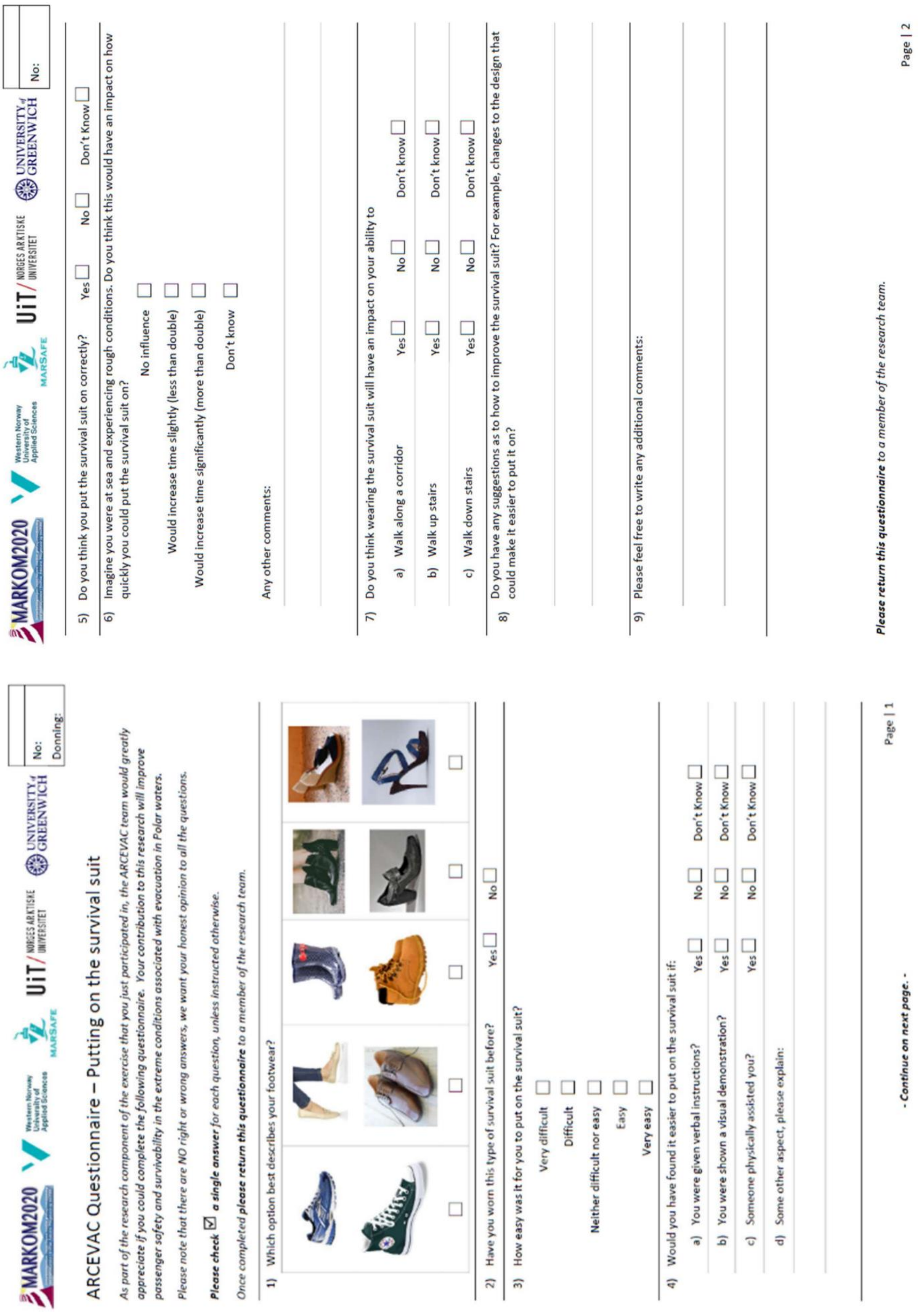
S5: Participants post-trial questionnaires

In total 96 participants donned the TPIS. Following the trials, 95 participants completed the post-trial questionnaires, representing a 99 % completion rate. Note, five participants were excluded from the DT and DC analysis due to incomplete data collection during their donning trial.

In total there were nine main questions in the post-trial questionnaire (see Figure SM 5, Figure SM 6 and [3] for the full details). Questions 1-7 were closed questions, with questions 3-7 relating to their experience while donning the TPIS, while questions 8 and 9 were open questions. Questions 4 and 6 also allowed participants to expand on their answers by writing an explanation. Some participants (N=18, 19 %) skipped questions resulting in missing values. The response rate for closed questions was minimum 80 %, i.e. minimum 8 out of 10 questions (including sub-questions) were answered. Missing values were judged to be completely at random and were removed pairwise. Out of the total of 95 completed questionnaires, 22 participants (23 %) left all open questions unanswered, with a remaining 73 (77 %) answering at least one of the open questions. The number of participants commenting on Q4, Q6, Q8, and Q9 was 22 (23 %), 14 (15 %), 71 (75 %), and 21 (22 %), respectively, see Sec. S2.7 for further details.

Note, that participants who saw the instruction video upfront, received a questionnaire with a slightly modified question number 4 (see Figure SM 6). Answer alternative (b) was modified to reflect the instruction methodology experienced. Thus, instead of answer alternative “b) you were shown a visual demonstration”, they were presented with “b) you were shown a live demonstration”.

The first question (Q1) asked the participants to identify the type of footwear they were wearing during the trial. This question, while not relevant for the donning trials was necessary for exploring the walking performance in subsequent trials as explored by Azizpour et al. [6] and Brünig et al. [7]. In the remainder of this section we present the results from the questionnaires.



No:
Donning:

ARCEVAC Questionnaire – Putting on the survival suit

As part of the research component of the exercise that you just participated in, the ARCEVAC team would greatly appreciate if you could complete the following questionnaire. Your contribution to this research will improve passenger safety and survivability in the extreme conditions associated with evacuation in Polar waters.

Please note that there are NO right or wrong answers, we want your honest opinion to all the questions.

Please check ☑ ***a single answer*** *for each question, unless instructed otherwise.*

Once completed ***please return this questionnaire*** *to a member of the research team.*

1) Which option best describes your footwear? ☐ ☐ ☐ ☐ ☐

2) Have you worn this type of survival suit before? Yes ☐ No ☐

3) How easy was it for you to put on the survival suit?
- Very difficult ☐
- Difficult ☐
- Neither difficult nor easy ☐
- Easy ☐
- Very easy ☐

4) Would you have found it easier to put on the survival suit if:
- a) You were given verbal instructions? Yes ☐ No ☐ Don't Know ☐
- b) You were shown a visual demonstration? Yes ☐ No ☐ Don't Know ☐
- c) Someone physically assisted you? Yes ☐ No ☐ Don't Know ☐
- d) Some other aspect, please explain:

Page | 1

*- **Continue on next page.** -*

No:

5) Do you think you put the survival suit on correctly? Yes ☐ No ☐ Don't Know ☐

6) Imagine you were at sea and experiencing rough conditions. Do you think this would have an impact on how quickly you could put the survival suit on?
- No influence ☐
- Would increase time slightly (less than double) ☐
- Would increase time significantly (more than double) ☐
- Don't know ☐

Any other comments:

7) Do you think wearing the survival suit will have an impact on your ability to
- a) Walk along a corridor Yes ☐ No ☐ Don't know ☐
- b) Walk up stairs Yes ☐ No ☐ Don't know ☐
- c) Walk down stairs Yes ☐ No ☐ Don't know ☐

8) Do you have any suggestions as to how to improve the survival suit? For example, changes to the design that could make it easier to put it on?

9) Please feel free to write any additional comments:

Please return this questionnaire *to a member of the research team.*

Page | 2

Figure SM 5 Post-trial questionnaires

4) Would you have found it easier to put on the survival suit if:

a) You were given verbal instructions? Yes ☐ No ☐ Don't Know ☐

b) You were shown a live demonstration? Yes ☐ No ☐ Don't Know ☐

c) Someone physically assisted you? Yes ☐ No ☐ Don't Know ☐

d) Some other aspect, please explain:

Figure SM 6 Modified question number 4 for participants in the video instruction (VI) group.

S5.1 Question 2, Prior Experience.

Question 2 enquired if the participant had previously worn this type of TPIS before and generated 91 valid data points (four did not respond). In total 13% (12) answered that they had worn this type of TPIS before, 87% (79) answered that they had not worn this type of TPIS before. The impact of prior experience on donning performance is further discussed in [1].

S5.2 Question 3, Ease of donning the TPIS.

Question 3 explored how easy the participants thought it was to don the TPIS using a five-point Likert scale and generated 95 valid data points. Taken across all instruction types, the majority of participants answered that it was easy/very easy to don the TPIS (49.5%, 47), whereas only 15.8% (15) answered it was difficult to don the TPIS, and none of the participant thought that it was very difficult (see Table SM 3). Interestingly, none of the participants in the Video Instruction (VI) group found it difficult/very difficult to don the TPIS. Furthermore, the number of participants finding it easy/very easy to don the TPIS in the Written Instruction (WI) and VI groups (WI: 66.7%, 8; VI: 83.3%, 10) were substantially larger than the corresponding number of participants in the No Instruction (NI) group (40.8%, 29). The difference between the instruction categories was determined to be significant using a one-way ANOVA ($F(2,91)=4.13$, $p=0.019$). Bonferroni Post-hoc tests showed a statistically significant difference between NI and VI group ($p=0.018$), whereas none of the other comparisons were statistically significant different ($p>0.562$). Comparing DT with the perceived difficulty of donning the TPIS (see Figure SM 7), participants that found it difficult to don the TPIS took longer on average than participants that found it easy/very easy. Interestingly, participants that found it very easy to don the TPIS took longer on average than participants that found it easy. This is counter-intuitive, as it would be expected that the easier the participant perceived the TPIS was to don, the faster they would be. However, considering the average number of DEs incurred, the easier to don the TPIS was perceived to be, the lower the average number of DEs (see Figure SM 7). Nevertheless, this trend in DEs was not found to be statistically significant (Independent-Samples Kruskal-Wallis Test, $H(3)=6.631$, $p=0.085$). However, as discussed in [1], the relationship between DT and number of DEs is complex, and this may mask the correlation between the participants perception of the donning ease and DT.

Table SM 3 Response to question 3 related to ease of donning

	Very difficult	Difficult	Neither difficult nor easy	Easy	Very easy
No instruction	0.0% (0)	18.3% (13)	40.8% (29)	35.2% (25)	5.6% (4)
Written instruction	0.0% (0)	16.7% (2)	16.7% (2)	58.3% (7)	8.3% (1)
Video instruction	0.0% (0)	0.0% (0)	16.7% (2)	66.7% (8)	16.7% (2)
Total	0.0% (0)	15.8% (15)	34.7% (33)	42.1% (40)	7.4% (7)
Male	0.0% (0)	20.8% (15)	34.7% (25)	38.9% (28)	5.6% (4)
Female	0.0% (0)	24.1% (7)	27.6% (8)	37.9% (11)	10.3% (3)
No instruction -Male	0.0% (0)	15.4% (8)	44.2% (23)	36.5% (19)	3.8% (2)
No instruction -Female	0.0% (0)	26.3% (5)	31.6% (6)	31.6% (6)	10.5% (2)
No instruction - AG1	0.0% (0)	7.1% (2)	53.6% (15)	28.6% (8)	10.7% (3)
No instruction - AG2	0.0% (0)	15.8% (3)	15.8% (3)	63.2% (12)	5.3% (1)
No instruction - AG3	0.0% (0)	33.3% (8)	45.8% (11)	20.8% (5)	0.0% (0)

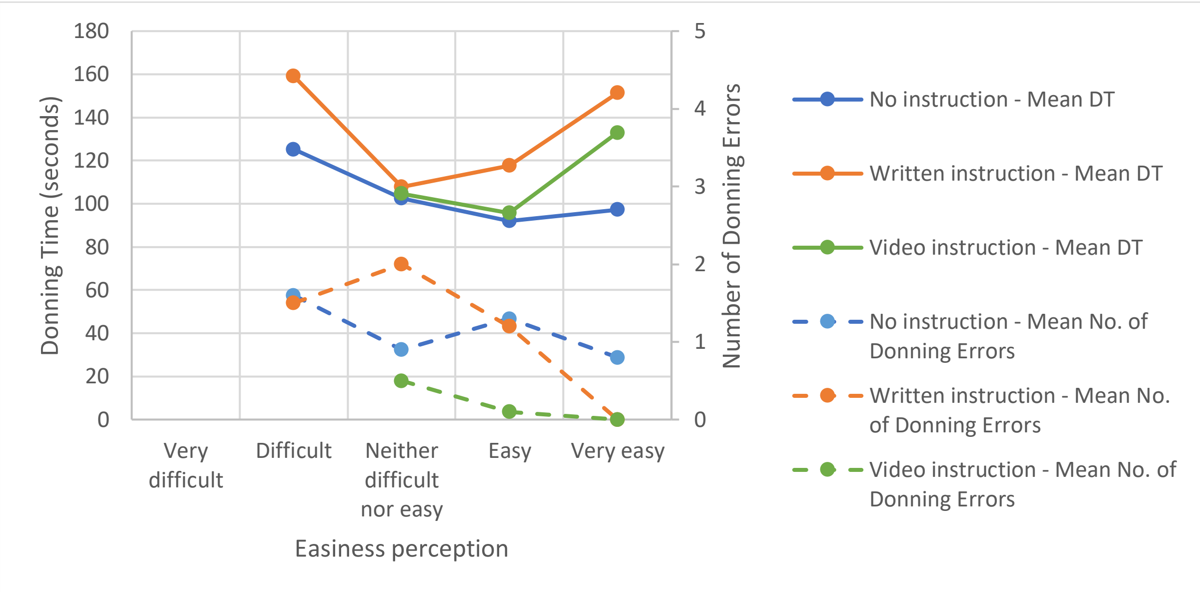


Figure SM 7 Mean Donning Time and Mean Number of Donning Errors for each Easiness perception level according to the instruction category

S5.3 Question 4, Impact of instruction type.

Question 4 explored whether participants thought it would have been easier to don the TPIS if they were given a) either a verbal instruction, b) shown a visual demonstration, or c) if someone physically assisted them.

As mentioned earlier, participants who experienced the instruction video upfront, received a questionnaire with a slightly modified question number 4 (see Figure SM 6). Answer alternative (b) was modified to reflect the instruction method experienced. Thus, instead of answer alternative "b) you were shown a visual demonstration", they were presented with "b) you were shown a live demonstration".

Question 4a), b), and c) generated 89, 94, and 87 valid data points, respectively. Across all instruction categories, 80 % (75) of all participants felt that providing a visual demonstration would have been most helpful (see Table SM *4* and Figure SM *8*). Had verbal instruction been provided, 68 % (44) and 42 % (5) of the NI and WI groups respectively thought that this would have been helpful, whereas the majority of the VI group (58 %, 7) did not feel it would be helpful. Interestingly, across all groups, 56 % (49) of participants felt that physical assistance would not have been helpful with only 30 % (26) suggesting that it would have been helpful. This suggests that the majority did not find the TPIS physically challenging to don. Understandably, participants in the NI group had the highest proportion of participants that thought some additional instruction would have been helpful, with 68 % (44) suggesting verbal instructions and 86 % (60) suggesting visual instructions would have been helpful. This high rate of request for additional instructions is reflected in the measured donning performances, which showed a higher DE rate for the NI compared to the VI group. Similarly, for the WI group, 42% (5) suggested verbal instructions and 67% (8) suggested visual instruction would have been helpful and this is reflected in the higher DE rate and longer DTs incurred compared to the VI group. Thus, it is desirable to provide a visual demonstration of the donning process during the actual donning. However, this is likely to be impractical onboard a vessel during an actual incident.

Table SM 4 Response to question 4 related to alternative methods of instruction in percentage and (number) of participants

	No Instructions			Written Instruction			Video Instruction			Total		
	Yes	**No**	**Don't know**	**Yes**	**No**	**Don't know**	**Yes**	**No**	**Don't know**	**Yes**	**No**	**Don't know**
a) Verbal instructions	68 % (44)	20 % (13)	12 % (8)	42% (5)	33% (4)	25% (3)	42% (5)	58% (7)	0% (0)	61 % (54)	27 % (24)	12 % (11)
b) Visual demonstration	86 % (60)	13% (9)	1% (1)	67% (8)	33% (4)	0% (0)	N/A			80 % (75)	19% (18)	1% (1)
b)* A demonstration by a person (not video)	N/A						58% (7)	42% (5)	0% (0)			
c) Physical assistance	32 % (20)	54 % (34)	14 % (9)	17% (2)	75% (9)	8% (1)	33% (4)	50% (6)	17% (2)	30% (26)	56 % (49)	14 % (12)

*different answer option for VI group

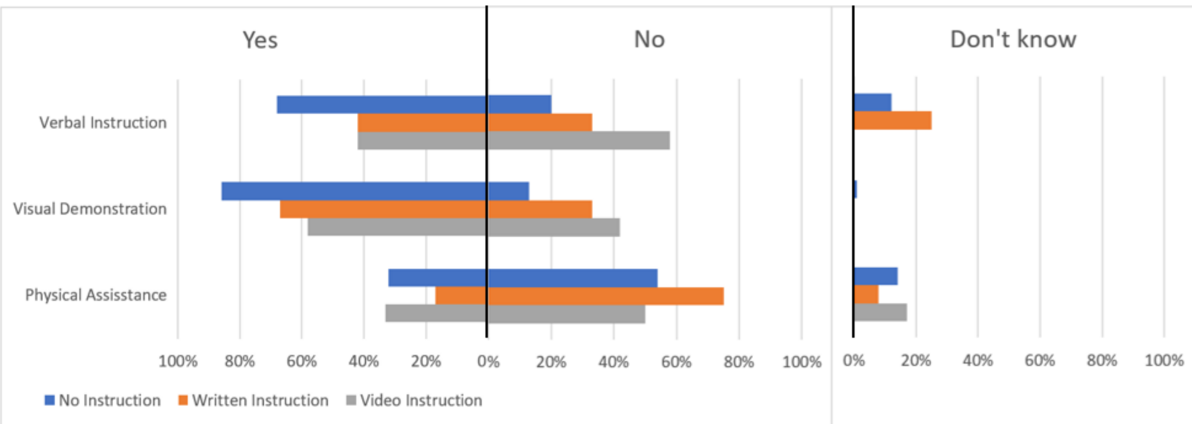


Figure SM 8 Response to question 4 related to alternative methods of instruction in percentage of participants

S5.4 Question 5, Donning Correctness.

Question 5 explored participant perception of whether they had donned the TPIS correctly. All participants with a broken zipper are excluded in this section since the zipper malfunction may have impacted their response, thus resulting in 72 valid data points. Taken across all instruction groups, 71% (50) believed they had donned the TPIS correctly, while 21% (15) felt as though they had made some error(s). This high level of confidence in donning the TPIS correctly suggests that the TPIS was considered, rightly or wrongly, relatively easy to don. Participants in the NI and WI group were reasonably confident in having donned the TPIS correctly with 67% (34) and 67% (6) respectively, believing they had. Nevertheless, all participants in the VI group believed they had donned the TPIS correctly (see Table SM 5).

The actual ease of donning the TPIS is reflected in the incurred DE rate, in particular for the NI group, as these participants had no guidance or instructions. For those in the NI group, 35% (12) of those who thought they had donned the TPIS correctly did so without error while 29% (15) of the NI group as a whole donned the TPIS without error. That so many, even without any form of instruction could don the suit without error suggests that the suit was relatively easy and somewhat intuitive to don. Furthermore, 39% (27) of the entire sample (across all instruction groups) managed to don the TPIS without error.

Of those who thought that the TPIS was correctly donned, 65% (22), 50% (3) and 10% (1) in the NI, WI and VI groups respectively incurred at least one DE. Furthermore, all participants in the WI and VI group were confident in judging their donning correctness, whereas a small number of participants in the NI group stated that they didn't know if they have donned the TPIS correctly.

This suggests that providing donning instructions improves participant confidence in being able to self-evaluate the donning correctness. Furthermore, the self-evaluation achieved a higher percentual correctness when providing donning instructions, where 35% (12), 50% (3) and 90% (9) of those who thought they had donned the TPIS correctly the NI, WI and VI groups respectively, correctly evaluated that they had donned the TPIS without error. Having a faulty perception of an incorrectly donned TPIS is problematic, as passengers may not seek for help or information as to how the TPIS is worn correctly, and thus, might be less likely to correct the error. Incorrectly donning the TPIS may significantly impact evacuation performance and survivability. Ideally, in emergency situations passengers should achieve complete donning correctness, or have the ability to correctly self-evaluate the correctness of their donning and seek assistance or further information to rectify the situation.

Table SM 5 Response to question 5 related to perception of donning correctness

Instruction category	**Number of donning errors**	**Number of participants**		
		Yes	**No**	**Don't know**
No Instructions (NI)	0	12	1	2
	≥1	22	11	3
	All *percentage of NI group*	**34** *67 %*	**12** *24 %*	**5** *10 %*
Written Instructions (WI)	0	3	0	0
	≥1	3	3	0
	All *percentage of WI group*	**6** *67 %*	**3** *33 %*	**0** *0 %*
Video Instructions (VI)	0	9	0	0
	≥1	1	0	0

	All	**10**	**0**	**0**
	percentage of VI group	*100 %*	*0 %*	*0 %*

S5.5 Question 6, Impact of environment on donning time.

Question 6 explored participant perception of whether environmental conditions, e.g., the sea state, would impact their DT and generated 95 valid data points. A ship in distress that requires passengers to don a TPIS is likely to experience adverse movements, such as pitch (heel or trim), heave, or roll, either due to the sea state or to the damage sustained by the vessel. However, the donning trials took place in ideal conditions on land with no movement. For this question there were four possible replies, there would be no impact, a small impact (less than double the DT), a large impact (more than double the DT) or don't know. Only 5 % (5) thought there would be no influence and 4 % (4) didn't know. Over 90 % (86) thought that adverse environmental conditions would increase their DT, with 35 % (33) suggesting it would increase time significantly and 56 % (53) suggesting it would increase slightly. This suggests that the DT reported in this study should be considered optimistic as they were undertaken in the equivalent of dead calm conditions.

Participants were also given the opportunity to comment on their answer, of which 9 (9.5 %) provided a commentary. Three comments suggested that they were likely to make greater DEs in rough conditions. Two pointed out that mental stress, fear, or panic would impact or increase DT and the DE made. One stated that an unstable surface would make it more challenging to get the legs in, but not significantly. And three comments referred to that sitting down on a chair or bed, or sitting on the floor would impact DT negatively, as well as positively. The perception held by most participants that rough sea conditions would only slightly increase DT, reflects on the relative ease of donning this particular TPIS. The comments also suggest that it is necessary to consider adverse environmental conditions as part of the experimental protocol for DTs. Furthermore, the possibility for participants to sit on a chair, bench or similar should be evaluated in the experimental set-up as a control variable, as this may impact DT.

S5.6 Question 7, Impact of TPIS on walking ability.

Question 7 explored participant perception of whether wearing the TPIS would impact their ability to walk along a corridor, ascend and descend stairs, and generated 95, 90 and 90 valid data points, respectively, summing up to a total of 275 responses. There were three possible answers to each question, yes, no, or don't know. Overall, 53% (147) of the responses suggested that the TPIS would impact walking performance, while 38% (105) of the responses said it wouldn't and 8% (23) didn't know. The majority of participants suggested that wearing the TPIS would impact both ascending (60%, 54) and descending (55%, 50) stairs, whereas there was no clear trend in the response to walking along a corridor, where (45%,43) said it would impact and (43%, 41) said it would not impact walking speed.

The TPIS used in these trials was a light weight and spacious suit as one size is intended to be suitable for all adult users. However, the only means by which to adjust the fit of the suit was through the ankle straps. These could be used to gather excess suit material from around the feet so that excess material does not impede walking. It is suggested that the weight of the TPIS would not impact walking speeds in a corridor, but the fit might, especially the fit around the ankles and feet. The fit of the TPIS is dependent on the size and shape of the participant and whether the participant correctly fastened the ankle straps. This is supported by the study by Azizpour et al. [6] which quantifies walking speed along a corridor when participants wear the TPIS. The experimental procedure required that participants correctly wear the TPIS, including having the ankle straps correctly fastened. The study found that at 0° heel, the TPIS did not impact walking speeds along the corridor. The impact of the TPIS on stair ascent/descent speeds is currently being assessed [7].

Table SM 6 Response to question 7 related to participants perception of the impact of the TPIS on walking speed

	Yes	No	Don't know
Walking along a corridor	45 % (43)	43 % (41)	11 % (11)
Walking up stairs	60 % (54)	34 % (31)	6 % (5)
Walking down stairs	55 % (50)	37 % (33)	8 % (7)
Sum	53 % (147)	38 % (105)	8% (23)

S5.7 Questions 8 and 9, suggestions to improve the TPIS and any other comments.

Questions 8 and 9 were open questions, allowing the participants the opportunity to make suggestions as to how they would improve the TPIS suit design and to provide additional general feedback. Questions 4 and 6 also provided the participants the opportunity to comment. A total of 175 comments were recorded. During the analysis of the post-trial questionnaires, it became apparent that it would be beneficial to analyse all the comments from these four questions together, as the comments all tended to describe struggles/difficulties experienced by the participants and/or suggestions to improve the suit design and experimental set-up. Furthermore, the comments were generally concerned with specific topics relating to: instructions, zipper, gusset, ankle straps, gloves, size/fit, TPIS material, and other. The participant's feedback can be used to support donning performance findings and in the evaluation of TPIS design. The responses are summarised below:

a) Instructions (22 comments): Most participants did not have instructions available, and thus, it is not surprising that many (7 out of 15 comments) of these participants (NI group) suggested instructions would be useful. The type of instructions requested varied from instructions with pictures, labelling on the TPIS, to visual demonstration. Also, an explanation for the gusset strap was requested. One participant commented that it was not obvious that one needed to close the gusset before closing the zipper and one commented that they took some time to find the opening and understand how to don the suit.
Participants, that had written instructions, commented on the nature and quality of the instructions (5 comments), e.g., one participant suggested that written instructions describing the donning sequence should only include one key point with each image to make it clearer what was intended by the picture. However, this is in fact already the case, the only difference is that each key point is written in four languages. Another participant commented that it was easy to forget about the hood and that it would be useful to have a reminder. However, image 5 in the instructions specifically relates to donning the hood. One participant pointed out that there was no specific information relating to whether the shoes should be on or off. Several participants (5 comments), with and without instructions, suggested that the TPIS should include specific labelling to assist in correctly donning the TPIS. For example, it was suggested that labelling to identify the front and back of the TPIS, torso opening in strong/visible colour, or colour coding of inside and outside would be useful. And one participant commented that they should have tied the hair back as it was in the way when closing the zipper. Clearly, the participants felt that the quality of the instructions could be improved.

b) **Zipper (31 comments):** Many comments (19 comments) regarding the zipper referred to the poor quality and the broken zippers. Participants suggested improved/better zipper, functioning zipper, stronger/robust zipper. Furthermore, participants commented (10 comments) that it was difficult to grab the zipper with the gloves on and suggested to use a longer/bigger/better puller on the zipper, to make it easier to grab with the gloves on.

c) **Gusset (22 comments)**: Participants commented (9 comments) that there was too much gusset fabric, and that it got in the way while donning, both when inserting their legs into the suit and when closing the zipper. Also, it was suggested that the gusset strap was too long, and that it got in the way when closing the zipper (6 comments). Furthermore, the tightening mechanism received comments (5 comments) on being too difficult to operate while wearing gloves, and participants requested an easier way/a better solution to tighten the gusset strap. One suggested a large toggle on the drawstring. Some participants found it unintuitive/not obvious as to how to close the gusset strap.

d) **Ankle strap (9 comments)**: Participants criticised the use of velcro and the small size of the ankle straps (7 comments). One participant suggested using the same closing mechanism as for the gusset strap. Others suggested using wider/more robust straps. Generally, it was commented that the ankle straps needed improvement, and to be extended so that the leg material also did not hang loose. Furthermore, it was commented that the ankle strap was located behind the foot and so not visible.
e) **Gloves (18 comments):** Most comments (12 comments) indicated that the gloves were too large and that it was difficult to grab small things like the zip puller or getting the gusset strap inside the suit when closing the zipper. Participants suggested to provide removable gloves (6 comments).
f) **Size/fit of suit (22 comments):** The size and/or fit of the suit was perceived to be too big in general. Also, reference was made to the gloves, area around the waist, and arm sleeves, to be too large. Participants suggested more tightening arrangements, like straps around the wrist or waist. It was also suggested that different sizes should be provided. However, this would not be practical as distribution during an emergency would be difficult.
g) **Material (11 comments):** A few comments were made regarding materials used in the construction of the TPIS. Most comments (7 comments) related to the sole of the TPIS sock. It was suggested to add a sole with anti-slip properties to reduce the danger of slipping. One participant reacted to the strong chemical smell of the suit and had difficulties to breath.
h) **Other (28 comments):** Some participants (15 comments) commented that they had no suggestions to improve the TPIS and thought that it was easy/intuitive to put on and that it felt light. A general suggestion was made to minimise closing mechanisms to the essential. One participant commented that when wearing the suit with the hood on and zipper all the way up, that their glasses hurt the bridge of their nose. This participant also commented that their glasses tended to fog up. Another participant commented that their vision was impaired when the suit was fully zipped up.

S6: Improvement suggestions for the written instructions

Presented in Figure 10 in [1] is the frequency for the various types of DEs incurred by the NE+WI (Figure 10b) and the NE+VI group (Figure 10c). For example, by far the most frequent DE for the NE+WI group concerned neglecting to fasten the ankle straps. The donning instructions do not involve a diagram showing someone fastening the ankle straps. In the first pictogram, the ankle straps are clearly unfastened, and in the second and subsequent pictograms the ankle straps are fastened, but there is no pictogram depicting this important action. Furthermore, in the post-trial questionnaire (see Sec. S5), several participants complained about the design of the ankle straps. One participant commented that the straps were fastened to the back of the TPIS leg and so were not necessarily visible. Thus, in addition to improving the provided instructions, the design of the leg straps should be reconsidered.

S7: Comparison of donning performance data of two substantially different TPIS

The TPIS used in the study described in [1] was a lightweight survival suit produced by Hansen Protection (Sea Pass passenger suit). In a similar study by Azizpour et al. [2] a similar trial methodology was employed, however, a significantly different type of TPIS was used, i.e., an immersion suit with fully integrated buoyancy and thermal insulation produced by Viking (Yousafe Blizzard PS5002). While there are some similarities in the key findings for these two different types of suits, there are also significant differences. The differences in performance of the two types of TPIS are important in demonstrating the variation in donning performance that can be achieved by different types of TPIS and in highlighting that it is not possible to generalise donning performance for different types of TPIS. Furthermore, it is important to understand key differences in the trial methodologies used to quantify the TPIS performance. Presented in Table SM 6 is a comparison of the methodologies used in [1] and [2], while presented in Table SM 7 is a comparison of the key TPIS performance results for the two types of TPIS.

Table SM 7 Description of methodology and main parameters described in the TPIS donning trials in [1] and [2]

		Description of methodology and main parameters in [1]	**Description of methodology and main parameters in [2]**
Definitions	Preparation phase (PP)	Represented as Preparation Time (PT) and is only applicable for participants with a new TPIS. Starts from the trial start until the participant attempts to tear open the packaging. It is anticipated that the participant takes some time to read the instructions, that are available on the face of the packaging. Note that for some participants no distinct PT could be determined, and a combined PT and XT was defined.	Represented as Preparation Time (PT). Starts from trial start until participant touches the zipper. It is anticipated that the participant takes some time to read the instructions, that are available on the face of the carrier bag.
	Extraction phase (XP)	Represented as Extraction Time (XT) and is only applicable for participants with a new TPIS. Starts when the participant starts to tear open the packaging until the TPIS is fully extracted. Note that for some participants no distinct XT could be determined, and a combined PT and XT was defined.	Represented as Extraction Time (XT) and is only applicable for participants with a new TPIS. Starts when the participant touches the plastic bag until the TPIS is fully extracted. Note that the XT does not refer to the time it takes to extract the TPIS from the carrier bag.
	Donning phase (DP)	Represented as Donning Time (DT). Starts at the end of XP for participants with a new TPIS, or from trial start with participants with a used TPIS.	Represented as Net Donning Time (NDT). Starts at the end of the PP until the end of the trial and less the XT, if applicable. Note that the NDT includes the time it takes to extract the bag from the carrier bag.
	Donning error	Incorrect completion of a key donning feature, which is used to describe donning correctness (DC). Five key donning features were defined referring to the state of the shoes, hood, ankle straps, gusset straps, and zipper.	Incorrect completion of a key donning feature. Seven key donning features were defined referring to the state of the shoes, hood, ankle straps, interior length straps, zipper, chest buckle, and gloves.
	Method of instruction	Three groups: • Written instructions (WI) • Video instructions (VI) • No instructions (NI)	Two groups: • Written instructions (WI) • Video instructions (VI)
Used analysis	Assessing factors influencing donning correctness	Kruskal-Wallis test and Mann-Whitney test	Poisson regression
	Assessing factors influencing donning time	Mann-Whitney test and Log-linear regression	Log-linear regression and Kruskal-Wallis test

Table SM 8 Comparison of key donning performance parameters for the two different types of TPIS described in [1] and [2]

		Light weight survival suit without integrated buoyancy explored in [1] (Hansen Protection, Sea Pass)	**Immersion suit with fully integrated buoyancy and thermal insulation of the type explored in [2] (Viking Yousafe Blizzard PS5002)**
Results	Preparation time (PT)	Instructions: info graphics, covers the entire surface of package, large images, short text. WI: 2.0 s – 20.3 s VI: 1.3 s - 3.6 s WI vs VI: **Significant** Participants exposed to VI had a shorter PT than participants in the WI group. NE+WI group: 5 of the 10 participants (50 %) were observed to look at the instructions during PP, and 8 of the 10 participants (80 %) were observed to look at the instructions during the trial (i.e. during PP and DP, or DP only).	Instructions: info graphics, but lots of additional information also displayed on the package. WI: 1.0 s – 35.0 s VI: Not reported WI vs VI: Not reported. WI group: For 14 of the 89 participants (16 %) an extended PT was defined, which is anticipated to include time to study the instructions.
	Extraction time (XT)	Packaging: Vacuum sealed package. WI: 4.8 s – 13.1 s VI: 2.8 s – 11.7 s	Packaging: Sealed plastic bag withing a zippered bag. XT refers to the suit extraction of the sealed plastic bag, excluding the extration from the zippered bag (which is included in NDT). VI: 9.8 s – 31.5 s
	(Net) Donning time (DT/NDT)	Best represented by a log-normal distribution. Overall: 55 s – 186 s	Male and female data is best represented by a log-normal distribution. Overall: 65 s – 365 s
	Number of donning	Max possible types of errors: 5	Max possible types of errors: 7

		Light weight survival suit without integrated buoyancy explored in [1] (Hansen Protection, Sea Pass)	Immersion suit with fully integrated buoyancy and thermal insulation of the type explored in [2] (Viking Yousafe Blizzard PS5002)
	errors	Overall: 39% no errors, Mode: 0, Average: 1.0 NE+NI: 28 % no errors, Mode: 1, Average: 1.3 NE+WI: 33 % no errors, Mode: 0,1, Average: 1.1 NE+VI: 80 % no errors, Mode: 0, Average: 0.2	Overall: 4 % no errors, Mode: 2, Average: 2.2 NI: Mode: -, Average: - NE+WI: 6 % no errors, Mode: 3, Average: 2.5 NE+VI: 0 % no errors, Mode: 2, Average: 1.6
Factors influencing donning correctness	Location of data collection	No indication that location influenced the number of donning errors.	Not reported.
	Gender	NE+NI group: **Not significant**.	All: **Not significant**.
	Age	NE+NI group: **Significant**. Males under the age of 50 years made significantly fewer donning errors than males over the age of 50.	All: **Not significant**.
	Experience	NI group: **Not significant**. Tested for those under the age of 50 years. Note that the definition of experience is not accounting for the quality, frequency and how the donning experience was gained.	All: **Significant**. Previous donning experience reduces the expected average number of donning errors by 27 %. However, considered indicative due to the lack of accounting for the quality, frequency and how the donning experience was gained.
	Preparation time	Insufficient data for a meaningful statistical analysis.	Analysis not conclusive: Using all data points: PT is a **significant** predictor. PT reduces number of errors by 4 %/sec preparation time. However, when excluding outlier, PT was **not** found to be a **significant** predictor.

		Light weight survival suit without integrated buoyancy explored in [1] (Hansen Protection, Sea Pass)	**Immersion suit with fully integrated buoyancy and thermal insulation of the type explored in [2] (Viking Yousafe Blizzard PS5002)**
	Method of instruction	WI vs VI (NE group, <50 years): **Significant**. Receiving VI significantly reduces the number of donning errors compared to WI. NI vs WI (NE group): **Not significant**.	WI vs VI (All): **Significant**. However due to small sample size, analysis judged as not conclusive. Receiving VI reduces the expected average number of donning errors by 41 % compared to WI. No NI data.
Factors influencing DT/NDT	Location of data collection	No indication that location influenced the DT.	No indication that location influenced the NDT.
	Gender	NE+NI group: **Not significant**.	All: **Significant***. Females take about 33 % longer to don than males.
	Age	NE+NI group: **Significant**. Increase in DT of 7.5 % per every 10-years increase in age.	All: **Significant***. Increase in NDT of 6.6 % per every 10-years increase in age.
	Experience	NI group: **Not significant**. Note that definition of experience is not accounting for the quality, frequency and how the donning experience was gained.	All: **Significant***. About 17 % reduction in NDT with having previous donning experience. Note that definition of experience is not accounting for the quality, frequency and how the donning experience was gained.
	Preparation time	Insufficient data.	All: **Not significant***.
	Number of errors	Insufficient data. Complex relationship.	Insufficient data. Complex relationship.
	Method of instruction	WI vs VI (NE group, <50 years): **Not significant** However, average for VI is 22 s (about 18 %) lower than for WI.	WI vs VI: **Significant***. About 21 % increase when receiving VI.

		Light weight survival suit without integrated buoyancy explored in [1] (Hansen Protection, Sea Pass)	Immersion suit with fully integrated buoyancy and thermal insulation of the type explored in [2] (Viking Yousafe Blizzard PS5002)
		NI vs WI (NE group, <50 years): Significant. NE+NI group in AG1+AG2 donned the TPIS 33 s faster on average (about 26 %) than WI group.	NI vs WI: No data on NI group available.
Failure rates	Donning correctness	Percentage of participants that made at least one donning error: WI: 70 % NE+WI: 66 % VI: 18 % NE+VI: 20 % NI: 69 % NE+NI: 72 %	Percentage of participants that made at least one donning error: WI: 96 % NE+WI: 95 % VI: 100 % NE+VI: 100 % NI: - NE+NI: -
	2-minutes criteria	NE+VI group: Referring to total donning time (including PT and XT but excluding DT of a lifejacket): 30 %. Referring to total donning time (including PT and XT) and including an average DT for a lifejacket: 100 %. NE+WI group: Referring to total donning time (including PT and XT but excluding DT of a lifejacket): 80 %. Referring to total donning time (including PT and XT) and including an average DT for a lifejacket: 100 %. NE+NI group: Referring to total donning time (including PT and XT but excluding DT of a lifejacket): 44 %. Referring to total donning time (including PT and XT) and including an average DT for a lifejacket: 91 %.	NE+VI group: Failure rates not reported. Calculating NDT based on Eq (14) for 1000 pax, and PT = 1s, Gender = random, Age = random ∈ (18, 72), failure to remove shoes prior to donning = 0, a failure rate of about 93 % can be estimated. NE+WI group: Referring to total donning time (including PT and XT): 96 %. Referring to NDT (excluding PT and XT): 89 %.

*Based on Model 2.

S8: Example of donning time generation for evacuation modelling: Utilizing probability density function for DT and PXT

Here an example is presented demonstrating the generation of Total Donning Times (TDTs) for evacuation modelling purposes based on the recommended Donning Time (DT) and Preparation and Extraction Time (PXT) distributions as outlined in Sec. 6.2 in [1].

The DT is defined by the PDF distributions given by equations 9, 10 and 11 in [1]. Each passenger is randomly assigned a unique DT from the PDF distribution based on their age.

The PXT is defined by the PDF distribution given by equation 12 in [1]. Each passenger is randomly assigned a unique PXT from the PDF distribution.

And the TDT for each unique passenger is defined by, TDT = PXT + DT (see equation (3) in [1]).

The example presented here is based on the requirements of the evacuation framework for passenger ships described in IMO MSC.1/Circ 1533 [8]. The make-up of the ship's population is defined in Table 3.1. in IMO MSC.1/Circ 1533 [8]. An arbitrary total ship population of 3000 passengers was chosen, which is typical of modern large ocean-going cruise ships. For the TPIS explored in this study, DT was found not to be dependent on gender and so the male and female sub-populations are combined. Table SM *9* provides details on population groups and the corresponding passenger numbers according to IMO guidelines for a total of 3000 passengers. As can be seen from Table SM *9*, IMO mandates that 40 % of the ship's population be mobility impaired. However, it is important to note that the DT was quantified based on data collected from healthy, non-disabled participants. Consequently, no data is available regarding donning performance for individuals with mobility impairments. Thus, it is not known if and how the mobility impairments impact the DT until relevant data has been collected.

To showcase the generation of TDT in this section, the population with mobility impairments (1200 passengers) are allocated among the remaining groups using the following distribution: 23% (276) for passengers under 30 years of age, 23% (276) for passengers between 30 and 50 years of age, and 54% (648) for passengers over 50 years of age. This results in a total population of 696 passengers for the males and females younger than 30 years, 696 passengers for the males and females 30-50 years old, and 1608 passengers for males and females older than 50 years.

Table SM 9 Example population make-up

Population groups - passengers	Percentage of passengers (%)	Number of passengers
Males and females younger than 30 years	14	420
Males and females 30-50 years old	14	420
Males and females older than 50	32	960
Males and females older than 50, mobility impaired (1) and (2)	40	1200

Using equation (9), (10), and (11) from [1], random DTs according to the respective PDF were created for the various age groups (AG). Similarly, using equation (12) from [1], a random time for PXT was generated from the PDF for each DT. The TDT for each passenger is then determined by adding DT and PXT (refer to equation (3) in [1]). Furthermore, for display purposes, integers were generated randomly within the limits of the specific age groups for the specific number of passengers. The integers represent the age of the passenger.

Presented in Figure SM 5 and Figure SM 6 are the DT and TDT, respectively, of the randomly generated data points for each age group. For comparison, the DTs collected in the trial for the NE+NI group are also displayed (red dots). As can be seen, the actual DT (and inferred TDT) data points collected from the trials fall within the range of predicted DT/TDTs.

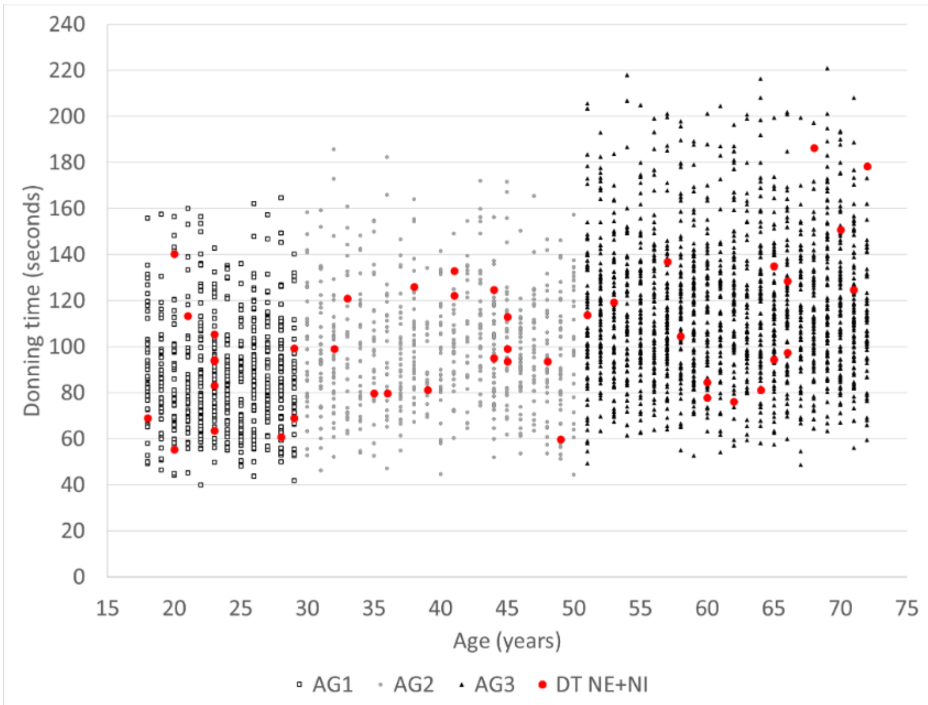


Figure SM 9 Example of randomly created DT for AG1, AG2, and AG3 for 3000 passengers utilising suggested PDFs for evacuation modelling and experimental DTs of NE+NI group (red dot)

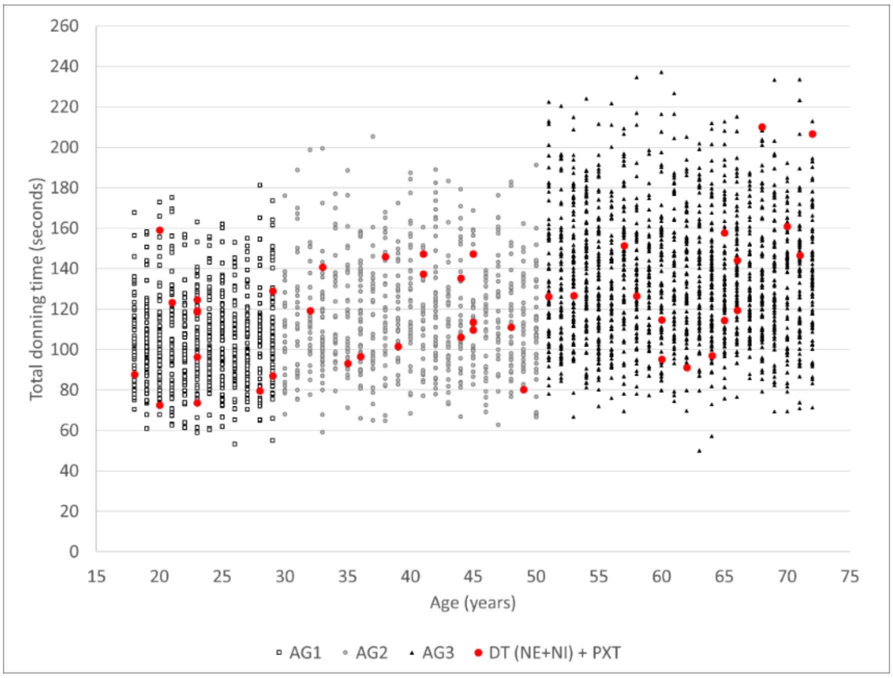
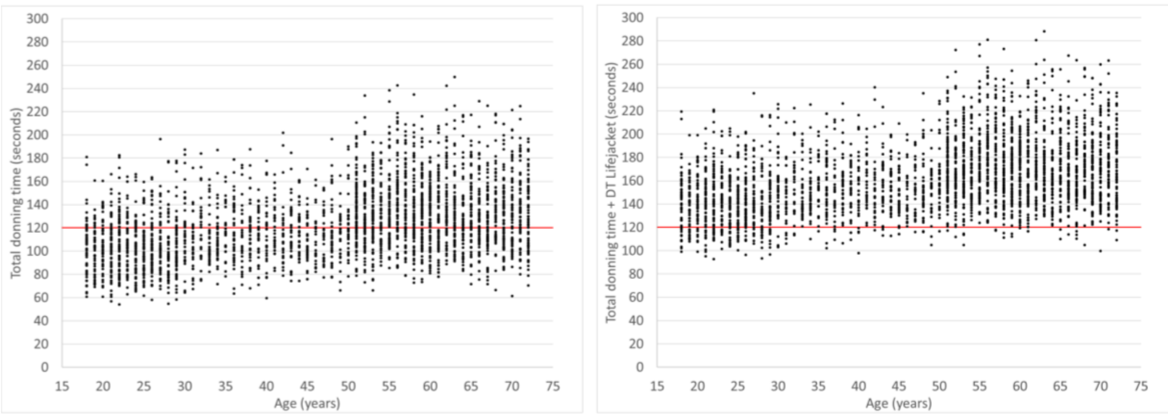


Figure SM 10 Examples of randomly created TDTs for AG1, AG2, and AG3 for 3000 passengers utilising suggested PDFs for DTs and PXTs and experimental DTs of the NE+NI group (red dot) including an additional factor for PXT

Presented in Figure SM 11 are the distribution of 3000 TDTs for the randomly generated data points as a function of age that would be utilised in an evacuation simulation according to the IMO evacuation simulation guidelines [8]. Note that for each repeated simulation for a given IMO required scenario, a different random distribution would be generated based on equations 9 – 12 of [1]. The horizontal red line indicates the maximum 120 s DT specified in IMO/ISO guidelines [9], [10]. As the TPIS investigated in this study did not include a buoyancy device, the TDT presented in figure SM 7a is incomplete. To satisfy the requirements of the IMO [9] the donning process must also include the donning of a lifejacket (see Sec. 6.1 in [1]). As described in Sec. 6.1 of [1] an average lifejacket DT of 38.5 s [11] can be added to the predicted TDT to represent the time required to don the lifejacket, as shown in Figure SM 7b. Clearly, adding an average lifejacket donning of 38.5 s results in virtually all the TDTs exceeding 120 s.

Finally, it is noted that no data is currently available to represent the DTs for mobility impaired passengers, and so they have not been included in the suggested analysis. It is reasonable to assume that DTs for mobility impaired passengers will be longer than those measured in the trials for a non-disabled passenger of the same age. For example, a mobility impaired passenger may require physical assistance from crew or other passengers to don the TPIS or may require to be seated while donning the TPIS. For modelling purposes, rather than ignoring mobility impaired passengers completely from the analysis, as a first approximation it is suggested that irrespective of the age of the mobility impaired passenger, they be randomly assigned the DTs associated with the eldest passengers. In this way, they will be assigned the longest DTs derived from this study. It is accepted that this is a crude approximation but is perhaps better than ignoring mobility impaired passengers completely. However, if this approach is used, it is essential that this approximation is highlighted as part of the simulation limitations.



(a) (b)

Figure SM 11 Examples of 3000 randomly created (a) TDTs utilising suggested PDFs for evacuation modelling and (b) TDTs including lifejacket donning time, with the IMO/ISO 2-minute criteria indicated by the red line

References

[1] R. Brünig, E. R. Galea, S. Erland, B.-M. Batalden, S. Deere, and H. A. Oltedal, "Passenger donning time and donning correctness for a non-insulated immersion suit - an experimental study," *Fire Technol.*, 2025.

[2] H. Azizpour, E. R. Galea, S. Erland, B.-M. Batalden, S. Deere, and H. Oltedal, "Factors influencing the time required to don thermal protective immersion suits correctly," *Saf. Sci.*, vol. 164, p. 106064, 2023.

[3] R. Brünig, E. R. Galea, S. Erland, B.-M. Batalden, and H. A. Oltedal, "A methodology for collecting donning times of thermal protective immersion suits intended to be worn by passengers on vessels operating in cold environment" in *Proceedings of 3rd COTech Conference*, H. G. Lemu, Ed., Stavanger: IOG Conference Series: Material Science and Engineering, 2021.

[4] IMO LSA Code, *International Life-Saving Appliance (LSA) Code*. London, 2017.

[5] International Maritime Organization (IMO), "International convention for the safety of life at sea (SOLAS), as amended," 1974.

[6] H. Azizpour, E. R. Galea, S. Erland, B.-M. Batalden, S. Deere, and H. Oltedal, "An experimental analysis of the impact of thermal protective immersion suit and angle of heel on individual walking speeds," *Saf. Sci.*, vol. 152, p. 105621, 2022.

[7] R. Brünig, E. R. Galea, S. Erland, B.-M. Batalden, S. Deere, and H. A. Oltedal, "The impact of thermal protective immersion suits on walking speeds while ascending and descending stairs (in preparation)".

[8] IMO MSC.1/Circ 1533, *Revised Guidelines for Evacuation Analysis for New and Existing Passenger Ships*. 6 June 2016, 2016.

[9] IMO, *Resolution MSC.81(70) Revised recommendation on testing of life-saving appliances*. 1998.

[10] International Organization for Standardization (ISO), "15027-2:2012 Immersion suits - Part 2: Abandonment suits, requirements including safety," 2012.

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